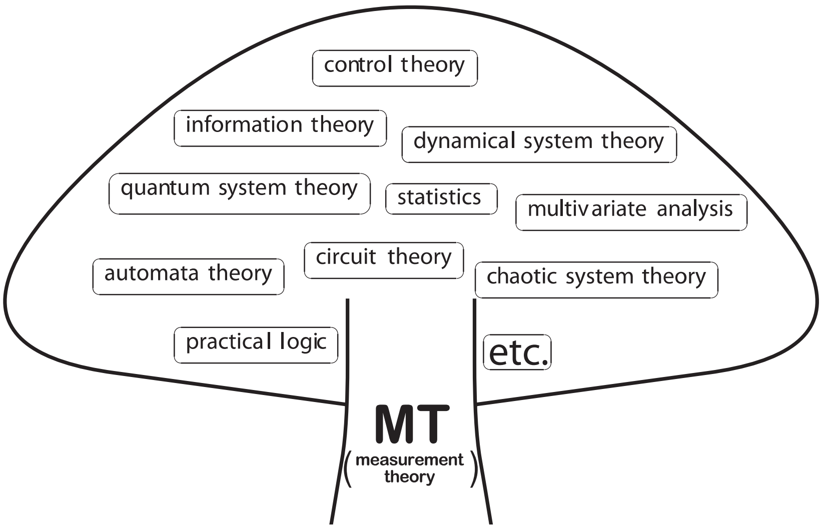
MATHEMATICAL FOUNDATIONS OF MEASUREMENT THEORY

「測定理論の数学的基礎」

Shiro ISHIKAWA
石川 史郎

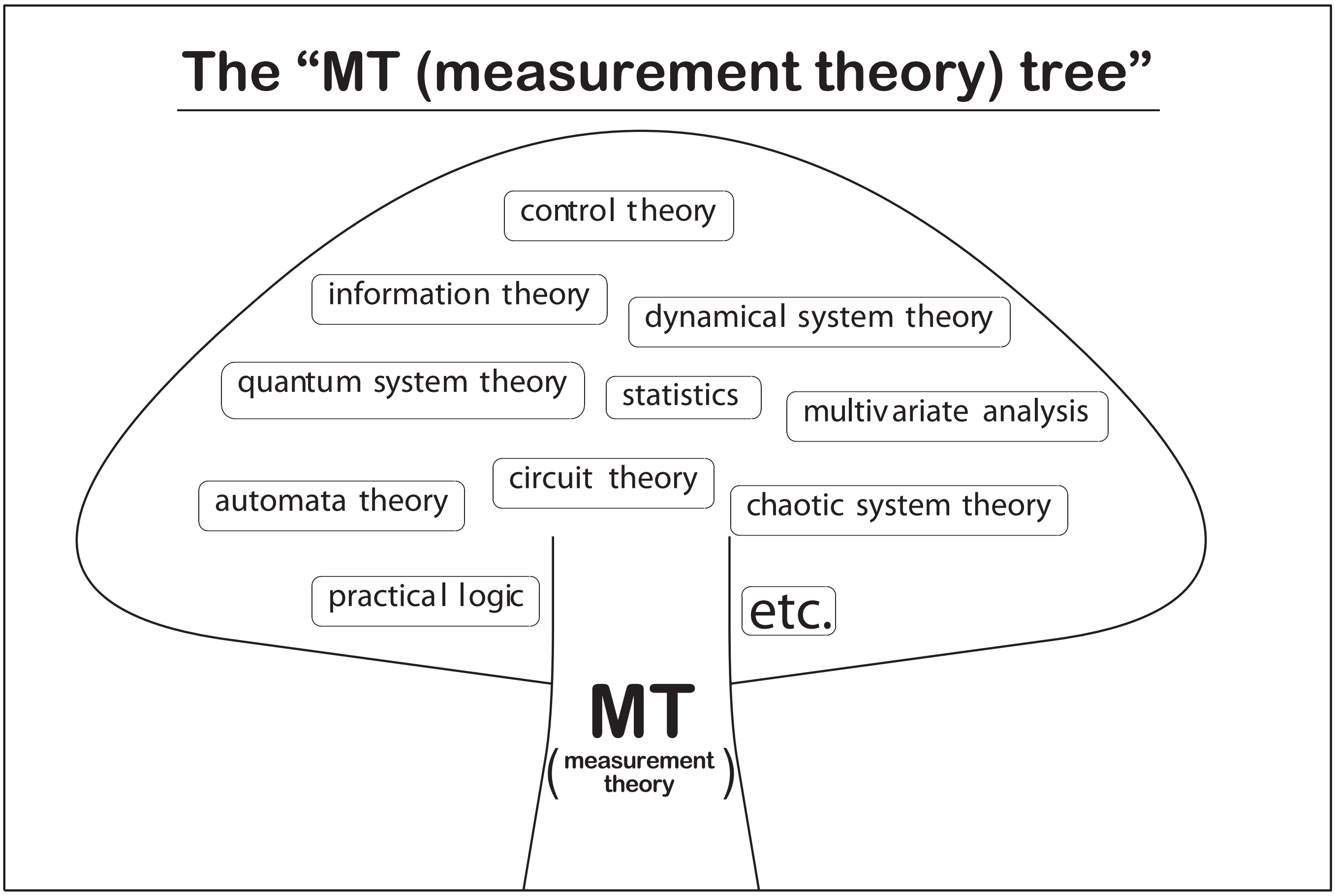


MATHEMATICAL
FOUNDATIONS
OF
MEASUREMENT
THEORY

Keio University Press Inc.

MATHEMATICAL FOUNDATIONS OF MEASUREMENT THEORY

(Japanese title: 測定理論の数学的基礎)



Shiro ISHIKAWA (石川史郎)

Keio University Press Inc. Tokyo

S. Ishikawa
Mathematical Foundations of Measurement Theory
Keio University Press (2006)

Includes bibliographical references and index
AMS Subject Classifications (2000) Primary 62A01, Secondary 93Axx
1. applied mathematics, 2. statistics, 3. system theory
Keio University Press, Tokyo, JAPAN (URL: http://www.keio-up.co.jp/eng/)

The manuscript of this book was written with LaTeX.
The PDF-file can be seen at http://www.keio-up.co.jp/eng/
For the details, also see http://www.keio-up.co.jp/kup/mfomt/

Published by the author,
distributed by Keio University Press Inc.
19-30, 2chome, Mita, Minato-ku
Tokyo 108-8346 Japan

Printed in Japan
First edition, 2006
ISBN978-4-7664-1398-4

Contents

Preface

1 The philosophy of measurement theory **1**
1.1 How to construct "measurement theory" . . . 1
1.2 What is measurement theory? . . . 4
1.3 Measurement theory in engineering . . . 10
1.4 The spirit of "the mechanical world view" . . . 12

2 Measurements (Axiom 1) **15**
2.1 Mathematical preparations . . . 15
2.2 Observables . . . 21
2.3 The meanings of observables and crisp observables . . . 26
2.4 Measurement (Axiom 1) . . . 29
2.5 Remarks . . . 30
2.6 Examples . . . 33
2.7 Operations of observables . . . 42
2.8 Frequency probabilities . . . 45
2.9 Appendix (Bell's thought experiment) . . . 49
2.9.1 EPR thought experiment . . . 49
2.9.2 Bell's thought experiment . . . 51

3 The relation among systems (Axiom 2) **55**
3.1 Newton Equation and Schrödinger equation . . . 55
3.2 The relation among systems (Definition) . . . 58
3.3 Examples (Several tree structures) . . . 60
3.4 The relation among systems (Axiom 2) . . . 64
3.5 Heisenberg picture and Schrödinger picture . . . 66
3.6 Measurability theorem . . . 67
3.7 Appendix (Bell's inequality) . . . 69
3.7.1 Deterministic evolution or Stochastic evolution? . . . 69
3.7.2 Generalized Bell's inequality in mathematics . . . 70
3.7.3 Generalized Bell's inequality in Measurements . . . 73

4 Boltzmann's equilibrium statistical mechanics **75**
4.1 Introduction . . . 75
4.2 Dynamical aspects of equilibrium statistical mechanics . . . 78
4.3 Probabilistic aspects of equilibrium statistical mechanics . . . 86
4.4 Conclusions . . . 88
4.5 Appendix (The law of large numbers) . . . 90

5 Fisher's statistics I (under Axiom 1) **91**
5.1 Introduction . . . 91
5.2 Fisher's maximum likelihood method . . . 92
5.2.1 Fisher's maximum likelihood method . . . 92
5.2.2 Monty Hall problem in PMT . . . 105

5.3 Inference interval . . . 107
5.4 Testing statistical hypothesis . . . 110
5.5 Measurement error model in PMT . . . 117
5.6 Appendix (Iterative likelihood function method) . . . 121

6 Fisher's statistics II (related to Axioms 1 and 2) 127
6.1 Regression analysis I . . . 127
6.1.1 Introduction . . . 127
6.1.2 Regression analysis I in measurements . . . 130
6.2 Bayes operator, Schrödinger picture, and S-states . . . 134
6.3 Regression analysis II in measurements . . . 142
6.4 Conclusions . . . 147

7 Practical logic 149
7.1 Measurement, inference, control and practical logic . . . 150
7.2 Quasi-product observables with dependence . . . 151
7.3 Consistency and syllogism . . . 156
7.3.1 Consistent condition . . . 156
7.3.2 Practical syllogism . . . 162
7.4 Conclusion . . . 168
7.5 Appendix (Zadeh's fuzzy sets theory) . . . 169
7.5.1 What is Zadeh's fuzzy sets theory? . . . 169
7.5.2 Why is Zadeh's paper cited frequently? . . . 172

8 Statistical measurements in C^*-algebraic formulation 175
8.1 Statistical measurements (C^*-algebraic formulation) . . . 176
8.1.1 General theory of statistical measurements . . . 176
8.1.2 Examples of statistical measurements . . . 182
8.1.3 Problems (statistical measurements) . . . 188
8.2 General statistical system (Example) . . . 193
8.3 Bayes theorem in statistical MT . . . 195
8.4 Kalman filter in noise . . . 199
8.4.1 The measurement theoretical formulation of Figure (8.45) . . . 200
8.4.2 Kalman filter in Noise . . . 202
8.5 Information and entropy . . . 204
8.6 Belief measurement theory (=BMT) . . . 208
8.6.1 The general argument about BMT . . . 209
8.6.2 The principle of equal weight . . . 212
8.6.3 Is BMT necessary? . . . 216
8.7 Appendix (Bertrand's paradox) . . . 218
8.7.1 Review (Bertrand's paradox) . . . 218
8.7.2 Bertrand's paradox in measurement theory . . . 219

9 Statistical measurements in W^*-algebraic formulation 225
9.1 Statistical measurements (W^*-algebraic formulation) . . . 225
9.2 The relation among systems (Proclaim$^{W^*}$2 in SMT$^{W^*}$) . . . 236
9.3 Quantum mechanics in $B(L^2(\mathbf{R}))$ . . . 244
9.3.1 Schrödinger equation and Heisenberg kinetic equation . . . 244
9.3.2 A simplest example of Schrödinger equation . . . 245
9.3.3 The de Broglie paradox . . . 247
9.4 The method of moments . . . 248
9.4.1 The moment method . . . 248
9.4.2 Example 1 [Normal distribution (= Gaussian distribution)] . . . 250
9.4.3 Example 2 (measurement error model in SMT) . . . 252
9.5 Principal components analysis in MT . . . 255

10 Newtonian mechanics in measurement Theory **261**
10.1 Kolmogorov's extension theorem in W^*-algebra . . . 261
10.2 The definition of "trajectories" . . . 264
10.3 Trajectories and Newtonian mechanics . . . 266
10.4 Brownian motions . . . 269
10.5 Conclusions . . . 273

11 Measurement error **275**
11.1 Approximate measurements for quantities . . . 275
11.2 The estimation under loss function in statistics . . . 277
11.3 Random observable . . . 282
11.4 The principle of equal probability . . . 288

12 Heisenberg's uncertainty relation **293**
12.1 Introduction . . . 294
12.2 Example due to Arthurs-Kelley . . . 297
12.3 Approximate simultaneous measurement . . . 299
12.4 Lemmas . . . 306
12.5 Existence theorem . . . 308
12.6 Uncertainty relations . . . 311
12.7 EPR-experiment and Heisenberg's uncertainty relation . . . 315

Postscript **319**

Bibliography **325**

Index **336**

PREFACE

It is well known that the dynamical system theory (DST) starts from the following equations:

$$\boxed{\text{DST}} = \begin{cases} \frac{dx(t)}{dt} = f(x(t), u_1(t), t), \ x(0) = x_0 & \cdots (\text{state equation}), \\ y(t) = g(x(t), u_2(t), t) & (\underline{\underline{\text{measurement}}} \text{ equation}) \end{cases} \qquad \text{(D)}$$

where u_1 and u_2 are external forces (or noises). Also recall that quantum mechanics is formulated as the following form:

$$\boxed{\text{quantum mechanics}} = \underset{(\text{Schrödinger equation})}{[\text{the rule of time evolution}]} + \underset{(\text{Born's quantum measurements})}{\underline{\underline{[\text{measurement}]}}} \qquad \text{(Q)}$$

The above two theories (D) and (Q) are, of course, fundamental and famous. Thus, a quarter of a century ago, I already knew them. However, about fifteen years ago, I was suddenly surprised by the similarity between (D) and (Q), particularly, the fact that:

(F) the term "measurement" is common to both dynamical system theory (D) and quantum system theory (Q).

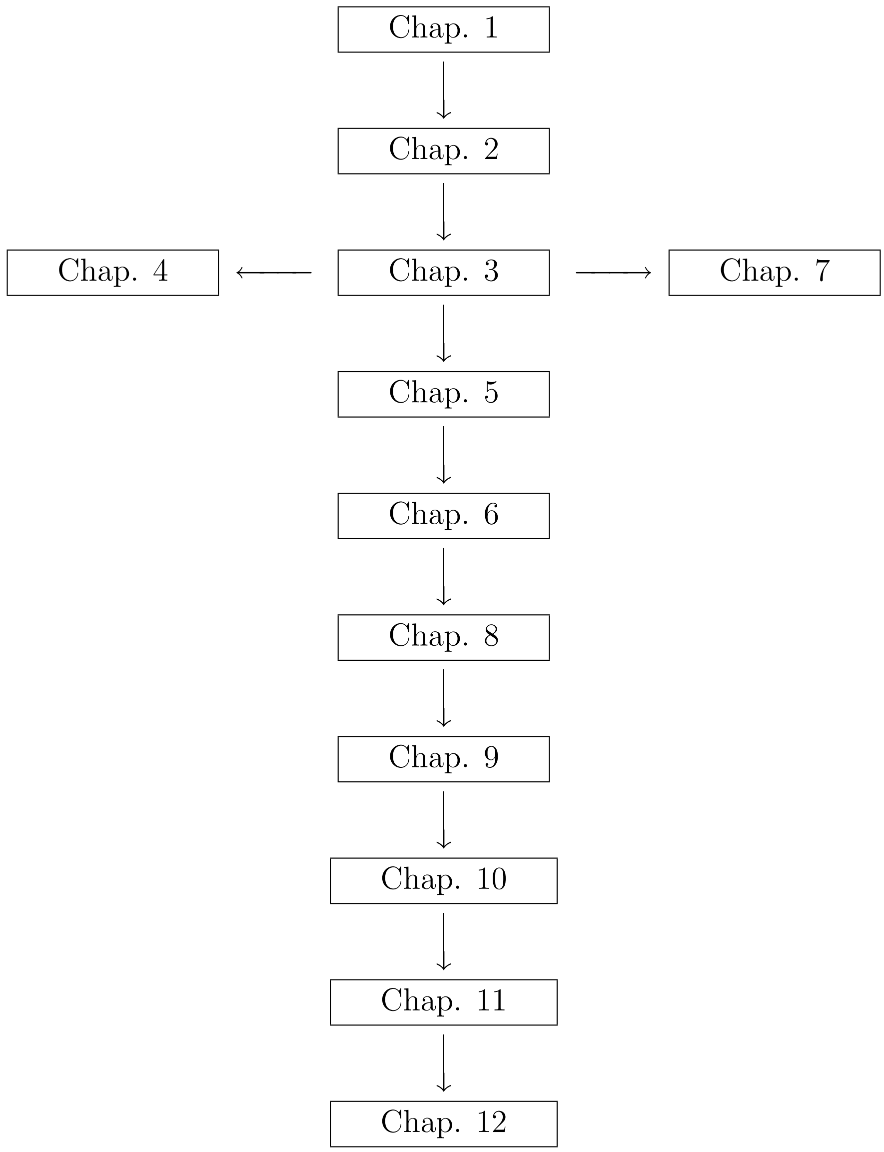
This surprise urged me to propose "measurement theory". I want to share my surprise with all people.[1] This is the reason for this book.

Shiro ISHIKAWA[2]

21st, October, 2006

[1]Some sections of this book were lectured in the master-course program: "Advanced study of mathematics A" at Keio university (three-hour lecture every week from April to July in 2006).
[2]For the further information of our theory, see "http://www.keio-up.co.jp/kup/mfomt/"

It is recommended to read this book as follows:



Chapter 1

The philosophy of measurement theory

The purpose of this book is to propose "mathematical foundations of measurement theory". The statement:

$$\text{\textit{“There is no science without measurements”}} \tag{1.1}$$

is an old famous saying, which of course emphasizes the importance of "measurement". We believe in the saying, i.e., the concept of "measurements" should be the most fundamental in science. However, it is certain that we do not have an authorized "measurement theory" in science yet. Thus, we think that it is worthwhile proposing the mathematical foundations of "measurement theory" [1]:

Chapters 2, 3, 8	···Mathematical foundations of measurement theory
Chapter 4	···An application (of measurement theory) to statistical mechanics
Chapters 5∼12	···Several theories (e.g., statistics, classical and quantum system theories, etc.) in measurement theory

It should be noted that "measurement theory" and "theoretical physics" are different. In particular, their philosophies are completely different. Although it is a matter of course that it is impossible to understand the philosophy of measurement theory without the complete knowledge of measurements (i.e., the contents of Chapters 2 ∼ 12), the philosophy of measurement theory is also indispensable for the understanding of measurement theory. Therefore, in this first chapter, we devote ourselves to the philosophy of measurement theory.

1.1 How to construct "measurement theory"

It is well known that the dynamical system theory (DST, classical system theory)

[1] The measurement theory is proposed in the references [41]∼[48],[55] in this book. We devote ourselves to the mathematical aspect of "measurement theory". For the other aspects (e.g., practical and general aspects), see [30], which is educational and enlightening.

starts from the following equations:

$$\boxed{\text{DST}} = \begin{cases} \frac{dx(t)}{dt} = f(x(t), u_1(t), t), \ x(0) = x_0 & \cdots \text{((stochastic) state equation)}\ ^{2} \\ y(t) = g(x(t), u_2(t), t) & \text{(measurement equation)} \end{cases} \tag{1.2a}$$

where u_1 and u_2 are external forces (or noises),

or more precisely,

$$= \text{"Apply (1.2}a\text{) to every phenomenon by an analogy of Newtonian mechanics and the coin-tossing problem"}\ ^{3} \tag{1.2b}$$

That is, DST is usually believed to be a kind of epistemology called "the mechanical world view", namely, an epistemology to understand and analyze (moreover, control) every phenomenon — economics, psychology, engineering and so on — by an analogy of Newtonian mechanics (and coin-tossing).

Also recall that quantum mechanics is formulated as the following form (*cf.* von Neumann [84]):

$$\boxed{\text{quantum mechanics}} = \underset{\text{(Born's quantum measurement)}}{[\text{measurement}]} + \underset{\text{(Schrödinger equation)}}{[\text{the rule of time evolution}]} \tag{1.3}$$

which was discovered by W. Heisenberg, E. Schrödinger, M. Born in between 1924 and 1926.

Here, it should be noted that the term "measurement" appears in both (1.2) and (1.3). Thus, our proposal, i.e., "measurement theory (=MT)", is constructed as follows:

(I_1) Quantum mechanics (1.3) is formulated in $B(H)$, the algebra composed of all bounded linear operators on a Hilbert space H (*cf.* von Neumann (1932: [84])). Thus it is easy to generalize quantum mechanics in C^*-algebra $\mathcal{A}$ ($\subseteq B(H)$, *cf.* Definition 2.1 in §2.1) such that it includes DST (1.2) as a special case. Namely, (1.2)+(1.3)$\subset$"MT".

That is, as a kind of generalization of quantum mechanics (1.3), we can propose as follows:

---

[2] A stochastic differential equation (or stochastic difference equation) in dynamical system theory is usually called a *stochastic state equation.*

[3] That is, DST is, from the mathematical point of view, based on "the theory of differential equations" and "probability theory". Thus, I think that I.Newton (*cf.* [66]) and A.Kolmogorov (*cf.* [56]) are greatest in DST.

$$\boxed{\text{“measurement theory (or in short, MT)”}}$$

$$= \underset{\text{“Axiom 1 (2.37)”}}{[\text{measurement}]} + \underset{\text{“Axiom 2 (3.26)”}}{[\text{“the rule of the relation among systems”}]} \qquad \text{in } C^*\text{-algebra } \mathcal{A} \tag{1.4a}$$

or more precisely,

$$= \text{“Apply (1.4}a\text{) to every phenomenon by an analogy of quantum mechanics”} \tag{1.4b}$$

(For the details, see Chapter 2 [Axiom 1 (2.37)], and Chapter 3 [Axiom 2 (3.26)]). Here it should be noted that MT (= Axiom 1 + Axiom 2) is composed of a few key-words i.e.,

(I_2) *system, state, observable, measurement, measured-value, probability, Markov relation, sequential observable, Heisenberg picture, etc.*

and Axioms 1 and 2 explain how to use these words. Roughly speaking, Axioms 1 and 2 say *“Use these words by analogy of quantum mechanics”*.[4]

We have the classification of MT as follows:[5]

$$\boxed{\text{“MT”}} = \begin{cases} \boxed{\text{“classical MT”}} & \text{in a commutative } C^*\text{-algebra } C_0(\Omega) \\ \boxed{\text{“quantum MT”}} & \text{in a non-commutative } C^*\text{-algebra } B(H) \end{cases} \tag{1.5}$$

where a C^*-algebra is either commutative or non-commutative. Also, as mentioned in (I_1), we consider the following correspondence:

$$\boxed{\text{“MT”}} = \begin{cases} \boxed{\text{“classical MT” in (1.5)}} & \leftrightarrow \ \boxed{\text{DST in (1.2)}} \\ \boxed{\text{“quantum MT” in (1.5)}} & \leftrightarrow \ \boxed{\text{quantum theory in (1.3)}} \end{cases} \tag{1.6}$$

[4] Thus, our approach is, from the philosophical point of view, characterized as so called *foundationalism*.

[5] As seen later (i.e., Chapter 8), we also have the classification of MT, i.e., “(pure) measurement theory (= PMT)” and “statistical measurement theory (=SMT)”. That is,

$$\text{MT (=“measurement theory”)} \begin{cases} \text{PMT (=“(pure) measurement theory”) in Chapters } 2 \sim 7 \\ \text{SMT (=“statistical measurement theory”) in Chapters } 8 \sim) \end{cases}$$

PMT is essential. That is, we can say that there is no SMT without PMT. (Cf. Chapter 8.)

1.2 What is measurement theory?

We think that the question "What is measurement theory?" is much more difficult than the question "How is measurement theory constructed?".

As mentioned in (I_1) in §1.1, MT is the mathematical generalization of quantum mechanics (1.3). That is, MT is not quantum mechanics but "something beyond mechanics". Thus, we can assert that

(I_3) MT is the mathematical representation of the epistemology called "the mechanical world view" (just like DST(1.2) is).

Also, it should be noted that MT is quite a wide theory, that is, we assert:

(I_4) MT is the most fundamental theory of so-called "theoretical informatics", including *dynamical system theory, quantum system theory, practical logic, statistics, circuit theory, control theory, chaotic system theory, multivariate analysis, information theory, automata theory, OR, game theory, etc.*

This will be discussed in Chapters 5 $\sim$ 12. Also, note that the above (I_4) should be regarded as the same as the following assertion:

(I_5) The term: "theoretical informatics" is defined as the academic discipline that is composed of all theories understood in MT. That is, "theoretical informatics" = "MT".

We assert that

$$\begin{cases} \text{the most fundamental theory of theoretical physics} \Longrightarrow \text{`TOE (string theory(?))'}^{6} \\ \text{the most fundamental theory of theoretical informatics} \Longrightarrow \text{MT} \end{cases}$$

And therefore, we can present the following table, which indicates where MT is in science.

[6] The string theory (*cf.* [28]) is not necessarily authorized yet. Thus, in this book, the term 'TOE (Theory of Everything)' is used as the symbol of the most fundamental theory of theoretical physics. As emphasized in this section, the philosophy of theoretical physics is different from that of theoretical informatics. And thus, the meanings of "the most fundamental theory" are respectively different in theoretical physics and in theoretical informatics.

Table (1.7)

- **Sci. Theor.**
  - **Mathematics** "Cantor's sets theory" — mathematical logic, algebra, geometry, analysis, etc. (C_1)
  - **Math. Sci. Theor.**
    - **Theor. Physics** 'TOE' — (C_2)
      - Newtonian mechanics
      - quantum mechanics
      - Maxwell's electromagnetic theory
      - Einstein's relativity theory
      - Weinberg-Salam theory
      - quantum chromodynamics
      - etc.
    - [My proposal in this book] **Theor. Informatics** MT (=measurement theory) — (C_3)
      - dynamical system theory
      - quantum system theory
      - practical logic
      - statistics, circuit theory
      - control theory
      - multivariate analysis
      - information theory
      - chaotic system theory
      - automata theory
      - OR, game theory, etc.
    - **The third mathematical scientific theory** (undiscovered) (C_4)
  - **Usual Sci. Theor.** — economics, chemistry, biology, medicine, psychology, statistical mechanics, fluid mechanics, engineering (also, see (I_7) and (I_8) later), etc. (C_5)

That is, the mathematical structures of all theories in (C_3) are common, and thus, they are discussed in the framework of MT.

We add the following remark.

Remark 1.1. (About Table (1.7)).

(a). Note that the class (C_3) $(= (I_4))$ is usually called "applied mathematics". In this sense, we think that MT is the main part of so-called applied mathematics.

(b). For example, if electromagnetic theory and relativity theory can not be unified, we must consider two categories (e.g., "Theoretical physics (I)" and "Theoretical physics (II)") in theoretical physics. However, most physicists believe that physics consists of only one category, that is, the theories in (C_2) must be unified in the most fundamental theory (= 'TOE'). The purpose of this book is, of course, to show that the theories in (C_3) are mathematically understood in MT. Also, in this book, "Newtonian mechanics" [resp. "quantum mechanics"] in MT is called "classical system theory (= dynamical system theory)" [resp. "quantum system theory"] (though the addition of "measurement equation" to DTS(1.2a) should be regarded as the act of genius (since there is no concept of "measurement" in Newtonian mechanics)). That is, the two (i.e., Newtonian mechanics and quantum mechanics) are common in both "theoretical physics" and "theoretical informatics" (*cf.* §10.5).

(c). The purpose of theoretical physics is to represent "natural forces" in terms of mathematics. On the other hand, as mentioned in (I_3), MT is a kind of epistemology called "the mechanical world view", namely, an epistemology to understand and analyze (moreover, control) every phenomenon — economics, psychology, engineering and so on — by an analogy of mechanics. That is, MT is the mathematical representation of "the mechanical world view". Or, precisely speaking, the definition of "the mechanical world view" is given by MT.

(d). From the mathematical point of view, the difference between "theoretical physics" and "theoretical informatics" is that of "differential geometry" and "the theory of Hilbert spaces (or operator algebras)". Cf. Remark 8.26.

(e). It is a matter of course that the theories in theoretical physics (= (C_2) in (1.7)) should be tested by experiments. For example, the question: "Is electromagnetic theory experimentally true or not?" is meaningful. In fact, serious experiments have been often conducted as big projects (such as SERN, Kamioka Observatory (Japan), etc.) in theoretical physics. On the other hand, the experimental tests of the theories in theoretical informatics (= (C_3)) are nonsense. For example, the experimental test of statistics is meaningless just like that of mathematics (e.g., linear algebra) is obviously meaningless.[7] Thus, we think that the question: "Is statistics experimentally true or not?" is meaningless. However, it should be noted that the question: "Is statistics convenient (= useful)?" is meaningful.

(f). We hope that some will find and propose "The third mathematical scientific theory in (C_4)".

■

[7]There may be some truth in the assertion that statistics is a kind of mathematics. However, as mentioned in Table (1.7), we think that "statistics" = "mathematics + something".

Summing up, we assert the following table:

Table (1.8a)

	Theoretical Physics	Theoretical Informatics
(1). the theories in this field (*cf.* Remark 1.1 (b))	classical and quantum mechanics, electromagnetic theory, Weinberg-Salam theory, etc. (*cf.* (C_2))	dynamical system theory, statistics, logic, quantum system theory, information theory, etc. (*cf.* (C_3))
(2). the most fundamental theory (*cf.* Remark 1.1 (b))	'TOE (Theory of Everything)' (will be proposed in the future)	MT (measurement theory) (proposed in this book *cf.* [41]∼[48],[55])
(3). the purpose (*cf.* Remark 1.1 (c))	the mathematical representation of "force"	the mathematical representation of "the mechanical world view"
(4). mathematical language (*cf.* Remark 1.1 (d))	differential geometry (gauge theory)	operator algebra (functional analysis, real analysis)
(5). experimentally true or false (*cf.* Remark 1.1 (e))	meaningful	meaningless

Next let us consider the following problem.

Problem 1.2. ("experimentally true or false" and "theoretically true or false"). Consider the following problems (i) and (ii).

(i) Assume that someone proposes "psychokinetic theory" as a theory of theoretical physics. Determine whether his/her theory is true or false.

(ii) In [93], Zadeh proposed "the fuzzy sets theory" as a theory of theoretical informatics. Determine whether his theory is true or false.[8]

[*Answer* (i)]. The problem (i) is solved by two methods. One is the experimental test. If it is OK (i.e., if it is experimentally true), "psychokinetic theory" should be accepted as a physical theory. Also, if we have the most fundamental theory (= 'TOE'), we can determine whether "psychokinetic theory" is theoretically true in 'TOE'. If it is OK (i.e., if it can be understood in 'TOE'), the "psychokinetic theory" should be accepted as a physical theory. Of course, it always holds that "experimentally true" = "theoretically true".

[8] One of our motivations for this research may be inspired by the fashion of Zadeh's fuzzy sets theory (*cf.* [93], which is the most cited paper in all fields of 20th century science) in 1980s ∼ 1990s. We had a lot of arguments about "Is Zadeh's fuzzy sets theory true or false?" or "Can it be justified?" However, these arguments may be fruitless. That is because all controversies were engaged without the understanding of the meaning of "true" (or "justification"). It should be noted that we do not only have the answer to the question: "Is Zadeh's fuzzy sets theory (theoretical) true or false?" but also "Is Fisher's statistics (theoretically) true or false?". (These will be respectively answered in Chapter 5∼7.) In this sense, we can say that the purpose of this book is *to introduce the criterion: "theoretically true or false" into theoretical informatics.* (Cf. Declaration (1.11) later). Here, two criterions of "theoretically true or false (in theoretical informatics)" and "useful or not (in informatics-related engineering)" should not be confused. Throughout this book we are not concerned with "useful or not" but "theoretically true or false", though we, of course, know that the criterion "useful or not" is also quite important.

[*Answer* (ii)]. On the other hand, the problem (ii) is solved by one method. If we have the most fundamental theory (='measurement theory'), we can determine whether "Zadeh's fuzzy sets theory" is theoretically true or false in the most fundamental theory. If it can be understood in the most fundamental theory, "Zadeh's fuzzy sets theory" should be accepted as a theory of theoretical informatics. Our answer will be presented in Chapter 7. However, as mentioned in Remark 1.1 (e), it should be noted that the question: "Is Zadeh's fuzzy sets theory experimentally true or not?" is nonsense.

■

Remark 1.3. (What should measurement theory be applied to?). Recall that MT is a kind of epistemology called "the mechanical world view", namely, an epistemology to understand and analyze (moreover, control) *every* phenomenon by an analogy of mechanics. In this sense, MT may be applied to everything. However, it is certain that some problems (or phenomena) are fit for "the mechanical world view", but others are not. Thus, we have the following question.

(I_6) What phenomenon should measurement theory be applied to?

The following fields are generally believed to be fit for "the mechanical world view" to some degree.

(I_7) *the fields in informatics-related engineering, e.g., information engineering, administration engineering, mathematical psychology, statistical medicine, mathematical economics, financial engineering, cognitive engineering, quality control engineering, chaotic engineering, electrical circuit engineering* [9]*, etc.*

And further, we add

(I_8) *statistical mechanics, fluid mechanics, etc.*[10]

though the two are usually believed to belong to theoretical physics. As mentioned later (i.e., the footnote under (I_{13})), the theories in (C_5) in Table (1.7) should be studied by several methods (and not only by "the mechanical world view" (= MT)). Also, we say

[9] For example, the distinction between "electrical circuit engineering" in (I_7) and "circuit theory" in (I_4) may be ambiguous. However, we want to say "MT itself is not engineering but the mathematical representation of "the mechanical world view".

[10] Boltzmann's statistical mechanics will be discussed as one of applications (of MT) in Chapter 4. Therefore, there is a reason to call "theoretical physics" [resp. "theoretical informatics"] "the first physics" [resp. "the second physics"] .

(I_9) It is too optimistic to consider that the completely precise theory exists in (I_7) and (I_8). However, the theories in (I_7) and (I_8) may be "almost experimentally true" to such a degree that they are assured to be "useful". That is, every theory in (I_7) and (I_8) is, more or less, ambiguous. Although the challenge to make a precise theory should be worthy of praise, what is most important is not "precise" but "useful" in engineering.

■

Remark 1.4. (Aristotles and Plato). As mentioned before, theoretical physics must be always checked by experimental tests. That is, it is based on realism (i.e., the Aristotles spirit). On the other hand, recall that the experimental test for MT is nonsense. Therefore, we can not deny MT by any experimental tests.[11] Thus, we may agree to the opinion that

"MT is self-righteous".

In this sense, we cay say that MT is based on idealism (i.e., the Plato spirit).[12] However, it does not imply "unfair". That is because, if some want to deny MT, it suffices to propose another fundamental theory *better* than MT. Here,

(I_{10}) the question: "*Which is better*?" is decided by majority (or popularity).

Here it should be noted that to win popularity is as difficult as to find the truth. Also, as mentioned in (I_9), we can expect that every theory in (I_7) and (I_8) is "almost experimentally true". That is because, if it is not "almost experimentally true", it can never win popularity.

■

[11] Thus, I assume that MT itself is a kind of *metaphysics* (and not science in the sense of Popper [70], "falsifiability").

[12] If I were familiar with the history of philosophy, I could stress the correspondences: "theoretical physics (realism) ↔ Aristotles" and "theoretical informatics (idealism)↔Plato".

Summing up, we assert the following table:

Table (1.8b)

	Theoretical Physics	Theoretical Informatics
(6). important criterion (*cf.* Remark 1.1 (e), Remark 1.4)	experimentally true or false objective	useful or not, likes or dislikes popularity, economical, subjective
(7). theoretically true or false (*cf.* Problem 1.2 (i),(ii))	meaningful in 'TOE'	meaningful in MT
(8). what to be applied to (*cf.* Remark 1.3)	physical phenomena	all phenomena (particularly, appearing in (I_7) and (I_8))
(9). fundamental spirit (*cf.* Remark 1.4)	realism (due to Aristotles) Theory is dominated by experiment.	idealism (due to Plato) Theory is free from experiment.

1.3 Measurement theory in engineering

As mentioned in (I_7) in §1.2, MT plays an important role in engineering. The theoretical physics (= 'TOE') itself may be worthy even if it has no applications. However, MT is not so. Thus, in this section, we consider the relation between engineering and MT. Here, engineering is usually considered to be composed of physics-related engineering (e.g., laser engineering, etc.), chemistry-related engineering (e.g., chemistry engineering, etc.), informatics-related engineering (e.g., financial engineering, etc.), etc.

The area of physics-related engineering is clear. That is because the physics-related engineering is generally believed to be supported by "physics" as the theoretical backbone. We studied physics as one of the important subjects in high-school, and therefore, we believe that theoretical physics is only one discipline, i.e., classical mechanics, relativity theory, electromagnetic theory and so on that should be unified. That is, physics-related engineering has the authorized root (= physics). Also, note that the circumstance of chemistry-related engineering is similar to that of physics-related engineering.

On the other hand, the area of informatics-related engineering may be vague. This is due to the fact that we do not know the most fundamental root in informatics-related engineering. Note that there is a possibility that informatics-related engineering has two (or more than two) fundamental roots. If it is so, we must consider "informatics-related engineering (I)" and "informatics-related engineering (II)" (*cf.* Remark 1.1 (b)). Therefore, we must answer the following question:

(I_{11}) What subject is the most fundamental in informatics-related engineering? $\Big($Or, is theoretical informatics the only one?$\Big)$

Of course, our answer is

(I_{12}) MT is the most fundamental theoretical backbone in informatics-related engineering.

MT (or, theoretical informatics) is not studied in high-school. However, statistics and differential equations (which are closely related to MT (i.e., Axioms 1 and 2 in (1.4a))) are studied as mathematics in high-school. In this sense, theoretical informatics is not underestimated in high-school education.

Thus we have the following table.

Table (1.9)

	fundamental subject	area (applications)
physics-related engineering	physics (mathematical) experimental test is possible	semiconductor engineering, laser engineering, etc.
chemistry-related engineering	chemistry (non-mathematical) experimental test is possible	chemical engineering
informatics-related engineering	MT (mathematical) experimental test is meaningless	Cf. (I_7) and (I_8)

Thus we conclude that

(I_{13}) The area of informatics-related engineering is roughly[13]determined by MT (= "theoretical informatics"), just like the area of physics-related engineering is roughly determined by physics. Also, recall (I_5).

That is, we say:

$$
\begin{cases}
\boxed{\text{physical phenomena}} \overset{\text{M.R.}}{\underset{\text{ET}}{\Longrightarrow}} \boxed{\boxed{\text{theoretical physics (='TOE')}}} \overset{\text{Appl}}{\underset{\text{AET}}{\Longrightarrow}} \boxed{\text{physics-related engineering}} \\
\boxed{\text{mechanical world view}} \overset{\text{M.R.}}{\underset{\text{PP}}{\Longrightarrow}} \boxed{\boxed{\text{theoretical informatics (= MT)}}} \overset{\text{Appl}}{\underset{\text{AET}}{\Longrightarrow}} \boxed{\text{informatics-related engineering}}
\end{cases}
\tag{1.10}
$$

where "M.R." = "mathematical representation", "ET" = "experimental test" (*cf.* Problem 1.2 (i)), "Appl" = "Applications" (*cf.* Table (1.9)), "PP" = "popularity" (*cf.* (I_{10})), "AET" = "almost experimentally true" (*cf.* (I_9) in §1.3).

[13]For example, mechanical engineering is closely related to physics. However, control theory (in (C_3) of Table (1.7)) plays an important role in robot engineering (which is a kind of mechanical engineering). Also, electrical circuit engineering may be close to electromagnetic theory as well as dynamical system theory. Thus, such a classification of engineering (presented in Table (1.9)) is somewhat forcible. That is because "Use everything available" is the engineer's spirit. Thus we must say that physics (as well as measurement theory) is more or less influential to every field in (I_7) and (I_8). However we can, at least, assert that physics, chemistry and MT are the most fundamental subjects in the faculty of engineering.

Here again note that

(I_{14}) Theoretical physics has to be precise. On the other hand, engineering has to be useful rather than precise. Since ambiguous statements can not be tested "exactly", we use the term : "AET (= almost experimentally true)" in the above (1.10).

Thus we see

(I_{15}) There is a possibility that a phenomenon has two (or, more than two) explanations in MT. And moreover, in this case, we may not choose one from the two by experimental tests but a sense of beauty (≈ like or dislike).

1.4 The spirit of "the mechanical world view"

We think that "measurement", "its philosophy" and "its applications (≈ informatics-related engineering)" should be regarded as "the Trinity". And we assert the following declaration, which was essentially proposed in [Ishikawa, 2002, [48]].

Declaration (1.11)

We assert the following (i) ∼(iii), which should be understood as the different representations of the same thing:

(i) *MT is the most fundamental theory of theoretical informatics, which is regarded as the theoretical backbone of informatics-related engineering.*

(ii) *MT is the ultimately generalized form of the dynamical system theory (1.2). Thus, MT is regarded as the mathematical representation of the epistemology called "the mechanical world view". And thus, MT is sometimes called the general dynamical system theory* (or in short, GDST).

(iii) *MT is entitled to check all theories in theoretical informatics. In other words, we can, by using MT, introduce the criterion: "(theoretically) true or false" into theoretical informatics.*

Here, note that:

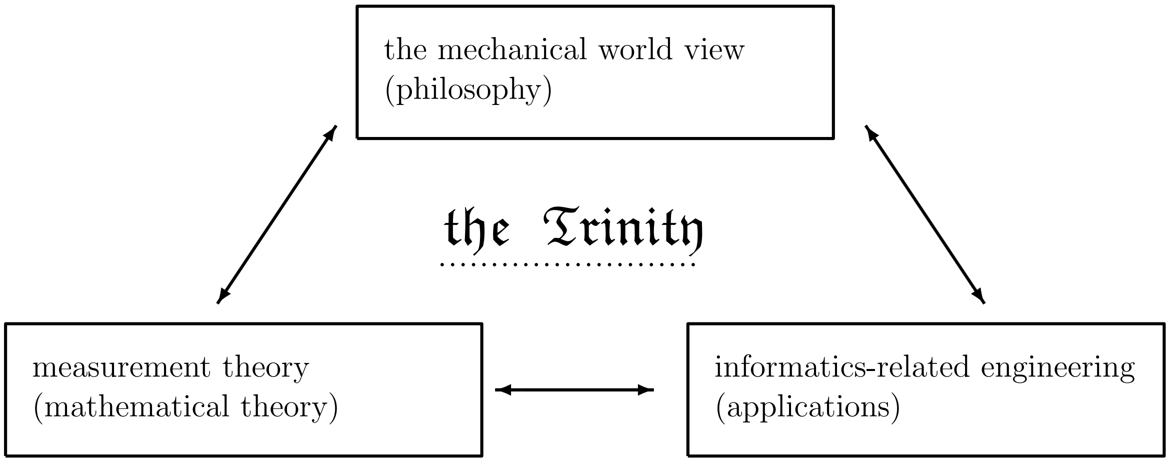
- in this book, "the mechanical world view" means "the quantum mechanical world view" and not "the Newtonian mechanical world view".

We might say too much in this chapter. It may suffice to say

The spirit of "the mechanical world view" (1.12)

- Mind Declaration (1.11) and Tables (1.7) and (1.8). And further, at any rate (= setting aside the reason), study every (physical or non-physical) problem in the framework of MT.[14]

Summing up, we have "the Trinity" as follows:



Here, again note that the philosophy of "theoretical informatics" is completely different from that of "theoretical physics", Although it is a matter of course that it is impossible to understand the philosophy of measurement theory without the complete knowledge of measurements (i.e., the contents of Chapters 2 $\sim$ 12), the philosophy of measurement theory is also indispensable for the understanding of measurement theory.

Remark 1.5 (Another important problem) The problem:

(I_{16}) "Propose [The third mathematical scientific theory] in (C_4) of Table (1.7)"

[14]As mentioned in Remark 1.4, we do not necessarily need a perfect reason in theoretical informatics. In this sense, the term: "extensive interpretation" is one of the most important terms in theoretical informatics.

may be the most important. I think that the above problem (I_{16}) is so difficult. Thus I may prefer waiting the appearance of a genius to doing it ourselves.[15]

[15]As mentioned in this chapter, our purpose may be, briefly speaking, to study all fields which can be understood in terms of "measurement (i.e., Axioms 1 and 2)". In this sense, Frieden's challenge [24] is also interesting. His purpose seems to study all fields (of physics) which can be understood in terms of "Fisher information". Although we do not completely understand his theory, we expect that his theory may be one of the candidates of The third mathematical scientific theory. We never hope that MT is the only one mathematical theory that belongs to the category of "idealism".

Chapter 2

Measurements (Axiom 1)

Measurement theory (MT) is classified two subjects, i.e., "(pure) measurement theory (PMT)" and "statistical measurement theory (SMT)". That is,

$$\text{MT (="measurement theory")} \begin{cases} \text{PMT (="(pure) measurement theory") in Chapters } 2 \sim 7 \\ \\ \text{SMT (="statistical measurement theory" in Chapters } 8 \sim) \end{cases} \tag{2.1}$$

PMT is essential, and it should be noted that *there is no SMT without PMT* (*cf.* Chapter 8). In Chapters 2 ∼ 7, we devote ourselves to PMT, which is formulated as follows:

$$\text{PMT} = \underset{[\text{Axiom 1 (2.37)}]}{\text{measurement}} + \underset{[\text{Axiom 2 (3.26)}]}{\text{the relation among systems}} \quad \text{in } C^*\text{-algebra} \, . \tag{2.2 (=(1.4))}$$

In this chapter we intend to explain "measurement (= Axiom 1)". (In Chapter 3 we will devote ourselves to Axiom 2 (i.e., "the relation among systems").)

2.1 Mathematical preparations

The theory of operator algebras (i.e, C^*-algebra and W^*-algebra) is a convenient mathematical tool to describe both classical and quantum mechanics (*cf.* [76]). Thus our theory is described in terms of C^*-algebras. Since our concern in this book is mainly concentrated on classical systems and not quantum systems, it may suffice to deal with only commutative C^*-algebras. In fact, most of our main results are related to classical systems. However, recall (1.4), that is:

$$\boxed{\text{PMT}} = \text{"Apply Axioms 1 and 2 to every phenomenon by an analogy of quantum mechanics"} \tag{2.3 (=(1.4))}$$

Thus we think that the essence of measurements can not be appreciated deeply without the knowledge of quantum measurements. In fact, the concept of measurements was first discovered and formulated by M. Born [13][1] as the most fundamental concept in quantum mechanics. Thus, we begin with general C^*-algebras, in which both classical and quantum systems are formulated.[2]

Let $\mathcal{A}$ be a linear associative algebra over the complex field $\mathbf{C}$. The algebra $\mathcal{A}$ is called a *Banach algebra* if it is associated to each element T a real number $\|T\|$, called the *norm* of T, with the properties:

$$
\begin{aligned}
&\text{(i) } \|T\| \geq 0, \quad \text{(ii) } \|T\| = 0 \text{ if and only if } T = 0, \text{(i.e., the 0-element in } \mathcal{A}),\\
&\text{(iii) } \|T+S\| \leq \|T\| + \|S\|, \quad \text{(iv) } \|\lambda T\| = |\lambda| \cdot \|T\|, \ \lambda \in \mathbf{C},\\
&\text{(v) } \|TS\| \leq \|T\| \cdot \|S\|, \quad \text{(vi) } \ \mathcal{A} \text{ is complete with respect to the norm } \|\cdot\|.
\end{aligned}
$$

A mapping $T \mapsto T^*$ of $\mathcal{A}$ into itself is called an *involution* (and T^* is called the *adjoint element* of T) if it satisfies the following conditions:

$$
\text{(i) } (T^*)^* = T, \qquad \text{(ii) } (T+S)^* = T^* + S^*, \qquad \text{(iii) } (TS)^* = S^*T^*,
$$

$$
\text{(iv) } (\lambda T)^* = \bar{\lambda} T^*, \ \lambda \in \mathbf{C}.
$$

A Banach algebra with an involution * is called a *Banach *-algebra.*

Definition 2.1. [C^*-algebra, identity, commutative C^*-algebra]. *A Banach *-algebra $\mathcal{A}$ (with the norm $\|\cdot\|_{\mathcal{A}}$) is called a C^*-algebra if it satisfies the C^*-condition, i.e., $\|T^*T\| = \|T\|^2$ for any $T \in \mathcal{A}$. A C^*-algebra $\mathcal{A}$ does not always have the identity element $I_{\mathcal{A}}$ (i.e., $I_{\mathcal{A}}T = TI_{\mathcal{A}} = T$ for all $T \in \mathcal{A}$), though in this book we usually suppose that a C^*-algebra $\mathcal{A}$ has the identity element $I_{\mathcal{A}}$. A C^*-algebra $\mathcal{A}$ is called unital, if it has the identity element $I_{\mathcal{A}}$. Also, a C^*-algebra $\mathcal{A}$ is called commutative, if it holds that $T_1T_2 = T_2T_1$ $(\forall T_1, T_2 \in \mathcal{A})$.*

■

An element F in $\mathcal{A}$ is called *self-adjoint* if it holds that $F = F^*$. A self-adjoint element F in $\mathcal{A}$ is called *positive* (and denoted by $F \geq 0$) if there exists an element F_0 in $\mathcal{A}$ such

[1] He proposed his theory in 1926, and he won the Nobel prize of physics in 1954.

[2] I am afraid that the mathematical preparation (in this section) discourages readers to want to read this book. Thus, it may be recommended to skip to Example 2.16 firstly. In order to read this book, it suffices to understand Example 2.16.

that $F = F_0^* F_0$ where F_0^* is the adjoint element of F_0. Also, a positive element F is called a *projection* if $F = F^2$ holds. Let $\mathcal{A}^*$ be the dual Banach space of $\mathcal{A}$. That is,

$$\mathcal{A}^* = \{\rho \mid \rho : \mathcal{A} \to \mathbf{C} \text{ is continuous linear } \}$$

with the norm $\|\rho\|_{\mathcal{A}^*}$ ($\equiv \sup\{|\rho(F)| \ : \ \|F\|_{\mathcal{A}} \le 1\}$). $\Big($The linear functional $\rho(F)$ is sometimes denoted by ${}_{\mathcal{A}^*}\Big\langle \rho, F \Big\rangle_{\mathcal{A}}$.$\Big)$ Define the *mixed state space* $\mathfrak{S}^m(\mathcal{A}^*)$ such that:

$$\mathfrak{S}^m(\mathcal{A}^*) = \{\rho \in \mathcal{A}^* \mid \|\rho\|_{\mathcal{A}^*} = 1 \text{ and } \rho(F) \ge 0 \text{ for all } F \ge 0\}. \tag{2.4}$$

A mixed state ρ ($\in \mathfrak{S}^m(\mathcal{A}^*)$) is called a *pure state* if it satisfies that "$\rho = \theta\rho_1 + (1-\theta)\rho_2$ for some $\rho_1, \rho_2 \in \mathfrak{S}^m(\mathcal{A}^*)$ and $0 < \theta < 1$" implies "$\rho = \rho_1 = \rho_2$". Define

$$\mathfrak{S}^p(\mathcal{A}^*) \equiv \{\rho^p \in \mathfrak{S}^m(\mathcal{A}^*) \mid \rho^p \text{ is a pure state}\}, \tag{2.5}$$

which is called a *state space* (or *pure state space, phase space*). Note that $\mathfrak{S}^m(\mathcal{A}^*)$ is convex and compact in the weak* topology $\sigma(\mathcal{A}^*; \mathcal{A})$. Also, $\mathfrak{S}^p(\mathcal{A}^*)$ is characterized as the set of the extreme points of $\mathfrak{S}^m(\mathcal{A}^*)$. (Cf. [92, 76]). Since $\mathfrak{S}^p(\mathcal{A}^*)$ is the closed set of $\mathfrak{S}^m(\mathcal{A}^*)$, the $\mathfrak{S}^p(\mathcal{A}^*)$ is also compact in the weak* topology.

The following Examples 2.2 and 2.3 will promote the understanding of Definition 2.1.

Example 2.2. [Commutative C^*-algebras; $C(\Omega)$, or generally, $C_0(\Omega)$]. When $\mathcal{A}$ is a commutative C^*-algebra, that is, $T_1 \cdot T_2 = T_2 \cdot T_1$ holds for all T_1, $T_2 \in \mathcal{A}$, by Gelfand theorem (*cf.* [74, 76]), we can put $\mathcal{A} = C(\Omega)$, the algebra composed of all continuous complex-valued functions on a compact Hausdorff space Ω. $\Big($If the commutative C^*-algebra $\mathcal{A}$ is not necessarily assumed to be unital, we can put $\mathcal{A} = C_0(\Omega)$, the algebra composed of all continuous complex-valued functions *vanishing at infinity* on a *locally* compact Hausdorff space Ω.$\Big)$ The norm $\|f\|_{C(\Omega)}$ is, of course, defined by $\|f\|_{C(\Omega)} = \max\{|f(\omega)| \ : \ \omega \in \Omega\}$ ($\forall f \in C(\Omega)$). Also, we can easily see that it satisfies the C^*-condition, i.e., $\|f \cdot f^*\|_{C(\Omega)} = \|f\|^2_{C(\Omega)}$ where $f^*(\omega)$ ($\equiv \overline{f(\omega)}$) is defined by the conjugate "$\mathfrak{Re}[f(\omega)] - \mathfrak{Im}[f(\omega)]i$" ($\forall \omega \in \Omega$) (where $\mathfrak{Re}$ is "real part", $\mathfrak{Im}$ is "imaginary part"). It is well known (i.e., Riese Theorem) that $C(\Omega)^* = \mathcal{M}(\Omega)$, i.e., the Banach space composed of all regular complex-valued measures on Ω. And therefore,

$$\mathfrak{S}^m(\mathcal{M}(\Omega)) = \{\rho \in \mathcal{M}(\Omega) \mid \rho \ge 0, \|\rho\|_{\mathcal{M}(\Omega)} = 1\}, \tag{2.6}$$

which is denoted by $\mathcal{M}^m_{+1}(\Omega)$. Also, it is clear that

$$\mathfrak{S}^p(\mathcal{M}(\Omega)) = \Big\{ \delta_\omega \in \mathcal{M}(\Omega) \mid \delta_\omega \text{ is a point measure at } \omega \in \Omega \Big\} \tag{2.7}$$

$\Big($i.e., ${}_{\mathcal{M}(\Omega)}\Big\langle \delta_\omega, f \Big\rangle_{C(\Omega)} = f(\omega)$ $(\forall f \in C(\Omega),\ \forall \omega \in \Omega)\Big)$, which is denoted by $\mathcal{M}^p_{+1}(\Omega)$, and called a *state space*. And therefore, we have the identification: $\Omega \approx \mathcal{M}^p_{+1}(\Omega)$ in the sense of

$$\Omega \ni \omega \longleftrightarrow \delta_\omega \in \mathcal{M}^p_{+1}(\Omega). \tag{2.8}$$

Thus the compact Hausdorff space Ω may be also called a *state space*.

■

Example 2.3. [Non-commutative C^*-algebras; $B(V)$ and $\mathcal{C}(V)$]. Let V be a (separable) Hilbert space with the inner product $\langle \cdot, \cdot \rangle_V$. Here we always assume that $\langle v_1, \alpha v_2 \rangle_V = \alpha \langle v_1, v_2 \rangle_V$ $(\forall v_1, v_2 \in V, \alpha \in \mathbf{C})$. (Cf. [4, 71].) Put

$$B(V) = \{T \ : \ T \text{ is a bounded linear operator from a Hilbert space } V \text{ into itself } \}.^3$$

Define $\|T\|_{B(V)} = \sup\{\|Tv\|_V \ : \ \|v\|_V = 1\}$, and $(T_1T_2)(v) = T_1(T_2 v)$ $(\forall v \in V)$. And T^* is defined by the adjoint operator of T. Note that it holds that $\|T^*T\|_{B(V)} = \|T\|^2$ $(\forall T \in B(V))$. Thus, we can easily see that the $B(V)$ is a non-commutative C^*-algebra. Also note that

$$\mathcal{C}(V) \equiv \{T \in B(V) : T \text{ is a compact operator } \} \tag{2.9}$$

is a C^*-subalgebra of $B(V)$. If the dimension of V is infinite, this C^*-algebra $\mathcal{C}(V)$ has no identity I. We see that

$$\mathcal{C}(V)^* = Tr(V)\Big(\equiv \{T \in B(V) \ : \ \|T\|_{tr} < \infty\}\Big). \tag{2.10}$$

Here $Tr(V)$ is the class of trace operators with the trace norm $\|\cdot\|_{tr}$ such that:

$$\|\rho\|_{tr} = \sum_{n=1}^{\infty} \langle e_n, \sqrt{\rho^* \rho}\, e_n \rangle_V$$

where $\{e_n\}_{n=1}^\infty$ is the complete orthonormal system in V. It is well known that the value $\|\rho\|_{tr}$ is independent of the choice of a complete orthonormal basis $\{e_\lambda | \lambda \in \Lambda\}$ in V. And we see

$$\mathfrak{S}^m(\mathcal{C}(V)^*) = Tr^m_{+1}(V) \equiv \{\rho \in Tr(V) : \rho \geq 0, \|\rho\|_{Tr(V)} = 1\}. \tag{2.11}$$

[3] "bounded linear operator" = "continuous linear operator" (*cf.* [92])

And further,

$$(Tr(V))^* = B(V).$$

Also, it is well known that

$$\text{“}\rho \in \mathfrak{S}^p(\mathcal{C}(V)^*)\text{”} \Leftrightarrow \text{“there exists } \psi \in V \ (\|\psi\|_V = 1) \text{ such that } \rho = |\psi\rangle\langle\psi|\text{”}, \tag{2.12}$$

where the Dirac notation "$|\psi_1\rangle\langle\psi_2|$" $\Big(\in B(V)\Big)$, $\psi_1, \psi_2 \in V$, is defined by

$$(|\psi_1\rangle\langle\psi_2|)\phi = \langle\psi_2, \phi\rangle_V \psi_1 \qquad \text{for all } \phi \in V.$$

The state space $\mathfrak{S}^p(\mathcal{C}(V)^*)$ is denoted by $Tr^p_{+1}(V)$, that is,

$$\mathfrak{S}^p(\mathcal{C}(V)^*) \equiv Tr^p_{+1}(V).$$

Also, it is well-known that "$\rho \in \mathfrak{S}^m(\mathcal{C}(V)^*)$" $\Leftrightarrow$ "there exists an orthonormal system $\{\psi_n\}_{n=1}^\infty$ in V and non-negative real numbers $\{\lambda_n\}_{n=1}^\infty$ (where $\sum_{n=1}^\infty \lambda_n = 1$) such that $\rho = \sum_{n=1}^\infty \lambda_n |\psi_n\rangle\langle\psi_n|$".

■

The following theorem is one of the most important theorems in the theory of operator algebras.

Theorem 2.4. [GNS-construction, Gelfand, Naimark, Siegel, *cf.* [50, 76]]. *Let $\mathcal{A}$ be a C^*-algebra. Then there exists a $B(V)$ such that:*

$$\mathcal{A} \subseteq B(V). \tag{2.13}$$

That is, $\mathcal{A}$ can be identified with the norm-closed C^*-subalgebra of a certain $B(V)$.

■

Example 2.5. [Commutative C^*-algebra $\mathrm{Mat}^{\mathrm{D}}(n; \mathbf{C})$ as the subalgebra of $B(\mathbf{C}^n)$]. Let $\mathbf{C}^n$ be an n-dimensional Hilbert space with the inner product $\langle\cdot,\cdot\rangle_{\mathbf{C}^n}$ (that is, $\|z\|_{\mathbf{C}^n} = \sqrt{\sum_{k=1}^n |z_k|^2}$ $(\forall z = (z_1, z_2, ..., z_n) \in \mathbf{C}^n)$). Consider the non-commutative C^*-algebra

$$B(\mathbf{C}^n) \equiv \{T \ : \ T \text{ is a (bounded) linear operator from a Hilbert space } \mathbf{C}^n \text{ into itself }\},$$

which is clearly equal to

$$\mathrm{Mat}(n; \mathbf{C}) \equiv \{T \ : \ T \text{ is a complex } (n \times n)\text{-matrix }\}. \tag{2.14}$$

That is,

$$B(\mathbf{C}^n) = \mathrm{Mat}(n; \mathbf{C}).$$

Put

$$\mathrm{Mat}^{\mathrm{D}}(n; \mathbf{C}) = \{T \ : \ T \text{ is a complex } (n \times n)\text{-diagonal matrix } \}, \tag{2.15}$$

which is clearly a commutative C^*-subalgebra of $B(\mathbf{C}^n)$. Also, it is obvious that the $\mathrm{Mat}^{\mathrm{D}}(n; \mathbf{C})$ is equivalent to $C(\Omega)$, where Ω is the finite state space $(\{1, 2, ..., n\})$ with the discrete topology.[4] That is, we see the following identifications:

$$\mathrm{Mat}^{\mathrm{D}}(n; \mathbf{C}) \approx C(\{\omega_1, \omega_2, ..., \omega_n\}) \approx \mathbf{C}^n$$

where $\mathbf{C}^n$ is assumed to have the max-norm $\|z\|^{\max}_{\mathbf{C}^n} \Big(= \max_{k=1,2,...,n} |z_k| \ (\forall z = (z_1, z_2, ..., z_n) \in \mathbf{C}^n) \Big)$. Also, the multiplication $(z^1_1, z^1_2, ..., z^1_n) \cdot (z^2_1, z^2_2, ..., z^2_n)$ is defined by $(z^1_1 z^2_1, z^1_2 z^2_2, ..., z^1_n z^2_n)$.

■

Remark 2.6. [(i): The identity]. Let $\mathcal{A}_0$ be a non-unital C^*-algebra. Theorem 2.4 (The GNS-construction) says that there exists a $B(V)$ such that $\mathcal{A}_0 \subseteq B(V)$. That is, $\mathcal{A}_0$ can be identified with the norm-closed subalgebra of $B(V)$. Thus we can define the C^*-algebra $\mathcal{A}_I$ such that it is the smallest algebra that includes $\{I\} \cup \mathcal{A}_0 \ (\subseteq B(V))$. Therefore, we can always add the identity I to $\mathcal{A}_0$, and construct the new unital C^*-algebra $\mathcal{A}_I$. This argument implies that the "unital condition" is not so strict. Thus, throughout this book, we usually deal with a unital C^*-algebra, though the $C_0(\Omega)$ is sometimes used.
[(ii): Minimal tensor C^*-algebras]. Here consider the minimal tensor C^*-algebra as follows: Let $\widehat{\mathcal{A}} \ \big(= \bigotimes_{k=1}^n \mathcal{A}_k \ \big)$ be the tensor product C^*-algebra of $\{\mathcal{A}_k \ : k = 1, 2, ..., n\}$. This can be easily constructed as follows: Since we can see, by Theorem 2.4 (GNS-construction), that

$$\mathcal{A}_k \subseteq B(V_k) \qquad (k = 1, 2, ..., n), \tag{2.16}$$

we can define $\bigotimes_{k=1}^n \mathcal{A}_k$ such that the smallest norm-closed sub-algebra (of $B(\bigotimes_{k=1}^n V_k)$) that contains

$$\Big\{ \bigotimes_{k=1}^n F_k \Big(\in B(\bigotimes_{k=1}^n V_k) \Big) \ \Big| \ F_k \in \mathcal{A}_k, k = 1, 2, ..., n \Big\} \tag{2.17}$$

[4]Throughout this book, we assume that a finite state space Ω $(\equiv \{\omega_1, \omega_2, ..., \omega_n\})$ has the discrete metric d_D (i.e., $d_D(\omega_1, \omega_2) = 1 \ (\omega_1 \neq \omega_2), = 0 \ (\omega_1 = \omega_2)$).

where $\bigotimes_{k=1}^{n} V_k$ is the tensor Hilbert space of $\{V_k \mid k = 1, 2, ..., n\}$. Though the general theory of tensor product C^*-algebras $\bigotimes_{k=1}^{n} \mathcal{A}_k$ is not easy, we only use the following properties (i)∼(iii) of the tensor C^*-algebras:

(i) $T_1 \otimes T_2 \otimes \cdots \otimes T_n \in \widehat{\mathcal{A}}$ for any $T_k \in \mathcal{A}_k$, $k = 1, 2, ..., n$,

(ii) $\rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_n \in \mathfrak{S}^p(\widehat{\mathcal{A}}^*)$ for any $\rho_k \in \mathfrak{S}^p(\mathcal{A}_k^*)$, $k = 1, 2, ..., n$,

(iii) $(\rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_n)\ (T_1 \otimes T_2 \otimes \cdots \otimes T_n) = \prod_{k=1}^{n} \rho_k(T_k)$ for any $\rho_k \in \mathcal{A}_k^*$ and any $T_k \in \mathcal{A}_k$, $k = 1, 2, ..., n$.

If we focus on only commutative cases, it is sufficient to know the fact that

$$\bigotimes_{k=1}^{n} C(\Omega_k) = C(\mathop{\times}_{k=1}^{n} \Omega_k) \qquad \text{and} \qquad \bigotimes_{k=1}^{n} \mathcal{M}(\Omega_k) = \mathcal{M}(\mathop{\times}_{k=1}^{n} \Omega_k), \tag{2.18}$$

where $\times_{k=1}^{n} \Omega_k$ is the product topological space of Ω_1,...,Ω_n. Therefore, for example, the above (iii) implies the elementary property of product measure (Fubini's theorem), i.e.,

$$\int_{\Omega_1 \times \Omega_2} f_1(\omega_1) \cdot f_2(\omega_2)(\rho_1 \otimes \rho_2)(d\omega_1 d\omega_2) = \int_{\Omega_1} f_1(\omega_1)\rho_1(d\omega_1) \cdot \int_{\Omega_2} f_2(\omega_2)\rho_2(d\omega_2)$$
$$(\forall f_1 \in C(\Omega_1), \quad \forall f_2 \in C(\Omega_2)). \tag{2.19}$$

For the deep studies of "tensor C^*-algebra", see [50].

■

2.2 Observables

Let X be a set. Let 2^X (or, $\mathcal{P}(X)$) be the power set of X. i.e., $2^X = \{\Xi \mid \Xi \subseteq X\}$. A set $\mathcal{F}(\subseteq 2^X)$ is called a *field* if the $\mathcal{F}$ is closed under the intersection (i.e., $\cap$) and the compliment (i.e., $[\,\cdot\,]^c$), that is, if "Ξ_1, $\Xi_2 \in \mathcal{F}$" implies "$\Xi_1 \cap \Xi_2 \in \mathcal{F}$" and "$\Xi_1^c \in \mathcal{F}$", where $\Xi_1^c = X \setminus \Xi_1 = \{x \mid x \in X \wedge x \notin \Xi_1\}$. Note that $\Xi_1 \cup \Xi_2 = (\Xi_1^c \cap \Xi_2^c)^c$, $\Xi_1 \setminus \Xi_2 = \Xi_1 \cap \Xi_2^c$ and $\Xi_1 \triangle \Xi_2 = (\Xi_1 \cup \Xi_2) \setminus (\Xi_1 \cap \Xi_2)$. Thus the field $\mathcal{F}$ is also closed under the operations $\cup$, $\setminus$ and $\triangle$.

Also, a set $\mathcal{R}(\subseteq 2^X)$ is called a *ring* if the $\mathcal{R}$ is closed under the intersection (i.e., $\cap$) and the symmetric difference (i.e., $\triangle$), that is, if "Ξ_1, $\Xi_2 \in \mathcal{R}$" implies "$\Xi_1 \cap \Xi_2 \in \mathcal{R}$" and

"$\Xi_1 \triangle \Xi_2 \in \mathcal{R}$". Note that $\Xi_1 \cup \Xi_2 = (\Xi_1 \triangle \Xi_2) \triangle (\Xi_1 \cap \Xi_2)$, $\Xi_1 \setminus \Xi_2 = \Xi_1 \triangle (\Xi_1 \cap \Xi_2)$. Thus the ring $\mathcal{R}$ is also closed under the operations $\cup$ and $\setminus$ (*cf.* [29]).

Motivated by the Davies' idea (in quantum mechanics, *cf.* [17]), we propose the following definition.

Definition 2.7. [C^*-observables in a unital $\mathcal{A}$]. *A C^*-observable (or in short, observable, fuzzy observable) $\mathbf{O} \equiv (X, \mathcal{F}, F)$ in a unital C^*-algebra $\mathcal{A}$ is defined such that it satisfies that*

(i) [field]. *X is a set (called a "measured value set" or "label set"), and $\mathcal{F}$ is the subfield of the power set $\mathcal{P}(X)$ ($\equiv \{\Xi : \Xi \subseteq X\}$),*

(ii) *for every $\Xi \in \mathcal{F}$, $F(\Xi)$ is a positive element in $\mathcal{A}$ such that $F(\emptyset) = 0$ and $F(X) = I_{\mathcal{A}}$ (where 0 is the 0-element in $\mathcal{A}$),*

(iii) *for any countable decomposition $\{\Xi_1, \Xi_2, ..., \Xi_n, ...\}$ of Ξ, $\Big($i.e., $\Xi, \Xi_n \in \mathcal{F}, \cup_{n=1}^{\infty} \Xi_n = \Xi$, $\Xi_n \cap \Xi_m = \emptyset (\text{if } n \neq m)\Big)$, it holds that $\rho\Big(F(\Xi)\Big) = \lim_{N \to \infty} \rho\Big(\sum_{n=1}^{N} F(\Xi_n)\Big)$ $(\forall \rho \in \mathfrak{S}^m(\mathcal{A}^*))$.*

Also, if $F(\Xi)$ is a projection for every Ξ ($\in \mathcal{F}$), a C^*-observable $(X, \mathcal{F}, F)$ is called a crisp C^*-observable (or, *a crisp observable, an idea*).

■

Remark 2.8. [(1): The case that X is finite]. In chapters 2∼8, we will usually deal with the case that X is finite. When we want to stress that X is finite, the $(X, \mathcal{F}, F)$ is often denoted by $(X, 2^X, F)$ or $(X, \mathcal{P}(X), F)$. Thus, in this case, the (iii) in Definition 2.7 means

$$F(\Xi_1 \cup \Xi_2) = F(\Xi_1) + F(\Xi_2) \quad (\forall \Xi_1, \forall \Xi_2 (\in 2^X) \text{ such that } \Xi_1 \cap \Xi_2 = \emptyset)).$$

[(2): C^*-observables in general C^*-algebras]. Although we are usually concerned with unital C^*-algebras, we add the generalization of Definition 2.7 as follows: *Let $\mathcal{A}$ be a C^*-algebra, which does not necessarily have the identity I. A C^*-observable (or in short, observable, fuzzy observable)* $\mathbf{O} \equiv (X, \mathcal{R}, F)$ in a C^*-algebra $\mathcal{A}$ *is defined such that it satisfies that*

(i) *X is a set, and $\mathcal{R}$ is the subring of the power set $\mathcal{P}(X)$ ($\equiv \{\Xi : \Xi \subseteq X\}$), that is, "$\Xi_1, \Xi_2 \in \mathcal{R}$" implies "$\Xi_1 \cap \Xi_2 \in \mathcal{R}$" and "$\Xi_1 \triangle \Xi_2 \in \mathcal{R}$",*

(ii) *for every* $\Xi \in \mathcal{R}$, $F(\Xi)$ *is a positive element in* $\mathcal{A}$ *such that* $F(\emptyset) = 0$ *(where* 0 *is the* 0*-element in* $\mathcal{A}$*),*

(iii) *for any countable decomposition* $\{\Xi_1, \Xi_2, ..., \Xi_n, ...\}$ *of* Ξ, $(\Xi, \Xi_n \in \mathcal{R})$, *it holds that* $\rho\Big(F(\Xi)\Big) = \lim_{N\to\infty} \rho\Big(\sum_{n=1}^N F(\Xi_n)\Big) \quad (\forall \rho \in \mathfrak{S}^m(\mathcal{A}^*))$,

(iv) *there exists a sequence* $\{\Xi_n^0\}_{n=1}^\infty$ *in* $\mathcal{R}$ *such that* $\Xi_1^0 \subseteq \Xi_2^0 \subseteq \cdots$ *and* $X = \cup_{n=1}^\infty \Xi_n^0$ *and* $\lim_{n\to\infty} \rho\Big(F(\Xi_n^0)\Big) = 1 \quad (\forall \rho \in \mathfrak{S}^m(\mathcal{A}^*))$.

Also, if $F(\Xi)$ *is a projection for every* Ξ $(\in \mathcal{R})$, *a* C^**-observable* $(X, \mathcal{R}, F)$ *is called a crisp* C^**-observable.*

■

Definition 2.9. [Image observable]. *Let* $\mathbf{O} \equiv (X, \mathcal{F}, F)$ *be an observable in a* C^**-algebra* $\mathcal{A}$. *Let* $\mathcal{G}$ *be a subfield of* 2^Y. *Let* $h : X \to Y$ *be a measurable map, i.e.,* $h^{-1}(\Gamma) \in \mathcal{F}$ $(\forall \Gamma \in \mathcal{G})$. *Then, we can define the observable* $\mathbf{O}_{[h]}$ $(\equiv (Y, \mathcal{G}, F \circ h^{-1}))$ *in* $\mathcal{A}$ *such that:*

$$(F \circ h^{-1})(\Gamma) = F(h^{-1}(\Gamma)) \qquad (\Gamma \in \mathcal{G}). \tag{2.20}$$

The $\mathbf{O}_{[h]} \equiv (X, \mathcal{F}, G \circ h^{-1})$ *is called the image observable of* $\mathbf{O} \equiv (Y, \mathcal{G}, G)$ *(in a* C^**-algebra* $\mathcal{A}$*) concerning the map* $h : X \to Y$. *The image observable* $\mathbf{O}_{[h]}$ *is also denoted by* $h(\mathbf{O})$.

■

Definition 2.10. [Quasi-product observable]. *For each* $k = 1, 2, ..., n$, *consider an observable* $\mathbf{O}_k \equiv (X_k, \mathcal{F}_k, F_k)$ *in a* C^**-algebra* $\mathcal{A}$. *Define the field* $\bigotimes_{k=1}^n \mathcal{F}_k$ $(\subseteq 2^{\times_{k=1}^n X_k})$ *such as the smallest field (on* $\times_{k=1}^n X_k$*) that contains* $\times_{k=1}^n \Xi_k$, $\Xi_k \in \mathcal{F}_k$. *The product field* $\bigotimes_{k=1}^n \mathcal{F}_k$ *is usually denoted by* $\times_{k=1}^n \mathcal{F}_k$. $\Big($*Throughout this book, the notation* $\times_{k=1}^n \mathcal{F}_k$ *does not mean the set* $\{\times_{k=1}^n \Xi_k : \Xi_k \in \mathcal{F}_k\}$.$\Big)$ *An observable* $\widehat{\mathbf{O}} \equiv (\times_{k=1}^n X_k, \times_{k=1}^n \mathcal{F}_k, \widehat{F})$ *in* $\mathcal{A}$ *is called the quasi-product observable of* $\{\mathbf{O}_k : k = 1, 2, ..., n\}$ $\Big($*or, quasi-product observable with marginal observables* $\{\mathbf{O}_k : k = 1, 2, ..., n\}\Big)$ *if it holds that*

$$\widehat{F}(X_1 \times \cdots \times X_{k-1} \times \Xi_k \times X_{k+1} \times \cdots \times X_n) = F_k(\Xi_k) \quad (\forall \Xi_k \in \mathcal{F}_k, \forall k = 1, ..., n). \tag{2.21}$$

The quasi-product observable $\widehat{\mathbf{O}}$ *(of* $\{\mathbf{O}_k\}_{k=1}^n$*) is denoted by*

$$\mathop{\times}^{\text{qp}}_{k=1,2,...,n} \mathbf{O}_k, \text{ or, } \Big(\mathop{\times}_{k=1}^{n} X_k, \mathop{\times}_{k=1}^{n} \mathcal{F}_k, \mathop{\times}^{\text{qp}}_{k=1,2,...,n} F_k\Big), \text{ or } \Big(\mathop{\times}_{k=1}^{n} X_k, \bigotimes_{k=1}^{n} \mathcal{F}_k, \mathop{\times}^{\text{qp}}_{k=1,2,...,n} F_k\Big), \tag{2.22}$$

i.e., $\widehat{\mathbf{O}} = \overset{\text{qp}}{\times}_{k=1,2,...,n}\mathbf{O}_k$, $\widehat{F} = \overset{\text{qp}}{\times}_{k=1,2,...,n}F_k$. *Also,* $\overset{\text{qp}}{\times}_{k=1,2,...,n}F_k$ *is sometimes written by* $\overset{\widehat{\mathbf{O}}}{\times}_{k=1,2,...,n}F_k$.

■

Note that the existence and the uniqueness of the quasi-product observable of $\{\mathbf{O}_k : k = 1, 2, ..., n\}$ are not guaranteed in general. However, when $\mathbf{O}_k$, $k = 1, 2, ..., n$, commute, i.e.,

$$F_k(\Xi_k)F_{k'}(\Xi_{k'}) = F_{k'}(\Xi_{k'})F_k(\Xi_k) \quad \text{for all } \Xi_k \in \mathcal{F}_k,\ \Xi_{k'} \in \mathcal{F}_{k'} \text{ such that } k \neq k' \ , \tag{2.23}$$

we can construct the quasi-product observable $(\times_{k=1}^n X_k, \times_{k=1}^n \mathcal{F}_k, \widetilde{F})$ in $\mathcal{A}$ such that:

$$\widetilde{F}(\Xi_1 \times \Xi_2 \times \cdots \times \Xi_n) = F_1(\Xi_1)F_2(\Xi_2)\cdots F_n(\Xi_n). \tag{2.24}$$

This kind of quasi-product observable is called a *product observable* and denoted by

$$\mathop{\times}_{k=1}^{n} \mathbf{O}_k \quad \Big(= \mathop{\times}_{k=1,2,...,n} \mathbf{O}_k, \text{ or, } (\mathop{\times}_{k=1}^{n} X_k, \mathop{\times}_{k=1}^{n} \mathcal{F}_k, \mathop{\times}_{k=1}^{n} F_k), \text{ or, } (\mathop{\times}_{k=1}^{n} X_k, \bigotimes_{k=1}^{n} \mathcal{F}_k, \mathop{\times}_{k=1}^{n} F_k), \Big). \tag{2.25}$$

$\times_{k=1}^n$ is sometimes written by $\prod_{k=1}^n$, and thus, we write: $\times_{k=1}^n \mathbf{O}_k = \prod_{k=1}^n \mathbf{O}_k$, $\times_{k=1}^n X_k = \prod_{k=1}^n X_k$, etc. Also, note that the product observable $\times_{k=1}^n \mathbf{O}_k$ always exists for any $\mathbf{O}_k$ in a commutative C^*-algebra $C(\Omega)$.

Summing up the above arguments, we can state the following theorem.

Theorem 2.11. *For each* $k \in K \equiv \{1, 2, ..., |K|\}$, *consider an observable* $\mathbf{O}_k \equiv (X_k, \mathcal{F}_k, F_k)$ *in a* C^**-algebra* $\mathcal{A}$. *If the commutativity condition:*

$$F_{k_1}(\Xi_{k_1})F_{k_2}(\Xi_{k_2}) = F_{k_2}(\Xi_{k_2})F_{k_1}(\Xi_{k_1}) \quad (\forall \Xi_{k_1} \in \mathcal{F}_{k_1},\ \forall \Xi_{k_2} \in \mathcal{F}_{k_2},\ k_1 \neq k_2) \tag{2.26}$$

holds, then we can construct a product observable $\widehat{\mathbf{O}} \equiv (\times_{k \in K} X_k, \times_{k \in K} \mathcal{F}_k, \widetilde{F} \equiv \times_{k \in K} F_k)$ *such that:*

$$\widetilde{F}(\Xi_1 \times \Xi_2 \times \cdots \times \Xi_{|K|}) = F_1(\Xi_1)F_2(\Xi_2)\cdots F_{|K|}(\Xi_{|K|}). \tag{2.27}$$

Note that the uniqueness (of quasi-product observables) is not guaranteed even under the above commutativity condition. Also, note that the product observable $\times_{k=1}^n \mathbf{O}_k$ *always exists for any* $\mathbf{O}_k$ *in a commutative* C^**-algebra* $C(\Omega)$.

■

Theorem 2.12. *Let $\mathbf{O} \equiv (X, \mathcal{R}, F)$ be a C^*-observable in a general C^*-algebra $\mathcal{A}$ (i.e., it does not necessarily have the identity). Let $\mathcal{A}_1$ be a C^*-algebra with the identity I (generated by the $\mathcal{A}$ such as in Remark 2.6(i)). Then, there uniquely exists the observable $(X, \mathcal{F}, \widetilde{F})$ be a C^*-observable in $\mathcal{A}_1$ such that:*

$$\text{(i)} \qquad \mathcal{F} = \mathcal{R} \cup \{X \setminus \Gamma \mid \Gamma \in \mathcal{R}\}$$

$$\text{(ii)} \qquad \widetilde{F}(\Xi) = \begin{cases} F(\Xi) & (\Xi \in \mathcal{R}) \\ I - F(\Xi^c) & (\Xi^c = (X \setminus \Xi) \in \mathcal{R}). \end{cases}$$

Proof. It suffices to show that $\mathcal{F}$ is the field. Let $\Xi_1 \in \mathcal{R}$ and $\Xi_2 \in \{X \setminus \Gamma \mid \Gamma \in \mathcal{R}\}$. Thus $\Xi_2 = X \setminus \Gamma$ (for some $\Gamma \in \mathcal{R}$). Then, we see $\Xi_1 \cap \Xi_2 = \Xi_1 \cap (X \setminus \Gamma) = \Xi_1 \cap (\Xi_1 \setminus \Gamma) \in \mathcal{F}$. Also, $\Xi_1 \cup \Xi_2 = (\Xi_1^c \cap \Xi_2^c)^c = (\Xi_1^c \cap \Gamma)^c = (\Gamma \setminus \Xi_1)^c \in \mathcal{F}$. Also, it is clear that "$\Xi \in \mathcal{F}$" $\Longrightarrow$ "$\Xi^c \in \mathcal{F}$". Thus, we see that $\mathcal{F}$ is the field. □

The following theorem (and Theorem 9.8) will be often used throughout this book.

Theorem 2.13. [*cf.* [42]]. *Let $\mathcal{A}$ be a C^*-algebra. Let $\mathbf{O}_1 \equiv (X_1, \mathcal{F}_1, F_1)$ and $\mathbf{O}_2 \equiv (X_2, \mathcal{F}_2, F_2)$ be C^*-observables in $\mathcal{A}$ such that at least one of them is crisp. $\big($So, without loss of generality, we assume that $\mathbf{O}_2$ is crisp$\big)$. Then, the following statements are equivalent:*

(i) *There exists a quasi-product observable $\mathbf{O}_{12} \equiv (X_1 \times X_2, \mathcal{F}_1 \times \mathcal{F}_2, F_1 \overset{\text{qp}}{\times} F_2)$ with marginal observables $\mathbf{O}_1$ and $\mathbf{O}_2$.*

(ii) *$\mathbf{O}_1$ and $\mathbf{O}_2$ commute, that is, $F_1(\Xi_1)F_2(\Xi_2) = F_2(\Xi_2)F_1(\Xi_1)$ $(\forall \Xi_1 \in \mathcal{F}_1, \forall \Xi_2 \in \mathcal{F}_2)$.*

Furthermore, if the above statements (i) and (ii) hold, the uniqueness of the quasi-product observable $\mathbf{O}_{12}$ of $\mathbf{O}_1$ and $\mathbf{O}_2$ is guaranteed.

Proof. It suffices to prove it in the case that $\mathcal{A}$ has the identity. When $\mathbf{O}_1 \equiv (X_1, \mathcal{F}_1, F_1)$ and $\mathbf{O}_2 \equiv (X_2, \mathcal{F}_2, F_2)$ are both crisp observables, it is proved in [17]. By the same way, we can prove this theorem. It is clear that (ii) $\Longrightarrow$ (i) since we can construct a C^*-observable $(X_1 \times X_2, \mathcal{F}_1 \times \mathcal{F}_2, H)$ such that:

$$H(\Xi_1 \times \Xi_2) = F_1(\Xi_1)F_2(\Xi_2) \quad (\forall \Xi_1 \in \mathcal{F}_1, \forall \Xi_2 \in \mathcal{F}_2).$$

Thus, it suffices to prove that (i) $\Longrightarrow$ (ii). Assume that (i) holds. Let Ξ_1 and Ξ_2 be any element in $\mathcal{F}_1$ and $\mathcal{F}_2$ respectively. Put $\Xi_1^1 = \Xi_1$, $\Xi_1^2 = X_1 \setminus \Xi_1$, $\Xi_2^1 = \Xi_2$ and $\Xi_2^2 = X_2 \setminus \Xi_2$. Put $H = F_1 \overset{\text{qp}}{\times} F_2$. Note that:

$$0 \le H(\Xi_1^i \times \Xi_2^j) \le H(X_1 \times \Xi_2^j) \equiv F_2(\Xi_2^j) \ \ (= \text{"projection"}). \tag{2.28}$$

This implies that $H(\Xi_1^i \times \Xi_2^j)$ and $F_2(\Xi_2^j)$ commute, and so, $H(\Xi_1^i \times \Xi_2^j)$ and $I - F_2(\Xi_2^j)$ commute. Hence, $F_1(\Xi_1)$ $(= H(\Xi_1^1 \times \Xi_2^1) + H(\Xi_1^1 \times \Xi_2^2))$ and $F_2(\Xi_2)$ $(= F_2(\Xi_2^1))$ commute. Therefore, we get that (i) $\Longrightarrow$ (ii).

Next we prove the uniqueness of H under the assumption (i) (and so (ii)). Note that $0 \le H(\Xi_1^i \times \Xi_2^j) \le H(\Xi_1^i \times X_2) \equiv F_1(\Xi_1^i)$. This implies, by the commutativity condition (ii) and (2.28), that

$$0 \le H(\Xi_1^i \times \Xi_2^j) \le F_2(\Xi_2^j) F_1(\Xi_1^i) F_2(\Xi_2^j) = F_1(\Xi_1^i) F_2(\Xi_2^j). \tag{2.29}$$

Therefore we see that $I = \sum_{i,j=1,2} H(\Xi_1^i \times \Xi_2^j) \le \sum_{i,j=1,2} F_1(\Xi_1^i) F_2(\Xi_2^j) = I$. Then, we obtain that $H(\Xi_1 \times \Xi_2) = F_1(\Xi_1) F_2(\Xi_2)$, that is, H is unique. Therefore, we finish the proof. □

2.3 The meanings of observables and crisp observables

In the conventional classical [resp. quantum] mechanics, the term "observable" usually means a *real valued continuous function on a state space* Ω [resp. *a self-adjoint operator in* $B(V)$]. Thus, the "observable" (defined in Definition 2.7) should be a kind of generalization of the above conventional "observable". In what follows we will see it.

Now we shall consider the several aspects (and properties) of the observable $\mathbf{O} \equiv (X, \mathcal{F}, F)$ in a C^*-algebra $\mathcal{A}$. Examining Definition 2.7, we can easily see

(A_1) An observable $\mathbf{O}$ $\big(\equiv (X, \mathcal{F}, F) \big)$ in $\mathcal{A}$ can be regarded as the $\mathcal{A}$-valued probability space[5], i.e., the additive set-function:

$$\mathcal{F} \ni \Xi \mapsto F(\Xi) \in \mathcal{A}.$$

Also, we may find the similarity between an observable $\mathbf{O}$ and *the resolution of the identity* I in what follows. Assume, for simplicity, that X is countable (i.e., $X \equiv \{x_1, x_2, ...\}$). Then, it is clear that

[5]In this book, the term "probability space" is used as "a positive measure space whose total measure is equal to 1". That is, the term "probability space" is used as the pure mathematical concept, and thus, it is not always assured to be related to the concept of "probability".

(i) $F(\{x_k\}) \geq 0$ for all $k = 1, 2, ...$

(ii) $\sum_{k=1}^{\infty} F(\{x_k\}) = I_{\mathcal{A}}$ in the sense of weak topology of $\mathcal{A}$,

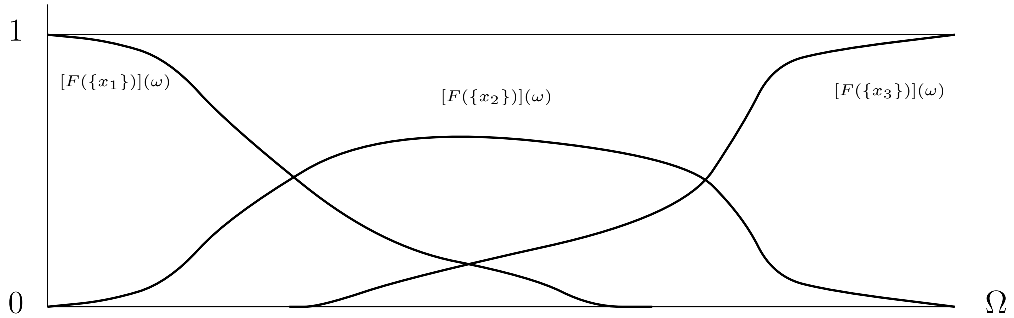
which imply that the $[F(\{x_k\}) \ : \ k = 1, 2, ..., n]$ can be regarded as *the resolution of the identity element* $I_{\mathcal{A}}$. Thus we say that

(A$_2$) An observable $\mathbf{O}$ $\big(\equiv (X, \mathcal{F}, F) \big)$ in $\mathcal{A}$ can be regarded as

$$\text{"the fuzzy decomposition"} \tag{2.30}$$

that is, *the resolution of the identity* $I_{\mathcal{A}}$, i.e., $[F(\{x_k\}) \ : \ k = 1, 2, ..., n]$.

"The figure of $\mathbf{O} \equiv (\{x_1, x_2, x_3\},\ 2^{\{x_1,x_2,x_3\}},\ F)$ in $C(\Omega)$"



Also, we note that

(A$_3$) An observable $\mathbf{O}$ $\big(\equiv (X, \mathcal{F}, F) \big)$ in $\mathcal{A}$ can be characterized as a kind of generalization of a self-adjoint element in $\mathcal{A}$.

This is shown as follows: For simplicity, assume that $\mathcal{A} = B(\mathbf{C}^N)$. And put

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \cdots, \quad e_N = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \tag{2.31}$$

Thus we see that

$$|e_1\rangle\langle e_1| = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}, \quad |e_2\rangle\langle e_2| = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}, \cdots, |e_N\rangle\langle e_N| = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

The spectral theorem says that a self-adjoint matrix $\widehat{F}$ ($\in B(\mathbf{C}^N)$) can be represented by

$$\begin{aligned}\widehat{F} &= U \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_N \end{bmatrix} U^* \\ &= U\Big(\lambda_1 |e_1\rangle\langle e_1| + \lambda_2 |e_2\rangle\langle e_2| + \dots + \lambda_N |e_N\rangle\langle e_N|\Big)U^* \\ &= \sum_{n=1}^N \lambda_n |Ue_n\rangle\langle Ue_n| \end{aligned} \tag{2.32}$$

where $\lambda_n \in \mathbf{R}$ $(\forall n = 1, 2, ..., N)$ and U is a unitary matrix in $B(\mathbf{C}^N)$. For any Ξ ($\in \mathcal{B}_{\mathbf{R}} =$ "Borel field")[6], put

$$F(\Xi) = \sum_{\lambda_n \in \Xi} |Ue_n\rangle\langle Ue_n|. \tag{2.33}$$

Here it should be noted that $F(\Xi)$ is a projection for all Ξ ($\in \mathcal{B}_{\mathbf{R}}$). This implies the the following identification:

$$\underset{\text{(self-adjoint operator)}}{\widehat{F}} \quad \longleftrightarrow \quad \underset{\text{(crisp observable)}}{(\mathbf{R}, \mathcal{B}_{\mathbf{R}}, F)} \quad \text{in } B(\mathbf{C}^N). \tag{2.34}$$

That is because $\widehat{F}$ is represented by (2.32), i.e.,

$$\widehat{F} = \int_{\mathbf{R}} \lambda F(d\lambda).$$

Next assume that $\mathcal{A} = C(\Omega)$, where Ω is, for simplicity, assumed to be the finite set $\{\omega_1, \omega_2, \omega_3, ..., \omega_N\}$ with the discrete topology. Consider a real valued continuous function $\widehat{F} : \Omega \to \mathbf{R}$. Define the observable $(\mathbf{R}, \mathcal{B}_{\mathbf{R}}, F)$ in $C(\Omega)$ such that:

$$[F(\Xi)](\omega) = \begin{cases} 1 & \text{if } \omega \in \widehat{F}^{-1}(\Xi) \\ \\ 0 & \text{if } \omega \notin \widehat{F}^{-1}(\Xi) \end{cases} \qquad (\forall \omega \in \Omega, \quad \forall \Xi \in \mathcal{B}_{\mathbf{R}}). \tag{2.35}$$

Note that

$$\widehat{F}(\omega) = \sum_{n=1}^N \widehat{F}(\omega_n)\Big([F(\{\omega_n\})](\omega)\Big) = \sum_{\lambda \in \mathbf{R}} \lambda [F(\{\lambda\})](\omega) \ \Big(= [\int_{\mathbf{R}} \lambda F(d\lambda)](\omega)\Big) \qquad (\forall \omega \in \Omega).$$

This implies the the following identification:

$$\underset{\text{(real valued function on } \Omega)}{\widehat{F}} \quad \longleftrightarrow \quad \underset{\text{(crisp observable)}}{(\mathbf{R}, \mathcal{B}_{\mathbf{R}}, F)} \quad \text{in } C(\Omega). \tag{2.36}$$

Therefore, we say, by (2.34) and (2.36), that

[6] "Borel field" = "the smallest σ-field that contains all open sets"

(A_4) "crisp observable $(\mathbf{R}, \mathcal{B}_{\mathbf{R}}, F)$" in $\mathcal{A}$ $\underset{\text{identification}}{\longleftrightarrow}$ "self-adjoint element" in $\mathcal{A}$.

$\Big($where $\mathcal{A} = B(\mathbf{C}^n)$ or $\mathcal{A} = C(\{\omega_1, \omega_2, ..., \omega_N\})\Big)$. Here, the "self-adjoint element" in $\mathcal{A}$ $\Big($ i.e., "crisp observable $(\mathbf{R}, \mathcal{B}_{\mathbf{R}}, F)$" in $\mathcal{A}\Big)$ is sometimes called a *"quantity* (or, *system theoretical quantity") in* $\mathcal{A}$.

Remark 2.14. [OR (= operation research) and game theory]. In OR [resp. game theory [85]], we are mainly concerned with the problem: "Study the maximal point [resp. the saddle point] of $\widehat{F}$!"

■

2.4 Measurement (Axiom 1)

Under the mathematical preparations in the previous sections, now we can describe the fundamental concepts of measurement theory (2.2) (=(1.4)).

With any *system* S, a C^*-algebra $\mathcal{A}$ can be associated in which measurement theory of that system can be formulated. A *state* of the system S is represented by a pure state ρ^p $(\in \mathfrak{S}^p(\mathcal{A}^*)$, i.e., a *state space*). Also, an *observable* is represented by a C^*-observable $\mathbf{O}$ $\equiv (X, \mathcal{F}, F)$ in the C^*-algebra $\mathcal{A}$.[7] The *measurement of an observable* $\mathbf{O}$ *for the system* S *with (or, in) the state* ρ^p *is represented by* $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[\rho^p]})$ *in the* C^**-algebra* $\mathcal{A}$*. Also, we can obtain a measured value* x $(\in X)$ *by the measurement* $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[\rho^p]})$.

The axiom presented below is analogous to (or, a kind of generalizations of) Born's probabilistic interpretation of quantum mechanics [13]. We of course assert that the axiom is a principle for all measurements, i.e., classical and quantum measurements. Cf. [41, 42].

AXIOM 1. [Measurement axiom]. *Consider a measurement* $\mathbf{M}_{\mathcal{A}}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[\rho^p]})$ *formulated in a* C^**-algebra* $\mathcal{A}$*. Assume that the measured value* x $(\in X)$ *is obtained by the measurement* $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[\rho^p]})$*. Then, the probability that the* x $(\in X)$ *belongs to a set* Ξ $(\in \mathcal{F})$ *is given by* $\rho^p(F(\Xi))$ $\big(\equiv {}_{\mathcal{A}^*}\big\langle \rho^p, F(\Xi)\big\rangle_{\mathcal{A}}\big)$. (2.37)

[7] I like to image the following correspondence (measurement theory and philosophy):

"state" ↔ "matter" "observable" ↔ "idea" (= "form")

We introduce the following classification in measurement theory:

$$\text{measurement theory} \begin{cases} \text{classical measurement theory (for classical systems)} \\ \\ \text{quantum measurement theory (for quantum systems)} \end{cases} \tag{2.38}$$

where a C^*-algebra $\mathcal{A}$ is commutative or non-commutative.

Recall the (1.3), that is, quantum mechanics (*cf.* [71]) is formulated by

$$\boxed{\text{“quantum mechanics”}} = \underset{\text{(“Born's quantum measurements”)}}{\text{measurement}} + \underset{\text{(“Schrödinger equation”)}}{\text{the rule of time evolution}} \tag{1.3}$$

Of course, Axiom 1 corresponds to “Born's quantum measurements”. Note that quantum measurement theory is well authorized as a principle of quantum mechanics (*cf.* [17, 34, 84]). Our interest in this book is mainly concentrated on classical systems. Therefore, in most cases, it suffices to assume that $\mathcal{A} = C(\Omega)$.

2.5 Remarks

In this section we add some remarks concerning Axiom 1.

[**(I): Probability**]. It should be noted that the term “probability” appears in Axiom 1. Following the common knowledge of quantum mechanics (*cf.* [71, 84]), we believe that any scientific statement including the term “probability” is meaningless without the concept of “measurement”. That is, we say that

($\sharp$) *“There is no probability without measurements”*.

Throughout this book, the above spirit ($\sharp$) is quite important.

[**(II): It is prohibited to take measurements twice**]. The quasi-product observable (or, the product observable) is used to represent *“the measurement of (more than one) observables”* as follows: For example, consider *“the measurement of* $\mathbf{O}_1$ *and* $\mathbf{O}_2$ *for the system with the state* ρ^p $(\in \mathfrak{S}^p(\mathcal{A}^*))$*.”* If the quasi-product observable $\mathbf{O}_1 \overset{\text{qp}}{\times} \mathbf{O}_2$ of $\mathbf{O}_1$ and $\mathbf{O}_2$ exists, the measurement is represented by $\mathbf{M}_{\mathcal{A}}(\mathbf{O}_1 \overset{\text{qp}}{\times} \mathbf{O}_2, S_{[\rho^p]})$ $\Big($and not “$\mathbf{M}_{\mathcal{A}}(\mathbf{O}_1, S_{[\rho^p]})$ $+ \mathbf{M}_{\mathcal{A}}(\mathbf{O}_2, S_{[\rho^p]})$”$\Big)$. If the quasi-product observable $\mathbf{O}_1 \overset{\text{qp}}{\times} \mathbf{O}_2$ does not exist, the measurement does not also exist. That is, the symbol “$\mathbf{M}_{\mathcal{A}}(\mathbf{O}_1, S_{[\rho^p]}) + \mathbf{M}_{\mathcal{A}}(\mathbf{O}_2, S_{[\rho^p]})$” is nonsense. Thus we can say that

(♯) *only one measurement is permitted to be conducted even in the classical measurement theory.*

which is the well-known fact in quantum mechanics. The measurement $\mathbf{M}_{\mathcal{A}}\big(\mathbf{O}_1 \overset{\text{qp}}{\times} \mathbf{O}_2, S_{[\rho^p]}\big)$ is sometimes called a *simultaneous measurement (or iterated measurement)* of two observables $\mathbf{O}_1$ and $\mathbf{O}_2$. That is, it is prohibited to take measurements twice in measurement theory. For example, the following statement:

- "Take two measurements $\mathbf{M}_{\mathcal{A}}\big(\mathbf{O}_1, S_{[\rho^p]}\big)$ and $\mathbf{M}_{\mathcal{A}}\big(\mathbf{O}_2, S_{[\rho^p]}\big)$."

is prohibited.

[(III): Sample space]. Let ρ^m be a mixed state, i.e., $\rho^m \in \mathfrak{S}^m(\mathcal{A}^*)$. Applying Hopf extension theorem (*cf.* [92]), we can get the measure space $(X, \overline{\mathcal{F}}, \overline{\rho^m(F(\,\cdot\,))}\,)$ such that $\overline{\rho^m(F(\Xi))} = \rho^m(F(\Xi))$ for all $\Xi \in \mathcal{F}$ where $\overline{\mathcal{F}}$ is the smallest σ-field that contains $\mathcal{F}$. For simplicity, the $\overline{\rho^m(F(\,\cdot\,))}$ is also denoted by $\rho^m(F(\,\cdot\,))$ or ${}_{\mathcal{A}^*}\Big\langle \rho^m, F(\,\cdot\,)\Big\rangle_{\mathcal{A}}$. Axiom 1 makes us call the measure space $(X, \overline{\mathcal{F}}, \overline{\rho^p(F(\,\cdot\,))}\,)$ $\Big($or in short, $(X, \mathcal{F}, \rho^p(F(\,\cdot\,))\,)\Big)$ a *sample space* concerning a measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[\rho^p]}\,)$.

[(IV): Conditional probability]. Let $\mathbf{O} \equiv (X, \mathcal{F}, F)$ and $\mathbf{O}' \equiv (Y, \mathcal{G}, G)$ be observables in $\mathcal{A}$. Let $\widehat{\mathbf{O}}$ be a quasi-product observable of $\mathbf{O}$ and $\mathbf{O}'$, that is, $\widehat{\mathbf{O}} \equiv \mathbf{O} \overset{\text{qp}}{\times} \mathbf{O}' = (X \times Y, \mathcal{F} \times \mathcal{G}, F \overset{\text{qp}}{\times} G)$. Assume that we know that the measured value (x, y) $(\in X \times Y)$ obtained by a measurement $\mathbf{M}_{\mathcal{A}}(\widehat{\mathbf{O}}, S_{[\rho^p]})$ belongs to $\Xi \times Y$ $(\in \mathcal{F} \times \mathcal{G})$. Then, it is clear that the unknown measured value y $(\in Y)$ is distributed under the conditional probability $P_\Xi(\cdot)$, where

$$P_\Xi(\Gamma) = \frac{{}_{\mathcal{A}^*}\langle \rho^p, F(\Xi) \overset{\text{qp}}{\times} G(\Gamma) \rangle_{\mathcal{A}}}{{}_{\mathcal{A}^*}\langle \rho^p, F(\Xi) \rangle_{\mathcal{A}}} \left(= \frac{\rho^p(F(\Xi) \overset{\text{qp}}{\times} G(\Gamma))}{\rho^p(F(\Xi))} \right) \quad (\forall \Gamma \in \mathcal{G}).$$

[(V): Commutativity and simultaneous measurability]. Let ρ^p be a pure state, i.e., $\rho^p \in \mathfrak{S}^p(\mathcal{A}^*)$. Let $\mathbf{O} \equiv (X, \mathcal{F}, F)$ and $\mathbf{O}' \equiv (Y, \mathcal{G}, G)$ be crisp observables in $\mathcal{A}$. Now we have the following problem:

- What is the simultaneous measurability condition of $\mathbf{O}$ and $\mathbf{O}'$ for the fixed ρ^p?

This is answered in [39] as follows:

- ρ^p-commutativity, i.e., $F(\Xi)G(\Gamma)\rho^p = G(\Gamma)F(\Xi)\rho^p$ for all $\Xi \in \mathcal{F}, \Gamma \in \mathcal{G}$.

However, in this book we are not concerned with such arguments.

[**(VI): Schrödinger's cat paradox**]. Note that Schrödinger's cat does not appear in the world of MT. Let us explain it as follows: In 1935 (*cf.* [77]) Schrödinger published an essay describing the conceptual problems in quantum mechanics. A brief paragraph in this essay described the cat paradox.

- Suppose we put a cat in a cage with a radioactive atom, a Geiger counter, and a poison gas bottle; further suppose that the atom in the cage has a half-life of one hour, a fifty-fifty chance of decaying within the hour. If the atom decays, the Geiger counter will tick; the triggering of the counter will get the lid off the poison gas bottle, which will kill the cat. If the atom does not decay, none of the above things happen, and the cat will be alive. Now the question:

  (Q) We then ask: What is the state of the cat after the hour?

  The answer according to quantum mechanics is that

  (A) the cat is in a state which can be thought of as half-alive and half-dead, that is, the state such as $\frac{\text{“Fig.(a)”} + \text{“Fig.(b)”}}{2}$

Fig.(a)

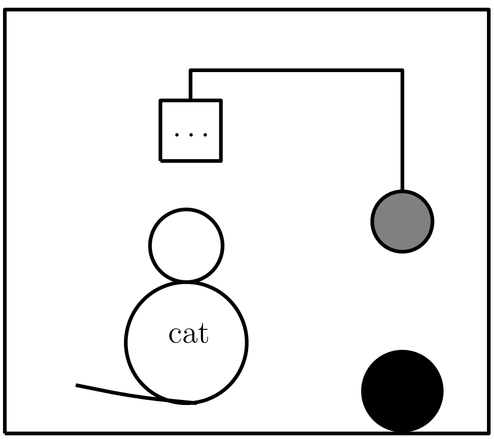
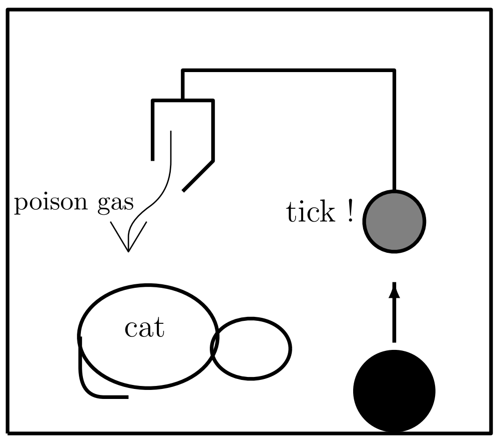


Fig.(b)



Of course, this answer (A) is curious. This is the so-called Schrödinger's cat paradox. This paradox is due to the fact that micro mechanics and macro mechanics are mixed in the above situation. On the other hand, as seen in (2.38), micro mechanics (= quantum measurement theory) and macro mechanics (= classical measurement theory) are always separated in MT. Therefore, Schrödinger's cat does not appear in the world of MT, though this may be a surface solution of Schrödinger's cat paradox.

2.6 Examples

Again recall the (1.4), i.e.,

$$\boxed{\text{“measurement theory (or in short, PMT)”}}$$

$$=\underbrace{[\text{measurements}]}_{\text{“Axiom 1 (2.37)”}} + \underbrace{[\text{the relation among systems}]}_{\text{[Axiom 2 (3.26)]}} \quad \text{in } C^*\text{-algebra } \mathcal{A} \qquad \begin{matrix}(2.39a)\\ (=(1.4a))\end{matrix}$$

or more precisely,

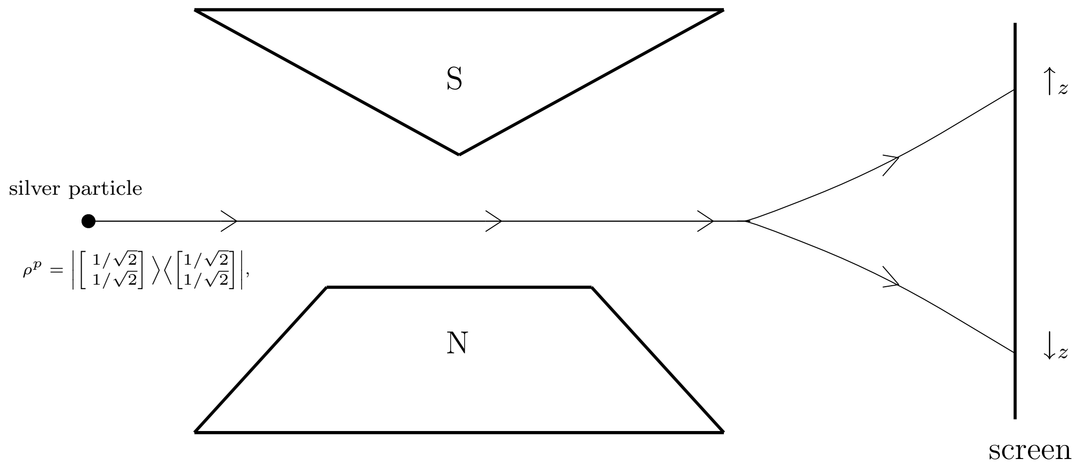
$$= \text{“Apply (2.39}a\text{) to every phenomenon by an analogy of quantum mechanics”} \qquad \begin{matrix}(2.39b)\\ (=(1.4b))\end{matrix}$$

Thus, in order to understand PMT, we need a little knowledge of quantum mechanics.

The following example is enough tested ;[8] and thus, it is the most firm in PMT

Example 2.15. [(i): The spin observable concerning the z-axis, Stern and Gerlach's experiment]. Assume that we examine the beam (of silver particles) after passing through the magnetic field. Then, as seen in the following figure, we see that all particles are deflected either equally upwards or equally downwards in a 50:50 ratio.

“Stern and Gerlach's experiment (1922)”



Consider the two dimensional Hilbert space $V = \mathbf{C}^2$, And therefore, we get the non-commutative C^*-algebra $\mathcal{A} = B(V)$, that is, the algebra composed of all 2×2 matrices.

[8] A lot of tests of quantum mechanics have been conducted. Especially Aspect's experiment [8] is well authorized. (Cf. §2.9 Bell's inequality) Recall that “quantum system theory” $\subset$ “PMT”. Thus, quantum mechanics must be enough tested though the experimental test of PMT is generally meaningless. (Cf. Remark 1.1(e).)

Note that $\mathcal{A} = B(V) = \mathcal{C}(V) = \mathcal{C}_I(V)$ (*cf.* Example 2.3 and Remark 2.6 (i)) since the dimension of V is finite. Define $\mathbf{O}^z \equiv (Z, 2^Z, F^z)$, the spin observable concerning the z-axis, such that, $Z = \{\uparrow_z, \downarrow_z\}$ and

$$F^z(\{\uparrow_z\}) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad F^z(\{\downarrow_z\}) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \tag{2.40}$$

$$F^z(\emptyset) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad F^z(\{\uparrow_z, \downarrow_z\}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

For example, consider the measurement $\mathsf{M}_{B(\mathbf{C}^2)}\big(\mathbf{O}^z \equiv (Z = \{\uparrow_z, \downarrow_z\}, 2^Z, F^z), S_{[\rho^p]}\big)$, where $\rho^p = \Big|\begin{bmatrix}\alpha \\ \beta\end{bmatrix}\Big\rangle\Big\langle\begin{bmatrix}\alpha \\ \beta\end{bmatrix}\Big|$, $|\alpha|^2 + |\beta|^2 = 1$. That is, consider

the measurement $\mathsf{M}_{B(\mathbf{C}^2)}\big(\mathbf{O}^z \equiv (Z = \{\uparrow_z, \downarrow_z\}, 2^Z, F^z), S_{[\rho^p]}\big)$
$\Big($= "the measurement of the observable $\mathbf{O}^z$ for a particle with the state ρ^p"$\Big)$.

Then, the probability that the measured value "$\uparrow_z$" [resp. "$\downarrow_z$"] is obtained by the measurement $\mathsf{M}_{B(\mathbf{C}^2)}\big(\mathbf{O}^z, S_{[\rho^p]}\big)$ is given by $\rho^p(F^z(\{\uparrow_z\})) = |\alpha|^2$ [resp. $\rho^p(F^z(\{\downarrow_z\})) = |\beta|^2$]. Thus, if $\rho^p = \Big|\begin{bmatrix}1/\sqrt{2} \\ 1/\sqrt{2}\end{bmatrix}\Big\rangle\Big\langle\begin{bmatrix}1/\sqrt{2} \\ 1/\sqrt{2}\end{bmatrix}\Big|$, we see that $\rho^p(F^z(\{\uparrow_z\})) = 1/2$ [resp. $\rho^p(F^z(\{\downarrow_z\})) = 1/2$]. For the further argument, see §2.9 (Bell's thought experiment).

[(ii): The other spin observables]. Also, we can define $\mathbf{O}^x \equiv (X, 2^X, F^x)$, the spin observable concerning the x-axis, such that, $X = \{\uparrow_x, \downarrow_x\}$ and

$$F^x(\{\uparrow_x\}) = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}, \quad F^x(\{\downarrow_x\}) = \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix}. \tag{2.41}$$

And furthermore, we can define $\mathbf{O}^y \equiv (Y, 2^Y, F^y)$, the spin observable concerning the y-axis, such that, $Y = \{\uparrow_y, \downarrow_y\}$ and

$$F^y(\{\uparrow_y\}) = \begin{bmatrix} 1/2 & i/2 \\ -i/2 & 1/2 \end{bmatrix}, \quad F^y(\{\downarrow_y\}) = \begin{bmatrix} 1/2 & -i/2 \\ i/2 & 1/2 \end{bmatrix}, \tag{2.42}$$

where $i = \sqrt{-1}$.

■

The following example (= "urn problem") is the most important in the classical PMT, though it is somewhat artificial. That is, we believe that it is not too much to say that

- the probability in Axiom 1 for classical systems is essentially the same as the probability in the following urn problem. (2.43)

However, it should be noted that no serious test for the urn problem has been conducted.[9] It is generally considered to be self-evident without serious experiments. Recall that theoretical informatics does not require serious experiments (*cf.* §1.4).

Example 2.16. [The urn problem (i)]. There are three urns U_1, U_2 and U_3. The urn U_1 [resp. U_2, U_3] contains 8 white and 2 black balls [resp. 4 white and 6 black balls, 1 white and 9 black balls]. That is,

	white balls	black balls
urn U_1	8	2
urn U_2	4	6
urn U_3	1	9

(2.44)

Here, consider the following measurement M_2^c:

$M_2^c :=$ "Pick out one ball from the urn U_2, and recognize the color of the ball"

In measurement theory, the "measurement" M_2^c is formulated as follows: Define the state space Ω by $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Here,

$$\omega_1 = [8:2], \quad \omega_2 = [4:6], \quad \omega_3 = [1:9].$$

Thus, we see that

$U_1 \quad \cdots \quad$ "the urn with the state ω_1"

$U_2 \quad \cdots \quad$ "the urn with the state ω_2"

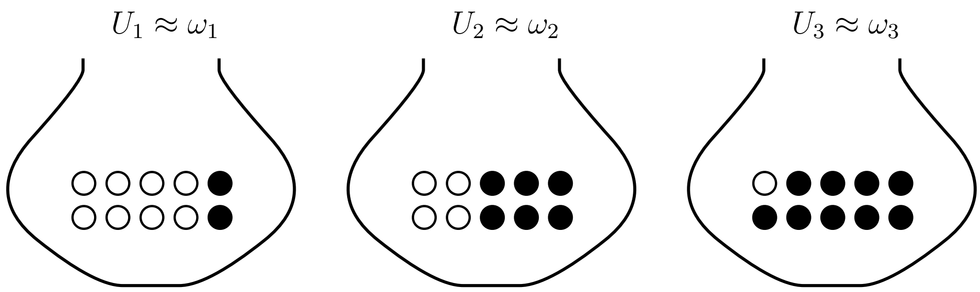
$U_3 \quad \cdots \quad$ "the urn with the state ω_3"

In this sense, we have the identification;

$$U_1 \approx \omega_1, \quad U_2 \approx \omega_2, \quad U_3 \approx \omega_3.$$

That is,

[9][Fuzzy statement and precise statement]. Such a test (i.e., the experimental test of an urn problem) is usually considered to be no more than the good theme of a child's homework. However, the question *"Why is a serious test (concerning the urn problem) not required?"* may be profound. The reason can be understood if we think that the urn problem is a model within theoretical informatics. Cf. §1.4. That is, any model, represented by a precise statement, must be tested in theoretical physics. On the other hand, a model in theoretical informatics is not required to be tested, that is, it suffices to be useful. Cf. (I_{14}) in §1.3. We can say that the urn problem is as true as the statement "A cat is stronger than a mouse". It should be noted that the statement "A cat is stronger than a mouse" is "almost experimentally true" (*cf.* (I_9)) in §1.2, though it is ambiguous, fuzzy, vague, etc.



And further, define the observable $\mathbf{O} = (\{w, b\}, 2^{\{w,b\}}, F)$ in $C(\Omega)$ such that

$$\begin{aligned} F(\{w\})(\omega_1) &= 0.8, \qquad F(\{b\})(\omega_1) = 0.2, \\ F(\{w\})(\omega_2) &= 0.4, \qquad F(\{b\})(\omega_2) = 0.6, \\ F(\{w\})(\omega_3) &= 0.1, \qquad F(\{b\})(\omega_3) = 0.9, \end{aligned} \tag{2.45}$$

where 'w' and 'b' mean white and black respectively. Then, we see that

$$M_2^c = \mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_2}]}). \tag{2.46}$$

Of course, the probability that a measured value w [resp. b] is obtained is, by Axiom 1, given by

$$F(\{w\})(\omega_2) = 0.4 \qquad [\text{ resp. } F(\{b\})(\omega_2) = 0.6] \tag{2.47}$$

[The urn problem (ii)] Further, assume that the (white or black) balls in the urns U_1, U_2 and U_3 are also made of "stone" or "metal". For example, assume that the urn U_1 [resp. U_2, U_2] contains 4 stone and 6 metal balls [resp. 5 stone and 5 metal balls, 1 stone and 9 metal balls]. That is,

	stone balls	metal balls
urn U_1	4	6
urn U_2	5	5
urn U_3	7	3

(2.48)

Here, consider the following measurement M_2^m:

$M_2^m :=$ "Pick out one ball from the urn U_2, and recognize the materials of the ball"

The measurement M_2^m is formulated as follows: Define the state space Ω by $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Here,

$$\omega_1 = [4 : 6], \quad \omega_2 = [5 : 5], \quad \omega_3 = [7 : 3].$$

Thus, we see that

$$U_1 \quad \cdots \quad \text{"the urn with the state } \omega_1\text{"}$$
$$U_2 \quad \cdots \quad \text{"the urn with the state } \omega_2\text{"}$$
$$U_3 \quad \cdots \quad \text{"the urn with the state } \omega_3\text{"}$$

In this sense, we have the identification;

$$U_1 \approx \omega_1, \quad U_2 \approx \omega_2, \quad U_3 \approx \omega_3.$$

And further, define the observable $\mathbf{O}' = (\{s, m\}, 2^{\{s,m\}}, G)$ in $C(\Omega)$ such that

$$\begin{aligned} G(\{s\})(\omega_1) &= 0.4, & G(\{m\})(\omega_1) &= 0.6, \\ G(\{s\})(\omega_2) &= 0.5, & G(\{m\})(\omega_2) &= 0.5, \\ G(\{s\})(\omega_3) &= 0.7, & G(\{m\})(\omega_3) &= 0.3. \end{aligned} \tag{2.49}$$

Thus, we see:

$$M_2 = \mathbf{M}_{C(\Omega)}(\mathbf{O}', S_{[\delta_{\omega_2}]}). \tag{2.50}$$

For example, the probability that a measured value s [resp. m] is obtained is, by Axiom 1, given by

$$G(\{s\})(\omega_2) = 0.5 \qquad [\text{ resp. } G(\{m\})(\omega_2) = 0.5]. \tag{2.51}$$

[The urn problem (iii)] However, it should noted that some information is not represented in the tables (2.44) and (2.48). That is, the situation is, for example, stated precisely as follows:

(1) the urn U_1 contains 10 balls such as

	stone balls	metal balls
white balls	4	4
black balls	0	2

(2.52)

(2) the urn U_2 contains 10 balls such as

	stone balls	metal balls
white balls	4	0
black balls	1	5

(2.53)

(3) the urn U_3 contains 10 balls such as

	stone balls	metal balls
white balls	1	0
black balls	6	3

(2.54)

Here, consider the following measurement M_2^{cm}:

M_2^{cm} := "Pick out one ball from the urn U_2, and recognize the color and materials of the ball".

The measurement M_{12} is formulated as follows: Put $\Omega = \{\omega_1, \omega_2, \omega_2\}$. Define the state space Ω by $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Here,

$$\omega_1 = \begin{bmatrix} 4 & 4 \\ 0 & 2 \end{bmatrix}, \quad \omega_2 = \begin{bmatrix} 4 & 0 \\ 1 & 5 \end{bmatrix}, \quad \omega_3 = \begin{bmatrix} 1 & 0 \\ 6 & 3 \end{bmatrix}.$$

Thus, we see that

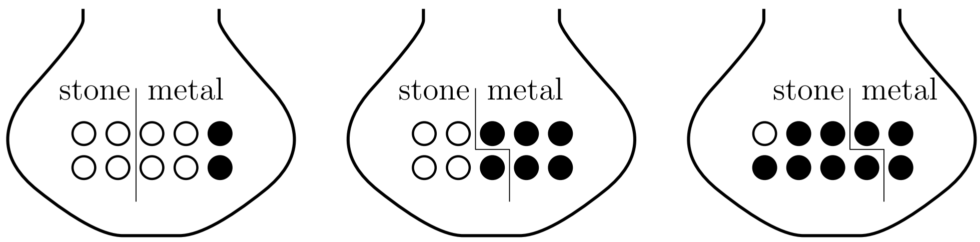
$$U_1 \quad \cdots \quad \text{"the urn with the state } \omega_1\text{"}$$
$$U_2 \quad \cdots \quad \text{"the urn with the state } \omega_2\text{"}$$
$$U_3 \quad \cdots \quad \text{"the urn with the state } \omega_3\text{"}$$

In this sense, we have the identification;

$$U_1 \approx \omega_1, \quad U_2 \approx \omega_2, \quad U_3 \approx \omega_3.$$

That is,

$U_1 \approx \omega_1$ $\qquad$ $U_2 \approx \omega_1$ $\qquad$ $U_3 \approx \omega_1$



And further, define the observable $\widehat{\mathbf{O}} = (\{w, b\} \times \{s, m\}, 2^{\{w,b\} \times \{s,m\}}, H(\equiv F \overset{\text{qp}}{\times} G))$ in $C(\Omega)$ such that

$$H(\{(w,s)\})(\omega_1) = 0.4, \; H(\{(w,m)\})(\omega_1) = 0.4, \; H(\{(b,s)\})(\omega_1) = 0.0, \; H(\{(b,m)\})(\omega_1) = 0.2,$$
$$H(\{(w,s)\})(\omega_2) = 0.4, \; H(\{(w,m)\})(\omega_2) = 0.0, \; H(\{(b,s)\})(\omega_2) = 0.1, \; H(\{(b,m)\})(\omega_2) = 0.5,$$
$$H(\{(w,s)\})(\omega_3) = 0.1, \; H(\{(w,m)\})(\omega_3) = 0.0, \; H(\{(b,s)\})(\omega_3) = 0.6, \; H(\{(b,m)\})(\omega_3) = 0.3, \tag{2.55}$$

which is, of course, constructed by (2.52) + (2.53) + (2.54) . Then, we see that

$$M_{12} = \mathbf{M}_{C(\Omega)}(\widehat{\mathbf{O}}, S_{[\delta_{\omega_2}]}). \tag{2.56}$$

Of course, the probability that a measured value (w, s) [resp. (w, m), (b, s), (b, m)] is obtained is, by Axiom 1, given by

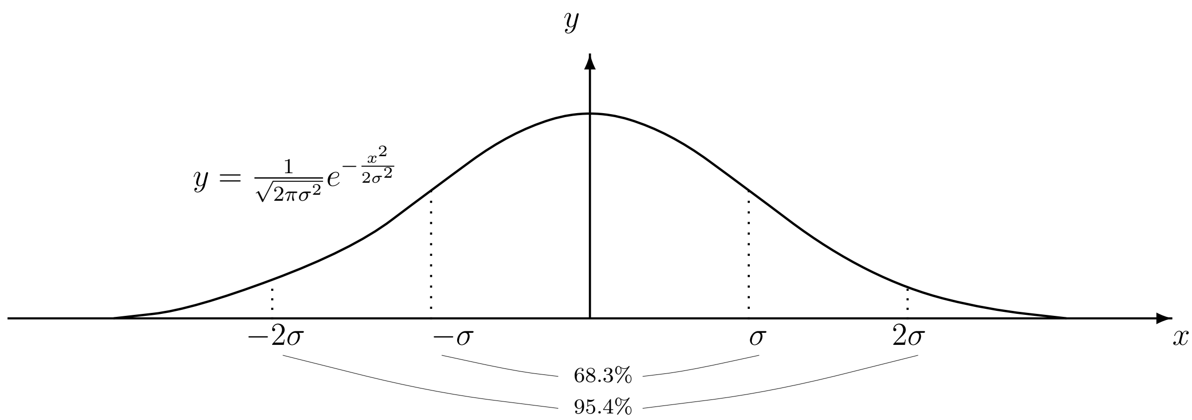
$$\begin{aligned} &F(\{(w,s)\})(\omega_2) = 0.4 \\ &[\text{ resp. } F(\{(w,m)\})(\omega_2) = 0.0, F(\{(b,s)\})(\omega_2) = 0.1, F(\{(b,m)\})(\omega_2) = 0.5]. \end{aligned} \tag{2.57}$$

■

Example 2.17. [Gaussian observable[10]]. [(i): Gaussian observable in $C(\Omega)$]. Put $\Omega = [a, b]$ ($\subseteq \mathbf{R}$, the real line), i.e., the closed interval And let σ be a fixed positive real. Define the *normal observable* (or *Gaussian observable*) $\mathbf{O}_{G^\sigma} \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, G^\sigma)$ in $C(\Omega)$ such that:

$$[G^\sigma(\Xi)](\omega) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_\Xi e^{-\frac{(x-\omega)^2}{2\sigma^2}} dx \quad (\forall \Xi \in \mathcal{B}_{\mathbf{R}}, \forall \omega \in \Omega \equiv [a, b]),$$

which will be often used in this book.



Here, $\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\sigma}^{\sigma} e^{-\frac{x^2}{2\sigma^2}} dx = 0.683...$ and $\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-2\sigma}^{2\sigma} e^{-\frac{x^2}{2\sigma^2}} dx = 0.954...$ Also, note that

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-1.96\sigma}^{1.96\sigma} e^{-\frac{x^2}{2\sigma^2}} dx \approx 0.95, \quad \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{1.65\sigma} e^{-\frac{x^2}{2\sigma^2}} dx \approx 0.95 \tag{2.58}$$

[10]Why is the Gaussian observable fundamental? We should not be too serious with the question. That is because we do not necessarily need a complete reason in theoretical informatics (*cf.* Chapter 1), though the differential geometrical reason must be indispensable for theoretical physics. In informatics, what is important is "useful or not". And we know that the Gaussian observable is quite useful. Also recall that every equation (e.g., Boltzmann's kinetic equation, Navier-Stokes equation, etc.) in theoretical informatics is somewhat empirical. As mentioned in (I_9) in §1.2, we think that "useful" $\Longrightarrow$ "almost experimentally true".

[(ii).Gaussian observable in $C_0(\mathbf{R}^d)$]. Consider a commutative C^*-algebra $C_0(\mathbf{R}^d)$ and the Borel ring $(\mathbf{R}^d, \mathcal{B}^{\rm bd}_{\mathbf{R}^d})$, where $\mathcal{B}^{\rm bd}_{\mathbf{R}^d} = \{\Xi \in \mathcal{B}_{\mathbf{R}^d} : \Xi$ is a bounded Borel set in $\mathbf{R}^d$ $\}$. And define the *d-dimensional Gaussian observable* $\mathbf{O}_\Sigma \equiv (\mathbf{R}^d, \mathcal{B}^{\rm bd}_{\mathbf{R}^d}, F^\Sigma)$ in $C_0(\mathbf{R}^d)$ such that:

$$[F^\Sigma(\Xi)](\vec{\omega}) = \frac{1}{\sqrt{2\pi}^d |\Sigma|^{1/2}} \int_\Xi \exp[-\frac{1}{2}(\vec{x}-\vec{\omega})^t \Sigma^{-1}(\vec{x}-\vec{\omega})]d\vec{x} \qquad (\forall \Xi \in \mathcal{B}^{\rm bd}_{\mathbf{R}^d}, \quad \forall \vec{\omega} \in \mathbf{R}^d), \tag{2.59}$$

where the Σ is a covariance $(d\times d)$-matrix, i.e., a positive definite $(d\times d)$-matrix. Of course, the probability that a measured value obtained by the measurement $\mathbf{M}_{C_0(\mathbf{R}^d)}(\mathbf{O}_\Sigma, S_{[\delta_{\vec{\omega}_0}]})$ belongs to Ξ $(\in \mathcal{B}^{\rm bd}_{\mathbf{R}^d})$ is given by $[F^\Sigma(\Xi)](\vec{\omega}_0)$.

■

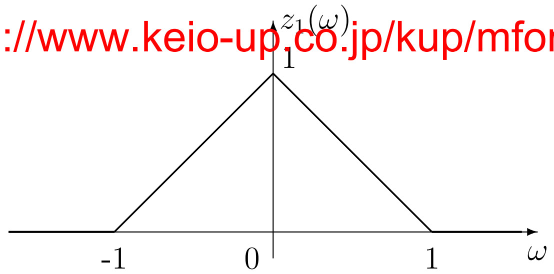
Example 2.18. [Discrete Gaussian observable]. Put $\Omega \equiv [a,b]$ ($\subseteq \mathbf{R}$, the real line), the closed interval. Let $\sigma > 0$. And let N be a sufficiently large fixed integer. Put $X_N \equiv \{\frac{k}{N} \mid k = 0, \pm 1, \pm 2, ..., \pm N^2\}$. And define the *discrete Gaussian observable* $\mathbf{O}_{\sigma^2,N}$ $\equiv (X_N,\ 2^{X_N}, F_{\sigma,N})$ in the commutative C^*-algebra $C([a,b])$ such that:

$$[F_{\sigma,N}(\{k/N\})](\omega)$$
$$= \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{N-\frac{1}{2N}}^{\infty} \exp[-\frac{(x-\omega)^2}{2\sigma^2}]dx & (k = N^2, \forall \omega \in [a,b]), \\ \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\frac{k}{N}-\frac{1}{2N}}^{\frac{k}{N}+\frac{1}{2N}} \exp[-\frac{(x-\omega)^2}{2\sigma^2}]dx & (\forall k = 0, \pm 1, \pm 2, ..., \pm(N^2-1), \quad \forall \omega \in [a,b]), \\ \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{-N+\frac{1}{2N}} \exp[-\frac{(x-\omega)^2}{2\sigma^2}]dx & (k = -N^2, \forall \omega \in [a,b]). \end{cases} \tag{2.60}$$

And thus, for any Ξ ($\subseteq X_N$), we define $[F_{\sigma,N}(\Xi)](\omega) = \sum_{\frac{k}{N}\in\Xi}[F_{\sigma,N}(\{k/N\})]\ (\omega)$. This $\mathbf{O}_{\sigma^2,N}$, as well as the d-dimensional Gaussian observable $\mathbf{O}_\Sigma$ (in Example 2.17), is the most important observable in classical measurements. ■

Example 2.19. [Fuzzy numbers observable (= triangle observable = round error observable)]. Let Δ be any positive number. Define the membership function (i.e., triangle fuzzy number) $\mathcal{Z}_\vartriangle$ $\Big(\in C_0(\mathbf{R})$, where $\mathbf{R}$ is the real line with the usual topology$\Big)$ such that:

$$\mathcal{Z}_\vartriangle(\omega) = \begin{cases} 1 - \frac{\omega}{\Delta} & 0 \le \omega \le \Delta \\ \frac{\omega}{\Delta} + 1 & -\Delta \le \omega \le 0 \\ 0 & \text{otherwise}. \end{cases}$$



Put $\mathbb{Z}_\Delta \equiv \{\Delta k \;\; : \;\; k \in \mathbb{Z} \equiv \{0, \pm 1, \pm 2, ...\}\}$. Define the C^*-observable $\mathbf{O}_{\mathcal{Z}_\Delta} \equiv (\mathbb{Z}_\Delta, \mathcal{P}_0(\mathbb{Z}_\Delta), \zeta^\Delta_{(\cdot)})$ in the commutative C^*-algebra $C_0(\mathbf{R})$ such that $\zeta^\Delta_\Xi(\omega) = \sum_{x\in\Xi} \mathcal{Z}_\Delta(\omega - x)$ $(\forall \Xi \in \mathcal{P}_0(\mathbb{Z}_\Delta), \forall\omega \in \mathbf{R})$. This C^*-observable is called a *fuzzy numbers observable* in $C_0(\mathbf{R})$. Putting $\Delta = 1$, we frequently use the fuzzy numbers observable $\mathbf{O}_{\mathcal{Z}} \equiv (\mathbb{Z}, \mathcal{P}_0(\mathbb{Z}), \zeta_{(\cdot)})$ in this book.

■

Example 2.20. [(i): Exact observable]. Let $\mathbb{Z}$ be the set of all integers, i.e., $\mathbb{Z} = \{0, \pm 1, \pm 2, ...\}$. And put $\mathcal{P}_0(\mathbb{Z}) = \{A(\subseteq \mathbb{Z}) \mid \; A \text{ is finite }\}$. Consider a commutative C^*-algebra $C_0(\mathbb{Z})$. And define the exact observable $\mathbf{O}_{\text{EXA}} \equiv (\mathbb{Z}, \mathcal{P}_0(\mathbb{Z}), E_{(\cdot)})$ in $C_0(\mathbb{Z})$ such that:

$$E_\Xi(n) = \begin{cases} 1 & n \in \Xi(\in \mathcal{P}_0(\mathbb{Z})) \\ \\ 0 & n \notin \Xi(\in \mathcal{P}_0(\mathbb{Z})) \end{cases} \tag{2.61}$$

which is called *the exact observable* (or, *fundamental observable*) in $C_0(\mathbb{Z})$. Of course we want to define the exact observable in $C_0(\mathbf{R})$ (or, $C([a, b])$). However, it is impossible in the C^*-algebraic formulation. For this, we must prepare the W^*-algebraic formulation (*cf.* Chapter 9).

[(ii): Approximate exact observable]. Though the exact observable in $C([0, 1])$ can not be defined, we have the approximate exact observable $\mathbf{O}^A_{\text{EXA}}$ in $C([0, 1])$ as follows: Let N be a sufficiently large integer. Put $X_N = \{\frac{1}{N}, \frac{2}{N}, \frac{3}{N}, ..., \frac{N}{N}(\equiv 1)\}$. Define the approximate exact observable $\mathbf{O}^A_{\text{EXA}} \equiv (X_N, \mathcal{P}(X_N), F)$ in $C([0, 1])$ such that:

$$[F(\{\frac{1}{N}\})](\omega) = \begin{cases} 1 & (0 \le \omega \le \frac{1}{N} - \frac{1}{N^2}) \\ -\frac{N^2}{2}(\omega - \frac{1}{N}) + \frac{1}{2} & (\frac{1}{N} - \frac{1}{N^2} \le \omega \le \frac{1}{N} + \frac{1}{N^2}) \\ 0 & (\frac{1}{N} + \frac{1}{N^2} \le \omega \le 1) \end{cases}$$

$$[F(\{\frac{N}{N}\})](\omega) = \begin{cases} 0 & (0 \le \omega \le \frac{N-1}{N} - \frac{1}{N^2}) \\ \frac{N^2}{2}(\omega - \frac{N-1}{N}) + \frac{1}{2} & (\frac{N-1}{N} - \frac{1}{N^2} \le \omega \le \frac{N-1}{N} + \frac{1}{N^2}) \\ 1 & (\frac{N-1}{N} + \frac{1}{N^2} \le \omega \le \frac{N}{N} - \frac{1}{N^2}) \end{cases}$$

For $n = 2, 3, ..., N-1$,

$$[F(\{\frac{n}{N}\})](\omega) = \begin{cases} 0 & (0 \le \omega \le \frac{n-1}{N} - \frac{1}{N^2}) \\ \frac{N^2}{2}(\omega - \frac{n-1}{N}) + \frac{1}{2} & (\frac{n-1}{N} - \frac{1}{N^2} \le \omega \le \frac{n-1}{N} + \frac{1}{N^2}) \\ 1 & (\frac{n-1}{N} + \frac{1}{N^2} \le \omega \le \frac{n}{N} - \frac{1}{N^2}) \\ -\frac{N^2}{2}(\omega - \frac{n}{N}) + \frac{1}{2} & (\frac{n}{N} - \frac{1}{N^2} \le \omega \le \frac{n}{N} + \frac{1}{N^2}) \\ 0 & (\frac{n}{N} + \frac{1}{N^2} \le \omega \le 1) \end{cases}$$

Note that the observable (i.e., fuzzy numbers observable) in Example 2.19 is also regarded as "approximate exact observable", if Δ is sufficiently small.

■

Example 2.21. [Null observable]. Define the observable $\mathbf{O}^{(\mathrm{nl})} \equiv (\{0,1\}, 2^{\{0,1\}}, F^{(\mathrm{nl})})$ in $\mathcal{A}$ such that:

$$F^{(\mathrm{nl})}(\emptyset) \equiv 0, \quad F^{(\mathrm{nl})}(\{0\}) \equiv 0, \quad F^{(\mathrm{nl})}(\{1\}) \equiv 1_{\mathcal{A}}, \quad F^{(\mathrm{nl})}(\{0,1\}) \equiv 1_{\mathcal{A}} \quad \text{in } \mathcal{A}, \tag{2.62}$$

which may be called the *null observable* (or, *existence observable*). Then, we have the measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}^{(\mathrm{nl})} \equiv (\{0,1\}, 2^{\{0,1\}}, F^{(\mathrm{nl})}), S_{[\rho^p]})$. Note that:

($\sharp$) the probability that measured value (by $\mathbf{M}_{\mathcal{A}}(\mathbf{O}^{(\mathrm{nl})}, S_{[\rho^p]})$) is equal to 1 ($\in \{0,1\}$) is given by 1. That is, the measured value is always equal to 1 ($\in \{0,1\}$).

Thus, we think that "to take the measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}^{(\mathrm{nl})}, S_{[\rho^p]})$" is the same as "to assure the existence of the system".

■

2.7 Operations of observables

Recall the identification (2.36), that is, we have the following identification:

$$\underset{\text{(real valued function on }\Omega)}{\widehat{F}_k} \longleftrightarrow \underset{\text{(crisp observable)}}{\mathbf{O}_k = (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, F_k) \text{ in } C(\Omega)} \qquad (k = 1, 2, ..., n). \tag{2.63}$$

Note that $\widehat{F}_1 + \widehat{F}_2$, $\widehat{F}_1 \cdot \widehat{F}_2$, etc. are meaningful in the ordinary sense since $\widehat{F}_1$ and $\widehat{F}_2$ are real-valued functions. This makes us ask the following question.

- For each $k = 1, 2, ..., n$, consider an observable $\mathbf{O}_k \equiv (X_k, \mathcal{F}_k, F_k)$ in a C^*-algebra $\mathcal{A}$. Are $\mathbf{O}_1 + \mathbf{O}_2$, $\mathbf{O}_1 \cdot \mathbf{O}_2$, etc. meaningful in general? Or, how the operations of observables are defined?

This will be answered in what follows.

For each $k = 1, 2, ..., n$, consider an observable $\mathbf{O}_k \equiv (X_k, \mathcal{F}_k, F_k)$ in a C^*-algebra $\mathcal{A}$. Put $\mathbf{O} = \overset{\text{qp}}{\times}_{k=1,2,...,n} \mathbf{O}_k$. Let $g : \times_{k=1}^n \to Y$ be a measurable map, where Y has the subfield $\mathcal{G}$ of 2^Y. Then we can define the observable $(Y, \mathcal{G}, G)$, which is symbolically represented by $g(\mathbf{O}_1, \mathbf{O}_2, ..., \mathbf{O}_n)$, as follows:

- the $(Y, \mathcal{G}, G)$ is the image observable of the quasi-product observable $\mathbf{O} \equiv (\times_{k=1}^n X_k, \times_{k=1}^n \mathcal{F}_k, \widehat{F})$ concerning g (if it exists). That is,

$$(Y, \mathcal{G}, G) = g(\mathbf{O}) \tag{2.64}$$

i.e.,

$$G(\Gamma) = \widehat{F}(g^{-1}(\Gamma)) \qquad (\forall \Gamma \in \mathcal{G}). \tag{2.65}$$

Example 2.22. [The addition of triangle observables]. Let $\mathbf{O}_{\mathbb{Z}} \equiv (\mathbb{Z}, \mathcal{P}_0(\mathbb{Z}), \zeta_{(\cdot)})$ be the fuzzy numbers observable in $C_0(\mathbf{R})$ (*cf.* Example 2.19). Now let us calculate $\mathbf{O}_{\mathbb{Z}} + \mathbf{O}_{\mathbb{Z}}$ as follows: Note that the product observable $\mathbf{O}_{\mathbb{Z}} \times \mathbf{O}_{\mathbb{Z}} \equiv (\mathbb{Z}^2, \mathcal{P}_0(\mathbb{Z}^2), \zeta_{(\cdot)} \times \zeta_{(\cdot)})$ is represented by

(i) $|m - n| \geq 2$

$$[\zeta_{\{m\}} \times \zeta_{\{n\}}](\omega) = 0 \tag{2.66}$$

(ii) $|m - n| = 1$

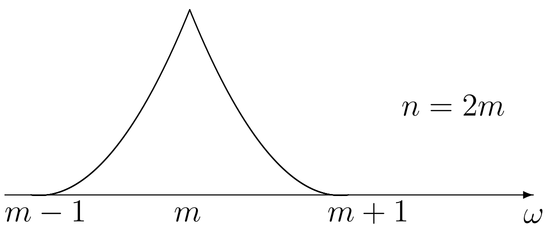
$$[\zeta_{\{m\}} \times \zeta_{\{n\}}](\omega) = \begin{cases} 0 & \omega \leq \min\{m, n\} \\ \frac{(x-m)(x-n)}{2} & \min\{m, n\} \leq \omega \leq \max\{m, n\} \\ 0 & \omega \leq \min\{m, n\} \end{cases}$$

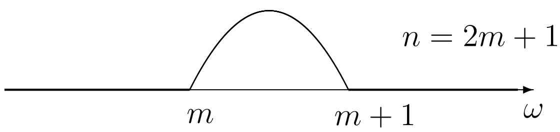
(iii) $m = n$

$$[\zeta_{\{m\}} \times \zeta_{\{m\}}](\omega) = \begin{cases} 0 & \omega \leq m - 1 \\ (x - (m-1))^2 & m - 1 \leq \omega \leq m \\ (x - (m+1))^2 & m \leq \omega \leq m + 1 \\ 0 & m + 1 \leq \omega \end{cases} \tag{2.67}$$

Thus we see

$$
\begin{aligned}
&(\zeta+\zeta)_{\{n\}}(\omega) \\
&= \begin{cases}
\underset{\text{(when } n=2m)}{(\zeta+\zeta)_{\{2m\}}(\omega)} = \begin{cases} 0 & \omega \le m-1 \\ (\omega-(m-1))^2 & m-1 \le \omega \le m \\ (\omega-(m+1))^2 & m \le \omega \le m+1 \\ 0 & m+1 \le \omega \end{cases} \\
\underset{\text{(when } n=2m+1)}{(\zeta+\zeta)_{\{2m+1\}}(\omega)} = \begin{cases} 0 & \omega \le m \\ -(\omega-(2m+1)/2)^2+1/2 & m \le \omega \le m+1 \\ 0 & m+1 \le \omega \end{cases}
\end{cases}
\end{aligned}
\tag{2.68}
$$





Therefore we get the $\mathbf{O}_{\mathbb{Z}} + \mathbf{O}_{\mathbb{Z}} \equiv (\mathbb{Z}, \mathcal{P}_0(\mathbb{Z}), (\zeta+\zeta)_{(\cdot)})$ in $C_0(\mathbf{R})$, where

$$(\zeta+\zeta)_{\Xi}(\omega) = \sum_{n\in\Xi}(\zeta+\zeta)_{\{n\}}(\omega) \qquad (\Xi \in \mathcal{P}_0(\mathbb{Z}), \omega \in \Omega).$$

■

Example 2.23. (χ^2-observable). Consider the (1-dimensional) Gaussian observable $\mathbf{O}_{\sigma^2} \equiv (\mathbf{R}, \mathcal{B}^{\rm bd}_{\mathbf{R}}, G^\sigma)$ in $\mathcal{A} \equiv C_0(\mathbf{R})$ such that:

$$[G^\sigma(\Xi)](\mu) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_\Xi e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \qquad (\forall \mu \in \mathbf{R}\ \forall \Xi \in \mathcal{B}^{\rm bd}_{\mathbf{R}}),$$

(where σ^2 is a variance). And further, for each ϕ $(=0,1,2,\ldots)$, define the product observable $(\mathbf{O}_{\sigma^2})^{\phi+1}$ such that

$$(\mathbf{O}_{\sigma^2})^{\phi+1} = (\mathbf{R}^{\phi+1}, \mathcal{B}^{\rm bd}_{\mathbf{R}^{\phi+1}}, (G^\sigma)^{\phi+1}) \qquad (\text{ in } \mathcal{A} \equiv C_0(\mathbf{R})$$

where

$$(G^\sigma)^{\phi+1}(\Xi_1 \times \Xi_2 \times \cdots \times \Xi_{\phi+1}) = G^\sigma(\Xi_1) \times G^\sigma(\Xi_2) \times \cdots \times G^\sigma(\Xi_{\phi+1}).$$

Define the map $g : \mathbf{R}^{\phi+1} \to \mathbf{R}$ such that

$$\mathbf{R}^{\phi+1} \ni (x_1, x_2, x_3, ..., x_{\phi+1}) \mapsto \sum_{k=1}^{\phi+1} \frac{(x_k - \frac{\sum_{j=1}^{\phi+1} x_j}{\phi+1})^2}{\sigma^2} \in \mathbf{R}.$$

The image observable $g((\mathbf{O}_{\sigma^2})^{\phi+1})$ is called the χ^2-*observable with* ϕ, *the degree of freedom.*

■

2.8 Frequency probabilities

The meaning of "probability" in Axiom 1 seems to be a matter of common knowledge in quantum mechanics. However, we, in this section, study the relation between "the probability in Axiom 1" and "frequency probability".

For each $k = 1, 2, ..., n$, consider a measurement $\mathbf{M}_{\mathcal{A}_k}(\mathbf{O}_k \equiv (X, \mathcal{P}(X), F_k), S_{[\rho_k^p]})$ in a C^*-algebra $\mathcal{A}_k$, where we assume, for simplicity, that X is finite. Put $\widehat{\mathcal{A}} = \bigotimes_{k=1}^n \mathcal{A}_k$, i.e., the tensor product C^*-algebra of $\{\mathcal{A}_k : k = 1, 2, ..., n\}$. Here, consider the tensor-product C^*-observable $\bigotimes_{k=1}^n \mathbf{O}_k \equiv (X^n, \mathcal{P}(X^n), \widehat{F} \equiv \bigotimes_{k=1}^n F_k)$ in $\widehat{\mathcal{A}}$ ($\equiv \bigotimes_{k=1}^n \mathcal{A}_k$) such that:

$$\widehat{F}(\Xi_1 \times \Xi_2 \times \cdots \times \Xi_n) = F_1(\Xi_1) \otimes F_2(\Xi_2) \otimes \cdots \otimes F_n(\Xi_n) \quad (\forall \Xi_k \in \mathcal{P}(X), \ k = 1, 2, ..., n). \tag{2.69}$$

Therefore, we get the measurement $\mathbf{M}_{\otimes \mathcal{A}_k}(\bigotimes_{k=1}^n \mathbf{O}_k, S_{[\bigotimes_{k=1}^n \rho_k^p]})$ in $\bigotimes_{k=1}^n \mathcal{A}_k$, which is also denoted by $\bigotimes_{k=1}^n \mathbf{M}_{\mathcal{A}_k}(\mathbf{O}_k, S_{[\rho_k^p]})$ and called the *repeated measurement* (or, *"parallel measurement"*) of $\mathbf{M}_{\mathcal{A}_k}(\mathbf{O}_k, S_{[\rho_k^p]})$'s. Put $\mathcal{M}_{+1}^m(X) = \{\nu : \nu$ is a positive measure on X such that $\nu(X) = 1\}$ and define the map $g : X^n \to \mathcal{M}_{+1}^m(X)$ such that:

$$[g(x_1, x_2, ..., x_n)](\Xi) = \frac{\sharp[\{k : x_k \in \Xi\}]}{n} \quad (\forall \Xi \in \mathcal{P}(X)), \tag{2.70}$$

where $\sharp[B]$ = "the number of the elements of a set B".

Then we have the following proposition.

Proposition 2.24. [The weak law of large numbers, cf [56]]. *Suppose the above notations. For any* $\epsilon > 0$ *and any* Ξ ($\in \mathcal{P}(X)$), *define* $\widehat{D}_{\Xi,\epsilon}$ ($\in \mathcal{P}(X^n)$) *by*

$$\widehat{D}_{\Xi,\epsilon} = \Big\{ \widehat{x} = (x_1, x_2, ..., x_n) \in X^n \ : \ \Big| [g(\widehat{x})](\Xi) - \frac{1}{n} \sum_{k=1}^n \rho_k^p(F_k(\Xi)) \Big| < \epsilon \Big\}. \tag{2.71}$$

Then we see that

$$1 - \frac{1}{4\epsilon^2 n} \le (\otimes_{k=1}^n \rho_k^p)\Big(\widehat{F}(\widehat{D}_{\Xi,\epsilon})\Big) \le 1, \quad (\forall \Xi \in \mathcal{P}(X), \forall \epsilon > 0, \forall n). \tag{2.72}$$

Proof. We easily see that $[g(\widehat{x})](\Xi) = \frac{1}{n}\sum_{k=1}^n \chi_\Xi(\pi_k(\widehat{x}))$ $(\forall \widehat{x} = (x_1, x_2, ..., x_n) \in X^n)$, where $\pi_k : X^n \to X$ is defined by $\pi_k(\widehat{x}) \equiv \pi_k(x_1, x_2, ..., x_k, ..., x_n) = x_k$ and $\chi_\Xi : X \to \mathbf{R}$ is the characteristic function of Ξ (i.e., $\chi_\Xi(x) = 1$ $(x \in \Xi)$, $= 0$ $(x \notin \Xi)$). Using the terms in Kolmogorov's probability theory, we can say that $\chi_\Xi(\pi_k(\cdot))$, $k = 1, 2, ..., n$, are independent variables on a probability space $\Big(X^n, \mathcal{P}(X^n), \widehat{P}(\cdot) \equiv (\otimes_{k=1}^n \rho_k^p)(\widehat{F}(\cdot))\Big)$. Also it is clear that $\int_{X^n} \chi_\Xi(\pi_k(\widehat{x}))\widehat{P}(d\widehat{x}) = \int_{X^n}[\chi_\Xi(\pi_k(\widehat{x}))]^2 \widehat{P}(d\widehat{x}) = \rho_k^p(F_k(\Xi))$ $(k = 1, 2, ..., n)$. Therefore, by Čebyšev inequality, we see

$$\begin{aligned}
\widehat{P}\Big(X^n \setminus \widehat{D}_{\Xi,\epsilon}\Big) &= \widehat{P}\Big(\Big\{\widehat{x} \in X^n \; : \; \Big|\frac{\sum_{k=1}^n \chi_\Xi(\pi_k(\widehat{x}))}{n} - \frac{\sum_{k=1}^n \rho_k^p(F_k(\Xi))}{n}\Big| \ge \epsilon\Big\}\Big) \\
&\le \frac{1}{\epsilon^2 n^2}\int_{X^n} \Big|\sum_{k=1}^n \Big(\chi_\Xi(\pi_k(\widehat{x})) - \rho_k^p(F_k(\Xi))\Big)\Big|^2 \widehat{P}(d\widehat{x}) \\
&= \frac{1}{\epsilon^2 n^2}\sum_{k=1}^n \int_{X^n} |\chi_\Xi(\pi_k(\widehat{x})) - \rho_k^p(F_k(\Xi))|^2 \widehat{P}(d\widehat{x}) \\
&\le \frac{1}{\epsilon^2 n}\max_{1\le k\le n}\big[\rho_k^p(F_k(\Xi))(1 - \rho_k^p(F_k(\Xi)))\big] \le \frac{1}{4\epsilon^2 n},
\end{aligned} \tag{2.73}$$

which implies (2.72). This completes the proof. □

Now we can show the following theorem as an immediate consequence of Proposition 2.24. It clarifies the "probability" in Axiom 1 from the statistical point of view.

Theorem 2.25. [Frequency probability, *cf.* [42]]. *Put* $\mathcal{A}_k = \mathcal{A}$, $\rho_k^p = \rho^p$ *and* $\mathbf{O}_k = \mathbf{O}$ $\equiv (X, \mathcal{P}(X), F)$, $k = 1, 2, ..., n$, *in Proposition 2.24. Consider the repeated measurement* $\mathbf{M}_{\otimes\mathcal{A}}(\bigotimes_{k=1}^n \mathbf{O}, S_{[\otimes_{k=1}^n \rho^p]})$ *in* $\bigotimes_{k=1}^n \mathcal{A}$. *Then, we see that*

$$1 - \frac{1}{4\epsilon^2 n} \le (\otimes_{k=1}^n \rho^p)\Big((\bigotimes_{k=1}^n F)\big(\{\widehat{x} \in X^n : \big|\rho^p(F(\Xi)) - \frac{\sharp[\{k : x_k \in \Xi\}]}{n}\big| < \epsilon\}\big)\Big) \le 1,$$

$$(\forall \Xi \in \mathcal{P}(X), \forall \epsilon > 0, \forall n).$$

Here note, by Axiom 1, that $(\otimes_{k=1}^n \rho^p)\Big((\bigotimes_{k=1}^n F)(\widehat{\Xi})\Big)$ *is the probability that a measured value by* $\mathbf{M}_{\otimes\mathcal{A}}(\bigotimes_{k=1}^n \mathbf{O}, S_{[\otimes_{k=1}^n \rho^p]})$ *belongs to* $\widehat{\Xi}$. *Therefore, if* n *is sufficiently large, for a measured value* $\widehat{x}$ $(= (x_1, x_2, ..., x_n) \in X^n)$ *by* $\mathbf{M}_{\otimes\mathcal{A}}(\bigotimes_{k=1}^n \mathbf{O}, S_{[\otimes_{k=1}^n \rho^p]})$, *we can consider* *(in the sense of (2.72)) that*

$$\rho^p(F(\Xi)) \approx \frac{\sharp[\{k : x_k \in \Xi\}]}{n}. \tag{2.74}$$

■

The (2.74) says that

- "probability in Axiom 1" = "frequency probability".

Thus, there is a reason that the probability space $(X, \mathcal{F}, \rho^p(F(\cdot)))$ is called *a sample space obtained by a measurement* $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[\rho^p]})$.

Remark 2.26. ["repeated measurement = iterated measurement" for $S_{[\delta_{\omega_0}]}$]. As seen in this section, we think that

"take a measurement M_{ω_0} N times" $\Leftrightarrow$ "take a measurement $\mathbf{M}_{\otimes_{n=1}^N C(\Omega)}(\otimes_{n=1}^N \mathbf{O}, S_{[\otimes_{n=1}^N \delta_{\omega_0}]})$"

Thus, in classical measurements, we have the following identification:

"take a measurement $\mathbf{M}_{\otimes_{n=1}^N C(\Omega)}(\otimes_{n=1}^N \mathbf{O}, S_{[\otimes_{n=1}^N \delta_{\omega_0}]})$" $\Leftrightarrow$ "take a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}^N, S_{[\delta_{\omega_0}]})$"

That is because it holds that

$${}_{\otimes_{n=1}^N \mathcal{M}(\Omega)}\Big\langle \otimes_{n=1}^N \delta_{\omega_0}, \otimes_{n=1}^N F(\Xi_n) \Big\rangle_{\otimes_{n=1}^N C(\Omega)} = {}_{\mathcal{M}(\Omega)}\Big\langle \delta_{\omega_0}, \times_{n=1}^N F(\Xi_n) \Big\rangle_{C(\Omega)}.$$

However, it should be noted that it does not always hold that "repeated measurement = iterated measurement" in statistical measurement theory (mentioned in Chapter 8) and quantum measurement theory.

■

Definition 2.27. [Semi-distance, moment method (inference for a pure state in repeated measurement)].

[(i): Semi-distance]. *Let Y be a set. If the map $\Delta : Y \times Y \to \mathbf{R}$ satisfies the following (a)~(d):*

(a): $\Delta(x, y) \geq 0$ $(\forall x, y \in Y)$, *(b):* "$x = y$" $\Rightarrow \Delta(x, y) = 0$,

(c): $\Delta(x, y) = \Delta(y, x)$ $(\forall x, y \in Y)$, *(d):* $\Delta(x, y) \leq \Delta(x.z) + \Delta(z, y)$ $(\forall x, y, z \in Y)$,

then, the Δ is called a semi-distance on Y. In addition, if "(b'): $x = y \Leftrightarrow \Delta(x, y) = 0$" is assumed, then the Δ is called a distance (or metric) on Y.

[(ii): Moment method]. *Assume the* ρ_0^p *(in* $\mathbf{M}_{\mathcal{A}}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[\rho_0^p]})$*) is unknown. And further, we get the sample space* $(X, \mathcal{F}, \nu_0)$ *from the measured value* $\widehat{x}$ $(= (x_1, x_2, ..., x_n) \in X^n)$ *obtained by* $\mathbf{M}_{\otimes\mathcal{A}}(\bigotimes_{k=1}^n \mathbf{O}, S_{[\otimes_{k=1}^n \rho_0^p]})$. *That is,* $\nu_0(\Xi) \approx \frac{\sharp[\{k : x_k \in \Xi\}]}{n}$. *Note, by (2.74), that* $\rho^p(F(\Xi)) \approx \nu_0(\Xi)$ $(\forall \Xi \in \mathcal{F})$. *Let* Δ *be a semi-distance on* $\mathcal{M}_{+1}^m(X)$.[11] *Then, there is a very reason to infer the unknown* ρ_0^p $(\in \mathfrak{S}^p(\mathcal{A}^*))$ *such that*

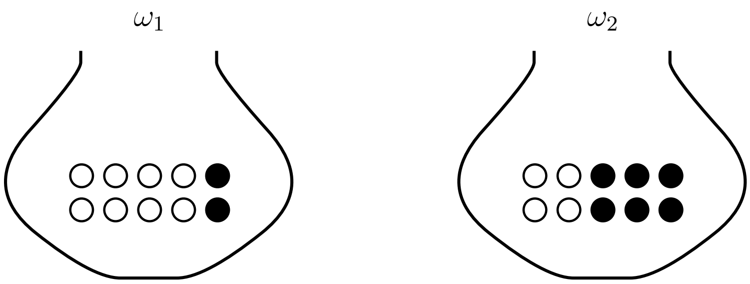
$$\Delta(\nu_0, \rho_0^p(F(\,\cdot\,))\,) = \min_{\rho^p \in \mathfrak{S}^p(\mathcal{A}^*)} \Delta(\nu_0, \rho^p(F(\,\cdot\,))\,).$$

This method is called "generalized moment method" or "moment method". Cf. §9.4. *Note that the "semi-distance* Δ *on* $\mathcal{M}_{+1}^m(X)$*" is not always unique. In this sense, the moment method is somewhat artificial.*

■

Example 2.28. [The urn problem by the moment method]. There are two urns ω_1 and ω_2. The urn ω_1 [resp. ω_2] contains 8 white and 2 black balls [resp. 4 white and 6 black balls]. Assume that they can not be distinguished in appearance. Choose one urn from the two. Assume that you do not know whether the chosen urn is ω_1 or ω_2. Now you sample, randomly, with replacement after each ball. In 7 samples, you get (w, b, b, w, b, w, b) in sequence where "w" = "white", "b" = "black".

(*Q*) *Which is the chosen urn,* ω_1 *or* ω_2*?*



[*Answer*]. We regard Ω $(\equiv \{\omega_1, \omega_2\})$ as the state space. And consider the observable $\mathbf{O}(\equiv (X \equiv \{w, b\}, 2^{\{w,b\}}, F))$ in $C(\Omega)$ where

$$[F(\{w\})](\omega_1) = 0.8, \qquad [F(\{b\})](\omega_1) = 0.2,$$
$$[F(\{w\})](\omega_2) = 0.4, \qquad [F(\{b\})](\omega_2) = 0.6.$$

Note that we have the real sample space $(X \equiv \{w, b\}, 2^{\{w,b\}}, \nu_0)$ such that:

$$\nu_0(\emptyset) = 0, \quad \nu_0(\{w\}) = 3/7, \quad \nu_0(\{b\}) = 4/7, \quad \nu_0(\{w, b\}) = 1.$$

[11] The definition of the semi-distance Δ may be too strong for the generalized moment method. However, in this book we focus on the above definition.

Also, note that the measurement

$$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_1}]}) \quad [\text{resp. } \mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_2}]})]$$

has the sample space

$$(X \equiv \{w, b\}, 2^{\{w,b\}}, [F(\,\cdot\,)](\omega_1)) \quad [\text{resp. } (X \equiv \{w, b\}, 2^{\{w,b\}}, [F(\,\cdot\,)](\omega_2))].$$

Thus, it suffices to compare

$$\Delta(\nu_0, [F(\,\cdot\,)](\omega_1)) \quad \text{and} \quad \Delta(\nu_0, [F(\,\cdot\,)](\omega_2)),$$

where Δ is a certain distance on $\mathcal{M}_{+1}^m(\{w, b\})$. For example define the distance Δ such that:

$$\Delta(\nu_1, \nu_2) = |\nu_1(\{w\}) - \nu_2(\{w\})| + |\nu_1(\{b\}) - \nu_2(\{b\})| \quad (\forall \nu_1, \nu_2 \in \mathcal{M}_{+1}^m(\{w, b\})).$$

Then, we see

$$\Delta(\nu_0, [F(\,\cdot\,)](\omega_1)) = |3/7 - 8/10| + |4/7 - 2/10| = 52/70$$

and

$$\Delta(\nu_0, [F(\,\cdot\,)](\omega_2)) = |3/7 - 4/10| + |4/7 - 6/10| = 10/70.$$

Thus, we can, by the moment method, infer that the unknown urn is ω_2.

■

2.9 Appendix (Bell's thought experiment)

(Continued from Example 2.15. Also see the footnote below[12])

2.9.1 EPR thought experiment

Although the original "EPR experiment (*cf.* [22])" was proposed in the framework of classical mechanics (*cf.* Chapter 12), the following argument is the quantum form of the "EPR experiment".[13]

---

[12] All appendixes in this book can be skipped.

[13] The argument in §2.9.1 is essentially the same as EPR-experiment (i.e., EPR-paradox, *cf.* [22]), which will be again discussed in §12.7.

Now consider the quantum system composed of two particles with the singlet state ρ_s (concerning z-axis) formulated in $B(\mathbf{C}^2 \bigotimes \mathbf{C}^2)$, where $\mathbf{C}^2 \bigotimes \mathbf{C}^2$ is the tensor Hilbert space of $\mathbf{C}^2$ and $\mathbf{C}^2$. The singlet state ρ_s is represented by $\rho_s = |\psi_s\rangle\langle\psi_s| \ \big(\in \mathfrak{S}^p(B(\mathbf{C}^2 \bigotimes \mathbf{C}^2)^*)\big)$, where

$$\psi_s = \frac{1}{\sqrt{2}}(\vec{e}_1 \otimes \vec{e}_2 - \vec{e}_2 \otimes \vec{e}_1) \quad (\in \mathbf{C}^2 \otimes \mathbf{C}^2). \quad \vec{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \in \mathbf{C}^2, \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \in \mathbf{C}^2. \tag{2.75}$$

And consider the measurement $\mathbf{M}_{B(\mathbf{C}^2)\otimes B(\mathbf{C}^2)}\big(\mathbf{O}^z \otimes \mathbf{O}^z \equiv (Z^2 = \{\uparrow_z, \downarrow_z\}^2, 2^{Z^2}, F^z \otimes F^z), S_{[\rho_s]}\big)$, where

$$F^z(\{\uparrow_z\}) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad F^z(\{\downarrow_z\}) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

$$F^z(\emptyset) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad F^z(\{\uparrow_z, \downarrow_z\}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Taking the measurement $\mathbf{M}_{B(\mathbf{C}^2)\otimes B(\mathbf{C}^2)}\big(\mathbf{O}^z \otimes \mathbf{O}^z \equiv (Z^2 = \{\uparrow_z, \downarrow_z\}^2, 2^{Z^2}, F^z \otimes F^z), S_{[\rho_s]}\big)$, we see that

(a) the probability that a measured value $(\uparrow_z, \uparrow_z)$ is obtained is equal to

$$\begin{aligned}
&=\rho_s\Big(F^z(\{\uparrow_z\}) \otimes F^z(\{\uparrow_z\})\Big) \\
&={}_{\mathbf{C}^2\otimes\mathbf{C}^2}\Big\langle \frac{1}{\sqrt{2}}(\vec{e}_1 \otimes \vec{e}_2 - \vec{e}_2 \otimes \vec{e}_1), [F^z(\{\uparrow_z\}) \otimes F^z(\{\uparrow_z\})] \frac{1}{\sqrt{2}}(\vec{e}_1 \otimes \vec{e}_2 - \vec{e}_2 \otimes \vec{e}_1) \Big\rangle_{\mathbf{C}^2\otimes\mathbf{C}^2} \\
&=0
\end{aligned}$$

(b) the probability that a measured value $(\uparrow_z, \downarrow_z)$ is obtained is equal to

$$\begin{aligned}
&=\rho_s\Big(F^z(\{\uparrow_z\}) \otimes F^z(\{\downarrow_z\})\Big) \\
&={}_{\mathbf{C}^2\otimes\mathbf{C}^2}\Big\langle \frac{1}{\sqrt{2}}(\vec{e}_1 \otimes \vec{e}_2 - \vec{e}_2 \otimes \vec{e}_1), [F^z(\{\uparrow_z\}) \otimes F^z(\{\downarrow_z\})] \frac{1}{\sqrt{2}}(\vec{e}_1 \otimes \vec{e}_2 - \vec{e}_2 \otimes \vec{e}_1) \Big\rangle_{\mathbf{C}^2\otimes\mathbf{C}^2} \\
&=1/2
\end{aligned}$$

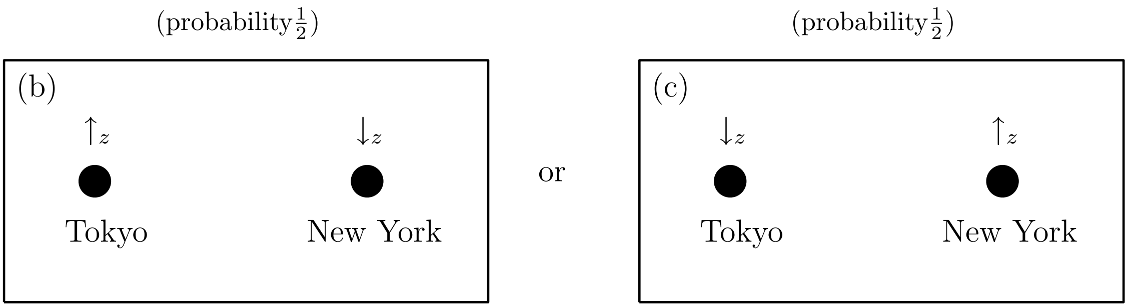
(c) the probability that a measured value $(\downarrow_z, \uparrow_z)$ is obtained is equal to

$$\begin{aligned}
&=\rho_s\Big(F^z(\{\downarrow_z\}) \otimes F^z(\{\uparrow_z\})\Big) \\
&={}_{\mathbf{C}^2\otimes\mathbf{C}^2}\Big\langle \frac{1}{\sqrt{2}}(\vec{e}_1 \otimes \vec{e}_2 - \vec{e}_2 \otimes \vec{e}_1), [F^z(\{\downarrow_z\}) \otimes F^z(\{\uparrow_z\})] \frac{1}{\sqrt{2}}(\vec{e}_1 \otimes \vec{e}_2 - \vec{e}_2 \otimes \vec{e}_1) \Big\rangle_{\mathbf{C}^2\otimes\mathbf{C}^2} \\
&=1/2
\end{aligned}$$

(d) the probability that a measured value $(\downarrow_z, \downarrow_z)$ is obtained is equal to

$$
\begin{aligned}
&= \rho_s\Big(F^z(\{\downarrow_z\}) \otimes F^z(\{\downarrow_z\}) \Big) \\
&= {}_{\mathbf{C}^2 \otimes \mathbf{C}^2}\Big\langle \frac{1}{\sqrt{2}}(\vec{e}_1 \otimes \vec{e}_2 - \vec{e}_2 \otimes \vec{e}_1), [F^z(\{\downarrow_z\}) \otimes F^z(\{\downarrow_z\})] \frac{1}{\sqrt{2}}(\vec{e}_1 \otimes \vec{e}_2 - \vec{e}_2 \otimes \vec{e}_1) \Big\rangle_{\mathbf{C}^2 \otimes \mathbf{C}^2} \\
&= 0.
\end{aligned}
$$

Here, it should be noted that we can assume that the x_1 and the x_2 (in $(x_1, x_2) \in \{ (\uparrow_z, \uparrow_z), (\uparrow_z, \downarrow_z), (\downarrow_z, \uparrow_z), (\downarrow_z, \downarrow_z)\}$) are respectively obtained in Tokyo and in New York (or, in the earth and in the polar star).



This fact is, figuratively speaking, explained as follows:

- Immediately after the particle in Tokyo is measured and the measured value $\uparrow_z$ [resp. $\downarrow_z$] is observed, the particle in Tokyo informs the particle in New York "Your measured value has to be $\downarrow_z$ [resp. $\uparrow_z$]".

Therefore, the above fact implies that quantum mechanics says that *there is something faster than light.* This is essentially the same as *the de Broglie paradox* (*cf.* [20]. Also see §9.3.3). That is,

- if we admit quantum mechanics, we must also admit the fact that there is something faster than light. (*cf.* [18, 78]). (2.76)

Of course we admit PMT, and therefore, we believe that there is something faster than light.

2.9.2 Bell's thought experiment

In this section, we review Bell's thought experiment in (quantum) measurement theory. (Cf. [9, 18, 78].) All the idea is, of course, owed to J.S. Bell [9]. Thus, we do not intend to assert our originality in this section. The argument is divided into two steps (i.e., [Step: I] and [Step: II]). [Step: I] is essentially the same as the previous section (i.e., §2.9.1).

[Step: I]. Let $a = (\alpha_1, \alpha_2)$ be any element in $\mathbf{R}^2$ such that $\|a\|_{\mathbf{R}^2} \equiv (|\alpha_1|^2 + |\alpha_2|^2)^{1/2} = 1$. Put

$$\sigma_a = \begin{bmatrix} 0 & \alpha_1 - \alpha_2\sqrt{-1} \\ \alpha_1 + \alpha_2\sqrt{-1} & 0 \end{bmatrix} \in B(\mathbf{C}^2), \quad \vec{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \in \mathbf{C}^2, \quad \vec{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \in \mathbf{C}^2.$$

It is easy to see that the self-adjoint matrix $\sigma_a : \mathbf{C}^2 \to \mathbf{C}^2$ has a unique spectral representation : $\sigma_a = F_a^{(1)} - F_a^{(-1)}$, where $F_a^{(1)}$ and $F_a^{(-1)}$ are orthogonal projections on $\mathbf{C}^2$ such that

$$F_a^{(1)} = \frac{1}{2}\begin{bmatrix} 1 & \alpha_1 - \alpha_2\sqrt{-1} \\ \alpha_1 + \alpha_2\sqrt{-1} & 1 \end{bmatrix}, \quad F_a^{(-1)} = \frac{1}{2}\begin{bmatrix} 1 & -\alpha_1 + \alpha_2\sqrt{-1} \\ -\alpha_1 - \alpha_2\sqrt{-1} & 1 \end{bmatrix}.$$

Define the observable $\mathbf{O}_a \equiv \big(X \equiv \{1, -1\}, \mathcal{P}(X), F_a\big)$ in $B(\mathbf{C}^2)$ such that $F_a(\{1\}) = F_a^{(1)}$ and $F_a(\{-1\}) = F_a^{(-1)}$

Now consider the quantum system composed of two particles with the singlet state ρ_s (concerning z-axis) formulated in $B(\mathbf{C}^2 \bigotimes \mathbf{C}^2)$, where $\mathbf{C}^2 \bigotimes \mathbf{C}^2$ is the tensor Hilbert space of $\mathbf{C}^2$ and $\mathbf{C}^2$. The singlet state ρ_s is represented by $\rho_s = |\psi_s\rangle\langle\psi_s| \ \big(\in \mathfrak{S}^p(B(\mathbf{C}^2 \bigotimes \mathbf{C}^2)^*)\big)$, where

$$\psi_s = \frac{1}{\sqrt{2}}(\vec{e}_1 \otimes \vec{e}_2 - \vec{e}_2 \otimes \vec{e}_1) \quad (\in \mathbf{C}^2 \otimes \mathbf{C}^2).$$

Put $a = (\alpha_1, \alpha_2)$, $b = (\beta_1, \beta_2) \in \mathbf{R}^2$ where $\|a\|_{\mathbf{R}^2} = \|b\|_{\mathbf{R}^2} = 1$. And define the tensor product observable $\mathbf{O}_{ab}$ ($\equiv \mathbf{O}_a \otimes \mathbf{O}_b) = (X^2, \mathcal{P}(X^2), F_a \bigotimes F_b)$ in $B(\mathbf{C}^2 \otimes \mathbf{C}^2)$ such that

$$(F_a \bigotimes F_b)(\{(x_1, x_2)\}) = F_a(\{x_1\}) \bigotimes F_b(\{x_2\}) \qquad (\forall (x_1, x_2) \in X^2 \equiv \{-1, 1\}^2).$$

Thus we get a measurement $\mathbf{M}_{B(\mathbf{C}^2 \otimes \mathbf{C}^2)}(\mathbf{O}_{ab}, S_{[\rho_s]})$ in $B(\mathbf{C}^2 \otimes \mathbf{C}^2)$. Axiom 1 says that the probability that a measured value x ($= (x_1, x_2)) \in X^2$ ($\equiv \{1, -1\}^2$) obtained by the measurement $\mathbf{M}_{B(\mathbf{C}^2 \otimes \mathbf{C}^2)}(\mathbf{O}_{ab}, S_{[\rho_s]})$ belongs to a set B ($\subseteq X^2$) is given by $\nu_{\text{EPR}}(B)$, where $\nu_{\text{EPR}}(B) = \sum_{x \equiv (x_1, x_2) \in B} \rho_s\big((F_a \otimes F_b)(\{(x_1, x_2)\})\big)$. Therefore, we see, for example, that

(♯) if we know that $x_1 = 1$, quantum mechanics says that the probability that $x_2 = 1$ [resp. $x_2 = -1$] is given by

$$\frac{\nu_{\text{EPR}}(\{1\} \times \{1\})}{\nu_{\text{EPR}}(\{1\} \times \{1, -1\})} \quad \Big[\text{resp.} \frac{\nu_{\text{EPR}}(\{1\} \times \{-1\})}{\nu_{\text{EPR}}(\{1\} \times \{1, -1\})}\Big]$$

and further, if we know that $x_1 = -1$, the probability that $x_2 = 1$ [resp. $x_2 = -1$] is given by

$$\frac{\nu_{\text{EPR}}(\{-1\} \times \{1\})}{\nu_{\text{EPR}}(\{-1\} \times \{1, -1\})} \quad \Big[\text{resp.} \frac{\nu_{\text{EPR}}(\{-1\} \times \{-1\})}{\nu_{\text{EPR}}(\{-1\} \times \{1, -1\})}\Big].$$

[Step: II]. Let $a^1(=(\alpha_1^1,\alpha_2^1))$, $a^2(=(\alpha_1^2,\alpha_2^2))$, $b^1(=(\beta_1^1,\beta_2^1))$ and $b^2(=(\beta_1^2,\beta_2^2))$ be elements in $\mathbf{R}^2$ such that $\|a^1\|_{\mathbf{R}^2}=\|a^2\|_{\mathbf{R}^2}=\|b^1\|_{\mathbf{R}^2}=\|b^2\|_{\mathbf{R}^2}=1$. Further, consider the parallel measurement $\bigotimes_{i,j=1,2}\mathbf{M}_{B(\mathbf{C}^2\otimes\mathbf{C}^2)}(\mathbf{O}_{a^ib^j},S_{[\rho_s]})$ in $\otimes_{i,j=1,2}B(\mathbf{C}^2\otimes\mathbf{C}^2)$ $(\equiv B(\otimes_{i,j=1,2}(\mathbf{C}^2\otimes\mathbf{C}^2)))$, that is,

$$\bigotimes_{i,j=1,2}\mathbf{M}_{B(\mathbf{C}^2\otimes\mathbf{C}^2)}(\mathbf{O}_{a^ib^j},S_{[\rho_s]})$$
$$=\mathbf{M}_{B(\otimes_{i,j=1,2}(\mathbf{C}^2\otimes\mathbf{C}^2))}\Big(\big(\mathop{\times}_{i,j=1,2}X^2,\mathcal{P}(\mathop{\times}_{i,j=1,2}X^2),\bigotimes_{i,j=1,2}(F_{a^i}\otimes F_{b^j})\big),S_{[\otimes_{i,j=1,2}\rho_s]}\Big).$$

Here note that $\otimes_{i,j=1,2}\rho_s=\rho_s\otimes\rho_s\otimes\rho_s\otimes\rho_s=|\psi_s\otimes\psi_s\otimes\psi_s\otimes\psi_s\rangle\langle\psi_s\otimes\psi_s\otimes\psi_s\otimes\psi_s|$ and

$$\mathop{\times}_{i,j=1,2}X^2\ni\big((x_1^{11},x_2^{11}),(x_1^{12},x_2^{12}),(x_1^{21},x_2^{21}),(x_1^{22},x_2^{22})\big)=x\in X^8\equiv\{-1,1\}^8.$$

Axiom 1 (2.37) says that the probability that a measured value $x\in X^8$ $(\equiv\{1,-1\}^8)$ obtained by the parallel measurement $\bigotimes_{i,j=1,2}\mathbf{M}_{B(\mathbf{C}^2\otimes\mathbf{C}^2)}(\mathbf{O}_{a^ib^j},S_{[\rho_s]})$ belongs to a set B $(\subseteq X^8)$ is given by $\nu_{\text{BTE}}(B)$, where $\nu_{\text{BTE}}(B)=\sum_{x\in B}\prod_{i,j=1,2}\rho_s\big((F_{a^i}\otimes F_{b^j})(\{(x_1^{ij},x_2^{ij})\})\big)$. That is, we have the sample space $(X^8,\mathcal{P}(X^8),\nu_{\text{BTE}})$, which is induced by the parallel measurement $\bigotimes_{i,j=1,2}\mathbf{M}_{B(\mathbf{C}^2\otimes\mathbf{C}^2)}\,(\mathbf{O}_{a^ib^j},\,S_{[\rho_s]})$.

Define the $\{-1,1\}$-valued functions g_k^{ij} on X^8, $(i,j,k=1,2)$, such that

$$g_k^{ij}((x_1^{11},x_2^{11}),(x_1^{12},x_2^{12}),(x_1^{21},x_2^{21}),(x_1^{22},x_2^{22}))=x_k^{ij}\qquad(\forall i,\forall j,\forall k\in\{1,2\}).\tag{2.77}$$

Note that it holds that

$$\nu_{\text{BTE}}\big((g_1^{11})^{-1}(\{1\})\big)=\nu_{\text{BTE}}\big((g_1^{12})^{-1}(\{1\})\big),\quad\nu_{\text{BTE}}\big((g_1^{21})^{-1}(\{1\})\big)=\nu_{\text{BTE}}\big((g_1^{22})^{-1}(\{1\})\big),$$
$$\nu_{\text{BTE}}\big((g_2^{11})^{-1}(\{1\})\big)=\nu_{\text{BTE}}\big((g_2^{12})^{-1}(\{1\})\big),\quad\nu_{\text{BTE}}\big((g_2^{21})^{-1}(\{1\})\big)=\nu_{\text{BTE}}\big((g_2^{22})^{-1}(\{1\})\big).$$

Here note that (*cf.* (3.42) in §3.7 later)

$$g_1^{11}\neq g_1^{12},\quad g_1^{21}\neq g_1^{22},\quad g_2^{11}\neq g_2^{21},\quad g_2^{12}\neq g_2^{22}.\tag{2.78}$$

Moreover, define the correlation functions $P(g_1^{ij},g_2^{ij})$ $(i,j=1,2)$ by

$$P(g_1^{ij},g_2^{ij})\equiv\int_{X^8}g_1^{ij}(x)\cdot g_2^{ij}(x)\nu_{\text{BTE}}(dx),\tag{2.79}$$

which may be also denoted by $P(a^i,b^j)$. A simple calculation shows that $P(a^i,b^j)=-(\alpha_1^i\beta_1^j+\alpha_2^i\beta_2^j)$. Thus, putting

$$a^1=(0,1),\quad b^1=(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}),\quad a^2=(1,0)\quad\text{ and }\quad b^2=(\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}}),$$

we see that

$$|P(a^1, b^1) - P(a^1, b^2)| \ + \ |P(a^2, b^1) + P(a^2, b^2)| = 2\sqrt{2}. \tag{2.80}$$

This is precisely Bell's calculation concerning Bell's thought experiment.

The (2.80) can be tested by the repeated measurement $\bigotimes_{k=1}^{K}\big(\bigotimes_{i,j=1,2} \mathbf{M}_{B(\mathbf{C}^2 \otimes \mathbf{C}^2)} (\mathbf{O}_{a^i b^j}, S_{[\rho_s]}) \big)$. Let $\widehat{x} = \{\big((x_{1,k}^{11}, x_{2,k}^{11}), (x_{1,k}^{12}, x_{2,k}^{12}), (x_{1,k}^{21}, x_{2,k}^{21}), (x_{1,k}^{22}, x_{2,k}^{22})\big)\}_{k=1}^{K}$ be a measured value of the repeated measurement. Then, we see that

$$P(a^i, b^j) \approx \frac{1}{K} \sum_{k=1}^{K} x_{1,k}^{ij} x_{2,k}^{ij}$$

for sufficiently large K. Thus, the experimental test: "$2\sqrt{2}$ or not?" is possible. In fact, Aspect's experiment [8] is generally believed to guarantee the (2.80). It is, of course, important since quantum mechanics must be always tested.

(Continued in §3.7 (Appendix(Bell's inequality)))

Chapter 3

The relation among systems (Axiom 2)

As mentioned in Chapter 1, (pure) measurement theory (PMT) is formulated as follows:

$$\text{PMT} = \underset{\text{[Axiom 1 (2.37)]}}{\text{measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} \quad \text{in } C^*\text{-algebra} . \qquad \underset{(=(1.4))}{(3.1)}$$

In Chapter 2 we studied "measurement (= Axiom 1)". In this chapter we intend to explain "the relation among systems (= Axiom 2)".

3.1 Newton Equation and Schrödinger equation

In this section, we review the Newton equation and Schrödinger equation.

[I: Newtonian Mechanics]

Put $\mathcal{A} = C_0(\mathbf{R}^s_q \times \mathbf{R}^s_p)$ and $\mathcal{A}^* = \mathcal{M}(\mathbf{R}^s_q \times \mathbf{R}^s_p)$, where $\mathbf{R}^s_q \times \mathbf{R}^s_p \equiv \{(q,p) = (q_1, q_2, \cdots, q_s, p_1, p_2, \cdots, p_s) \mid q_j, p_j \in \mathbf{R}, j = 1, 2, \cdots, s\}$ and $(\mathbf{R}^s_q \times \mathbf{R}^s_p)$ is the $2s$-dimensional space (*cf.* Example 2.2). It is well known that the Newton equation is mathematically equivalent to the following Hamilton equation:

$$\frac{d}{dt} q_j(t) = \frac{\partial \mathcal{H}}{\partial p_j}(q(t), p(t), t), \quad \frac{d}{dt} p_j(t) = -\frac{\partial \mathcal{H}}{\partial q_j}(q(t), p(t), t), \quad j = 1, 2, \cdots, s \qquad (3.2)$$

$$(q(0), p(0)) \in \mathbf{R}^s_q \times \mathbf{R}^s_p. \qquad (3.3)$$

where $\mathcal{H} : \mathbf{R}^s_q \times \mathbf{R}^s_p \times \mathbf{R} \to \mathbf{R}$ is a Hamiltonian. Using the solution of Newton equation (i.e., Hamilton equation (3.2)), we define the continuous map $\psi_{t_1,t_2} : \mathbf{R}^s_q \times \mathbf{R}^s_p \to \mathbf{R}^s_q \times \mathbf{R}^s_p$,

$\forall t_1 \le \forall t_2$, such that:

$$\psi_{t_1,t_2}(q(t_1), p(t_1)) = (q(t_2), p(t_2)) \qquad (\forall (q(t_1), p(t_1)) \in \mathbf{R}^s_q \times \mathbf{R}^s_p), \tag{3.4}$$

which is equivalent to (3.2).

Put $\Omega = \mathbf{R}^s_q \times \mathbf{R}^s_p$. Also, put $\Omega_t = \Omega$ ($\forall t \in \mathbf{R}$), and $\omega^0_0 = (q(0), p(0))$ ($\in \Omega_0$). Thus, the pair $[\omega^0_0, \{\psi_{t_1,t_2} : \Omega_{t_1} \to \Omega_{t_2}\}_{t_1 \le t_2}]$ can be considered to be equivalent to "(3.3)+(3.2)".

Using the continuous map $\psi_{t_1,t_2} : \Omega_{t_1} \to \Omega_{t_2}$ ($\forall t_1 \le \forall t_2$), we define the continuous linear operator $\Phi_{t_1,t_2} : C_0(\Omega_{t_2}) \to C_0(\Omega_{t_1})$ such that:

$$[\Phi_{t_1,t_2}(f_{t_2})](\omega_{t_1}) = f_{t_2}(\phi_{t_1,t_2}(\omega_{t_1})) \qquad (\forall f_{t_2} \in C_0(\Omega_{t_2}), \forall \omega_{t_1} \in \Omega_{t_1}).$$

And therefore, we can consider the following identifications:

$$\text{"(3.3)+(3.2)"} \Leftrightarrow [\omega^0_0, \{\psi_{t_1,t_2} : \Omega_{t_1} \to \Omega_{t_2}\}_{t_1 \le t_2}] \Leftrightarrow [\delta_{\omega^0_0}, \{\Phi_{t_1,t_2} : C_0(\Omega_{t_2}) \to C_0(\Omega_{t_1})\}_{t_1 \le t_2}]$$

where $\delta_{\omega^0_0}$ is the point measure at ω^0_0. The pair $[\delta_{\omega^0_0}, \{\Phi_{t_1,t_2} : C_0(\Omega_{t_2}) \to C_0(\Omega_{t_1})\}_{t_1 \le t_2}]$ will be called "general system" (*cf.* Definition 3.1), and will play an important role in our theory, that is, it is a special case of "the relation among systems" in (3.1).

[II:Quantum Mechanics in $\mathcal{C}(L^2(\mathbf{R}_q, dq))$]

We begin with the classical mechanics. For simplicity, consider the one dimensional case, i.e., $\mathbf{R}_q = \{q \mid q \in \mathbf{R}\}$. Thus $q(t)$, $-\infty < t < \infty$, means the particle's position at time t, and thus, $p(t)$ ($\equiv m\frac{dq(t)}{dt}$) means the particle's momentum at time t. Let $\mathbf{R}^2_{q,p}$ ($\equiv \{(q,p) \mid q, p \in \mathbf{R}\}$ be a phase space. Define a Hamiltonian $\mathcal{H} : \mathbf{R}^2_{q,p} \to \mathbf{R}$ such that:

$$\mathcal{H}(q,p) = \frac{p^2}{2m}\Big(=\text{kinetic energy}= \tfrac{1}{2}m(\tfrac{dq(t)}{dt})^2\Big) \quad + \quad V(q)\Big(=\text{potential energy}\Big). \tag{3.5}$$

Thus we see

$$\underset{\text{(total energy)}}{E} = \mathcal{H}(q,p) = \underset{\text{(kinetic energy)}}{\frac{p^2}{2m}} + \underset{\text{(potential energy)}}{V(q)}. \tag{3.6}$$

Put $H = L^2(\mathbf{R}_q, dq)$, that is, the Hilbert space composed of all complex valued L^2-functions f on $\mathbf{R}_q$, i.e., $\|f\|_{L^2(\mathbf{R}_q,dq)} \equiv [\int_{-\infty}^{\infty} |f(q)|^2 dq]^{1/2} < \infty$. And put $\mathcal{A} = \mathcal{C}(H) = \mathcal{C}(L^2(\mathbf{R}_q, dq))$, (i.e., the algebra composed of all compact operators on H, *cf.* Example 2.3). Applying the quantumization:

$$E \mapsto i\hbar\frac{\partial}{\partial t}, \quad p \mapsto -i\hbar\frac{\partial}{\partial q}, \quad q \mapsto q \quad (\text{where } i = \sqrt{-1},\ \hbar = \text{"Plank constant"} \ /2\pi) \tag{3.7}$$

to the (3.6), we obtain the Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t} = \mathcal{H}(q, -i\hbar\frac{\partial}{\partial q}) = -\frac{\hbar^2\partial^2}{2m\partial q^2} + V(q) \tag{3.8}$$

or precisely

$$i\hbar\frac{\partial}{\partial t}\psi(q,t) = -\frac{\hbar^2\partial^2}{2m\partial q^2}\psi(q,t) + V(q)\psi(q,t). \tag{3.9}$$

This solution is, formally, written by

$$\psi(q,t) = e^{-\frac{i}{\hbar}\mathcal{H}(q,-i\hbar\frac{\partial}{\partial q})t}\psi(q,0).$$

Put $U(t) = e^{-\frac{i}{\hbar}\mathcal{H}(q,-i\hbar\frac{\partial}{\partial q})t}$, and $\psi(\cdot,t) = \psi_t$. Then, we see,

$$\psi_t = U(t)\psi_0 \qquad (\|\psi_0\|_H = 1).$$

Thus, the time-evolution of the state $|\psi_t\rangle\langle\psi_t| \Big(\equiv (\Psi_t^0)^*(|\psi_0\rangle\langle\psi_0|)\Big)$ is represented by

$$|\psi_t\rangle\langle\psi_t| = (\Psi_t^0)^*\Big(|\psi_0\rangle\langle\psi_0|\Big) = |U(t)\psi_0\rangle\langle U(t)\psi_0| \quad \Big(\in Tr_{+1}^p(H)\Big).$$

Let $\Psi_t^0 : \mathcal{C}(H) \to \mathcal{C}(H)$ be the pre-adjoint operator of $(\Psi_t^0)^*$. Let $\mathbf{O}_0 = (X, \mathcal{F}, F_0)$ be a C^*-observable in $\mathcal{C}(H)$. Then, the time-evolution of the observable $\mathbf{O}_t = (X, \mathcal{F}, F_t)$ is represented by

$$(X, \mathcal{F}, F_t) = (X, \mathcal{F}, U(t)F_0U(t)^*) = (X, \mathcal{F}, \Psi_t^0 F_0).$$

Putting $\Phi_{t_1,t_2} = \Psi_{t_2-t_1}^0$, we get the pair $[|\psi_0\rangle\langle\psi_0|, \{\Phi_{t_1,t_2} : \mathcal{C}(H) \to \mathcal{C}(H)\}_{t_1\le t_2}]$. Also, it should be note that the above F_t is the solution of the following Heisenberg kinetic equation:

$$i\hbar\frac{dF_t}{dt} = F_t\mathcal{H} - \mathcal{H}F_t \qquad \text{in } \mathcal{C}(H)\ , \tag{3.10}$$

which is equivalent to the Schrödinger equation (3.9). (Cf. [84].) The pair $\Big[|\psi_0\rangle\langle\psi_0|, \{\Phi_{t_1,t_2} : \mathcal{C}(L^2(\mathbf{R}_q, dq)) \to \mathcal{C}(L^2(\mathbf{R}_q, dq))\}_{t_1\le t_2}\Big]$ will be called "general system" (*cf.* Definition 3.1), and will play an important role in our theory, that is, it is also a special case of "the relation among systems" in (3.1).

3.2 The relation among systems (Definition)

By the hint of the arguments in the previous section, we shall devote ourselves to "the relation among systems (i.e., Axiom 2)" in PMT (3.1) (=(1.4)).

Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be C^*-algebras. A continuous linear operator $\Psi_{1,2} : \mathcal{A}_2 \to \mathcal{A}_1$ is called a *Markov operator,* if it satisfies that

(i) $\Psi_{1,2}(F_2) \geq 0$ for any positive element F_2 in $\mathcal{A}_2$,

(ii) $\Psi_{1,2}(I_2) = I_1$, where I_k is the identity in $\mathcal{A}_k$ $(k = 1, 2)$.

Here note that, for any observable $(X, \mathcal{F}, F_2)$ in $\mathcal{A}_2$, the $(X, \mathcal{F}, \Psi_{1,2}F_2)$ is an observable in $\mathcal{A}_1$, which is denoted by $\Psi_{12}\mathbf{O}_2$. For example, it is easy to see that

$$\begin{aligned} &[\Psi_{1,2}F_2](\Xi \cup \Xi') = \Psi_{1,2}(F_2(\Xi \cup \Xi')) = \Psi_{1,2}(F_2(\Xi) + F_2(\Xi')) \\ =&[\Psi_{1,2}(F_2)](\Xi) + [\Psi_{1,2}(F_2)](\Xi') \quad (\text{for all } \Xi, \Xi'(\in \mathcal{F}) \text{ such that } \Xi \cap \Xi' = \emptyset). \end{aligned} \tag{3.11}$$

Also, a Markov operator $\Psi_{1,2} : \mathcal{A}_2 \to \mathcal{A}_1$ is called a *homomorphism* (or precisely, *C^*-homomorphism),* if it satisfies that

(i) $\Psi_{1,2}(F_2)\Psi_{1,2}(G_2) = \Psi_{1,2}(F_2G_2)$ for any F_2 and G_2 in $\mathcal{A}_2$,

(ii) $(\Psi_{1,2}(F_2))^* = \Psi_{1,2}(F_2^*)$ for any F_2 in $\mathcal{A}_2$.

Let $\Psi_{1,2}^* : \mathcal{A}_1^* \to \mathcal{A}_2^*$ be the dual operator[1] of a Markov operator $\Psi_{1,2} : \mathcal{A}_2 \to \mathcal{A}_1$, that is, it holds that

$${}_{\mathcal{A}_1^*}\Big\langle \rho_1, \Psi_{1.2}F_2 \Big\rangle_{\mathcal{A}_1} = {}_{\mathcal{A}_2^*}\Big\langle \Psi_{1.2}^*\rho_1, F_2 \Big\rangle_{\mathcal{A}_2} \quad (\forall \rho_1 \in \mathcal{A}_1^*, \forall F_2 \in \mathcal{A}_2). \tag{3.12}$$

Then the following mathematical results are well known (*cf.* [50, 76, 82]).

(a) $\Psi_{1,2}^*(\mathfrak{S}^m(\mathcal{A}_1^*)) \subseteq \mathfrak{S}^m(\mathcal{A}_2^*)$, (3.13)

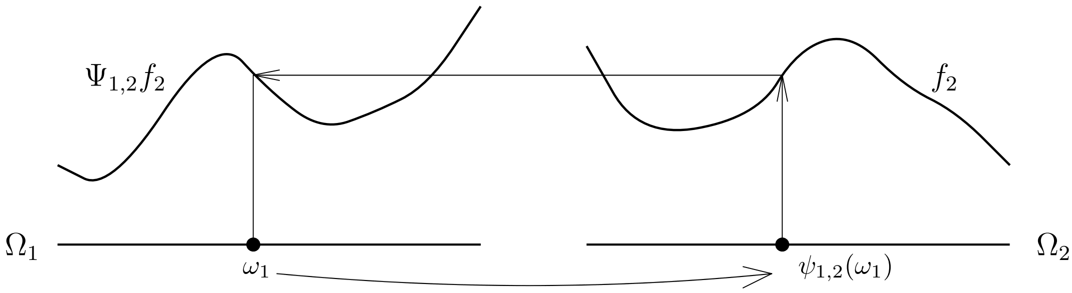
(b) $\Psi_{1,2}^*(\mathfrak{S}^p(\mathcal{A}_1^*)) \subseteq \mathfrak{S}^p(\mathcal{A}_2^*)$ if $\Psi_{1,2} : \mathcal{A}_2 \to \mathcal{A}_1$ is homomorphic.

Suppose that $\mathcal{A}_1$ and $\mathcal{A}_2$ are commutative unital C^*-algebras, i.e., $\mathcal{A}_1 = C(\Omega_1)$ and $\mathcal{A}_2 = C(\Omega_2)$. Then, under the identification that $\mathfrak{S}^p(\mathcal{A}_1^*) = \mathcal{M}_{+1}^p(\Omega_1) = \Omega_1$ and $\mathfrak{S}^m(\mathcal{A}_2^*) = \mathcal{M}_{+1}^m(\Omega_2)$ (*cf.* §2.1), the above (a) implies that the dual operator $\Psi_{1,2}^*$ of a Markov operator

[1] The symbol $*$ is used in the three following ways (i) ∼ (iii) in this book. (i) involution operator (e.g., F^*), (ii) dual operator (e.g., Ψ^*), (iii) dual space (e.g., $\mathcal{A}^*$).

Ψ_{12} can be identified with a *transition probability rule* $M(\omega_1, B_2)$, $(\omega_1 \in \Omega_1,\ B_2 \in \mathcal{B}_{\Omega_2})$, such that $M(\omega_1, B_2) = (\Psi^*_{1,2}(\delta_{\omega_1}))(B_2)$. Also, under the identification that $\mathcal{M}^p_{+1}(\Omega_1) = \Omega_1$ and $\mathcal{M}^p_{+1}(\Omega_2) = \Omega_2$, the above (b) implies that the dual operator $\Psi^*_{1,2}$ of a homomorphism $\Psi_{1,2}$ can be identified with a continuous map $\psi_{1,2}$ from Ω_1 into Ω_2 such that:

$$(\Psi_{1,2} f_2)(\omega_1) = f_2(\psi_{1,2}(\omega_1)) \qquad (\forall \omega_1 \in \Omega_1, \forall f_2 \in C(\Omega_2)). \tag{3.14}$$



Let $(T, \leq)$ be a tree-like partial ordered set, i.e., a partial ordered set such that "$t_1 \leq t_3$ and $t_2 \leq t_3$" implies "$t_1 \leq t_2$ or $t_2 \leq t_1$". Put $T^2_{\leq} = \{(t_1, t_2) \in T^2 : t_1 \leq t_2\}$. An element $t_0 \in T$ is called a *root* if $t_0 \leq t$ $(\forall t \in T)$ holds. Since we usually consider the subtree T_{t_0} $(\subseteq T)$ with the root t_0, we assume that the tree-like ordered set has a root. In this chapter, assume, for simplicity, that T is finite (though it is sometimes infinite in applications).

Definition 3.1. [Markov relation among systems, General systems, Sequential observable]. *The pair* $\mathbf{S}_{[\rho^p_{t_0}]} \equiv [S_{[\rho^p_{t_0}]}, \{\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2) \in T^2_{\leq}}]$ *is called a general system with an initial state* $\rho^p_{t_0}$ *if it satisfies the following conditions* (i)∼(iii).

(i) *With each* t $(\in T)$, *a* C^**-algebra* $\mathcal{A}_t$ *is associated.*

(ii) *Let* t_0 $(\in T)$ *be the root of* T. *And, assume that a system* S *has the state* $\rho^p_{t_0}$ $(\in \mathfrak{S}^p(\mathcal{A}^*_{t_0}))$ *at* t_0, *that is, the initial state is equal to* $\rho^p_{t_0}$.

(iii) *For every* $(t_1, t_2) \in T^2_{\leq}$, *a Markov operator* $\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}$ *is defined such that* $\Phi_{t_1,t_2}\Phi_{t_2,t_3} = \Phi_{t_1,t_3}$ *holds for all* (t_1, t_2), $(t_2, t_3) \in T^2_{\leq}$.

The family $\{\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2) \in T^2_{\leq}}$ *is also called a "Markov relation among systems". Let an observable* $\mathbf{O}_t \equiv (X_t, 2^{X_t}, F_t)$ *in a* C^**-algebra* $\mathcal{A}_t$ *be given for each* $t \in T$. *The pair* $[\{\mathbf{O}_t\}_{t \in T}, \{\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2) \in T^2_{\leq}}]$ *is called a "sequential observable", which is* *denoted by* $[\mathbf{O}_T]$, *i.e.*, $[\mathbf{O}_T] = [\{\mathbf{O}_t\}_{t\in T}, \{\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2)\in T^2_{\leq}}\ ]$.

■

3.3 Examples (Several tree structures)

Before we propose Axiom 2 (3.26), we prepare some notations and examples. For simplicity, assume that T is finite, or a finite subtree of a whole tree. Let T ($= \{0,1,...,N\}$) be a tree with the root 0. Define the *parent map* $\pi : T\setminus\{0\} \to T$ such that $\pi(t) = \max\{s \in T : s < t\}$. It is clear that the tree $(T \equiv \{0,1,...,N\}, \leq\ )$ can be identified with the pair $(T \equiv \{0,1,...,N\}, \pi : T\setminus\{0\} \to T)$. Also, note that, for any $t \in T\setminus\{0\}$, there uniquely exists a natural number $h(t)$ (called the *height* of t) such that $\pi^{h(t)}(t) = 0$. Here, $\pi^2(t) = \pi(\pi(t))$, $\pi^3(t) = \pi(\pi^2(t))$, etc. Also, put $\{0,1,...,N\}^2_{\leq} = \{(m,n) \mid 0 \leq m \leq n \leq N\}$. Thus, the general system $\mathbf{S}_{[\rho^p_0]} \equiv [S^0_{[\rho^p_0]}, \{\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2)\in\{0,1,...,N\}^2_{\leq}}]$ is sometimes represented by $[S^0_{[\rho^p_0]}, \mathcal{A}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{A}_{\pi(t)}$ ($t \in \{0,1,...,N\}\setminus\{0\})]$. Let $\mathbf{O}_t \equiv (X_t, \mathcal{F}_t, F_t)$ be an observable in $\mathcal{A}_t$ ($\forall t \in T$). The "measurement" of $\{\mathbf{O}_t : t \in T\}$ for the $\mathbf{S}_{[\rho^p_{t_0}]}$ is symbolically described by $\mathfrak{M}(\{\mathbf{O}_t\}_{t\in T}, \mathbf{S}_{[\rho^p_{t_0}]})$. The Markov relation $\{\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2)\in T^2_{\leq}}$ is also denoted by $\{\mathcal{A}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{A}_{\pi(t)}\}_{t\in T\setminus\{0\}}$

The following Examples 3.2, 3.3 and 3.4 will promote the understanding of Axiom 2 later.

Example 3.2. [Series structures[2]]. Suppose that a tree $(T \equiv \{0,1,...,N\}, \pi)$ has a "series" structure, i.e., $\pi(t) = t-1$ ($\forall t \in T\setminus\{0\}$). Consider a general system $\mathbf{S}_{[\rho^p_0]} \equiv [S^0_{[\rho^p_0]}, \mathcal{A}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{A}_{\pi(t)}$ ($t \in T\setminus\{0\})]$ with the initial system $S^0_{[\rho^p_0]}$, that is,

$$\mathcal{A}_0 \overset{\Phi_{0,1}}{\longleftarrow} \mathcal{A}_1 \overset{\Phi_{1,2}}{\longleftarrow} \mathcal{A}_2 \overset{\Phi_{2,3}}{\longleftarrow} \cdots\cdots\cdots \overset{\Phi_{N-2,N-1}}{\longleftarrow} \mathcal{A}_{N-1} \overset{\Phi_{N-1,N}}{\longleftarrow} \mathcal{A}_N. \tag{3.15}$$

For each $t \in T$, consider an observable $\mathbf{O}_t \equiv (X_t, \mathcal{F}_t, F_t)$ in a C^*-algebra $\mathcal{A}_t$. Thus, we have a sequential observable $[\{\mathbf{O}_t\}_{t\in T}, \{\Phi_{t,\pi(t)} : \mathcal{A}_t \to \mathcal{A}_{\pi(t)}\}_{t\in T\setminus\{0\}}\ ]$. Put $\widetilde{\mathbf{O}}_N$ ($\equiv (X_N, \mathcal{F}_N, \widetilde{F}_N)) = \mathbf{O}_N$ ($\equiv (X_N, \mathcal{F}_N, F_N))$. According to the Heisenberg picture (*cf.* §3.5), the observable $\mathbf{O}_N$ in $\mathcal{A}_N$ can be identified with the observable $\Phi_{N-1,N}\widetilde{\mathbf{O}}_N$ in $\mathcal{A}_{N-1}$. Thus, we can consider the quasi-product observable $\widetilde{\mathbf{O}}_{N-1} \equiv \mathbf{O}_{N-1} \overset{\text{qp}}{\times} \Phi_{N-1,N}\mathbf{O}_N \equiv (X_{N-1} \times$

[2]Most problems in dynamical system theory are formulated as the general systems with series trees (i.e., T="time") Cf. Kalman filter in §8.4.

$X_N, \mathcal{F}_{N-1} \times \mathcal{F}_n, \widetilde{F}_{N-1})$ in $\mathcal{A}_{N-1}$, that is,

$$\widetilde{F}_{N-1}(\Xi_{N-1} \times \Xi_N) = \big(F_{N-1} \overset{\text{qp}}{\times} (\Phi_{N-1,N} F_N)\big)(\Xi_{N-1} \times \Xi_N), \tag{3.16}$$

(though the existence and the uniqueness are not guaranteed in general). By a similar way, we can define the quasi-product observable $\widetilde{\mathbf{O}}_{N-2} \equiv \mathbf{O}_{N-2} \overset{\text{qp}}{\times} \Phi_{N-2,N-1} \widetilde{\mathbf{O}}_{N-1} \equiv (X_{N-2} \times X_{N-1} \times X_N, \mathcal{F}_{N-2} \times \mathcal{F}_{N-1} \times \mathcal{F}_n, \widetilde{F}_{N-2})$ in $\mathcal{A}_{N-2}$, that is,

$$\widetilde{F}_{N-2}(\Xi_{N-2} \times \Xi_{N-1} \times \Xi_N) = \big(F_{N-2} \overset{\text{qp}}{\times} (\Phi_{N-2,N-1} \widetilde{F}_{N-1})\big)(\Xi_{N-2} \times \Xi_{N-1} \times \Xi_N). \tag{3.17}$$

Iteratively we get as follows:

$$\begin{array}{ccccccccccc}
[\mathcal{A}_0] & \xleftarrow{\Phi} & [\mathcal{A}_1] & \xleftarrow{\Phi} \cdots \xleftarrow{\Phi} & [\mathcal{A}_{N-2}] & \xleftarrow{\Phi} & [\mathcal{A}_{N-1}] & \xleftarrow{\Phi} & [\mathcal{A}_N] \\
F_0 & & F_1 & \cdots & F_{N-2} & & F_{N-1} & & F_N \\
\downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\
\begin{smallmatrix}(F_0 \overset{\text{qp}}{\times} \Phi \widetilde{F}_1) \\ = \widetilde{F}_0\end{smallmatrix} & \xleftarrow{\Phi} & \begin{smallmatrix}(F_1 \overset{\text{qp}}{\times} \Phi \widetilde{F}_2) \\ = \widetilde{F}_1\end{smallmatrix} & \xleftarrow{\Phi} \cdots \xleftarrow{\Phi} & \begin{smallmatrix}(F_{N-2} \overset{\text{qp}}{\times} \Phi \widetilde{F}_{N-1}) \\ = \widetilde{F}_{N-2}\end{smallmatrix} & \xleftarrow{\Phi} & \begin{smallmatrix}(F_{N-1} \overset{\text{qp}}{\times} \Phi \widetilde{F}_N) \\ = \widetilde{F}_{N-1}\end{smallmatrix} & \xleftarrow{\Phi} & \begin{smallmatrix}(F_N) \\ = \widetilde{F}_N\end{smallmatrix}
\end{array}$$

And finally, we get the quasi-product observable $\widetilde{\mathbf{O}}_0 \equiv \mathbf{O}_0 \overset{\text{qp}}{\times} \Phi_{0,1} \widetilde{\mathbf{O}}_1 \equiv (\times_{t=0}^N X_t, \times_{t=0}^N \mathcal{F}_t, \widetilde{F}_0)$ in $\mathcal{A}_0$, that is,

$$\widetilde{F}_0(\Xi_0 \times \Xi_1 \times \Xi_2 \times \cdots \times \Xi_N) = \big(F_0 \overset{\text{qp}}{\times} (\Phi_{0,1} \widetilde{F}_1)\big)(\Xi_0 \times \Xi_1 \times \Xi_2 \times \cdots \times \Xi_N). \tag{3.18}$$

Here $\widetilde{\mathbf{O}}_0$ is a realization of the sequential observable $[\{\mathbf{O}_t\}_{t \in T}, \{\Phi_{t,\pi(t)} : \mathcal{A}_t \to \mathcal{A}_{\pi(t)}\}_{t \in T \setminus \{0\}}$]. Then, we have the "measurement" $\mathfrak{M}(\{\mathbf{O}_t\}_{t \in T}, \mathbf{S}_{[\rho_0^p]})$ such as

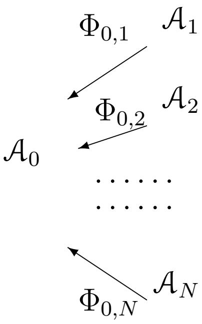
$$\mathfrak{M}(\{\mathbf{O}_t\}_{t \in T}, \mathbf{S}_{[\rho_0^p]}) = \mathbf{M}_{\mathcal{A}_0}(\widetilde{\mathbf{O}}_0 \equiv (\mathop{\times}_{t \in T} X_t, \mathop{\times}_{t \in T} \mathcal{F}_t, \widetilde{F}_0), S^0_{[\rho_0^p]}). \tag{3.19}$$

Also, note that the above arguments can be executed under the hypothesis that quasi-product observables (i.e., $\widetilde{\mathbf{O}}_n$, $n = 0, 1., , , .N$) exist. In other words, the existence of the "measurement" $\mathfrak{M}(\{\mathbf{O}\}_{t \in T}, \mathbf{S}_{[\rho_0^p]})$ is equivalent to that of the observable $\widetilde{\mathbf{O}}_0$.

■

Example 3.3. [Parallel structures[3]]. Suppose that a tree $(T \equiv \{0, 1, ..., N\}, \pi)$ has a "parallel" structure, i.e., $\pi(t) = 0$ $(\forall t \in T \setminus \{0\})$. Consider a general system $\mathbf{S}_{[\rho_0^p]} \equiv [S^0_{[\rho_0^p]}, \mathcal{A}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{A}_{\pi(t)}$ $(t \in T \setminus \{0\})]$ with the initial system $S^0_{[\rho_0^p]}$, that is,

[3]Most problems in statistics are formulated as the general systems with parallel trees. Cf. Figure (6.12) in regression analysis.



(3.20)

For each $t \in T$, consider an observable $\mathbf{O}_t \equiv (X_t, \mathcal{F}_t, F_t)$ in a C^*-algebra $\mathcal{A}_t$. Thus, we have a sequential observable $[\{\mathbf{O}_t\}_{t\in T}, \{\Phi_{t,\pi(t)} : \mathcal{A}_t \to \mathcal{A}_{\pi(t)}\}_{t\in T\setminus\{0\}}\ ]$. Then, we get the quasi-product observable $\widetilde{\mathbf{O}}_0 \equiv (\times_{t=0}^N X_t,\ \times_{t=0}^N \mathcal{F}_t,\ \widetilde{F}_0)$ in $\mathcal{A}_0$ such that:

$$\widetilde{F}_0(\Xi_0 \times \Xi_1 \times \Xi_2 \times \cdots \times \Xi_N) = \big(\overset{\text{qp}}{\underset{t\in T}{\times}} \Phi_{0,t}F_t)\big)(\Xi_0 \times \Xi_1 \times \Xi_2 \times \cdots \times \Xi_N). \tag{3.21}$$

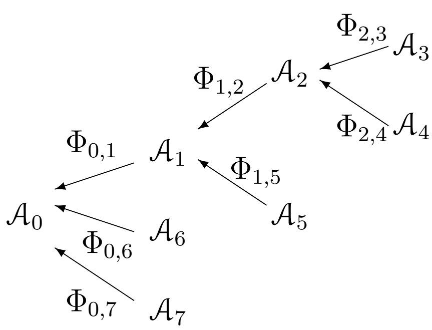
Here $\widetilde{\mathbf{O}}_0$ is a realization of the sequential observable $[\{\mathbf{O}_t\}_{t\in T}, \{\Phi_{t,\pi(t)} : \mathcal{A}_t \to \mathcal{A}_{\pi(t)}\}_{t\in T\setminus\{0\}}$]. Then, we have the "measurement" $\mathfrak{M}(\{\mathbf{O}_t\}_{t\in T},\ \mathbf{S}_{[\rho_0^p]})$ such as

$$\mathfrak{M}(\{\mathbf{O}_t\}_{t\in T}, \mathbf{S}_{[\rho_0^p]}) = \mathbf{M}_{\mathcal{A}_0}(\widetilde{\mathbf{O}}_0 \equiv (\underset{t\in T}{\times} X_t, \underset{t\in T}{\times} \mathcal{F}_t, \widetilde{F}_0), S^0_{[\rho_0^p]}). \tag{3.22}$$

Also, note that the above arguments can be executed under the hypothesis that quasi-product observables exist. In other words, the existence of the "measurement" $\mathfrak{M}(\{\mathbf{O}\}_{t\in T}$, $\mathbf{S}_{[\rho_0^p]})$ is equivalent to that of the observable $\widetilde{\mathbf{O}}_0$.

■

Example 3.4. [A simple general system, Heisenberg picture]. Suppose that a tree $(T \equiv \{0, 1, ..., 6, 7\}, \pi)$ has an ordered structure such that $\pi(1) = \pi(6) = \pi(7) = 0$, $\pi(2) = \pi(5) = 1$, $\pi(3) = \pi(4) = 2$. $\big($See the figure (3.23).$\big)$ Consider a general system $\mathbf{S}_{[\rho_0^p]} \equiv [S_{[\rho_0^p]},\ \{\mathcal{A}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{A}_{\pi(t)}\}_{t\in T\setminus\{0\}}]$ with the initial system $S_{[\rho_0^p]}$.



(3.23)

Also, for each $t \in \{0, 1, ..., 6, 7\}$, consider an observable $\mathbf{O}_t \equiv (X_t, 2^{X_t}, F_t)$ in a C^*-algebra $\mathcal{A}_t$. Thus, we have a sequential observable $[\{\mathbf{O}_t\}_{t\in T}, \{\Phi_{t,\pi(t)} : \mathcal{A}_t \to \mathcal{A}_{\pi(t)}\}_{t\in T\setminus\{0\}}\ ]$. Now we want to consider the following "measurement",

($\sharp$) for a system $S_{[\rho_0^p]}$, take a measurement of *"a sequential observable* $[\{\mathbf{O}_t\}_{t\in T}, \{\mathcal{A}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{A}_{\pi(t)}\}_{t\in T\setminus\{0\}}]$," i.e., take a measurement of an observable $\mathbf{O}_0$ at $0(\in T)$, and next, take a measurement of an observable $\mathbf{O}_1$ at $1(\in T)$, $\cdots\cdots$, and finally take a measurement of an observable $\mathbf{O}_7$ at $7(\in T)$,

which is symbolized by $\mathfrak{M}(\{\mathbf{O}_t\}_{t\in T}, \mathbf{S}_{[\rho_0^p]})$. Note that the $\mathfrak{M}(\{\mathbf{O}_t\}_{t\in T}, \mathbf{S}_{[\rho_0^p]})$ is merely a symbol since only one measurement is permitted (*cf.* §2.5 Remark(II)). In what follows let us describe the above ($\sharp$) ($= \mathfrak{M}(\{\mathbf{O}_t\}_{t\in T}, \mathbf{S}_{[\rho_0^p]})$) precisely. Put

$$\widetilde{\mathbf{O}}_t = \mathbf{O}_t \quad \text{and thus} \quad \widetilde{F}_t = F_t \quad (t = 3, 4, 5, 6, 7).$$

First we construct the quasi-product observable $\widetilde{\mathbf{O}}_2$ in $\mathcal{A}_2$ such as

$$\widetilde{\mathbf{O}}_2 = (X_2 \times X_3 \times X_4, 2^{X_2\times X_3\times X_4}, \widetilde{F}_2) \quad \text{where } \widetilde{F}_2 = F_2 \overset{\text{qp}}{\times} (\overset{\text{qp}}{\times}_{t=3,4} \Phi_{2,t}\widetilde{F}_t),$$

if it exists. Iteratively, we construct the following:

$$\begin{array}{ccccc}
\mathcal{A}_0 & \overset{\Phi_{0,1}}{\longleftarrow} & \mathcal{A}_1 & \overset{\Phi_{1,2}}{\longleftarrow} & \mathcal{A}_2 \\
F_0 \overset{\text{qp}}{\times} \Phi_{0,6}\widetilde{F}_6 \overset{\text{qp}}{\times} \Phi_{0,7}\widetilde{F}_7 & & F_1 \overset{\text{qp}}{\times} \Phi_{1,5}\widetilde{F}_5 & & \\
\downarrow & & \downarrow & & \\
\widetilde{F}_0 & \overset{\Phi_{0,1}}{\longleftarrow} & \widetilde{F}_1 & \overset{\Phi_{1,2}}{\longleftarrow} & \widetilde{F}_2 \\
(F_0 \overset{\text{qp}}{\times} \Phi_{0,6}\widetilde{F}_6 \overset{\text{qp}}{\times} \Phi_{0,7}\widetilde{F}_7 \overset{\text{qp}}{\times} \Phi_{0,1}\widetilde{F}_1) & & (F_1 \overset{\text{qp}}{\times} \Phi_{1,5}\widetilde{F}_5 \overset{\text{qp}}{\times} \Phi_{1,2}\widetilde{F}_2) & & (F_2 \overset{\text{qp}}{\times} \Phi_{2,3}\widetilde{F}_3 \overset{\text{qp}}{\times} \Phi_{2,4}\widetilde{F}_4)
\end{array} \qquad (3.24)$$

That is, we get the quasi-product observable $\widetilde{\mathbf{O}}_1 \equiv (\prod_{t=1}^5 X_t, 2^{\prod_{t=1}^5 X_t}, \widetilde{F}_1)$ of $\mathbf{O}_1$, $\Phi_{1,2}\widetilde{\mathbf{O}}_2$ and $\Phi_{1,5}\widetilde{\mathbf{O}}_5$, and finally, the quasi-product observable $\widetilde{\mathbf{O}}_0 \equiv (\prod_{t=0}^7 X_t, 2^{\prod_{t=0}^7 X_t}, \widetilde{F}_0)$ of $\mathbf{O}_0$, $\Phi_{0,1}\widetilde{\mathbf{O}}_1$, $\Phi_{0,6}\widetilde{\mathbf{O}}_6$ and $\Phi_{0,7}\widetilde{\mathbf{O}}_7$, if it exists. Here, $\widetilde{\mathbf{O}}_0$ is called *the realization* (or, *the Heisenberg picture representation*) of *a sequential observable* $[\{\mathbf{O}_t\}_{t\in T}, \{\mathcal{A}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{A}_{\pi(t)}\}_{t\in T\setminus\{0\}}]$. Then, we have the measurement

$$\mathbf{M}_{\mathcal{A}_0}(\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, 2^{\prod_{t\in T} X_t}, \widetilde{F}_0), S_{[\rho_0^p]}),$$

which is called *the realization* (or, *the Heisenberg picture representation) of the symbol* $\mathfrak{M}(\{\mathbf{O}_t\}_{t\in T}, \mathbf{S}_{[\rho_{t_0}^p]})$.

■

Remark 3.5. Let $(T \equiv \{0,1,...,N\}, \pi : T\setminus\{0\} \to T)$ be any tree with the root 0. Let τ be any element of T. Consider a series structure $\widetilde{T}_\tau$ such that $\widetilde{T}_\tau = \{\pi^k(\tau) \mid k = 0,1,2,...,h(\tau)\}$ ($\subseteq T$), where $h(\tau)$ is the height of τ, i.e., $\pi^{h(\tau)}(\tau) = 0$. Note that Example 3.4 (i.e, diagram (3.24)) means that any general system (with a tree structure T) can be regarded as a general system with a series structure $\widetilde{T}_\tau$.

■

3.4 The relation among systems (Axiom 2)

Examining Example 3.4, we see as follows: Let $(T \equiv \{0,1,...,N\}, \pi : T\setminus\{0\} \to T)$ be a tree with root 0 and let $\mathbf{S}_{[\rho_0^p]} \equiv [S_{[\rho_0^p]}, \mathcal{A}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{A}_{\pi(t)}\ (t \in T\setminus\{0\})]$ be a general system with the initial system $S_{[\rho_0^p]}$. And, let an observable $\mathbf{O}_t \equiv (X_t, \mathcal{F}_t, F_t)$ in a C^*-algebra $\mathcal{A}_t$ be given for each $t \in T$. Thus, we have a sequential observable $[\{\mathbf{O}_t\}_{t\in T}, \{\Phi_{t,\pi(t)} : \mathcal{A}_t \to \mathcal{A}_{\pi(t)}\}_{t\in T\setminus\{0\}}\ ]$. For each s ($\in T$), define the observable $\widetilde{\mathbf{O}}_s \equiv (\prod_{t\in T_s} X_t, \prod_{t\in T_s} \mathcal{F}_t, \widetilde{F}_s)$ in $\mathcal{A}_s$ such that:

$$\widetilde{\mathbf{O}}_s = \begin{cases} \mathbf{O}_s & (\text{if } s \in T\setminus\pi(T)) \\ \mathbf{O}_s \overset{\text{qp}}{\times} (\overset{\text{qp}}{\times}_{t\in\pi^{-1}(\{s\})} \Phi_{\pi(t),t}\widetilde{\mathbf{O}}_t) & (\text{if } s \in \pi(T)) \end{cases} \tag{3.25}$$

if possible. Then, if an observable $\widetilde{\mathbf{O}}_0$ (i.e., the Heisenberg picture representation of the sequential observable $[\{\mathbf{O}_t\}_{t\in T}, \{\Phi_{t,\pi(t)} : \mathcal{A}_t \to \mathcal{A}_{\pi(t)}\}_{t\in T\setminus\{0\}}\ ])$ in $\mathcal{A}_0$ exists (such as in Example 3.4), we have the measurement

$$\mathbf{M}_{\mathcal{A}_0}(\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, \prod_{t\in T} \mathcal{F}_t, \widetilde{F}_0), S_{[\rho_0^p]}),$$

which is called the Heisenberg picture representation of the symbol $\mathfrak{M}(\{\mathbf{O}_t\}_{t\in T}, \mathbf{S}_{[\rho_{t_0}^p]})$.

Summing up the essential part of the above argument, we can propose the following axiom, which corresponds to "the rule of the relation among systems" in PMT (1.4). Cf. [43, 44, 46].

AXIOM 2. [The Markov relation among systems, the Heisenberg picture] *The relation among systems is represented by a Markov relation* $\{\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2)\in T^2_{\leqq}}$. *Let* $\mathbf{O}_t$ ($\equiv (X_t, \mathcal{F}_t, F_t)$) *be an observable in* $\mathcal{A}_t$ *for each* t ($\in T$). *If the procedure (3.25) is possible, a sequential observable* $[\mathbf{O}_T] \equiv [\{\mathbf{O}_t\}_{t\in T}, \{\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2)\in T^2_{\leqq}}]$ *can be realized as the observable*

$$\widetilde{\mathbf{O}}_0 \equiv (\textstyle\prod_{t\in T} X_t, \prod_{t\in T} \mathcal{F}_t, \widetilde{F}_0) \text{ in } \mathcal{A}_0. \tag{3.26}$$

It is quite important to note that Axiom 2 is stated in terms of $\mathcal{A}$ (and not in terms of $\mathcal{A}^*$).[4] Also, we must add the following statement:

- Let $\mathbf{S}_{[\rho^p_{t_0}]} \equiv [S_{[\rho^p_{t_0}]}, \{\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2)\in T^2_{\leqq}}]$ be a general system with an initial state $\rho^p_{t_0}$ ($\in \mathfrak{S}^p(\mathcal{A}^*_{t_0})$). Then, a measurement represented by the symbol $\mathfrak{M}(\{\mathbf{O}_t\}_{t\in T}, \mathbf{S}_{[\rho^p_{t_0}]})$ can be realized by $\mathbf{M}_{\mathcal{A}_0}(\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, \prod_{t\in T} \mathcal{F}_t, \widetilde{F}_0), S_{[\rho^p_0]})$, if $\widetilde{\mathbf{O}}_0$ exists.

which explains the relation between Axiom 1 and Axiom 2.

Now we get the PMT (1.4). We have the following classification in PMT:

$$\begin{cases} \text{deterministic PMT} = \underset{\text{[Axiom 1 (2.37)]}}{\text{“measurement”}} + \underset{\text{[each } \Phi_{t_1,t_2} \text{ is homomorphic in Axiom 2 (3.26)]}}{\text{“the deterministic relation among systems”}}. \\ \\ \text{stochastic PMT} = \underset{\text{[Axiom 1 (2.37)]}}{\text{“measurement”}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{“the Markov relation among systems”}}. \end{cases} \tag{3.27}$$

Remark 3.6. (i). Roughly speaking, Axiom 2 asserts $\Phi_{0,1}\mathbf{O}_1$ is more fundamental than $\mathbf{O}_1$ in the following identification

$$\Phi_{0,1}\mathbf{O}_1 \text{ (in } \mathcal{A}_0) \longleftrightarrow \mathbf{O}_1 \text{ (in } \mathcal{A}_1)$$

where $\mathbf{O}_1$ is an observable in $\mathcal{A}_1$ and $\Phi_{0,1} : \mathcal{A}_1 \to \mathcal{A}_0$ is a Markov operator.
(ii). Also, it should be noted that Axiom 2 says that the time evolution of a system satisfies the Markov property. Thus, automata theory and circuit theory are characterized as special cases of measurement theory (especially, Axiom 2).
(iii). Axiom 2 has a great descriptive power. Note that “hysteresis” and “multiple Markov properties” can be described in the framework of Axiom 2.

■

[4] This fact makes us apply Axiom 2 to “statistical measurement theory” (in Chapter 8) as well as “PMT” (in this chapter).

3.5 Heisenberg picture and Schrödinger picture

Now let us mention something about the relation between Heisenberg picture and Schrödinger picture.

Suppose that a simplest tree $(T \equiv \{0,1\}, \pi)$ has a "series" structure, i.e., $\pi(1) = 0$. Consider a general system $\mathbf{S}_{[\rho_0^p]} \equiv [S_{[\rho_0^p]}, \mathcal{A}_1 \overset{\Phi_{0,1}}{\to} \mathcal{A}_0]$ with the initial system $S_{[\rho_0^p]}$, that is,

$$\mathcal{A}_0 \overset{\Phi_{0,1}}{\longleftarrow} \mathcal{A}_1 \tag{3.28}$$

Let $\mathbf{O}_1 = (X_1, \mathcal{F}_1, F_1)$ be an observable in $\mathcal{A}_1$. Now we consider

(M) the measurement of the observable $\mathbf{O}_1 = (X_1, \mathcal{F}_1, F_1)$ for the general system $\mathbf{S}_{[\rho_0^p]} \equiv [S_{[\rho_0^p]}, \mathcal{A}_1 \overset{\Phi_{0,1}}{\to} \mathcal{A}_0]$

Under the following identification:

$$\boxed{\Phi_{0,1}\mathbf{O}_1 \text{ in } \mathcal{A}_0} \longleftrightarrow \boxed{\mathbf{O}_1 \text{ in } \mathcal{A}_1} \tag{3.29}$$

we think that

$$(M) = \mathbf{M}_{\mathcal{A}_0}(\Phi_{0,1}\mathbf{O}_1, S_{[\rho_0^p]}). \tag{3.30}$$

This viewpoint is standard, and it is called the *Heisenberg picture representation* of (M). Axiom 1 says that

- the probability that the measured value of the measurement (M) (i.e.,$\mathbf{M}_{\mathcal{A}_0}(\Phi_{0,1}\mathbf{O}_1, S^0_{[\rho_0^p]})$) belongs to Ξ_1 $(\in \mathcal{F}_1)$ is given by

$$\rho_0^p(\Phi_{0,1}F(\Xi_1)) \big(\equiv {}_{\mathcal{A}_0^*}\Big\langle \rho_0^p, \Phi_{0,1}F(\Xi_1)\Big\rangle_{\mathcal{A}_0}\big). \tag{3.31}$$

On the other hand, under the following identification:

$$\boxed{\rho_0^p \text{ in } \mathfrak{S}(\mathcal{A}_0^*)} \longleftrightarrow \boxed{\Phi_{0,1}^*\rho_0^p \text{ in } \mathfrak{S}(\mathcal{A}_1^*)},$$

we also consider that

$$(M) = \mathbf{M}_{\mathcal{A}_1}(\mathbf{O}_1, S_{[\Phi_{0,1}^*\rho_0^p]}) \tag{3.32}$$

$\Big($though $\Phi_{0,1}^*\rho_0^p$ is not in $\mathfrak{S}^p(\mathcal{A}^*)$ but in $\mathfrak{S}^m(\mathcal{A}^*)$ if $\Phi_{0,1}$ is not homomorphic. Cf. Chapter 8 (statistical measurement theory),$\Big)$ This viewpoint is called *the Schrödinger picture representation* of (M). We of course think that

- the probability that the measured value of the measurement (M) (i.e.,$\mathbf{M}_{\mathcal{A}_1}(\mathbf{O}_1, S_{[\Phi^*_{0,1}\rho^p_0]})$) belongs to Ξ_1 is given by

$$\rho^p_0(\Phi_{0,1}F(\Xi_1))\big(\equiv {}_{\mathcal{A}_1^*}\Big\langle \Phi^*_{0,1}\rho^p_0, F(\Xi_1)\Big\rangle_{\mathcal{A}_1}\big). \tag{3.33}$$

It should be noted that (3.31) = (3.33) holds. Thus it is usually and roughly said that

- Heisenberg picture (i.e., observable moves) and Schrödinger picture (i.e., state moves) are equivalent,

though the Heisenberg picture is fundamental (and the Schrödinger picture representation should be regarded as a kind of prescription). For the further arguments, see §6.2.

3.6 Measurability theorem

The following theorem is the most fundamental in classical PMT.

Theorem 3.7. [The measurability theorem of a general system, *cf.* [43]]. *Let* $(T \equiv \{0, 1, ..., N\}, \pi : T \setminus \{0\} \to T)$ *be a tree with root* 0 *and let* $\mathbf{S}_{[\rho^p_0]} \equiv [S_{[\rho^p_0]}, \mathcal{A}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{A}_{\pi(t)}\ (t \in T\setminus\{0\})]$ *be a general system with the initial system* $S_{[\rho^p_0]}$. *And, let an observable* $\mathbf{O}_t \equiv (X_t, \mathcal{F}_t, F_t)$ *in a* C^**-algebra* $\mathcal{A}_t$ *be given for each* $t \in T$. *For each* $s\ (\in T)$, *define the observable* $\widetilde{\mathbf{O}}_s \equiv (\prod_{t\in T_s} X_t, \prod_{t\in T_s}\mathcal{F}_t, \widetilde{F}_s)$ *in* $\mathcal{A}_s$ *such that:*

$$\widetilde{\mathbf{O}}_s = \begin{cases} \mathbf{O}_s & (\textit{if } s \in T \setminus \pi(T)) \\ \mathbf{O}_s \overset{\mathrm{qp}}{\times} (\overset{\mathrm{qp}}{\times}_{t\in\pi^{-1}(\{s\})} \Phi_{\pi(t),t}\widetilde{\mathbf{O}}_t) & (\textit{if } s \in \pi(T)) \end{cases}$$

if possible. Then, if an observable $\widetilde{\mathbf{O}}_0$ *(i.e., the Heisenberg picture representation of the sequential observable* $[\{\mathbf{O}_t\}_{t\in T}, \{\Phi_{t,\pi(t)} : \mathcal{A}_t \to \mathcal{A}_{\pi(t)}\}_{t\in T\setminus\{0\}}\ ]$*) in* $\mathcal{A}_0$ *exists, we have the measurement*

$$\mathbf{M}_{\mathcal{A}_0}(\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, \prod_{t\in T}\mathcal{F}_t, \widetilde{F}_0), S_{[\rho^p_0]}), \tag{3.34}$$

$\Big(\bigotimes_{t\in T}\mathcal{F}_t$ *is sometimes denoted by* $\prod_{t\in T}\mathcal{F}_t$, *cf. Definition 2.10*$\Big)$, *which is called the Heisenberg picture representation of the symbol* $\mathfrak{M}(\{\mathbf{O}_t\}_{t\in T}, \mathbf{S}_{[\rho^p_{t_0}]})$. *If the system is classical, i.e.,* $\mathcal{A}_t \equiv C(\Omega_t)$ $(\forall t \in T)$, *then the measurement always exists, while the uniqueness is not always guaranteed. Also, it should be noted that, for each* $s(\in T)$, *it holds that* $\Phi_{\pi(s),s}\widetilde{F}_s(\prod_{t\in T_s}\Xi_t) = \widetilde{F}_{\pi(s)}((\Pi_{t\in T_{\pi(s)}\setminus T_s} X_t) \times (\prod_{t\in T_s}\Xi_t))$ $\quad(\forall \Xi_t \in \mathcal{F}_t\ (\forall t \in T))$.

Proof. It suffices to prove it in classical measurements. However it is clear since, in classical measurements, the product observable of any observables always exists. Therefore the construction mentioned in Example 3.4 is always possible in classical systems.

□

Example 3.8. [Random walk]. Suppose that a tree $(T \equiv \{0, 1, ..., N\}, \pi)$ has a "series" structure, i.e., $\pi(t) = t - 1$ $(\forall t \in T \setminus \{0\})$. Consider a general system $\mathbf{S}_{[\delta_0]} \equiv [S_{[\delta_0]}, \mathcal{A}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{A}_{\pi(t)}$ $(t \in T \setminus \{0\})]$ with the initial system $S_{[\delta_0]}$, that is,

$$\mathcal{A}_0 \overset{\Phi_{0,1}}{\longleftarrow} \mathcal{A}_1 \overset{\Phi_{1,2}}{\longleftarrow} \mathcal{A}_2 \overset{\Phi_{2,3}}{\longleftarrow} \cdots\cdots\cdots \overset{\Phi_{N-2,N-1}}{\longleftarrow} \mathcal{A}_{N-1} \overset{\Phi_{N-1,N}}{\longleftarrow} \mathcal{A}_N. \tag{3.35}$$

Let $\mathbb{Z}$ be the set of all integers, i.e., $\mathbb{Z} = \{0, \pm 1, \pm 2, ...\}$. Consider a commutative C^*-algebra $C_0(\mathbb{Z})$. Here, put

$$\mathcal{A}_t = C_0(\mathbb{Z}) \quad (\forall t \in \{0, 1, ..., N\})$$

and define a Markov operator $\Phi_{t-1,t}(\equiv \Phi) : \mathcal{A}_t(\equiv C_0(\mathbb{Z})) \to \mathcal{A}_{t-1}(\equiv C_0(\mathbb{Z}))$ such that:

$$(\Phi f)(n) = (\Phi_{t-1,t} f)(n) = \frac{f(n+1) + f(n-1)}{2} \qquad (\forall f \in \mathcal{A}_t(\equiv C_0(\mathbb{Z})), \forall n \in \mathbb{Z}).$$

Also, for each $t = 0, 1, 2, ..., N$, consider the exact observable $\mathbf{O}_t \equiv (X_t, \mathcal{R}_t, E) \equiv (\mathbb{Z}, \mathcal{P}_0(\mathbb{Z}), E)$ in $\mathcal{A}_t(= C_0(\mathbb{Z}))$ such that, (*cf.* Example 2.20),

$$[E(\Xi)](n) = \begin{cases} 1 & n \in \Xi(\in \mathcal{P}_0(\mathbb{Z})) \\ 0 & n \notin \Xi(\in \mathcal{P}_0(\mathbb{Z})). \end{cases} \tag{3.36}$$

Thus, we get the product observable $\widetilde{\mathbf{O}}_0 \equiv (\times_{t=0}^N X_t, \times_{t=0}^N \mathcal{F}_t, \widetilde{F}_0) \equiv (\mathbb{Z}^{N+1}, \mathcal{P}_0(\mathbb{Z}^{N+1}), \widetilde{F}_0)$ in $\mathcal{A}_0$ $(\equiv C_0(\mathbb{Z}))$, that is,

$$\widetilde{F}_0(\Xi_0 \times \Xi_1 \times \Xi_2 \times \cdots \times \Xi_N) = E(\Xi_0) \times \Phi(E(\Xi_1) \times \Phi(\cdots\cdots \Phi(E(\Xi_{N-1}) \times \Phi E(\Xi_N)) \cdots)).$$

Then, we have the "measurement" $\mathfrak{M}(\{\mathbf{O}_t\}_{t \in T}, \mathbf{S}_{[\delta_0]})$ such as

$$\mathfrak{M}(\{\mathbf{O}_t\}_{t \in T}, \mathbf{S}_{[\delta_0]}) = \mathbf{M}_{C(\mathbb{Z})}(\widetilde{\mathbf{O}}_0 \equiv (\mathbb{Z}^{N+1}, \mathcal{P}_0(\mathbb{Z}^{N+1}), \widetilde{F}_0), S_{[\delta_0]}).$$

where δ_0 is the point measure at 0 $(\in \mathbb{Z})$. The sample space $\Big(\mathbb{Z}^{N+1}, \mathcal{P}_0(\mathbb{Z}^{N+1}), [\widetilde{F}_0(\cdot)](0)\Big)$ is usually called a *random walk*.

■

For the further arguments, see §10.4 (Brown motion).

3.7 Appendix (Bell's inequality)

(Continued from §2.9 (Bell's Thought Experiment))[5]

3.7.1 Deterministic evolution or Stochastic evolution?

Recall the following classification (3.27) in PMT:

$$\begin{cases} \text{deterministic PMT} = \underset{[\text{Axiom 1 (2.37)}]}{\text{“measurement”}} + \underset{[\text{ each } \Phi_{t_1,t_2} \text{ is homomorphic in Axiom 2 (3.26)}]}{\text{“the deterministic relation among systems”.}} \\ \\ \text{stochastic PMT} = \underset{[\text{Axiom 1 (2.37)}]}{\text{“measurement”}} + \underset{[\text{Axiom 2 (3.26)}]}{\text{“the Markov relation among systems”.}} \end{cases}$$

However, we know that in classical (or quantum) mechanics, the general system $\mathbf{S}_{[\rho^p]} \equiv [S_{[\rho^p]}, \mathcal{A}_t \overset{\Psi_{\pi(t),t}}{\to} \mathcal{A}_{\pi(t)} \ (t \in T \setminus \{0\})]$ is always deterministic, that is, $\Psi_{\pi(t),t}$ is always homomorphic. (*cf.* "Newtonian mechanics and quantum mechanics" in §3.1.)

Recall (2.76), i.e., the de Broglie paradox (*cf.* [20]. Also see §9.3.3). That is,

- if we admit quantum mechanics $\Big($= "Axiom 1 + Axiom 2 (homomorphic time evolution)"$\Big)$, we must admit the fact that there is something faster than light. (*cf.* [18, 78]). (3.37) (=(2.76))

Of course we admit quantum mechanics, and therefore, we believe that there is something faster than light. However, most people may hope that quantum mechanics is not true rather than admit the fact that there is something faster than light. That is,

($\sharp$) Using the Schrödinger picture representation, they may assert that the singlet state ρ_s is not fixed, but the Markov time evolution $\Big($i.e., "the Markov relation among systems (Axiom 2)" and not "the homomorphic relation among systems Axiom 2)"$\Big)$:

$$\rho_s \overset{\Phi^*}{\rightsquigarrow} \rho_0^m \tag{3.38}$$

should be considered.

[5] Although Bell's inequality is generally said to be one of the most profound discoveries in 20-th century science, I could not understand the arguments (in [9, 18, 78, 8]), particularly, I had the question: "In what framework is Bell's inequality discussed (in [9, 18, 78])?". I wonder if these arguments are confusing physics with mathematics. Thus, I add this section, in which all arguments are discussed in the framework of PMT (Axioms 1 and 2).

The purpose of the following section (i.e., §3.7.2) is to show that we must admit that there is something faster than light, even under the above assumption (♯). That is, if we assert that PMT $\Big($= "Axiom 1 + Axiom 2 (Markov time evolution)", i.e., quantum mechanics with Markov (and not homomorphic) time evolution$\Big)$ is true, we must admit the fact that there is something faster than light.

3.7.2 Generalized Bell's inequality in mathematics

First we prepare some mathematical inequalities. Of course, what is most important is how to interpret these theorems in physics. This will be discussed in the next section. In order to avoid to confuse physical results and mathematical ones, in this §3.7.2, we devote ourselves to mathematical arguments.

Theorem 3.9. [Bell's inequality, *cf.* [9, 78]]. *Let $(Y, \mathcal{G}, m)$ be a probability space. Let g_1^1, g_1^2, g_2^1, g_2^2 be $\{-1,1\}$-valued measurable functions on Y. Define the correlation function $P'(g_1^i, g_2^j)$ such that:*

$$P'(g_1^i, g_2^j) = \int_Y g_1^i(y) g_2^j(y)\, m(dy). \tag{3.39}$$

Then, it holds that

$$|P'(g_1^1, g_2^1) - P'(g_1^1, g_2^2)| + |P'(g_1^2, g_2^1) + P'(g_1^2, g_2^2)| \le 2. \tag{3.40}$$

Proof. For completeness, we add the proof in what follows.

$$\begin{aligned}
&|P'(g_1^1, g_2^1) - P'(g_1^1, g_2^2)| + |P'(g_1^2, g_2^1) + P'(g_1^2, g_2^2)| \\
\le &\int_{X^4} |g_1^1(y)| \cdot |g_2^1(y) - g_2^2(y)| m(dy) + \int_Y |g_1^2(y)| \cdot |g_2^1(y) + g_2^2(y)| m(dy) \\
\le &\int_{X^4} |g_2^1(y) - g_2^2(y)| + |g_2^1(y) + g_2^2(y)| m(dy) = 2.
\end{aligned}$$

This completes the proof. □

Corollary 3.10. [Bell's inequality]. *Let $(Y, \mathcal{G}, m)$ be a probability space. Let g_1^{11}, g_1^{12}, g_1^{21}, g_1^{22}, g_2^{11}, g_2^{12}, g_2^{21} and g_2^{22} be $\{-1,1\}$-valued measurable functions on Y. Define the correlation function $P(g_1^{ij}, g_2^{ij})$ such that*

$$P(g_1^{ij}, g_2^{ij}) = \int_Y g_1^{ij}(y) g_2^{ij}(y)\, m(dy). \tag{3.41}$$

Further, assume that

$$g_1^{11} = g_1^{12}, \quad g_1^{21} = g_1^{22}, \quad g_2^{11} = g_2^{21}, \quad g_2^{12} = g_2^{22} \quad (\text{a.e. } m) \tag{3.42}$$

i.e., $m\big(\{y \in Y \;:\; g_1^{11}(y) = g_1^{12}(y)\}\big) = 1$, *etc. Then, it holds that*

$$|P(g_1^{11}, g_2^{11}) - P(g_1^{12}, g_2^{12})| + |P(g_1^{21}, g_2^{21}) + P(g_1^{22}, g_2^{22})| \le 2. \tag{3.43}$$

Proof. It immediately follows from Theorem 3.9. □

Next we present the following theorem, which can be regarded as a generalization of the above corollary (*cf.* Remark 3.12 later).

Theorem 3.11. [Generalized Bell's inequality]. *Let* $(Y, \mathcal{G}, m)$ *be a probability space. Let* g_1^{11}, g_1^{12}, g_1^{21}, g_1^{22}, g_2^{11}, g_2^{12}, g_2^{21} *and* g_2^{22} *be* $\{-1,1\}$*-valued measurable functions on* Y*. Assume that these satisfy*

$$m[(g_1^{ij}, g_2^{ij})^{-1}(B_1 \times B_2)] = \sum_{\ell \in L} \alpha_\ell \, \mu_{1,\ell}^i(B_1) \, \mu_{2,\ell}^j(B_2) \quad (\forall B_1, B_2 \subseteq \{-1,1\}, \ \forall i,j = 1,2) \tag{3.44}$$

for some probability measures $\mu_{k,\ell}^i$, $(k, i = 1, 2, \ell \in L)$, *on* $\{-1,1\}$ *and some nonnegative sequence* $\{\alpha_\ell\}_{\ell \in L}$ *such that* $\sum_{\ell \in L} \alpha_\ell = 1$*. Then, it holds that*

$$|P(g_1^{11}, g_2^{11}) - P(g_1^{12}, g_2^{12})| + |P(g_1^{21}, g_2^{21}) + P(g_1^{22}, g_2^{22})| \le 2, \tag{3.45}$$

where the correlation functions $P(g_1^{ij}, g_2^{ij})$ *are defined by (3.41).*

Proof. A simple calculation shows that

$$\begin{aligned} P(g_1^{ij}, g_2^{ij}) &= \sum_{\ell \in L} \alpha_\ell \,[\sum_{(x_1,x_2) \in \{-1,1\}^2} x_1 x_2 \; \mu_{1,\ell}^i(\{x_1\}) \mu_{2,\ell}^j(\{x_2\})] \\ &= \sum_{\ell \in L} \alpha_\ell (4\mu_{1,\ell}^i \mu_{2,\ell}^j + 1 - 2\mu_{1,\ell}^i - 2\mu_{2,\ell}^j), \end{aligned}$$

where $\mu_{k,\ell}^i = \mu_{k,\ell}^i(\{1\})$. Thus, we see that

$$
\begin{aligned}
&|P(g_1^{11}, g_2^{11}) - P(g_1^{12}, g_2^{12})| + |P(g_1^{21}, g_2^{21}) + P(g_1^{22}, g_2^{22})| \\
=&|\sum_{\ell \in L} \alpha_\ell (4\mu_{1,\ell}^1 \mu_{2,\ell}^1 + 1 - 2\mu_{1,\ell}^1 - 2\mu_{2,\ell}^1) - \sum_{\ell \in L} \alpha_\ell (4\mu_{1,\ell}^1 \mu_{2,\ell}^2 + 1 - 2\mu_{1,\ell}^1 - 2\mu_{2,\ell}^2)| \\
&\qquad + |\sum_{\ell \in L} \alpha_\ell (4\mu_{1,\ell}^2 \mu_{2,\ell}^1 + 1 - 2\mu_{1,\ell}^2 - 2\mu_{2,\ell}^1) + \sum_{\ell \in L} \alpha_\ell (4\mu_{1,\ell}^2 \mu_{2,\ell}^2 + 1 - 2\mu_{1,\ell}^2 - 2\mu_{2,\ell}^2)| \\
=&|\sum_{\ell \in L} \alpha_\ell (4\mu_{1,\ell}^1 \mu_{2,\ell}^1 - 2\mu_{2,\ell}^1 - 4\mu_{1,\ell}^1 \mu_{2,\ell}^2 + 2\mu_{2,\ell}^2)| \\
&\qquad + |\sum_{\ell \in L} \alpha_\ell (4\mu_{1,\ell}^2 \mu_{2,\ell}^1 + 2 - 4\mu_{1,\ell}^2 - 2\mu_{2,\ell}^1 + 4\mu_{1,\ell}^2 \mu_{2,\ell}^2 - 2\mu_{2,\ell}^2)| \equiv |A| + |B|,
\end{aligned}
$$

and consequently,

$$
\begin{aligned}
&= \begin{cases} |\sum_{\ell \in L} \alpha_\ell [2 - 4(\mu_{1,\ell}^2 + \mu_{2,\ell}^1 + \mu_{1,\ell}^1 \mu_{2,\ell}^2 - \mu_{1,\ell}^1 \mu_{2,\ell}^1 - \mu_{1,\ell}^2 \mu_{2,\ell}^1 - \mu_{1,\ell}^2 \mu_{2,\ell}^2)]| & (\text{if } A \cdot B \geq 0) \\ |\sum_{\ell \in L} \alpha_\ell [2 - 4(\mu_{1,\ell}^2 + \mu_{2,\ell}^2 + \mu_{1,\ell}^1 \mu_{2,\ell}^1 - \mu_{1,\ell}^1 \mu_{2,\ell}^2 - \mu_{1,\ell}^2 \mu_{2,\ell}^2 - \mu_{1,\ell}^2 \mu_{2,\ell}^1)]| & (\text{if } A \cdot B \leq 0) \end{cases} \\
&\leq \begin{cases} \sum_{\ell \in L} \alpha_\ell |2 - 4(\mu_{1,\ell}^2 + \mu_{2,\ell}^1 + \mu_{1,\ell}^1 \mu_{2,\ell}^2 - \mu_{1,\ell}^1 \mu_{2,\ell}^1 - \mu_{1,\ell}^2 \mu_{2,\ell}^1 - \mu_{1,\ell}^2 \mu_{2,\ell}^2)| & (\text{if } A \cdot B \geq 0) \\ \sum_{\ell \in L} \alpha_\ell |2 - 4(\mu_{1,\ell}^2 + \mu_{2,\ell}^2 + \mu_{1,\ell}^1 \mu_{2,\ell}^1 - \mu_{1,\ell}^1 \mu_{2,\ell}^2 - \mu_{1,\ell}^2 \mu_{2,\ell}^2 - \mu_{1,\ell}^2 \mu_{2,\ell}^1)| & (\text{if } A \cdot B \leq 0). \end{cases}
\end{aligned}
$$

Hence, it suffices to prove that $0 \leq C(x, y, z, w) \leq 1$ $(\forall (x, y, z, w) \in [0, 1]^4)$, where $C(x, y, z, w) = y + z + xw - xz - yz - yw$. This is shown as follows:
[Case 1; $w - z \geq 0$].

$$
\begin{aligned}
0 \leq y(1 - w) + z(1 - y) + x(w - z) \equiv C \leq C + (w - z)(1 - x) \\
= y(1 - w) + w - yz \leq 1 - yz \leq 1.
\end{aligned}
$$

[Case 2; $w - z \leq 0$].

$$
\begin{aligned}
0 &\leq y(1 - z) + w(1 - y) = y + z + (w - z) - yz - yw \\
&\leq y + z + x(w - z) - yz - yw \equiv C \leq y + z - yz - yw \leq y(1 - z) + z \leq 1.
\end{aligned}
$$

This completes the proof. □

Remark 3.12. It is interesting to see that Corollary 3.10 can be regarded as a particular case of Theorem 3.11. This can be easily shown as follows: Let $(Y, \mathcal{G}, m)$ and g_k^{ij} be as in Corollary 3.10. Thus, we assume that the condition (3.42) holds. Put $L = \{-1, 1\}^4$. For each ℓ $(\equiv (\ell_1^1, \ell_1^2, \ell_2^1, \ell_2^2) \in L)$, define the α_ℓ $(\in [0, 1])$ such that $\alpha_{(\ell_1^1, \ell_1^2, \ell_2^1, \ell_2^2)} = m\big((g_1^{11}, g_1^{22}, g_2^{11}, g_2^{22})^{-1}\ (\{(\ell_1^1, \ell_1^2,\ \ell_2^1, \ell_2^2)\})\big)$. Clearly it holds that $\sum_{\ell \in L} \alpha_\ell = 1$. Define the probability measures $\widehat{\mu}_1$ and $\widehat{\mu}_{-1}$ on $\{-1, 1\}$ such that $\widehat{\mu}_1(\{-1\}) = 0$, $\widehat{\mu}_1(\{1\}) = 1$

and $\widehat{\mu}_{-1} = 1 - \widehat{\mu}_1$. It is easy to see that $m\big((g_1^{11}, g_1^{22}, g_2^{11}, g_2^{22})^{-1}(\{(x_1^1, x_1^2, x_2^1, x_2^2)\})\big) = \sum_{\ell \in L} \alpha_\ell \ \widehat{\mu}_{\ell_1^1}(\{x_1^1\})\widehat{\mu}_{\ell_1^2}(\{x_1^2\})\widehat{\mu}_{\ell_2^1}(\{x_2^1\}) \ \widehat{\mu}_{\ell_2^2}(\{x_2^2\})$ $(\forall (x_1^1, x_1^2, x_2^1, x_2^2) \in \{-1,1\}^4)$. Thus, putting $\mu^i_{k,(\ell_1^1, \ell_1^2, \ell_2^1, \ell_2^2)} = \widehat{\mu}_{\ell_k^i}$, we can immediately see that the $\{\alpha_\ell\}_{\ell \in L}$ and the $\{\mu^i_{k,\ell} : i,k = 1,2, \ell \in L\}$ satisfy the condition (3.44).

■

3.7.3 Generalized Bell's inequality in Measurements

Put $X = \{-1, 1\}$. Consider a measurement $\mathbf{M}_{\mathcal{A}}(\ \mathbf{O} \equiv (X^8, \mathcal{P}(X^8), G),\ S_{[\rho_0]})$ formulated in arbitrary C^*-algebra $\mathcal{A}$. Putting $\nu^3_{\rm BI}(\,\cdot\,) = \rho_0(G(\,\cdot\,))$, we have the sample space $(X^8, \mathcal{P}(X^8), \nu^3_{\rm BI})$, which is induced by the measurement $\mathbf{M}_{\mathcal{A}}(\ \mathbf{O},\ S_{[\rho_0]})$. Consider the $\{-1,1\}$-valued functions g_k^{ij} on X^8, $(i,j,k = 1,2)$. And define the correlation functions $P(g_1^{ij}, g_2^{ij})$ $(i,j = 1,2)$ by (3.41). Assume the condition (3.44) in Theorem 3.11. Then, we see, by Theorem 3.11, that the following inequality holds:

$$|P(g_1^{11}, g_2^{11}) - P(g_1^{12}, g_2^{12})| + |P(g_1^{21}, g_2^{21}) + P(g_1^{22}, g_2^{22})| \le 2. \tag{3.46}$$

Therefore, it may be viable to compare the measurement $\mathbf{M}_{\mathcal{A}}(\ \mathbf{O},\ S_{[\rho_0]})$ with the measurement $\bigotimes_{i,j=1,2} \mathbf{M}_{B(\mathbf{C}^2 \otimes \mathbf{C}^2)}\ (\mathbf{O}_{a^i b^j}, S_{[\rho_s]})$ in Bell's thought experiment, though it is also sure that these are not connected with each other. For example, some may, by some reason, consider that the singlet state ρ_s in Bell's thought experiment (*cf.* the formula (2.75)) is reduced to a certain state ρ_0 $(\in \mathfrak{S}^p(B(\mathbf{C}^2 \otimes \mathbf{C}^2)^*))$ such as

$$\rho_s \rightsquigarrow \rho_0 = |\vec{e} \otimes \vec{f}\rangle\langle \vec{e} \otimes \vec{f}| \tag{3.47}$$

for some $\vec{e} \otimes \vec{f}$ $(\in \mathbf{C}^2 \otimes \mathbf{C}^2)$ such that $\|\vec{e}\|_{\mathbf{C}^2} = \|\vec{f}\|_{\mathbf{C}^2} = 1$. If so, instead of the measurement $\bigotimes_{i,j=1,2} \mathbf{M}_{B(\mathbf{C}^2 \otimes \mathbf{C}^2)}(\mathbf{O}_{a^i b^j}, S_{[\rho_s]})$, we must consider the measurement $\bigotimes_{i,j=1,2} \mathbf{M}_{B(\mathbf{C}^2 \otimes \mathbf{C}^2)}$ $(\mathbf{O}_{a^i b^j}, S_{[\rho_0]})$, which has the sample space $(X^8, \mathcal{P}(X^8), \nu)$ such that:

$$\nu(\{(x_1^{11}, x_2^{11}, x_1^{12}, x_2^{12}, x_1^{21}, x_2^{21}, x_1^{22}, x_2^{22})\}) = \prod_{i,j=1,2} \rho_0\big((F_{a^i} \otimes F_{b^j})(\{(x_1^{ij}, x_2^{ij})\})\big)$$
$$= \prod_{i,j=1,2} \big[\langle \vec{e}, F_{a^i}(\{x_1^{ij}\})\vec{e}\rangle \langle \vec{f}, F_{b^j}(\{x_2^{ij}\})\vec{f}\rangle\big].$$

Or more generally (or, in the sense of "ensemble"), using the adjoint operator Φ^* of a Markov operator $\Phi : B(\mathbf{C}^2 \otimes \mathbf{C}^2) \to B(\mathbf{C}^2 \otimes \mathbf{C}^2)$, we may consider the following Markov evolution:

$$\rho_s \overset{\Phi^*}{\rightsquigarrow} \rho_0^m = \sum_{n=1}^{2} \sum_{m=1}^{2} \alpha_{mn} |\vec{e}_m \otimes \vec{f}_n\rangle\langle \vec{e}_m \otimes \vec{f}_n|, \tag{3.48}$$

where $\{\vec{e}_m\}_{m=1}^2$ and $\{\vec{f}_m\}_{i=1}^2$ are respectively the complete orthonormal basis in $\mathbf{C}^2$, and $0 \le \alpha_{mn} \le 1$ such that $\sum_{n=1}^2 \sum_{m=1}^2 \alpha_{mn} = 1$. Thus we have the (statistical) measurement $\bigotimes_{i,j=1,2} \mathbf{M}_{B(\mathbf{C}^2 \otimes \mathbf{C}^2)} (\Phi \mathbf{O}_{a^i b^j}, S_{[\rho_s]})$. Thus, we may have the sample space $(X^8, \mathcal{P}(X^8), \nu)$ such that:

$$\nu(\{(x_1^{11}, x_2^{11}, x_1^{12}, x_2^{12}, x_1^{21}, x_2^{21}, x_1^{22}, x_2^{22})\}) = \prod_{i,j=1,2} \rho_s\big((\Phi F_{a^i} \otimes F_{b^j})(\{(x_1^{ij}, x_2^{ij})\})\big)$$

$$= \prod_{i,j=1,2} (\Phi^* \rho_s)\big((F_{a^i} \otimes F_{b^j})(\{(x_1^{ij}, x_2^{ij})\})\big) = \prod_{i,j=1,2} \rho_0^m \big((F_{a^i} \otimes F_{b^j})(\{(x_1^{ij}, x_2^{ij})\})\big)$$

$$= \prod_{i,j=1,2} [\sum_{m=1}^2 \sum_{n=1}^2 \alpha_{mn} \langle \vec{e}_m, F_{a^i}(\{x_1^{ij}\}) \vec{e}_m \rangle \langle \vec{f}_n, F_{b^j}(\{x_2^{ij}\}) \vec{f}_n \rangle].$$

Note that the probability space $(X^8, \mathcal{P}(X^8), \nu)$ and the g_k^{ij} defined by (2.77) satisfy the condition (3.44) in Theorem 3.11. That is because it suffices to put $L = \{-1, 1\}^2$ and

$$\mu_{1,(m,n)}^1(\,\cdot\,) = \langle \vec{e}_m, F_{a^1}(\,\cdot\,) \vec{e}_m \rangle, \qquad \mu_{1,(m,n)}^2(\,\cdot\,) = \langle \vec{e}_m, F_{a^2}(\,\cdot\,) \vec{e}_m \rangle,$$
$$\mu_{2,(m,n)}^1(\,\cdot\,) = \langle \vec{f}_n, F_{b^1}(\,\cdot\,) \vec{f}_n \rangle, \qquad \mu_{2,(m,n)}^2(\,\cdot\,) = \langle \vec{f}_n, F_{b^2}(\,\cdot\,) \vec{f}_n \rangle,$$

for each (m, n) $(\in L \equiv \{-1, 1\}^2)$. Thus, Theorem 3.11 says that such Markov evolution as the above (3.47) or (3.48) does not occur in Bell's thought experiment. Therefore we can conclude that

- if we admit PMT (= "Axiom 1 + Axiom 2 (Markov relation)"), we must also admit the fact that there is something faster than light. (3.49)

Of course we admit PMT, and therefore, we believe that there is something faster than light.

Chapter 4

Boltzmann's equilibrium statistical mechanics

As mentioned in Chapters 2 and 3, we see that (pure) measurement theory (= PMT) is formulated as follows:

$$\text{PMT} = \underset{[\text{Axiom 1 (2.37)}]}{\text{measurement}} + \underset{[\text{Axiom 2 (3.26)}]}{\text{the relation among systems}} \quad \text{in } C^*\text{-algebra} . \tag{4.1 (=(1.4))}$$

The purpose of this chapter[1] is to understand Boltzmann's equilibrium statistical mechanics [2] (i.e., "the principle of equal a priori probability" and "the ergodic hypothesis") as one of applications of PMT. We believe that our approach completely justifies the the thermodynamical weight method (i.e., the Gibbs method, *cf.* [26])[3].

4.1 Introduction

In spite that equilibrium statistical mechanics is generally believed to be based on Newtonian mechanics, the term "probability" frequently appears in equilibrium statistical mechanics. Therefore, if we want to understand equilibrium statistical mechanics in the framework of Newtonian mechanics, a certain rule concerning "probability" should be added. That is, we hope to understand equilibrium statistical mechanics such as:

[1]It may be recommended that this chapter is skipped if readers want to study statistics in the framework of PMT firstly (*cf.* Chapters 5 and 6).

[2]In this chapter readers are not required to have much knowledge of statistical mechanics.

[3]In this book, we think that statistical mechanics should be understood as one of applications of measurement theory and not theoretical physics, (*cf.* Table (1.7)). Thus, it should be noted that no serious test has been conducted in statistical mechanics. What we know is nothing but the fact that statistical mechanics is quite useful (*cf.* Table (1.8)). Or, statistical mechanics is "almost empirically true" to such a degree that statistical mechanics is assured to be useful in usual situations. Cf. the (I_9) in §1.2.

$$\underset{}{\text{“equilibrium statistical mechanics”}} = \underset{\text{[Axiom 2 (3.26)]}}{\text{“Newton equation”}} + \underset{\text{[Axiom 1 (2.37)]}}{\text{“probabilistic rule”}} \tag{4.2}$$

in PMT.

First we must answer the following question:

(Q_1) What is the "probabilistic rule" in (4.2)?

Recall Example 2.16 (the urn problem), which is the most fundamental in the classical measurement. Thus in order to understand "probabilistic rule (=Axiom 1) in (4.2)", it suffices to note the following simplest example:

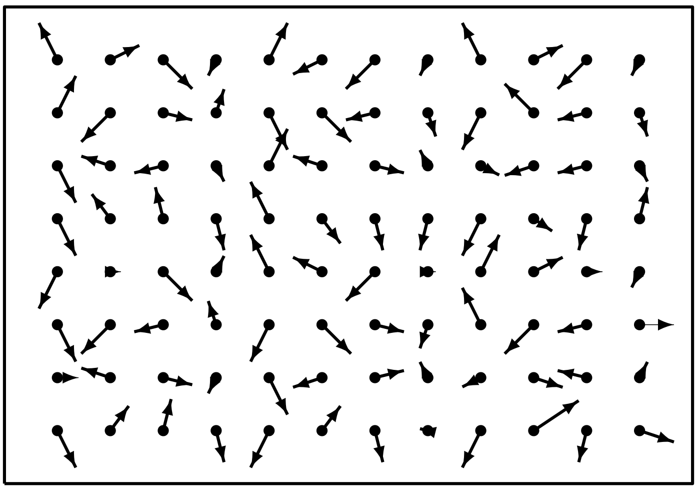
(A_1) *"Consider a box containing* 7×10^{23} *white balls and* 3×10^{23} *black balls, and choose a ball at random from the box. Then the probability that the ball is white is given as 0.7."*

Even without the knowledge of measurement theory (in Chapters 2 and 3), every reader surely agrees that the probability appearing in urn (i.e., box) problems is most typical in statistics.

Next we must refer to "Newtonian mechanics" in (4.2). Namely we must solve the following question.

(Q_2) What kinds of conditions are imposed on the Newton equation in (4.2)?

In equilibrium statistical mechanics, about 10^{24} ($\approx 6.02 \times 10^{23}$: "Avogadro constant") particles, of course, move hard in a box such as the following figure:



(4.3)

However it seems to be natural to think as follows:

(A_2^1) *All particles are even, or on a level.*

(A_2^2) *The motions of particles are (almost) independent of each other. In other words, the information about a subsystem composed of some particles is invalid for the inference of the state of another subsystem.*

This is our answer to the question (Q_2). In §4.2, the (A_2^1) and (A_2^2) will be represented in terms of PMT. Also, the (A_1) will be discussed in §4.3.

Summing up, we think that equilibrium statistical mechanics is formulated as follows:

$$\text{“equilibrium statistical mechanics”} = \underbrace{\underset{\text{(the probability such as in } (A_1))}{\text{“probabilistic rule”}} + \underset{\text{(the conditions } (A_2^1) \text{ and } (A_2^2))}{\text{“Newton equation”}}}_{\text{(+ “staying time interpretation”)}} \tag{4.4}$$

in PMT. Or, equivalently,

- An equilibrium statistical system can be regarded as an urn containing about 10^{24} particles. Also, the motions of particles are dominated by the Newtonian equation with the conditions (A_2^1) and (A_2^2). Also, the "staying time interpretation" implies the common sense such as it is almost impossible to find a rare event.

And moreover, two conventional principles (i.e., "the principle of equal a priori probability" and "the ergodic hypothesis") will be completely clarified in our proposal (4.4).

The first attempt to understand equilibrium statistical mechanics in the framework of PMT was executed in [45]. The content in [45] will be slightly modified and improved in this chapter.

Note, for completeness, that our purpose is to understand equilibrium statistical mechanics as one of applications of PMT and not to derive equilibrium statistical mechanics from Newtonian mechanics (*cf.* [75]). That is, we are in theoretical informatics and not in theoretical physics.[4]

[4] We have no experimental evidence that the ergodic approach to statistical mechanics is proper. However, in theoretical informatics, it suffices to find a reason that many people do not doubt.

4.2 Dynamical aspects of equilibrium statistical mechanics

In this section we shall devote ourselves to the mathematical description of the answers (A_2^1) and (A_2^2) mentioned in Section 4.1. Readers should note that all arguments in this section are within Newtonian mechanics. Namely, it should be noted that it is prohibited to use the term "probability" in this section. For example, Lemma 4.9 ("the law of large numbers" in §4.5 Appendix) is not only most important in Kolmogorov's probability theory but also in this section (i.e., the derivation of the ergodic hypothesis (= Theorem 4.6)). Therefore, readers will see that Lemma 4.9 is used independently of the concept of "probability". This is the reason that the term "normalized measure" (and not "probability measure") is used in Lemma 4.9.

Now let us begin with the well-known ergodic theorem (*cf.* [57, 83]). In Newtonian mechanics, any state of a system composed of $N(\approx 10^{24})$ particles is represented by a point (q,p) $\big(\equiv (q_{1n}, q_{2n}, q_{3n}, \ p_{1n}, p_{2n}, \ p_{3n})_{n=1}^N \ \big)$ in a phase (or state) space $\mathbf{R}^{6N}$ (*cf.* the formula (2.8)). Let $\mathcal{H} : \mathbf{R}^{6N} \to \mathbf{R}$ be a Hamiltonian, i.e., a positive continuous function on $\mathbf{R}^{6N}$. Define $V(E)$, $E \geq 0$, by "the volume of the set $\{(q,p) \in \mathbf{R}^{6N} \mid \mathcal{H}(q,p) \leq E\}$", and define the measure $\nu_{{}_E}$ on the energy surface $\mathcal{S}_{{}_E}$ $(\equiv \{(q,p) \in \mathbf{R}^{6N} \mid \mathcal{H}(q,p) = E\})$ such that

$$\nu_{{}_E}(B) = \int_B |\nabla \mathcal{H}(q,p)|^{-1} dm_{6N-1} \quad (\forall B \in \mathcal{B}_{\mathcal{S}_{{}_E}}, \text{ the Borel field of } \mathcal{S}_{{}_E})^5 \tag{4.5}$$

where dm_{6N-1} is the usual surface measure on $\mathcal{S}_{{}_E}$. Note that $\nu_{{}_E}(\mathcal{S}_{{}_E}) = \frac{dV(E)}{dE}$ holds. Let $\{\psi_t^E\}_{-\infty<t<\infty}$ be the flow on the energy surface $\mathcal{S}_{{}_E}$ induced by the Newton equation with the Hamiltonian $\mathcal{H}$. Liouville's theorem (*cf.* [11]) says that the measure $\nu_{{}_E}$ is invariant concerning the flow $\{\psi_t^E\}_{-\infty<t<\infty}$. Defining the normalized measure $\overline{\nu}_{{}_E}$ such that $\overline{\nu}_{{}_E} = \frac{\nu_E}{\nu_E(\mathcal{S}_E)}$, we have the normalized measure space $(\mathcal{S}_{{}_E}, \mathcal{B}_{\mathcal{S}_{{}_E}}, \overline{\nu}_{{}_E})$.

In order that equilibrium statistical mechanics must hold, we first assume that the Hamiltonian $\mathcal{H}$ satisfies the following *ergodic hypothesis* (EH):

(EH) The flow $\{\psi_t^E\}_{-\infty<t<\infty}$ on the $\mathcal{S}_{{}_E}$ is *ergodic*. That is, there uniquely exists an normalized invariant measure $\overline{\nu}_{{}_E}$ on $\mathcal{S}_{{}_E}$ such that $\overline{\nu}_{{}_E}(B) = \overline{\nu}_{{}_E}(\psi_t(B))$ $(-\infty < \forall t <$

[5] Or usually, $\nu_{{}_E}(B) = \frac{1}{h^{3N}N!} \int_B |\nabla \mathcal{H}(q,p)|^{-1} dm_{6N-1}$, where h is the Plank constant. In this book, for simplicity, the constant $\frac{1}{h^{3N}N!}$ will be omitted.

∞, $\forall B \in \mathcal{B}_{\mathcal{S}_E}$).

The ergodic theorem (*cf.* [11, 57]) says that the normalized measure $\overline{\nu}_E$ represents the *normalized averaging staying time,* i.e., it holds that

$$\overline{\nu}_E(B) = \lim_{K \to \infty} \frac{\sharp[\{k \mid \psi_{\epsilon k}\omega \in B, k = 1, 2, ..., K\}]}{K} \quad (\forall \omega \in \mathcal{S}_E, \forall \epsilon > 0).$$

or generally,

$$\underbrace{\int_\Omega f(\omega)\overline{\nu}_E(d\omega)}_{\text{(space average)}} = \underbrace{\lim_{T\to\infty} \frac{1}{T}\int_0^T f(\psi_t(\omega_0))dt}_{\text{(time average)}} \quad (\forall f \in C(\Omega), \quad \forall \omega_0 \in \Omega), \tag{4.6}$$

which is equivalent to the (EH). Thus the normalized measure space $(\mathcal{S}_E, \mathcal{B}_{\mathcal{S}_E}, \overline{\nu}_E)$ is called *the normalized averaging staying time space* (*cf.* Remark 4.1 later).

We assert that

(STI) [Staying time interpretation of statistical mechanics]. *Let $\mathcal{N} \in \mathcal{B}_{\mathcal{S}_E}$ such that the normalized averaging staying time $\overline{\nu}_E(\mathcal{N})$ is quite small (i.e., $\overline{\nu}_E(\mathcal{N}) \ll 1$). Then it is almost impossible (or precisely, quite rare) to see that the state $(q(t), p(t))$ belongs to the $\mathcal{N}$.*

We think that this (STI) is a common sense rather than a principle. The concept of "time" (or precisely "non-relativistic time") is within Newtonian mechanics, and therefore the statement (STI) (or "staying time") can be understood within Newtonian mechanics.

Remark 4.1. [The probabilistic interpretation of $(\mathcal{S}_E, \mathcal{B}_{\mathcal{S}_E}, \overline{\nu}_E)$]. The probabilistic interpretation is as follows:

(PI) [Probabilistic interpretation of statistical mechanics]. *The normalized averaging staying time space $(\mathcal{S}_E, \mathcal{B}_{\mathcal{S}_E}, \overline{\nu}_E)$ is regarded as Kolmogorov's probability space.*

That is, the probabilistic interpretation, which is usually called *"the principle of equal a priori probability"*, means that the probability that the state of the system belongs to $\Xi (\in \mathcal{B}_{\mathcal{S}_E})$ is given by $\overline{\nu}_E(\Xi)$. If the probabilistic interpretation (PI) is assumed, the (STI) obviously holds. However, the concept of "normalized staying time" is clearly different from that of "probability". Note that:

- the former (i.e., "the staying time interpretation") is within Newtonian mechanics, but the latter (i.e., "the probabilistic interpretation") is not so.

Thus, in this chapter we choose a common sense (i.e., "the staying time interpretation") rather than a principle (i.e., "the probabilistic interpretation").[6] This is the reason that the $(\mathcal{S}_E, \mathcal{B}_{\mathcal{S}_E}, \overline{\nu}_E)$ is not called the probability space in this chapter. Again note that all arguments in this section are within Newtonian mechanics. In this chapter the (STI) will be used instead of the (PI).

■

We introduce the following notation:

Notation 4.2. [In the sense of (STI)]. *Let* $\mathbf{P}(q,p)$ *be a proposition concerning a state* (q,p) $(\in \mathcal{S}_E)$ *such that* $\mathbf{P}(q(t),p(t))$ *is true for every* $t \in \mathcal{S}_E \setminus \mathcal{N}$ $(\equiv \{\omega \mid \omega \in \mathcal{S}_E, \omega \notin \mathcal{N}\})$. *Assume that the normalized averaging staying time* $\overline{\nu}_E(\mathcal{N})$ *is quite small (i.e.,* $\overline{\nu}_E(\mathcal{N}) \ll 1$*). Then we write it as*

$$\mathbf{P}(q(t),p(t)) \text{ is true} \quad (\text{almost every } t \text{ in the sense of (STI)}), \tag{4.7}$$
$$\Big(\text{Or, } \mathbf{P}(q(t),p(t)) \text{ is almost always true} \Big).$$

Also, when the probabilistic interpretation (cf. Remark 4.1) is added to the $(\mathcal{S}_E, \mathcal{B}_{\mathcal{S}_E}, \overline{\nu}_E)$*, we may write it as*

$$\mathbf{P}(q(t),p(t)) \text{ is true} \quad (\text{almost every } t \text{ in the sense of (PR)}).^7 \tag{4.8}$$

■

As seen in Remark 4.1, it holds that (4.8)⇒(4.7). Throughout this chapter we, of course, focus on the (4.7) and not (4.8).

Let $\epsilon > 0$, $f_1, f_2, ..., f_K \in C_0(\mathbf{R}^6)$. Define the 0-neighborhood U in $\mathcal{M}(\mathbf{R}^6)$ (in the sense of weak* topology of $\mathcal{M}(\mathbf{R}^6)$) such that:

$$U(= U^{\epsilon}_{f_1,\cdots f_K}) = \{\rho \in \mathcal{M}(\mathbf{R}^6)(= C_0(\mathbf{R}^6)^*) \ : \ |{}_{\mathcal{M}(\mathbf{R}^6)}\langle \rho, f_k \rangle_{C_0(\mathbf{R}^6)}| < \epsilon, k = 1,2,...,K\}. \tag{4.9}$$

[6] What is the most important is to recognize that statistical mechanics belongs to the category of theoretical informatics and not that of theoretical physics. (*cf.* Table (1.7)). Thus, the present situation is the same as the following situation. Two ready-made suits (A) and (B) are on sale. The (A) is somewhat big, and the (B) is somewhat small. Which do you choose, (A) or (B)? Cf. (I_{15}) in §1.3. We must choose one from "the staying time interpretation" and "the probabilistic interpretation". In theoretical informatics, it can not be decided by experimental test. What we can say is we believe that "the staying time interpretation" will win more popularity than "the probabilistic interpretation".

[7] Note that this notation is different from that of Kolmogorov's probability theory, in which we use the phrase "almost every t in the sense of (PR)" when $\bar{\nu}(\mathcal{N}) = 0$.

Put $D_N = \{1, 2, ..., N(\approx 10^{24})\}$. For each k ($\in D_N \equiv \{1, 2, ..., N(\approx 10^{24})\}$), define the map $X_k : \mathcal{S}_E(\subset \mathbf{R}^{6N}) \to \mathbf{R}^6$ such that

$$X_k((q_{1n}, q_{2n}, q_{3n}, p_{1n}, p_{2n}, p_{3n})_{n=1}^N) = (q_{1k}, q_{2k}, q_{3k}, p_{1k}, p_{2k}, p_{3k}) \tag{4.10}$$

for all $(q, p) = (q_{1n}, q_{2n}, q_{3n}, p_{1n}, p_{2n}, p_{3n})_{n=1}^N$ in $\mathcal{S}_E(\subset \mathbf{R}^{6N})$. For any subset D ($\subseteq D_N \equiv \{1, 2, ..., N(\approx 10^{24})\}$), define the map $R_D^{(\cdot)} : \mathcal{S}_E(\subset \mathbf{R}^{6N}) \to \mathcal{M}_{+1}^m(\mathbf{R}^6)$ ($\equiv \{\rho \in \mathcal{M}(\mathbf{R}^6) : \rho \geq 0, \ \rho(\mathbf{R}^6) = 1\}$) such that

$$R_D^{(q,p)} = \frac{1}{\sharp[D]} \sum_{k \in D} \delta_{X_k(q,p)} \qquad (\forall (q, p) \in \mathcal{S}_E(\subset \mathbf{R}^{6N})), \tag{4.11}$$

where $\sharp[D]$ is the number of the elements of D and δ_x is a point measure at x ($\in \mathbf{R}^6$).

Let U be a 0-neighborhood in $\mathcal{M}(\mathbf{R}^6)$ such as defined in (4.9). For any (p, q) ($\in \mathcal{S}_E$), put

$$H_U(p, q) = k_B \log \big[\nu_E(\{(p', q') \in S_E \mid R_{D_N}^{(p,q)} - R_{D_N}^{(p',q')} \in U\})\big], \tag{4.12}$$

(k_B is the Boltzmann constant, i.e., $k_B = 1.381 \times 10^{-23}$J/K), which is called the *U-entropy* of a state (p, q).

Let $D_0 \subseteq D_N$. Define $\overline{\nu}_E \circ ((X_k)_{k \in D_0})^{-1}$ ($\in \mathcal{M}_{+1}^m(\mathbf{R}^{6 \times \sharp[D_0]})$) by the image measure concerning the map $(X_k)_{k \in D_0} : \mathbf{R}^{6N} \to \mathbf{R}^{6 \times \sharp[D_0]}$, that is,

$$\overline{\nu}_E \circ ((X_k)_{k \in D_0})^{-1}(\bigtimes_{k \in D_0} A_k) = \overline{\nu}_E(\{(p, q) \in \mathcal{S}_E \mid X_k(p, q) \in A_k \ (k \in D_0)\}) \tag{4.13}$$

for any open set A_k ($\subseteq \mathbf{R}^6$) ($k \in D_0$).

In what follows we shall represent the conditions (A_2^1) and (A_2^2) (mentioned in §4.1) in terms of mathematics. Cf. [45].

Definition 4.3. [Thermodynamical condition, equilibrium state][8] *Let D_N be a set $\{1, 2, ..., N(\approx 10^{24})\}$. And let $\mathcal{H}$, E, ν_E, $\overline{\nu}_E$, $X_k : \mathcal{S}_E \to \mathbf{R}^6$ be as in the above. A Hamiltonian $\mathcal{H}$ on $\mathbf{R}^{6N}$ ($N \approx 10^{24}$) is said to be thermodynamical (concerning energy E) if the following condition (T) is satisfied:*

(T) $\{X_k : \mathcal{S}_E \to \mathbf{R}^6\}_{k=1}^N$ *is an almost independent sequence with the identical distribution.*

[8] Although this condition may be superficial and not fundamental, we believe, from the measurement theoretical point of view, that equilibrium statistical mechanics should start from this condition. Again note that our purpose is to understand equilibrium statistical mechanics as one of applications of PMT and not to derive equilibrium statistical mechanics from Newtonian mechanics.

In other words, there exists a normalized measure $\rho_{_E}$ *on* $\mathbf{R}^6$ *(i.e.,* $\rho_{_E} \in \mathcal{M}_{+1}^m(\mathbf{R}^6)$*) such that:*

(T^1) [*identical distribution, cf.* (A_2^1) *in* §4.1] *it holds that*

$$\rho_{_E} \approx \overline{\nu}_{_E} \circ X_k^{-1} \qquad (\forall k = 1, 2, ..., N(\approx 10^{24})), \tag{4.14}$$

(T^2) [*independence, cf.* (A_2^2) *in* §4.1] *it holds that*

$$\bigotimes_{k \in D_N} \rho_{_E}(: \text{product measure}) \approx \overline{\nu}_{_E} \circ ((X_k)_{k \in D_N})^{-1}, \tag{4.15}$$

though the condition (T^2) *is too strong to assume it literally, (see Remark 4.4).*

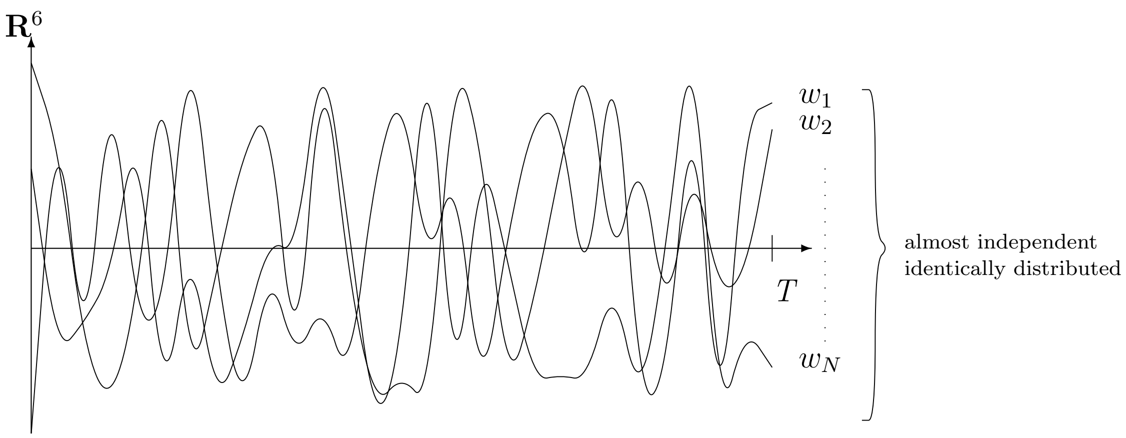
Here, a state (q, p) $(\in \mathcal{S}_E)$ *is called an equilibrium state if* $R_{D_N}^{(q,p)} \approx \rho_E$.[9]

■

Let T be a sufficiently large number. Assume that the closed interval $[0, T]$ has the measure: dt/T (thus, the total measure of $[0, T]$ is equal to 1). For each k $(\in D_N \equiv \{1, 2, ..., N(\approx 10^{24})\})$, define the map $w_k : [0, T] \to \mathbf{R}^6$ such that $w_k(t) = (q_{1k}(t), q_{2k}(t), q_{3k}(t), p_{1k}(t), p_{2k}(t), p_{3k}(t))$ for all t $(\in [0, T])$. Assume that

(♯) $\{w_k \mid k \in D_N\}$ is a set composed of almost independent functions with the identical distribution.

This assumption (♯) is essentially the same as (T) in Definition 4.3.



[9]In our formulation, we do not assume that the "equilibrium state" is defined by $\overline{\nu}_{_E}$ since $\overline{\nu}_{_E}$ is not assumed to have the probabilistic interpretation (*cf.* Remark 4.1).

Remark 4.4. As mentioned in Definition 4.3, the condition (T^2) is too strong. Thus, it should be understood symbolically and not literally. Therefore, we actually assume some hypotheses, which are weaker than the (T_2). For example we assume the following conditions $(T^2)'$ and $(T^2)''$:

$(T^2)'$ [*independence*] *it holds that*

$$\bigotimes_{k \in D_0} \rho_{_E} \approx \overline{\nu}_{_E} \circ ((X_k)_{k \in D_0})^{-1}, \tag{4.16}$$

$(\forall D_0 \subset \{1, 2, ..., N(\approx 10^{24})\}$ *such that* $1 \ll \sharp[D_0] \ll N)$.

This is needed for the derivation of the ergodic hypothesis (*cf.* Theorem 4.6 later). Also, we assume that

$(T^2)''$ [*independence*] *it holds that*

$$\Big(\overline{\nu}_{_E} \circ ((X_k)_{k \in D_1})^{-1}\Big) \bigotimes \Big(\overline{\nu}_{_E} \circ ((X_k)_{k \in D_2})^{-1}\Big) \approx \overline{\nu}_{_E} \circ ((X_k)_{k \in D_1 \cup D_2})^{-1}, \tag{4.17}$$

for any D_1, D_2 $(\subset D)$ *such that* $D_1 \cap D_2 = \emptyset$ *and* $1 \ll \sharp[D_1], \sharp[D_2] \leq N$.

That is because, in equilibrium statistical mechanics, we usually assume that the interaction between the subsystem composed of the particles D_1 and that of the particles D_2 can be neglected.

■

Remark 4.5. (i) If N_0 is arbitrarily large (and thus $N = \infty$) and if the approximation symbol "≈" is interpreted by the equality "=", then (4.4) and (4.16) imply that the sequence $\{X_k\}_{k=1}^{\infty}$ on the normalized averaging staying time space $(\mathcal{S}_{_E}, \mathcal{B}_{\mathcal{S}_{_E}}, \overline{\nu}_{_E})$ is an independent sequence with the identical distribution $\rho_{_E}$. Thus, Lemma 4.9 (i.e., the law of large numbers) says that

$$\lim_{N_0 (= \sharp[D]) \to \infty} R_D^{(q,p)} = \rho_{_E} \quad (\text{ in the sense of the weak}^* \text{ topology of } \mathcal{M}(\mathbf{R}^6)) \tag{4.18}$$

holds for almost every (q, p) in $(\mathcal{S}_{_E}, \mathcal{B}(\mathcal{S}_{_E}), \overline{\nu}_{_E})$. Note that Kolmogorov's probability theory [56] is mathematics, and therefore, it is valid even if the probabilistic interpretation (*cf.* Remark 4.1) is not added to the normalized averaging staying time measure space $(\mathcal{S}_{_E}, \mathcal{B}_{\mathcal{S}_{_E}}, \overline{\nu}_{_E})$. For completeness, again note that the terms: "identical distribution" in (T^1) and "independence" in (T^2) are not related to the concept of "probability" (but that of "staying time").

(ii) The reader may doubt if the concepts of "identical distribution" and "independence" are meaningful without the probabilistic interpretation. However, the following example shows that these concepts are not only meaningful on a measure space but also on a topological space. Let $f : \Omega \to \mathbf{R}$ be a continuous function on a topological space Ω. For each $n(=1,2,...,N)$, define the function $f_n : \Omega^N (=$ product topological space$) \to \mathbf{R}$ such that $\Omega^N \ni (\omega_1, \omega_2, ..., \omega_n, ..., \omega_N) \mapsto f(\omega_n) \in \mathbf{R}$. Then we may say that $\{f_n\}_{n=1}^N$ is *"an independent sequence with the identical distribution"*. In fact we often say "The motions of two particles are independent" in Newtonian mechanics (and not in statistical mechanics).

■

By an analogy of the arguments (i.e., the derivation of (4.18)) in the above Remark 4.5(i), we can assert that (4.14) and (4.16) imply that, if $1 \ll N_0(\approx \sharp[D_0]) \ll N(\approx 10^{24})$,

$$R_{D_0}^{(q(t),p(t))} \approx \overline{\nu}_{_E} \circ X_k^{-1} \ (\approx \rho_{_E}) \qquad (\text{ almost every time } t \text{ in the sense of (STI) }) \quad (4.19)$$

holds for any $k\ (=1,2,...,N(\approx 10^{24}))$. Here consider the decomposition $\{D_{(1)}, D_{(2)}, ..., D_{(L)}\}$ of $D_N\ (\equiv \{1,2,...,N(\approx 10^{24})\})$ such that $\sharp[D_{(l)}] \approx N_0$ $(l = 1,2,...,L)$. Then we see, by (4.19), that

$$R_{D_N}^{(q(t),p(t))} = \frac{\sum_{l=1}^{L} [\sharp[D_{(l)}] \times R_{D_{(l)}}^{(q(t),p(t))}]}{N} \approx \frac{\sum_{l=1}^{L} [\sharp[D_{(l)}] \times \rho_{_E}]}{N} = \overline{\nu}_{_E} \circ X_k^{-1} \ (\approx \rho_{_E})$$
$$(\text{ almost every time } t \text{ in the sense of (STI) })$$

holds for any $k\ (=1,2,...,N(\approx 10^{24}))$.

Summing up, we have the following theorem.

Theorem 4.6. (Ergodic hypothesis). *Assume the thermodynamical condition (i.e., (T_1) in Definition 4.3 and $(T^2)'$ in Remark 4.4). Then it holds that*

$$R_{D_N}^{(q(t),p(t))} \approx \overline{\nu}_{_E} \circ X_k^{-1} \ (\approx \rho_{_E}) \quad (k = 1,2,...,N(\approx 10^{24})) \quad (4.20)$$
$$(\text{ almost every time } t \text{ in the sense of (STI) })$$

Thus, the state of the system is almost always equal to the equilibrium state (cf. Definition 4.3). That is, we see:

- $R_{D_N}^{(q(t_1),p(t_1))} \approx R_{D_N}^{(q(t_2),p(t_2))}$ (*almost every time* t_1 *and* t_2 *in the sense of* (*STI*)).

(4.21)

■

This says that

$$\text{"the distribution of } N(\approx 10^{24}) \text{ particles at almost every time } t\text{" (in the sense of (STI))}$$
$$= \text{"normalized averaging staying time of the } k\text{-th particle } (\forall k = 1, 2, ..., N \approx 10^{24})\text{"} \tag{4.22}$$

We believe that this is just what should be represented by the *"ergodic hypothesis"* :[10]

$$\text{"population average of many particles"} = \text{"time average of one particle"}, \tag{4.23}$$

that is, we see that (4.20)=(4.22)=(4.23).

Remark 4.7. [Another formulation of equilibrium statistical mechanics]. For completeness, note that the condition $(T^2)'$ in (4.16) is assumed in order that (4.21) holds. Thus some may assert that it suffices to start from the $\mathcal{S}_E$ (with the measure ν_E which induces (STI)) and the (4.21). This formulation may be called the formulation without the ergodic hypothesis. Also, see the formula (4.29) later.

■

Remark 4.8. (i). If the probabilistic interpretation (i.e., the principle of equal a priori probability) is assumed, in (4.20) we can replace "almost every time t in the sense of (STI)" to "almost every time t in the sense of (PR)". However, if the (STI) is accepted as a common sense, we can do well without this replacement, that is, the replacement does not bring us any merit. Thus we think that the probabilistic interpretation is not needed. Cf. Remark 4.5

(ii). We may still have a question:

- *Why is the thermodynamical condition (i.e.,* (T^1) *and* (T^2)*) always satisfied in the usual circumstance of equilibrium statistical mechanics?*

Though we do not know the firm answer,[11] we can easily show, by (4.20), that the thermodynamical condition ((T^1) and (T^2)) explains the following law (i.e., *"the law of increasing*

[10] In this book, the term "ergodic hypothesis" has two meanings. One is used in the sense of the formula (4.6). And the other is used in the sense of the formula (4.23) (or, Theorem 4.6).

[11] If we think that statistical mechanics belongs to informatics and not physics (*cf.* in this book we consider so), the firm answer may not be needed. If the thermodynamical condition is useful, it is enough.

entropy").[12]

(IE) *the U-entropy $H_U(q(t), p(t))$, (cf. (4.12)), is increasing concerning t, that is*

$$H_U(q(t), p(t)) \uparrow \log[\nu(\mathcal{S}_E)] \quad (\text{ if } t \uparrow \infty) \quad \text{in the sense of (STI)} \tag{4.24}$$

for a suitable small 0-neighborhood U in $\mathcal{M}(\mathbf{R}^6)$.

That is because $H_U(q(t), p(t)) \approx \log[\nu(\mathcal{S}_E)]$ holds for almost every time t in the sense of (STI) if the neighborhood U is chosen suitably. (How to choose the U suitably is our future problem.) Therefore we consider that the law of increasing entropy is hidden behind the thermodynamical condition ((T^1) and (T^2)).

■

4.3 Probabilistic aspects of equilibrium statistical mechanics

In this section we shall study the probabilistic aspects of equilibrium statistical mechanics. Note that the (4.20) implies that the equilibrium statistical mechanical system at almost every time t (in the sense of the (STI)) can be regarded as:

(U) an urn including about 10^{24} particles such as the number of the particles whose states belong to Ξ ($\in \mathcal{B}_{\mathbf{R}^6}$) is given by $\rho_E(\Xi) \times 10^{24}$.

Recall the (A_1) in §4.1, that is, the probability appearing in classical systems (or particularly, in the probabilistic rule in (4.2)) is essentially the same as the probability appearing in urn problems. Therefore, we see, by the above (U),

(A_1') if we choose a particle at random from the urn (="box in Figure (4.3)") at time t, then the probability that the state of the particle belongs to Ξ ($\in \mathcal{B}_{\mathbf{R}^6}$) is given by $\rho_E(\Xi)$.

[12]If my memory serves me right, in some book A. Einstein says: *There is a possibility that someone will find his relativity theory is not true, but there is no possibility that someone will find that the law of increasing entropy is not true.* We can understand what he wants to say, if we think that statistical mechanics should be understood as an application of measurement theory, on the other hand, his relativity theory belongs to theoretical physics. That is, we think that the law of increasing entropy is as "true" as the statement "A cat is stronger than a mouse". (Cf. footnote[9] in Chapter 2.) It should be noted that the statement "A cat is stronger than a mouse" is ambiguous, fuzzy, vague, etc, though it is "almost experimentally true" (*cf.* (I_{14}) in §1.3).

In what follows, we shall represent this (A_1') in terms of measurements. Define the observable $\mathbf{O} = (\mathbf{R}^6, \mathcal{B}_{\mathbf{R}^6}, F)$ in $C(\mathcal{S}_E)$ such that, (*cf.* (4.11)),

$$[F(\Xi)](q,p) = [R_{D_N}^{(q,p)}](\Xi)\Big(\equiv \frac{\sharp[\{k \mid X_k(q,p) \in \Xi\}]}{\sharp[D_N]}\Big) \qquad (\forall \Xi \in \mathcal{B}_{\mathbf{R}^6}, \forall (q,p) \in \mathcal{S}_E(\subset \mathbf{R}^{6N})). \tag{4.25}$$

Thus, we have the measurement $\mathbf{M}_{C(\mathcal{S}_E)}(\mathbf{O} \equiv (\mathbf{R}^6, \mathcal{B}_{\mathbf{R}^6}, F), S_{[\delta_{\psi_t(q_0,p_0)}]})$. Then we see that

(B_1') the probability that the measured value obtained by the measurement $\mathbf{M}_{C(\mathcal{S}_E)}(\mathbf{O} \equiv (\mathbf{R}^6, \mathcal{B}_{\mathbf{R}^6}, F), S_{[\delta_{\psi_t(q_0,p_0)}]})$ belongs to $\Xi (\in \mathcal{B}_{\mathbf{R}^6})$ is given by $\rho_E(\Xi)$. That is because Theorem 4.6 says that

$$[F(\Xi)](\psi_t(q_0,p_0)) = \rho_E(\Xi) \quad \text{(almost every time } t \text{ in the sense of (STI))} \tag{4.26}$$

which is just the measurement theoretical representation of the (A_1').

Also, we see that

(A_1'') if we choose N' particles at random from the urn (="box in Figure (4.3)"), then statistics say that the distribution of the states of these particles is almost surely expected to be approximately equal to ρ_E, where $1 \ll N' \le N(\approx 10^{24})$.

Here, consider the product observable $\mathbf{O}^{N'} = (\mathbf{R}^{6N'}, \mathcal{B}_{\mathbf{R}^{6N'}}, F^{N'})$ in $C(\mathcal{S}_E)$. For each k $(\in K_{N'} \equiv \{1,2,...,N'\})$, define the map $X_k : \mathbf{R}^{6N'} \to \mathbf{R}^6$ such that

$$X_k((x_{1n}, x_{2n}, x_{3n}, x_{4n}, x_{5n}, x_{6n})_{n=1}^{N'}) = (x_{1k}, x_{2k}, x_{3k}, x_{4k}, x_{5k}, x_{6k})$$

for all $x = (x_{1n}, x_{2n}, x_{3n}, x_{4n}, x_{5n}, x_{6n})_{n=1}^{N'}$ in $\mathbf{R}^{6N'}$). Define the map $G_{N'} : \mathbf{R}^{6N'} \to \mathcal{M}_{+1}^m(\mathbf{R}^6)$ $(\equiv \{\rho \in \mathcal{M}(\mathbf{R}^6) : \rho \ge 0,\ \rho(\mathbf{R}^6) = 1\})$ such that

$$G_{N'}(x) = \frac{1}{N'} \sum_{n=1}^{N'} \delta_{X_n(x)} \qquad (\forall x \in \mathbf{R}^{6N'}). \tag{4.27}$$

Then we have the image observable $G_{N'}(\mathbf{O}^{N'}) \equiv (\mathcal{M}_{+1}^m(\mathbf{R}^6), \mathcal{B}_{\mathcal{M}_{+1}^m(\mathbf{R}^6)}, G_{N'}(F^{N'}))$. And we see, by Theorem 4.6, that

(B_1'') the measured value obtained by the measurement $\mathbf{M}_{C(\mathcal{S}_E)}(G_{N'}(\mathbf{O}^{N'}), S_{[\delta_{\psi_t(q_0,p_0)}]})$ is approximately equal to ρ_E.

which is just the measurement theoretical representation of the (A_1'').

4.4 Conclusions

In this chapter we assert that equilibrium statistical mechanics is formulated as follows:[13]

$$\text{"equilibrium statistical mechanics"} = \underbrace{\underset{((B_1'')(=\text{Axiom 1}))}{\text{"probabilistic rule"}} + \underset{((T^1)\text{ and }(T^2))\text{ under (EH))}}{\text{"Newton equation"}}}_{(+\text{ STI})} \tag{4.28}$$

(=(4.4))

in the framework of PMT.

It may be generally believed that the principle of equal a priori probability and the ergodic hypothesis are two basic principles of statistical mechanics. However, our formulation (4.28) says that the principle of equal a priori probability is not needed (*cf.* Remark 4.5 and Remark 4.8(i)), and moreover, the ergodic hypothesis is a consequence of the thermodynamical condition (i.e., (T^1) and (T^2) under the (EH)), *cf.* the formulas (4.20)∼(4.23).

However we may assert that the following formulation is also possible:

$$\text{"equilibrium statistical mechanics"} = \underbrace{\underset{((B_1'')(=\text{Axiom 1}))}{\text{"probabilistic rule"}} + \underset{((T^1)\text{ and }(T^2))\text{ under (EH))}}{\text{"Newton equation"}}}_{(+\text{ PI})} \tag{4.30}$$

which is, strictly speaking, related to SMT (*cf.* Chapter 8, Statistical measurement theory).

Thus we have the question:

- Which should be chosen, (4.28) or (4.30)?[14]

---

[13]Or simply (*cf.* Remark 4.7), we may consider that

$$\text{"equilibrium statistical mechanics"} = \underbrace{\underset{((B_1'')(=\text{Axiom 1}))}{\text{"probabilistic rule"}} + \underset{\nu_E\text{ (in (4.5)) and (4.21)}}{\text{"Newton equation"}}}_{(+\text{ STI})} \tag{4.29}$$

We believe that the term "economical" is one of the most important key-words of theoretical informatics (*cf.* Table (1.8b)). In this sense, the (4.29) should be also admitted though we did not focus on the (4.29) in this chapter.

[14]This situation is the same as the following situation. Two ready-made suits "the staying time interpretation" and "the probabilistic interpretation" are on sale. The former is too weak, and so somewhat ambiguous. The latter may be too strong. However, we must choose one from "the staying time interpretation" and "the probabilistic interpretation". In theoretical informatics, we believe that it can

The reason that we choose (4.28) is as follows: Recall quantum mechanics, in which it is often emphasized that the concept of "probability" is not related to "Schrödinger equation" but "Born's quantum measurements". Comparing quantum mechanics (1.3) and the above (4.28), we have the reason to emphasize that the concept of "probability" is not related to the thermodynamical condition but "probabilistic rule in (4.28)". That is because we want to believe in the spirit that the term of "probability" should be used commonly in both classical and quantum systems, or, that there is no probability without measurements. After all, we say that

- Our proposal (4.28) and quantum mechanics (1.3) are compatible.

On the other hand, the part "Newton equation ((T^1) and (T^2)) under (EH))" in (4.30) is related to the concept of "probability" under the assumption "probabilistic interpretation of $\overline{\nu}_E$". Thus, we think that

- The (4.30) and quantum mechanics (1.3) are not compatible.

Thus, we do not choose the (4.30). However, we may choose the following (4.18):

$$\text{“equilibrium statistical mechanics”} = \underbrace{\underset{((B_1'')(=\text{Axiom 1}))}{\text{“probabilistic rule”}}}_{(+\text{PI})} + \underbrace{\underset{((T^1)\text{ and }(T^2))\text{ under (EH))}}{\text{“Newton equation”}}}_{(+\text{STI})} \tag{4.31}$$

This (4.31)[15] and quantum mechanics (1.3) are compatible. Thus, the following question is meaningful in measurement theory.

- Which should be chosen, (4.28) or (4.31)?

This may be a matter of opinion (though it is not serious as statistical mechanics is assumed to belong to theoretical informatics in this chapter). If we are required to say something, we guess that the (4.28) will win more popularity than the (4.31). In fact,

---

not be decided by an experimental test. Or at least, we are convinced that it is not worthwhile deciding it by an experimental test. That is because we believe that nobody wants to challenge the following problem:

- Decide (4.28) or (4.30) (or (4.31)) by an experimental test!

Thus, "(4.28) or (4.30)" should be chosen from the philosophical point of view, if we are urged to choose one. Cf. (I_{15}) in §1.3.

[15]The part $\underbrace{\underset{((B_1'')(=\text{Axiom 1}))}{\text{“probabilistic rule”}}}_{(\text{PI})}$ in (4.31) is characterized as "Proclaim 1" in Chapter 8.

- we prefer (4.28) to (4.31),

since we do not want use the (PI) if possible.[16] This is our opinion, though, in theoretical informatics, we must admit the case that opinion is divided.

We hope that our proposal (4.28) (or, (4.29), (4.31)) will be accepted as the standard formulation of equilibrium statistical mechanics.

4.5 Appendix (The law of large numbers)

As a preparation of our main assertion (i.e., the derivation of the ergodic hypothesis (4.20)), we add the following well-known Lemma 4.9.

Lemma 4.9. [The strong law of large numbers, *cf.* [56]]. *Let* $(\mathcal{S}, \mathcal{B}_{\mathcal{S}}, \nu)$ *be a measure space such that* $\nu(\mathcal{S}) < \infty$. *Let* $\{X_n\}_{n=1}^{\infty}$ *be a sequence of bounded measurable (or generally,* L^1*) maps* $X_n : \mathcal{S} \to \mathbf{R}^6$ *such that there exists a normalized measure* ρ *on* $\mathbf{R}^6$ *(i.e.,* $\rho(\mathbf{R}^6) = 1$, $\rho(\Gamma) \geq 0$ $(\forall \Gamma \in \mathcal{B}_{\mathbf{R}^6})$*) such that:*

- (identical distribution)

$$\frac{\nu(\{x \in \mathcal{S} \mid X_n(x) \in \Gamma\})}{\nu(\mathcal{S})} = \rho(\Gamma) \quad (\forall n = 1, 2, ..., \quad \forall \Gamma \in \mathcal{B}_{\mathbf{R}^6}),$$

- (independence) *for any positive integer* N*, it holds that*

$$\frac{\nu(\{x \in \mathcal{S} \mid X_n(x) \in \Gamma_n \ \ (\forall n = 1, 2, ..., N)\})}{\nu(\mathcal{S})} = \mathop{\times}_{n=1}^{N} \rho(\Gamma_n) \qquad (\forall \Gamma_n \in \mathcal{B}_{\mathbf{R}^6}).$$

Then, there exists a measurable set $\mathcal{N}(\in \mathcal{B}_{\mathcal{S}})$ *such that* $\nu(\mathcal{N}) = 0$ *and*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n(x)} = \rho \quad \text{in the sense of weak}^* \text{ topology of } \mathcal{M}(\mathbf{R}^6),$$

for all $x \in \mathcal{S} \setminus \mathcal{N}$ $(\equiv \{x \mid x \in \mathcal{S}, x \notin \mathcal{N}\})$. *Here* $\delta_w(\in \mathcal{M}_{+1}^m(\mathbf{R}^6))$ *is a point measure at* $w(\in \mathbf{R}^6)$*, i.e.,* $\delta_w(\Gamma) = 1$ *(if* $w \in \Gamma \in \mathcal{B}_{\mathbf{R}^6}$*),* $= 0$ *(if* $w \notin \Gamma \in \mathcal{B}_{\mathbf{R}^6}$*).*

■

In the formula (4.18), readers should see that Lemma 4.9 is used in the part "Newton equation" (and not "probability rule") in our proposal (4.28), that is, Lemma 4.9 (the law of large numbers) is used independently of the concept of "probability".

[16] Also, recall "Occam's razor", that is, *"Given two equally predictive theories, choose the simplest".*

Chapter 5

Fisher's statistics I (under Axiom 1)

As mentioned in Chapters 2 and 3, measurement theory is formulated as follows:

$$\text{PMT} = \underset{\text{[Axiom 1 (2.37)]}}{\text{measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} \qquad \text{in } C^*\text{-algebra} \quad . \tag{5.1}$$
(=(1.4))

In this chapter we intend to understand Fisher's statistics in Axiom 1. The reader will see that Fisher's maximum likelihood method is a direct consequence of Axiom 1.[1] And further, we discuss "inference interval" and "testing statistical hypothesis" in Axiom 1. By the results obtained in this chapter (and in the next chapter), we conclude that Fisher's statistics is theoretically true. (Cf. "Declaration (1.11)" in §1.4.)[2]

5.1 Introduction

The first attempt of the measurement theoretical approach to statistics was proposed in [44]. Although the argument in [44] is not deep, at least it convinces us of the good possibility of the axiomatic formulation (i.e., the measurement theoretical formulation) of statistics.

Most statisticians consider that statistics is closely related to "measurements", or, statistics is the study to analyze "measured data" for some purpose. Therefore, PMT should be immediately examined in comparison with statistics. The purpose of this chapter is to execute it, in other words, to propose a measurement theoretical formulation of statistics. We think that statistics is mainly related to the following aspect of measurement theory:

[1]Readers are not required to have much knowledge of statistics.

[2]We believe that the philosophy of statistics should be more discussed in statistics, (Cf. [61]). That is because it is indispensable for the understanding of "statistics (= mathematics + something)". It should be noted that "to formulate statistics in the framework of MT" implies "to introduce the philosophy of MT into statistics".

(♯) how to derive some useful information from the measured data obtained by a measurement.

Let $\mathbf{M}_{\mathcal{A}}\big(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[\rho^p]}\big)$ be a measurement formulated in a C^*-algebra $\mathcal{A}$. Recall the (III) in §2.5 [Remarks], that is, the measurement $\mathbf{M}_{\mathcal{A}}\big(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[\rho^p]}\big)$ always determines the sample space $\Big(X, \overline{\mathcal{F}}, {}_{\mathcal{A}^*}\Big\langle \rho^p, F(\,\cdot\,)\Big\rangle_{\mathcal{A}}\Big)$. Here note that the mathematical structure of the sample space $\Big\{ {}_{\mathcal{A}^*}\Big\langle \rho^p, F(\Xi)\Big\rangle_{\mathcal{A}}\Big\}_{\rho^p \in \mathfrak{S}^p(\mathcal{A}^*), \Xi \in \overline{\mathcal{F}}}$ is the same as that of the conventional formulation of statistics $\Big($i.e, $\Big\{P(\Xi, \theta)\Big\}_{\theta \in \Theta, \Xi \in \overline{\mathcal{F}}}$, where, for each θ in a parameter space Θ, $P(\cdot, \theta)$ is a probability measure on a measurable space $(X, \overline{\mathcal{F}})$, *cf.* [86]$\Big)$. Therefore, there is good hope that statistics can be described in terms of measurements. Also, this is precisely our motivation in this chapter. Following the common knowledge of quantum mechanics, we believe that any scientific statement including the term "probability" is not meaningful without the concept of "measurement". (*cf.* §2.5. Remarks). As mentioned in the above, the term "state" in measurement theory corresponds to the term "parameter" in statistics. The reason that we use the term "state" is due to that we want to stress that PMT is constructed modeled on mechanics.[3]

5.2 Fisher's maximum likelihood method

The purpose of this section is to study and understand "Fisher's maximum likelihood method" completely under Axiom 1 (of measurement theory). The following Problem 5.1 is the most typical in all examples of "Fisher's maximum likelihood method".

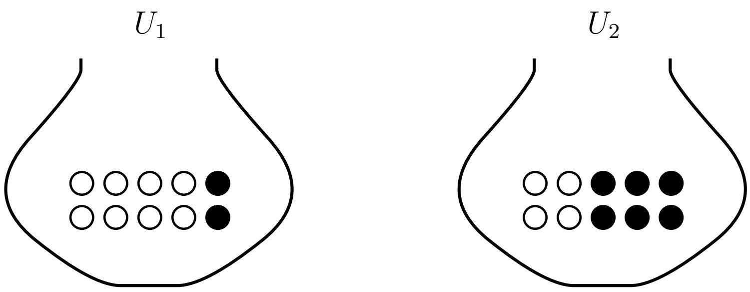
5.2.1 Fisher's maximum likelihood method

Problem 5.1. [The urn problem by Fisher's maximum likelihood method]. There are two urns U_1 and U_2. The urn U_1 [resp. U_2] contains 8 white and 2 black balls [resp. 4 white and 6 black balls].

---

[3]This means that we study statistics by an analogy of "mechanics". Note the following correspondence:

$$\underset{\text{[represented by pure state]}}{\text{system } S_{[\rho^p]} \text{ (in PMT)}} \iff \underset{\text{[represented by parameter]}}{\text{population (in the conventional statistics)}}$$

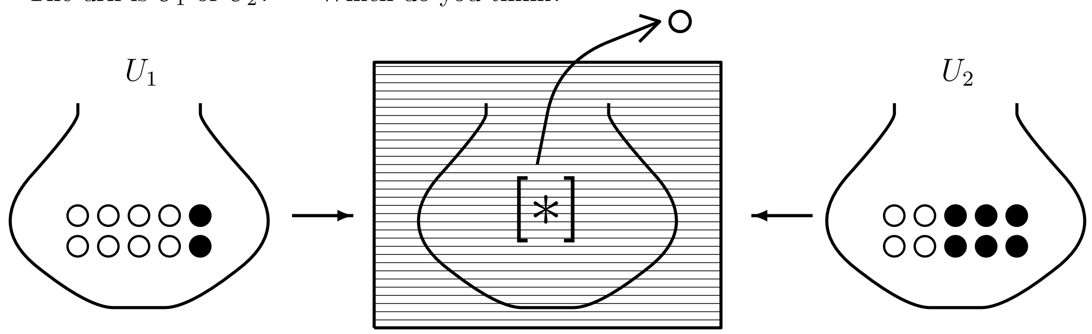


Here consider the following procedures (P_1) and (P_2).

(P_1) One of the two (i.e., U_1 or U_2) is chosen and is settled behind a curtain. Note, for completeness, that you do not know whether it is U_1 or U_2.

(P_2) Pick up a ball out of the urn chosen by the procedure (P_1). And you find that the ball is white.

You do not know which the urn behind the curtain is, U_1 or U_2.
Assume that you pick up a white ball from the urn.
The urn is U_1 or U_2? Which do you think?



Now we have the following question:

(Q) Which is the chosen urn (behind the curtain), U_1 or U_2?

This is quite easy. That is, everyone will immediately infer "the urn behind the curtain = U_1". However, it is just "Fisher's maximum likelihood method". Cf. Example 5.8.

■

We begin with the following definition.

Notation 5.2. $[\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[*]})]$. *Consider a measurement* $\mathbf{M}_{\mathcal{A}}(\mathbf{O} \equiv (X, \mathcal{F}, F),\ S_{[\rho^p]})$ *formulated in a* C^**-algebra* $\mathcal{A}$. *In most measurements, it is usual to think that the state* ρ^p $(\in \mathfrak{S}^p(\mathcal{A}^*))$ *is unknown. That is because the measurement* $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[\rho^p]})$ *may be taken* *in order to know the state ρ^p. Thus, when we want to stress that we do not know the state ρ^p, the measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[\rho^p]})$ is often denoted by $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[*]})$.*

■

By using this notation, we can state our present problem as follows:

(I) Infer the unknown state $[*]$ $(\in \mathfrak{S}^p(\mathcal{A}^*))$ from the measured data obtained by the measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[*]})$.

In order to answer this problem, in [44] we introduced Fisher's method (precisely, Fisher's max mum likelihood method) as follows: (Strictly speaking, Theorem 5.3 should not be called "theorem" but "assertion", since it is not a purely mathematical result but a consequence of Axiom 1.)

Theorem 5.3. [Fisher's maximum likelihood method in classical and quantum measurements, *cf.* [44]]. *Consider a measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[*]})$ formulated in a C^*-algebra $\mathcal{A}$. When we know that the measured value obtained by the measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[*]})$ belongs to Ξ $(\in \mathcal{F})$, there is a reason to infer that the state $[*]$ of the system S is equal to ρ_0^p $(\in \mathfrak{S}^p(\mathcal{A}^*))$ such that:*

$$ {}_{\mathcal{A}^*}\langle \rho_0^p, F(\Xi) \rangle_{\mathcal{A}} = \max_{\rho^p \in \mathfrak{S}^p(\mathcal{A}^*)} {}_{\mathcal{A}^*}\langle \rho^p, F(\Xi) \rangle_{\mathcal{A}}. \tag{5.2} $$

Here, note, for completeness, that the state $[*]$ (in $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[*]})$) is the state before the measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[*]})$. (Cf. Corollary 5.6 later.) Although the ρ_0^p in (5.2) is not generally determined uniquely, in this book we usually assume the uniqueness.

Proof (*or, Explanation*). Let ρ_1^p and ρ_2^p be elements in $\mathfrak{S}^p(\mathcal{A}^*)$. Assume that "$[\,*\,] = \rho_1^p$ " or "$[\,*\,] = \rho_2^p$ ". And assume that $\rho_1^p(F(\Xi)) < \rho_2^p(F(\Xi))$. Then, Axiom 1 says that the fact that the measured value obtained by the $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[\rho_1^p]})$ belongs to Ξ happens more rarely than the fact that the measured value obtained by the $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[\rho_2^p]})$ belongs to Ξ happens. Thus, there is a reason to regard the unknown state $[*]$ as the state ρ_2^p and not as the state ρ_1^p. Also, examining this proof, we can easily see that the state $[*]$ (in $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[*]})$) is the state before the measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[*]})$. This completes the proof. □

Remark 5.4. [Radon-Nikodým derivative]. Assume that there exists a measure ν on

$(X, \overline{\mathcal{F}})$ (*cf.* (III) in §2.5) and $f(\cdot, \rho^p) \in L^1(\Omega, \nu)$ $(\forall \rho^p \in \mathfrak{S}^p(\mathcal{A}^*))$ such that:

$$\rho^p(F(\Xi)) = \int_\Xi f(x, \rho^p)\nu(dx) \quad (\forall \Xi \in \overline{\mathcal{F}}, \forall \rho^p \in \mathfrak{S}^p(\mathcal{A}^*)). \tag{5.3}$$

Then, even if $\Xi = \{x_0\}$ and $\rho^p(F(\{x_0\}) = 0$ $(\forall \rho^p \in \mathfrak{S}^p(\mathcal{A}^*)$) in Theorem 5.3, we may calculate as follows:

$$\frac{\rho_1^p(F(\{x_0\}))}{\rho_2^p(F(\{x_0\}))} = \lim_{\Xi \to \{x_0\}} \frac{\rho_1^p(F(\Xi))}{\rho_2^p(F(\Xi))} = \frac{f(x_0, \rho_1^p)}{f(x_0, \rho_2^p)}. \tag{5.4}$$

In this sense (or, in the sense of "Radon-Nikodým derivative"), we can compare $\rho_1^p(F(\{x_0\}))$ with $\rho_2^p(F(\{x_0\}))$, even when $\rho_1^p(F(\{x_0\})) = \rho_2^p(F(\{x_0\})) = 0$. When we know that the measured value x_0 $(\in X)$ is obtained by the measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[*]})$, by the same reason in Theorem 5.3, we can infer that the state $[*]$ of the system S is equal to ρ_0^p $(\in \mathfrak{S}^p(\mathcal{A}^*))$ such that:

$$f(x_0, \rho_0^p) = \max_{\rho^p \in \mathfrak{S}^p(\mathcal{A}^*)} f(x_0, \rho^p).$$

Here, the map $E : X \to \mathfrak{S}^p(\mathcal{A}^*)$, (i.e., $X \ni x_0 \mapsto \rho_0^p \in \mathfrak{S}^p(\mathcal{A}^*)$), is called "Fisher's estimator".

■

We begin with the following corollary, which is used in the proof of Corollary 5.6 and our main assertion (i.e., Regression Analysis II (in Chapter 6)).

Corollary 5.5. [The conditional probability representation of Fisher's method, *cf.* [55]]. *Let* $\mathbf{O} \equiv (X, \mathcal{F}, F)$ *and* $\mathbf{O}' \equiv (Y, \mathcal{G}, G)$ *be observables in* $\mathcal{A}$. *Let* $\widehat{\mathbf{O}}$ *be a quasi-product observable of* $\mathbf{O}$ *and* $\mathbf{O}'$, *that is,* $\widehat{\mathbf{O}} \equiv \mathbf{O} \overset{\text{qp}}{\times} \mathbf{O}' = (X \times Y, \mathcal{F} \bigotimes \mathcal{G}, F \overset{\text{qp}}{\times} G)$. *Assume that we know that the measured value* (x, y) $(\in X \times Y)$ *obtained by a measurement* $\mathbf{M}_{\mathcal{A}}(\widehat{\mathbf{O}}, S_{[*]})$ *belongs to* $\Xi \times Y$ $(\in \mathcal{F} \bigotimes \mathcal{G})$. *Then, there is a reason to infer that the unknown measured value* y $(\in Y)$ *is distributed under the conditional probability* $P_\Xi(\cdot)$, *where*

$$P_\Xi(\Gamma) = \frac{{}_{\mathcal{A}^*}\langle \rho_0^p, F(\Xi) \overset{\text{qp}}{\times} G(\Gamma) \rangle_{\mathcal{A}}}{{}_{\mathcal{A}^*}\langle \rho_0^p, F(\Xi) \rangle_{\mathcal{A}}} \left(= \frac{\rho_0^p(F(\Xi) \overset{\text{qp}}{\times} G(\Gamma))}{\rho_0^p(F(\Xi))} \right) \quad (\forall \Gamma \in \mathcal{G}), \tag{5.5}$$

where ρ_0^p $(\in \mathfrak{S}^p(\mathcal{A}^*))$ *is defined by*

$$_{\mathcal{A}^*}\langle \rho_0^p, F(\Xi) \rangle_{\mathcal{A}} = \max_{\rho^p \in \mathfrak{S}^p(\mathcal{A}^*)} {}_{\mathcal{A}^*}\langle \rho^p, F(\Xi) \rangle_{\mathcal{A}}. \tag{5.6}$$

Proof. Since we know that the measured value (x, y) $(\in X \times Y)$ obtained by a measurement $\mathbf{M}_{\mathcal{A}}(\widehat{\mathbf{O}}, S_{[*]})$ belongs to $\Xi \times Y$ $(\in \mathcal{F} \bigotimes \mathcal{G})$, we can infer, by Theorem 5.3

(Fisher's method) and the equality $F(\Xi) = F(\Xi) \overset{\text{qp}}{\times} G(Y)$, that the $[\ast]$ (in $\mathbf{M}_{\mathcal{A}}(\widehat{\mathbf{O}}, S_{[\ast]})$) is equal to ρ_0^p $(\in \mathfrak{S}^p(\mathcal{A}^*))$. Thus, the conditional probability that $P_\Xi(\,\cdot\,)$ under the condition that we know that $(x, y) \in \Xi \times Y$ is given by

$$P_\Xi(\Gamma) = \frac{\rho_0^p(F(\Xi) \overset{\text{qp}}{\times} G(\Gamma))}{\rho_0^p(F(\Xi) \overset{\text{qp}}{\times} G(Y))} = \frac{\rho_0^p(F(\Xi) \overset{\text{qp}}{\times} G(\Gamma))}{\rho_0^p(F(\Xi))}. \tag{5.7}$$

This completes the proof. □

The following corollary is the most essential in classical measurements. That is because what we want to infer is usually the state after the measurement (*cf.* Theorem 5.3).

Corollary 5.6. [Fisher's maximum likelihood method in classical measurements, *cf.* [55]]. *Let* $\mathbf{O} \equiv (X, \mathcal{F}, F)$ *be an observable in a commutative* C^**-algebra* $C(\Omega)$*. Assume that we know that the measured value obtained by a measurement* $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\ast]})$ *belongs to* Ξ $(\in \mathcal{F})$*. Then, we can assert the following (i) and (ii):*

(i) *there is a reason to infer that the state* $[\ast]$ *of the system* S $\Big($*i.e., "the state before the measurement* $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\ast]})$*" cf. Fisher's method in classical and quantum measurements*$\Big)$ *is equal to* δ_{ω_0} $(\in \mathcal{M}_{+1}^p(\Omega))$*, where*

$$[F(\Xi)](\omega_0) = \max_{\omega \in \Omega} [F(\Xi)](\omega), \tag{5.8}$$

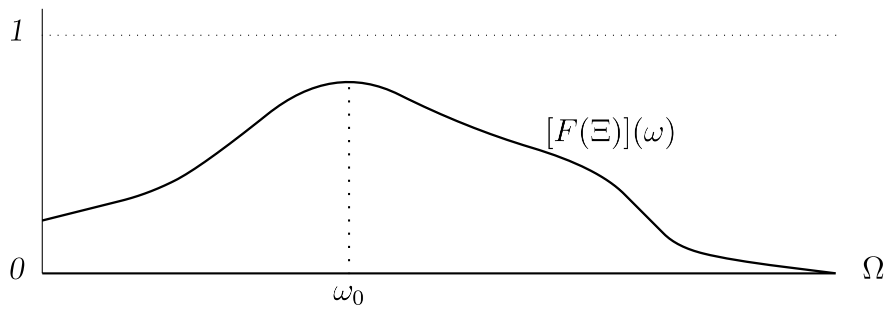
and,

(ii) *there is a reason to infer that the state after the measurement* $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\ast]})$ *is also regarded as the same* δ_{ω_0} $(\in \mathcal{M}_{+1}^p(\Omega))$*.*

Summing up the above (i) and (ii), we see that

(iii) *there is a reason to infer that*

$$[\ast] = \text{"the state after the measurement } \mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\ast]})\text{"} = \delta_{\omega_0}. \tag{5.9}$$



Proof. The (i) is the special case of Fisher's maximum likelihood method (*cf.* Theorem 5.3), i.e., $\mathcal{A} = C(\Omega)$. Thus it suffices to prove (ii) as follows: $\Big($This (ii) will be, under the definition of "S-state" (*cf.* Definition 6.7), proved in Remark 6.12 as a special case of Lemma 6.11 later. In this sense, the proof mentioned here is temporary.$\Big)$ Let $\mathbf{O}' \equiv (Y, \mathcal{G}, G)$ be any observable in $C(\Omega)$. Let $\widehat{\mathbf{O}}$ be the product observable of $\mathbf{O}$ and $\mathbf{O}'$, that is, $\widehat{\mathbf{O}} \equiv \mathbf{O} \times \mathbf{O}' = (X \times Y, \mathcal{F} \bigotimes \mathcal{G}, F \times G)$. Consider a measurement $\mathbf{M}_{C(\Omega)}(\widehat{\mathbf{O}} \equiv (X \times Y, \mathcal{F} \bigotimes \mathcal{G}, F \times G), S_{[*]})$. And assume

(A) we know that the measured value (x, y) $(\in X \times Y)$ obtained by the measurement $\mathbf{M}_{C(\Omega)}(\widehat{\mathbf{O}} \equiv (X \times Y, \mathcal{F} \bigotimes \mathcal{G}, F \times G), S_{[*]})$ belongs to $\Xi \times Y$.

Corollary 5.5 says that there is a reason to infer that the unknown measured value y $(\in Y)$ is distributed under the conditional probability $P_\Xi(\cdot)$, where

$$P_\Xi(\Gamma)[F(\Xi)](\omega_0) = [G(\Gamma)](\omega_0) \quad (\forall \Gamma \in \mathcal{G}), \tag{5.10}$$

where ω_0 $(\in \Omega)$ is defined in (5.8). Also note that the above (A) can be represented by the following two steps (A_1) and (A_2) $\Big($i.e., $(A) = (A_1) + (A_2)\Big)$:

(A_1) we know that the measured value by a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[*]})$ belongs to Ξ $(\in \mathcal{F})$.

and

(A_2) And successively, we take a measurement of the observable $\mathbf{O}' \equiv (Y, \mathcal{G}, G)$, and get a measured value y $(\in Y)$.

$\Big($The above is somewhat metaphorical since "two measurements" seem to appear (*cf.* §2.5[Remarks (II)]).$\Big)$ Comparing (A) and "$(A_1) + (A_2)$", we see, by (5.10), that

$$\text{"the probability that } y \text{ belongs to } \Gamma \ (\in \mathcal{G}) \text{ in } (A_2)\text{"} = [G(\Gamma)](\omega_0) \quad (\forall \Gamma \in \mathcal{G}) \tag{5.11}$$

That is, we get the sample space $(Y, \mathcal{G}, [G(\cdot)](\omega_0))$. Therefore, we say, from the arbitrariness of $\mathbf{O}' \equiv (Y, \mathcal{G}, G)$, that

(A_3) the state after the (A_1) (i.e., the state after the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$) is equal to δ_{ω_0}.

This completes the proof. $\Big($This corollary does not hold in quantum measurements, since the product observable $\widehat{\mathbf{O}} \equiv \mathbf{O} \times \mathbf{O}' = (X \times Y, \mathcal{F} \bigotimes \mathcal{G}, F \times G)$ does not always exist. That is, the concept of "the state after a measurement" is not always meaningful in quantum theory.$\Big)$ □

The "Bayes operator (in the following Remark 5.7)" is hidden in the above proof. This will be more clarified in Remark 6.12 later.

Remark 5.7. [Bayes operator]. Let $\mathbf{O} \equiv (X, \mathcal{F}, F)$ be an observable in $C(\Omega)$. For each $\Xi\ (\in \mathcal{F})$, define the continuous linear operator $B_\Xi^{(0,0)}$ (or, $B_\Xi^{\mathbf{O}}$, $B_\Xi^{\mathbf{O},(0,0)}$) $: C(\Omega) \to C(\Omega)$ such that:

$$B_\Xi^{(0,0)}(g)\Big(\equiv B_\Xi^{\mathbf{O}}(g) \equiv B_\Xi^{\mathbf{O},(0,0)}(g)\Big) = F(\Xi)\ \cdot\ g \qquad (\forall g \in C(\Omega)), \tag{5.12}$$

which is called the *Bayes operator* (or, *the simplest Bayes operator*). Note that it clearly holds that

(i) for any observable $\mathbf{O}_1 \equiv (Y, \mathcal{G}, G)$, there exists an observable $\widehat{\mathbf{O}} \equiv (X \times Y, \mathcal{F} \bigotimes \mathcal{G}, \widehat{F})$ in $C(\Omega)$ such that:

$$\widehat{F}(\Xi \times \Gamma) = B_\Xi^{(0,0)}(G(\Gamma)) \qquad (\Xi \in \mathcal{F}, \Gamma \in \mathcal{G}).$$

That is because it suffices to define $\widehat{\mathbf{O}}$ by the product observable $\mathbf{O} \times \mathbf{O}_1$. Define the map $R_\Xi^{(0,0)} : \mathcal{M}_{+1}^m(\Omega) \to \mathcal{M}_{+1}^m(\Omega)$ (called *"normalized Bayes dual operator"*) such that:

$$R_\Xi^{(0,0)}(\nu) = \frac{[B_\Xi^{(0,0)}]^*(\nu)}{\|[B_\Xi^{(0,0)}]^*(\nu)\|_{\mathcal{M}(\Omega)}} \qquad (\forall \nu \in \mathcal{M}_{+1}^m(\Omega),$$

where $[B_\Xi^{(0,0)}]^* : \mathcal{M}(\Omega) \to \mathcal{M}(\Omega)$ is the dual operator of $[B_\Xi^{(0,0)}]$, that is,

$$[R_\Xi^{(0,0)}(\nu)](D_0) = \frac{\int_{D_0}[F(\Xi)](\omega)\nu(d\omega)}{\int_\Omega [F(\Xi)](\omega)\nu(d\omega)} \qquad (\forall D_0 \in \mathcal{B}_\Omega). \tag{5.13}$$

Thus, we can describe the well known Bayes theorem (*cf.* [86]) such as

$$\mathcal{M}_{+1}^m(\Omega) \ni \nu\ (= \textit{pretest state}) \mapsto\ (\textit{posttest state} =) R_\Xi^{(0,0)}(\nu) \in \mathcal{M}_{+1}^m(\Omega)^4 \tag{5.14}$$

Note that this says that (i)⇒(ii) in Corollary 5.6. That is because a simple calculation shows that $R_\Xi^{(0,0)}(\delta_{\omega_0}) = \delta_{\omega_0}$ in the case of Corollary 5.6. In §6.2, the reader will again

[4] The pretest state [resp. posttest state] may be usually called "priori state" [resp. "posterior state"].

study the Bayes operator in more general situations.

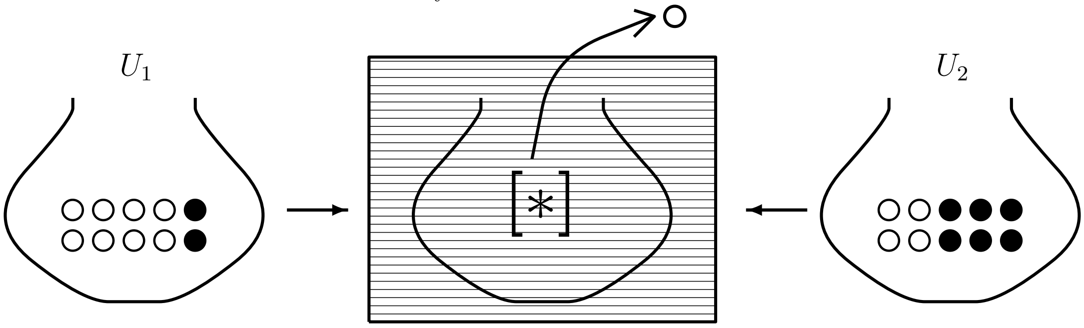
■

Example 5.8. [Continued from Problem 5.1 (Urn problem)[5]]. Recall Example 5.1. That is, consider the following procedures (P_1) and (P_2).

(P_1) One of the two (i.e., U_1 or U_2) is chosen and is settled behind a curtain. Note, for completeness, that you do not know whether it is U_1 or U_2.[6]

(P_2) Pick up a ball out of the urn chosen by the procedure (P_1). And you find that the ball is white.

You do not know which the urn behind the curtain is, U_1 or U_2.
Assume that you pick up a white ball from the urn.
The urn is U_1 or U_2? Which do you think?



Now we have the following question:

(Q) Which is the chosen urn (behind the curtain), U_1 or U_2?

[*Answer*]. Put $\Omega = \{\omega_1, \omega_2\}$. Here,

$$\begin{cases} \omega_1 & \cdots\cdots \text{the state that the urn } U_1 \text{ is behind the curtain} \\ \omega_2 & \cdots\cdots \text{the state that the urn } U_2 \text{ is behind the curtain.} \end{cases} \tag{5.15}$$

In this sense, we frequently use the following identification:

$$U_1 \approx \omega_1, \qquad U_2 \approx \omega_2. \tag{5.16}$$

[5]As mentioned in Example 2.16, we believe that "urn problem" is the most fundamental in all examples of statistics.

[6]Here we are not concerned with SMT_{PEP} (i.e., the principle of equal probability, *cf.* §11.4)

And define the observable $\mathbf{O}\big(\equiv (X \equiv \{w,b\}, 2^{\{w,b\}}, F)\big)$ in $C(\Omega)$ where

$$[F(\{w\})](\omega_1) = 0.8, \qquad [F(\{b\})](\omega_1) = 0.2,$$
$$[F(\{w\})](\omega_2) = 0.4, \qquad [F(\{b\})](\omega_2) = 0.6.$$

Since we do not know whether the state is ω_1 or ω_2, we have the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$. Thus, out situation is

- a measured value "w" is obtained by the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$.

Then, we conclude, by Fisher's maximum likelihood method, that

- the urn behind the curtain is U_1.

That is because

$$[F(\{w\})](\omega_1) = 0.8 = \max\{[F(\{w\})](\omega_1), [F(\{w\})](\omega_2)\}.$$

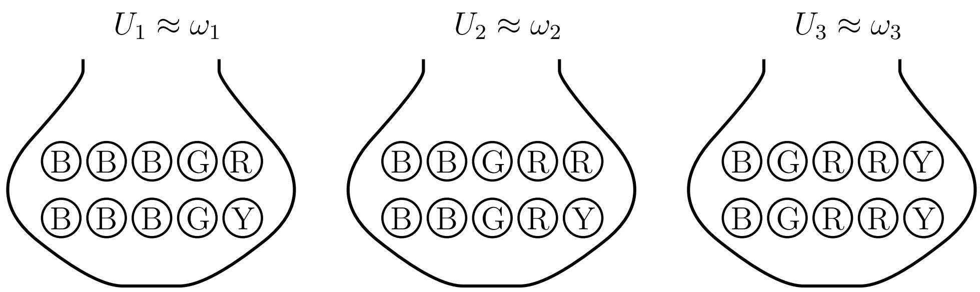
■

Example 5.9. [Urn problem]. Let U_j, $j = 1, 2, 3$, be urns that contain sufficiently many colored balls as follows:

	blue balls	green balls	red balls	yellow balls
urn U_1	60%	20%	10%	10%
urn U_2	40%	20%	30%	10%
urn U_3	20%	20%	40%	20%

(5.17)

Put $\mathbf{U} = \{U_1, U_2, U_3\}$. We consider the state space $\Omega \; \big(\equiv \{\omega_1, \omega_2, \omega_3\}\big)$ with the discrete topology, which is identified with $\mathbf{U}$, that is, $\mathbf{U} \ni U_j \leftrightarrow \omega_j \in \Omega \approx \mathcal{M}^p_{+1}(\Omega)$.[7]



[7]Strictly speaking, we must consider the identification as (5.15).

Define the observable $\mathbf{O} \equiv (X = \{b, g, r, y\}, \mathcal{P}(X), F_{(\cdot)})$ in $C(\Omega)$ by the usual way. That is,

$$\begin{array}{llll} F_{\{b\}}(\omega_1) = 6/10 & F_{\{g\}}(\omega_1) = 2/10 & F_{\{r\}}(\omega_1) = 1/10 & F_{\{y\}}(\omega_1) = 1/10 \\ F_{\{b\}}(\omega_2) = 4/10 & F_{\{g\}}(\omega_2) = 2/10 & F_{\{r\}}(\omega_2) = 3/10 & F_{\{y\}}(\omega_2) = 1/10 \\ F_{\{b\}}(\omega_3) = 2/10 & F_{\{g\}}(\omega_3) = 2/10 & F_{\{r\}}(\omega_3) = 4/10 & F_{\{y\}}(\omega_3) = 2/10. \end{array} \tag{5.18}$$

Then we have the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$.

[I] Now we consider the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$. And assume that we get the measured value 'b' by the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$. Then Fisher's maximum likelihood method (i.e., Corollary 5.6) says that there is a reason to infer that

$$[*] = \omega_1$$

since

$$F_{\{b\}}(\omega_1) = 0.6 = \max_{\omega \in \Omega} F_{\{b\}}(\omega) = \max\{0.6,\ 0.4,\ 0.2\}.$$

That is, the unknown urn $[*]$ is U_1.

[II] Also, consider the (iterated) measurement $\mathbf{M}_{C(\Omega)}(\times_{k=1}^2 \mathbf{O} \equiv (X^2, \mathcal{P}(X^2), \times_{k=1}^2 F), S_{[*]})$ where $(\times_{k=1}^2 F)_{\Xi_1 \times \Xi_2}(\omega) = F_{\Xi_1}(\omega) \cdot F_{\Xi_2}(\omega)$. Also, assume that

- the measured value (b, r) is obtained by the iterated measurement $\mathbf{M}_{C(\Omega)}(\times_{k=1}^2 \mathbf{O}, S_{[*]})$.

Applying Fisher's method (= Corollary 5.6), we get the conclusion as follows: Put

$$E(\omega) = F_{\{b\}}(\omega) F_{\{r\}}(\omega).$$

Clearly it holds that $E(\omega_1) = 6 \cdot 1/10^2 = 0.06$, $E(\omega_2) = 4 \cdot 3/10^2 = 0.12$ and $E(\omega_3) = 2 \cdot 4/10^2 = 0.08$. Therefore, there is a very reason to think that $[\,*\,] = \delta_{\omega_2}$, that is, the unknown urn $[*]$ is U_2.

[III; Remark (moment method)]. Here, let us consider the above [II] by the moment method (*cf.* Definition 2.27). Define the distance Δ on $\mathcal{M}_{+1}^m(X)$ such that:

$$\begin{aligned} \Delta(\nu_1, \nu_2) = & \sum_{x \in X \equiv \{b,g,r,y\}} |\nu_1(\{x\}) - \nu_2(\{x\})| \\ = & |\nu_1(\{b\}) - \nu_2(\{b\})| + |\nu_1(\{g\}) - \nu_2(\{g\})| + |\nu_1(\{r\}) - \nu_2(\{r\})| + |\nu_1(\{y\}) - \nu_2(\{y\})|. \end{aligned}$$

Note that ${}_{\mathcal{M}(\Omega)}\langle \delta_{\omega_1}, F_{\{b\}} \rangle_{C(\Omega)} = \delta_{\omega_1}(F_{\{b\}}) = F_{\{b\}}(\omega_1) = 6/10$, and similarly (*cf.* (5.18)),

$$
\begin{array}{llll}
\delta_{\omega_1}(F_{\{b\}}) = 6/10 & \delta_{\omega_1}(F_{\{g\}}) = 2/10 & \delta_{\omega_1}(F_{\{r\}}) = 1/10 & \delta_{\omega_1}(F_{\{y\}}) = 1/10 \\
\delta_{\omega_2}(F_{\{b\}}) = 4/10 & \delta_{\omega_2}(F_{\{g\}}) = 2/10 & \delta_{\omega_2}(F_{\{r\}}) = 3/10 & \delta_{\omega_2}(F_{\{y\}}) = 1/10 \\
\delta_{\omega_3}(F_{\{b\}}) = 2/10 & \delta_{\omega_3}(F_{\{g\}}) = 2/10 & \delta_{\omega_3}(F_{\{r\}}) = 4/10 & \delta_{\omega_3}(F_{\{y\}}) = 2/10.
\end{array}
$$

Since the measured value (b, r) is obtained, we have the sample space $(X, 2^X, \nu)$ such that

$$
\nu(\{b\}) = 1/2, \qquad \nu(\{g\}) = 0, \qquad \nu(\{r\}) = 1/2, \qquad \nu(\{y\}) = 0.
$$

Then, we see that

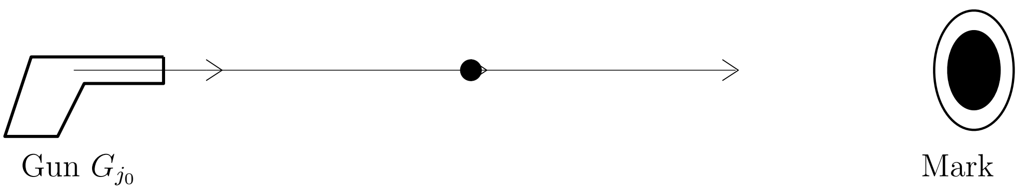
$$
\begin{aligned}
\Delta(\delta_{\omega_1}(F_{\{\cdot\}}), \nu) &= |6/10 - 1/2| + |2/10 - 0| + |1/10 - 1/2| + |1/10 - 0| = 8/10 \\
\Delta(\delta_{\omega_2}(F_{\{\cdot\}}), \nu) &= |4/10 - 1/2| + |2/10 - 0| + |3/10 - 1/2| + |1/10 - 0| = 6/10 \\
\Delta(\delta_{\omega_3}(F_{\{\cdot\}}), \nu) &= |2/10 - 1/2| + |2/10 - 0| + |4/10 - 1/2| + |2/10 - 0| = 8/10.
\end{aligned}
$$

Thus, the moment method says that the unknown urn [$*$] is U_2.

■

Example 5.10. [At a gun shop, [44]]. Let $\mathbf{G} \equiv \{G_1, ..., G_{50}\}$ be a set of guns in a gun shop. Assume that

$$
\text{the percentage of "hits of a gun } G_j\text{"} = \begin{cases} 80\% & \text{if} \quad 1 \le j \le 30, \\ 70\% & \text{if} \quad 31 \le j \le 40, \\ 10\% & \text{if} \quad 41 \le j \le 50. \end{cases} \tag{5.19}
$$



Assume the following situation (i)+(ii):

(i) Some one picks up a certain gun G_{j_0} from $\mathbf{G}$. He does not know the information concerning the j_0.

(ii) He shoots the gun G_{j_0} three times. First and second he hits the mark, and third he misses the mark.

Our present problem is to formulate the measurement (i)+(ii).

The above example is solved in what follows. Let Ω be a state space, which is identified with the set $\mathbf{G}$. That is, we have the identification: $\mathbf{G} \ni G_j \leftrightarrow \omega_j \in \Omega$. Define the observable $\mathbf{O} \equiv (X = \{0,1\}, \mathcal{P}(X), F_{(\cdot)})$ in $C(\Omega)$ such that:

$$F_{\{1\}}(\omega_j) = \begin{cases} 0.8 & \text{if} \quad 1 \le j \le 30, \\ 0.7 & \text{if} \quad 31 \le j \le 40, \\ 0.1 & \text{if} \quad 41 \le j \le 50 \end{cases} \tag{5.20}$$

and $F_{\{0\}}(\omega_j) = 1 - F_{\{1\}}(\omega_j)$. Of course we think that

($\sharp$) "hit the mark by a gun G_{j_0}" $\Leftrightarrow$ "get the measured value 1 by the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_{j_0}}]})$"

Here, consider the (three times) iterated measurement $\mathbf{M}_{C(\Omega)}(\times_{k=1}^{3} \mathbf{O} = (X^3, \mathcal{P}(X^3), \times_{k=1}^{3} F), S_{[\delta_{\omega_{j_0}}]})$ in $C(\Omega)$ such that:

$$(\mathop{\times}_{k=1}^{3} F)_{\Xi_1 \times \Xi_2 \times \Xi_3}(\omega) = F_{\Xi_1}(\omega) F_{\Xi_2}(\omega) F_{\Xi_3}(\omega) \qquad (\forall \Xi_1 \times \Xi_2 \times \Xi_3 \in \mathcal{P}(X^3), \forall \omega \in \Omega). \tag{5.21}$$

Clearly, the above statement (ii) implies that the measured value $(1,1,0)$ is obtained by $\mathbf{M}_{C(\Omega)}(\times_{k=1}^{3} \mathbf{O}, S_{[*]})$. (The observer does not know that $[*] = \delta_{\omega_{j_0}}$.) By a simple calculation, we see

$$F_{\{1\}}(\omega_j) F_{\{1\}}(\omega_j) F_{\{0\}}(\omega_j) = \begin{cases} 0.128 & \text{if} \quad 1 \le j \le 30, \\ 0.147 & \text{if} \quad 31 \le j \le 40, \\ 0.009 & \text{if} \quad 41 \le j \le 50. \end{cases} \tag{5.22}$$

Therefore, by Fisher's method (= Corollary 5.6), there is a very reason to consider that $31 \le j_0 \le 40$.

■

Example 5.11. [(i): Gaussian observable]. Consider a commutative C^*-algebra $C_0(\mathbf{R})$. And define the Gaussian observable $\mathbf{O}_{\sigma^2} \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}^{\rm bd}, F_{(\cdot)}^{\sigma^2})$ in $C_0(\mathbf{R})$ such that:

$$F_{\Xi}^{\sigma^2}(\mu) = \frac{1}{\sqrt{2\pi}\sigma} \int_{\Xi} \exp[-\frac{1}{2\sigma^2}(x-\mu)^2] dx \qquad (\forall \Xi \in \mathcal{B}_{\mathbf{R}}^{\rm bd}, \quad \forall \mu \in \mathbf{R}). \tag{5.23}$$

Further, consider the product observable $\mathbf{\times}_{k=1}^{3} \mathbf{O}$ (or in short, $\mathbf{O}_{\sigma^2}^3) \equiv (\mathbf{R}^3, \mathcal{B}_{\mathbf{R}^3}^{\rm bd}, F_{(\cdot)}^{\sigma^2, 3})$ in $C_0(\mathbf{R})$ such that:

$$F^{\sigma^2,3}_{\Xi_1\times\Xi_2\times\Xi_3}(\mu) = F^{\sigma^2}_{\Xi_1}(\mu)\cdot F^{\sigma^2}_{\Xi_2}(\mu)\cdot F^{\sigma^2}_{\Xi_3}(\mu)$$
$$= \frac{1}{(\sqrt{2\pi}\sigma)^3}\int_{\Xi_1\times\Xi_2\times\Xi_3} \exp[-\frac{(x_1-\mu)^2+(x_2-\mu)^2+(x_3-\mu)^2}{2\sigma^2}]dx_1dx_2dx_3$$
$$(\forall \Xi = k \in \mathcal{B}^{\rm bd}_{\mathbf{R}}, k=1,2,3, \quad \forall \mu \in \mathbf{R}). \tag{5.24}$$

Here consider the measurement $\mathbf{M}_{C_0(\mathbf{R})}(\mathbf{O}^3_{\sigma^2}, S_{[*]})$. And assume that

- the measured value (x_1^0, x_2^0, x_3^0) $(\in \mathbf{R}^3)$ is obtained by the $\mathbf{M}_{C_0(\mathbf{R})}(\mathbf{O}^3_{\sigma^2}, S_{[*]})$.

Then, Fisher's method (=Corollary 5.6) and Remark 5.4 say that there is a reason to think that the unknown state $[\,*\,] = \mu_0$, where

$$\frac{1}{(\sqrt{2\pi}\sigma)^3}\exp[-\frac{(x_1^0-\mu_0)^2+(x_2^0-\mu_0)^2+(x_3^0-\mu_0)^2}{2\sigma^2}]$$
$$= \max_{\mu\in\mathbf{R}}\Big[\frac{1}{(\sqrt{2\pi}\sigma)^3}\exp[-\frac{(x_1^0-\mu)^2+(x_2^0-\mu)^2+(x_3^0-\mu)^2}{2\sigma^2}]\Big], \tag{5.25}$$

which is equivalent to

$$(x_1^0-\mu_0)^2+(x_2^0-\mu_0)^2+(x_3^0-\mu_0)^2$$
$$= \min_{\mu\in\mathbf{R}}[(x_1^0-\mu)^2+(x_2^0-\mu)^2+(x_3^0-\mu)^2] \tag{5.26}$$

and moreover, equivalently,

$$\mu_0 = (x_1^0+x_2^0+x_3^0)/3. \tag{5.27}$$

[(ii): Gaussian observable]. Consider a commutative C^*-algebra $C([0,100])$, where $[0,100] \equiv \{\mu \in \mathbf{R} \mid 0 \le \mu \le 100\}$. And define the Gaussian observable $\mathbf{O}_{\sigma^2} \equiv (\mathbf{R}, \mathcal{B}^{\rm bd}_{\mathbf{R}}, F^{\sigma^2}_{(\cdot)})$ in $C([0,100])$ such that:

$$F^{\sigma^2}_{\Xi}(\mu) = \frac{1}{\sqrt{2\pi}\sigma}\int_{\Xi}\exp[-\frac{1}{2\sigma^2}(x-\mu)^2]dx \qquad (\forall \Xi \in \mathcal{B}^{\rm bd}_{\mathbf{R}}, \quad \forall \mu \in [0,100]). \tag{5.28}$$

Further, consider the product observable $\mathbf{O}^3_{\sigma^2} \equiv (\mathbf{R}^3, \mathcal{B}^{\rm bd}_{\mathbf{R}^3}, F^{\sigma^2,3}_{(\cdot)})$ in $C_0([0,100])$ such that:

$$F^{\sigma^2,3}_{\Xi_1\times\Xi_2\times\Xi_3}(\mu)$$
$$= \frac{1}{(\sqrt{2\pi}\sigma)^3}\int_{\Xi_1\times\Xi_2\times\Xi_3} \exp[-\frac{(x_1-\mu)^2+(x_2-\mu)^2+(x_3-\mu)^2}{2\sigma^2}]dx_1dx_2dx_3$$
$$(\forall \Xi = k \in \mathcal{B}^{\rm bd}_{\mathbf{R}}, k=1,2,3, \quad \forall \mu \in [0,100]). \tag{5.29}$$

Here consider the measurement $\mathbf{M}_{C([0,100])}(\mathbf{O}^3_{\sigma^2}, S_{[*]})$. And assume that

- the measured value (x_1^0, x_2^0, x_3^0) $(\in \mathbf{R}^3)$ is obtained by the $\mathbf{M}_{C([0,100])}(\mathbf{O}^3_{\sigma^2}, S_{[*]})$

Then, Fisher's method and Remark 5.4 say that there is a reason to think that the unknown state $[\,*\,] = \mu_0$, where $[\,*\,] = \mu_0$, where

$$\begin{aligned} &\frac{1}{(\sqrt{2\pi}\sigma)^3} \exp[-\frac{(x_1^0-\mu_0)^2+(x_2^0-\mu_0)^2+(x_3^0-\mu_0)^2}{2\sigma^2}] \\ =& \max_{\mu\in[0,100]} \Big[\frac{1}{(\sqrt{2\pi}\sigma)^3} \exp[-\frac{(x_1^0-\mu)^2+(x_2^0-\mu)^2+(x_3^0-\mu)^2}{2\sigma^2}]\Big] \end{aligned} \tag{5.30}$$

which is equivalent to

$$\begin{aligned} &(x_1^0-\mu_0)^2+(x_2^0-\mu_0)^2+(x_3^0-\mu_0)^2 \\ =& \min_{\mu\in[0,100]} [(x_1^0-\mu)^2+(x_2^0-\mu)^2+(x_3^0-\mu)^2] \end{aligned} \tag{5.31}$$

and moreover, equivalently,

$$\mu_0 = \begin{cases} 0 & \text{if } x_1^0+x_2^0+x_3^0 < 0 \\ (x_1^0+x_2^0+x_3^0)/3 & \text{if } 0 \le x_1^0+x_2^0+x_3^0 \le 100 \\ 100 & \text{if } x_1^0+x_2^0+x_3^0 > 100. \end{cases} \tag{5.32}$$

■

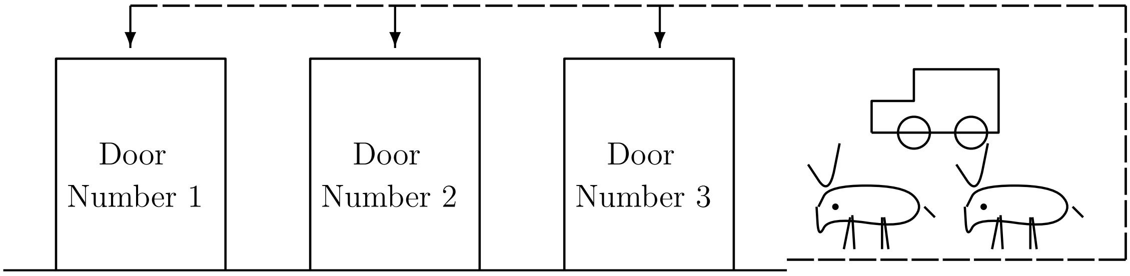
5.2.2 Monty Hall problem in PMT

Problem 5.12. [Monty Hall problem, *cf.*[33]].

The Monty Hall problem is as follows:

(P) Suppose you are on a game show, and you are given the choice of three doors (i.e., "number 1", "number 2", "number 3"). Behind one door is a car, behind the others, goats.

You pick a door, say number 1, and the host, who knows what's behind the doors, opens another door, say "number 3", which has a goat. He says to you, "Do you want to pick door number 2?" Is it to your advantage to switch your choice of doors?



[*Answer*]. Put $\Omega = \{\omega_1, \omega_2, \omega_3\}$, where

$\omega_1 \cdots\cdots$ the state that the car is behind the door number 1

$\omega_2 \cdots\cdots$ the state that the car is behind the door number 2

$\omega_3 \cdots\cdots$ the state that the car is behind the door number 3.

Define the observable $\mathbf{O} \equiv (\{1,2,3\}, 2^{\{1,2,3\}}, F)$ in $C(\Omega)$ such that

$$
\begin{array}{lll}
[F(\{1\})](\omega_1) = 0.0, & [F(\{2\})](\omega_1) = 0.5, & [F(\{3\})](\omega_1) = 0.5,^{8} \\
[F(\{1\})](\omega_2) = 0.0, & [F(\{2\})](\omega_2) = 0.0, & [F(\{3\})](\omega_2) = 1.0, \\
[F(\{1\})](\omega_3) = 0.0, & [F(\{2\})](\omega_3) = 1.0, & [F(\{3\})](\omega_3) = 0.0.
\end{array}
$$

Thus we have a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$. Here, note that

(1) :"measured value 1 is obtained" $\Longleftrightarrow$ The host says "Door (number 1) has a goat",

(2) :"measured value 2 is obtained" $\Longleftrightarrow$ The host says "Door (number 2) has a goat",

(3) :"measured value 3 is obtained" $\Longleftrightarrow$ The host says "Door (number 3) has a goat".

The host said "Door (number 3) has a goat". This implies that you get the measured value "3" by the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$. Therefore, Fisher's maximum likelihood method says that *you should pick door number 2.* That is because we see that

$$
\begin{aligned}
[F(\{3\})](\omega_2) = 1.0 &= \max\{0.5, \ \ 1.0, \ \ 0.0\} \\
&= \max\{[F(\{3\})](\omega_1), [F(\{3\})](\omega_2), [F(\{3\})](\omega_3)\},
\end{aligned}
$$

and thus, $[*] = \delta_{\omega_2}$. However, this is not all of the Monty Hall problem. See Remark 5.13, Problem 8.8 and Problem 11.13 later.

■

[8] Strictly speaking, $F(\{1\})(\omega_1) = 0.5$ and $F(\{2\})(\omega_1) = 0.5$ should be assumed in the problem (P).

Remark 5.13. [Monty Hall problem by the moment method (*cf.* Definition 2.27)]. Here, consider Problem 5.12 by the moment method. Since you get measured value 3, you get the sample space $(\{1,2,3\}, 2^{\{1,2,3\}}, \nu_s)$ such that $\nu_s(\{1\}) = 0$, $\nu_s(\{2\}) = 0$ and $\nu_s(\{3\}) = 1$. For example define the distance Δ such that: for any $\nu_1, \nu_2 \in \mathcal{M}^m_{+1}(\{1,2,3\})$,

$$\Delta(\nu_1, \nu_2) = |\nu_1(\{1\}) - \nu_2(\{1\})| + |\nu_1(\{2\}) - \nu_2(\{2\})| + |\nu_1(\{3\}) - \nu_2(\{3\})|.$$

Then, we see

$$\Delta(\nu_s, [F(\,\cdot\,)](\omega_1)) = |0-0| + |0-0.5| + |1-0.5| = 1,$$
$$\Delta(\nu_s, [F(\,\cdot\,)](\omega_2)) = |0-0| + |0-0| + |1-1| = 0$$

and

$$\Delta(\nu_s, [F(\,\cdot\,)](\omega_3)) = |0-0| + |0-1| + |1-0| = 2.$$

Thus, we can, by the moment method, infer that ω_2 is most possible, that is, the car is behind the door number 2.

■

5.3 Inference interval

Let $\mathbf{O}(\equiv (X, \mathcal{F}, F))$ be an observable formulated in a C^*-algebra $\mathcal{A}$. Assume that X has a metric d_X. And assume that the state space $\mathfrak{S}^p(\mathcal{A}^*)$ has the metric $d_{\mathfrak{S}}$, which induces the weak* topology $\sigma(\mathcal{A}^*, \mathcal{A})$. Let $E : X \to \mathfrak{S}^p(\mathcal{A}^*)$ be a continuous map, which is called *"estimator".* Let γ be a real number such that $0 \ll \gamma < 1$, for example, $\gamma = 0.95$. For any $\rho^p (\in \mathfrak{S}^p(\mathcal{A}^*))$, define the positive number $\eta^\gamma_{\rho^p}\ (>0)$ such that:

$$\eta^\gamma_{\rho^p} = \inf\{\eta > 0 : {}_{\mathcal{A}^*}\Big\langle \rho^p, F(E^{-1}(B(\rho^p;\eta)))\Big\rangle_{\mathcal{A}} \geq \gamma\} \tag{5.33}$$

where $B(\rho^p;\eta) = \{\rho^p_1 (\in \mathfrak{S}^p(\mathcal{A}^*)) : d_{\mathfrak{S}}(\rho^p_1, \rho^p) \leq \eta\}$. For any $x\ (\in X)$, put

$$D^\gamma_x = \{\rho^p (\in \mathfrak{S}^p(\mathcal{A}^*)) : d_{\mathfrak{S}}(E(x), \rho^p) \leq \eta^\gamma_{\rho^p}\}. \tag{5.34}$$

The D^γ_x is called *the* (γ)*-inference interval of the measured value* x.

Note that,

(A) *for any* $\rho_0^p (\in \mathfrak{S}^p(\mathcal{A}^*))$, *the probability, that the measured value* x *obtained by the measurement* $\mathbf{M}_{\mathcal{A}}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[\rho_0^p]})$ *satisfies the following condition* ($\flat$), *is larger than* γ *(e.g.,* $\gamma = 0.95$*).*

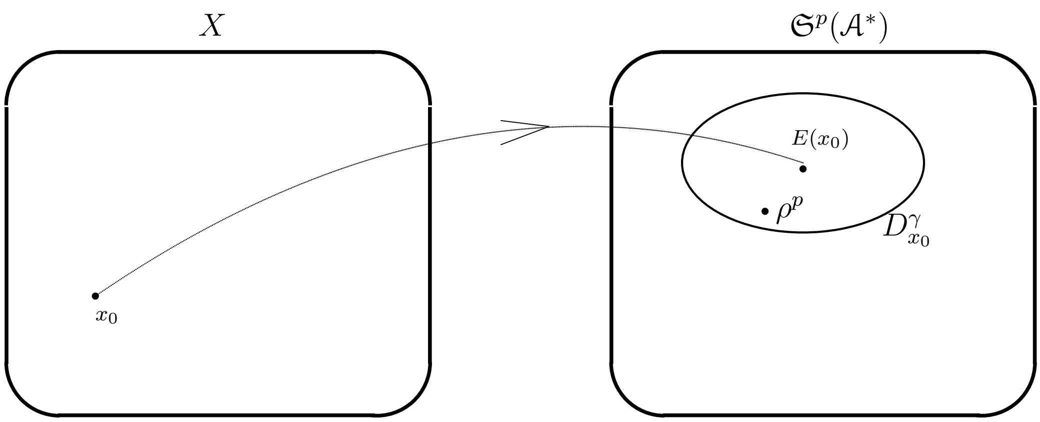
$$(\flat) \qquad E(x) \in B(\rho_0^p; \eta^\gamma_{\rho_0^p}) \qquad \text{or equivalently} \qquad d(E(x), \rho_0^p) \leq \eta^\gamma_{\rho_0^p}.$$

Assume that

(B) we get a measured value x_0 by the measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[\rho_0^p]})$.

Then, we see the following equivalences:

$$(\flat) \quad \Longleftrightarrow \quad d_{\mathfrak{S}}(E(x_0), \rho_0^p) \leq \eta^\gamma_{\rho_0^p} \quad \Longleftrightarrow \quad D^\gamma_{x_0} \ni \rho_0^p.$$



Summing the above argument, we have the following theorem.

Theorem 5.14. [Inference interval]. *Let* $\mathbf{O} \equiv (X, \mathcal{F}, F)$ *be an observable in* $\mathcal{A}$. *Let* ρ_0^p *be any fixed state, i.e.,* $\rho_0^p \in \mathfrak{S}^p(\mathcal{A}^*)$, *Consider a measurement* $\mathbf{M}_{\mathcal{A}}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[\rho_0^p]})$. *Let* $E : X \to \mathfrak{S}^p(\mathcal{A}^*)$ *be an estimator. Let* γ *be such as* $0 \ll \gamma < 1$ *(e.g.,* $\gamma = 0.95$*). For any* $x (\in X)$, *define* D^γ_x *as in (5.34). Then, we see,*

($\sharp$) *the probability that the measured value* $x_0 (\in X)$ *obtained by the measurement* $\mathbf{M}_{\mathcal{A}}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[\rho_0^p]})$ *satisfies the condition that*

$$D^\gamma_{x_0} \ni \rho_0^p \,, \tag{5.35}$$

is larger than γ.

■

Example 5.15. [The urn problem]. Put $\Omega = [0,1]$, i.e., the closed interval in $\mathbf{R}$. We assume that each ω $(\in \Omega \equiv [0,1])$ represents an urn that contains a lot of red balls and white balls such that:

$$\frac{\text{the number of white balls in the urn } \omega}{\text{the total number of red and white balls in the urn } \omega} \approx \omega \quad (\forall \omega \in [0,1] \equiv \Omega). \tag{5.36}$$

Define the observable $\mathbf{O} = (X \equiv \{r, w\}, 2^{\{r,w\}}, F)$ in $C(\Omega)$ such that where

$$F(\emptyset)(\omega) = 0, \quad F(\{r\})(\omega) = \omega, \quad F(\{w\})(\omega) = 1 - \omega, \quad F(\{r,w\})(\omega) = 1 \qquad (\forall \omega \in [0,1] \equiv \Omega). \tag{5.37}$$

Here, consider the following measurement M_ω:

$$M_\omega := \text{"Pick out one ball from the urn } \omega\text{, and recognize the color of the ball"} \tag{5.38}$$

That is, we consider

$$M_\omega = \mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_\omega]}). \tag{5.39}$$

Moreover we define the product observable $\mathbf{O}^N \equiv (X^N, \mathcal{P}(X^N), F^N)$, such that:

$$\begin{aligned} &[F^N(\Xi_1 \times \Xi_2 \times \cdots \times \Xi_{N-1} \times \Xi_N)](\omega) \\ =&[F(\Xi_1)](\omega) \cdot [F(\Xi_2)](\omega) \cdots [F(\Xi_{N-1})](\omega) \cdot [F(\Xi_N)](\omega) \\ &(\forall \omega \in \Omega \equiv [0,1], \quad \forall \Xi_1, \Xi_2, \cdots, \Xi_N \subseteq X \equiv \{r, w\}). \end{aligned} \tag{5.40}$$

As mentioned in Definition 2.27, we think that

$$\text{"take a measurement } M_\omega \text{ N times"} \Leftrightarrow \text{"take a measurement } \mathbf{M}_{C(\Omega)}(\mathbf{O}^N, S_{[\delta_\omega]})\text{"} \tag{5.41}$$

Define the estimator $E : X^N(\equiv \{r, w\}^N) \to \Omega(\equiv [0,1])$

$$E(x_1, x_2, \cdots, x_{N-1}, x_N) = \frac{\sharp[\{n \in \{1, 2, \cdots, N\} \mid x_n = r\}]}{N} \qquad (\forall x = (x_1, x_2, \cdots, x_{N-1}, x_N) \in X^N \equiv \{r, w\}^N). \tag{5.42}$$

For each $\omega(\in [0,1] \equiv \Omega)$, define the positive number η_ω^γ such that:

$$\begin{aligned} &\eta_\omega^\gamma \\ =&\inf \left\{ \eta > 0 \;\middle|\; [F^N(\{(x_1, x_2, \cdots, x_N) \mid \omega - \eta \le E(x_1, x_2, \cdots, x_N) \le \omega + \eta\})](\omega) > 0.95 \right\} \\ =&\inf_{[F^N(\{(x_1,x_2,\cdots,x_N) \mid \omega-\eta \le E(x_1,x_2,\cdots,x_N) \le \omega+\eta\})](\omega) > 0.95} \eta. \end{aligned} \tag{5.43}$$

Put

$$D_x^\gamma = \{\omega(\in \Omega) : |E(x) - \omega| \leq \eta_\omega^\gamma\}. \tag{5.44}$$

For example, assume that N is sufficiently large and $\gamma = 0.95$. Then we see, by (2.58), that

$$\eta_\omega^{0.95} \approx 1.96\sqrt{\frac{\omega(1-\omega)}{N}}$$

and

$$D_x^{0.95} = [E(x) - \eta_-, E(x) + \eta_+] \tag{5.45}$$

where

$$\eta_- = \eta_{E(x)-\eta_-}^{0.95}, \quad \eta_+ = \eta_{E(x)+\eta_+}^{0.95}. \tag{5.46}$$

Under the assumption that N is sufficiently large, we can consider that

$$\eta_- \approx \eta_+ \approx \eta_{E(x)}^{0.95} \approx 1.96\sqrt{\frac{E(x)(1-E(x))}{N}}.$$

Then we can conclude that

- *for any urn $\omega(\in \Omega \equiv [0,1])$, the probability, that the measured value $x = (x_1, x_2, \cdots, x_N)$ obtained by the measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}^N, S_{[\delta_\omega]})$ satisfies the following condition $(\sharp)$, is larger than γ (e.g., $\gamma = 0.95$).*

  $(\sharp)$ $$E(x) - 1.96\sqrt{\frac{E(x)(1-E(x))}{N}} \leq \omega \leq E(x) + 1.96\sqrt{\frac{E(x)(1-E(x))}{N}}.$$

  where E is defined by (5.42).

■

5.4 Testing statistical hypothesis

Now we study "testing statistical hypothesis", that is, answer the following question.

Problem 5.16. [Testing statistical hypothesis]. *Consider a measurement* $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[*]})$ *formulated in* $C(\Omega)$. *Let* $E : X \to \Omega$ *be Fisher's estimator. Assume the following hypothesis:*

(H) the unknown state $[*]$ *belongs to a closed set* C_H $(\subseteq \Omega)$.

And further assume that we see the following fact:

(F) a measured value $x_0 (\in X)$ *is obtained by measurement* $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[*]})$.

Here, our present purpose is to propose an algorithm that decides whether the above hypothesis (H) can be denied by the fact (F). This algorithm is called "the testing statistical hypothesis."

In the above problem, it is usually expected that the hypothesis (H) is not true. In this sense, the above (H) is called *the null hypothesis.*

Now we provide two answers (i.e., Answer 1 and Answer 2). Answer 1 (likelihood ratio test) is, of course, well-known and authorized. Also, in order to solve the question: "Is there another answer?", we add Answer 2 after Answer 1.

Answer 1. [Likelihood ratio test]. Consider a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[*]})$ formulated in $C(\Omega)$. Let $E : X \to \Omega$ be Fisher's estimator, i.e., it is defined by

$$E(x) = \lim_{\Xi_n \to \{x\}} \omega_n \qquad (\forall x \in X),$$

where ω_n $(\in \Omega)$ is chosen such that it satisfies

$$\frac{[F(\Xi_n)](\omega_n)}{\max_{\omega \in \Omega} [F(\Xi_n)](\omega)} = 1.$$

$\Big($For the exact argument, see Remark 5.4 (Radon-Nikodým derivative).$\Big)$ Assume both (H) and (F) in Problem 5.16. Consider a real number α such that $0 < \alpha \ll 1$ (e.g. $\alpha = 0.05$, which may be called *a significance level.* Let ω be in Ω. Then, by Axiom 1, we have a sample probability measure P_ω on X (of the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[\delta_\omega]}))$ such that:

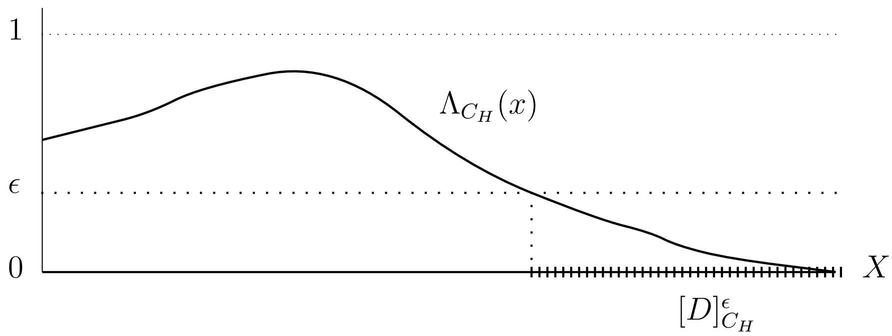
$$P_\omega(\Xi) = [F(\Xi)](\omega) \quad (\forall \Xi \in \mathcal{F}). \tag{5.47}$$

Here define the function $\Lambda_{C_H} : X \to [0, 1]$ such that:

$$\Lambda_{C_H}(x) = \lim_{\Xi \to \{x\}} \frac{\sup_{\omega \in C_H} P_\omega(\Xi)}{\sup_{\omega \in \Omega} P_\omega(\Xi)} \quad (\forall x \in X). \tag{5.48}$$

Also, for any ϵ $(0 < \epsilon \leq 1)$, define $[D]^{\epsilon}_{C_H}$ $(\in \mathcal{F})$ such that:

$$[D]^{\epsilon}_{C_H} = \{x \in X \mid \Lambda_{C_H}(x) < \epsilon\}. \tag{5.49}$$



Thus we can define $\epsilon^{0.05}_{\max}$ such that:

$$\epsilon^{0.05}_{\max} = \sup\{\epsilon \mid \sup_{\omega_0 \in C_H} P_{\omega_0}([D]^{\epsilon}_{C_H}) \leq 0.05\}. \tag{5.50}$$

Now we can conclude that

Answer 1

$$\begin{cases} \text{if } x_0 \in [D]^{\epsilon^{0.05}_{\max}}_{C_H}, \text{ then the hypothesis (H) can be denied} \\ \text{if } x_0 \notin [D]^{\epsilon^{0.05}_{\max}}_{C_H}, \text{ then the hypothesis (H) can not be denied} \end{cases} \tag{5.51}$$

□

Next we shall propose "Answer 2". Before this, we must prepare the following well-known lemma.

Lemma 5.17. [Neyman-Pearson theorem, α-influential domain of ν_1 for ν_2, *cf.* [59]]. *Let $(X, \mathcal{F})$ be a measurable space. Let ν_1 and ν_2 be probability measures on X. Define the Radon-Nikodým derivative $\frac{d\nu_1}{d\nu_2} : X \to [0, \infty)$ such that:*

$$\frac{d\nu_1}{d\nu_2}(x) = \lim_{\Xi \to x} \frac{\nu_1(\Xi)}{\nu_2(\Xi)} \quad (x \in X). \tag{5.52}$$

Put

$$[D](\epsilon, \frac{d\nu_1}{d\nu_2}) = \{x \in X \mid \frac{d\nu_1}{d\nu_2}(x) < \epsilon\}, \qquad (0 \leq \epsilon \leq \infty). \tag{5.53}$$

Thus we can define $\breve{\epsilon}^{0.05}_{\max}$ such that:

$$\breve{\epsilon}^{0.05}_{\max}\Big(\equiv \breve{\epsilon}^{\alpha=0.05}_{\max}\Big) = \sup\{\epsilon \mid \nu_1([D](\epsilon, \frac{d\nu_1}{d\nu_2})) \le 0.05\}. \tag{5.54}$$

Now we have the

$$[D](\breve{\epsilon}^{0.05}_{\max}, \frac{d\nu_1}{d\nu_2}), \tag{5.55}$$

which is called *"the* 0.05*-influential domain of* ν_1 *for* ν_2*."*

■

Answer 2. [A test using Neyman-Pearson theorem]. Consider a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[*]})$ formulated in $C(\Omega)$. Let $E : X \to \Omega$ be Fisher's estimator. Assume both (H) and (F) in Problem 5.16. Consider a real number α such that $0 < \alpha \ll 1$ (e.g. $\alpha = 0.05$ which may be also called *a significance level.* Let ω be in C_H. Consider a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[\delta_\omega]})$. Let x be in X. Then, we have two sample probability measures P_ω and $P_{E(x)}$ on X such that:

$$\nu_\omega(\Xi) = P_\omega(\Xi) = [F(\Xi)](\omega) \qquad (\forall \Xi \in \mathcal{F})$$

and

$$\nu_{E(x)} = P_{E(x)}(\Xi) = [F(\Xi)](E(x)) \qquad (\forall \Xi \in \mathcal{F}). \tag{5.56}$$

Thus, we have "the 0.05-influential domain of ν_1 for ν_2" such that:

$$[D](\breve{\epsilon}^{0.05}_{\max}, \frac{d\nu_\omega}{d\nu_{E(x)}}). \tag{5.57}$$

Put

$$[D]^{\breve{\epsilon}^{0.05}_{\max}}_{C_H,x} = \cap_{\omega \in C_H} [D](\breve{\epsilon}^{0.05}_{\max}, \frac{d\nu_\omega}{d\nu_{E(x)}}). \tag{5.58}$$

Lastly, we put

$$[D]^{\breve{\epsilon}^{0.05}_{\max}}_{C_H} = \{x \in X \mid x \in [D]^{\breve{\epsilon}^{0.05}_{\max}}_{C_H,x}\}. \tag{5.59}$$

Now we can conclude that

Answer 2

$$\begin{cases} \text{if } x_0 \in [D]^{\breve{\epsilon}^{0.05}_{\max}}_{C_H}, \text{ then the hypothesis (H) can be denied} \\ \text{if } x_0 \notin [D]^{\breve{\epsilon}^{0.05}_{\max}}_{C_H}, \text{ then the hypothesis (H) can not be denied} \end{cases} \tag{5.60}$$

Remark 5.18. [*Answers 1 and 2*]. We believe that the above two answers 1 and 2 are proper though the meanings of "significant level" is different in each answer (*cf.* [II; $C_H = [0, \infty)$] in Examples 5.16 and 5.17). We do not know whether there is another proper answer.

■

Example 5.19. [Likelihood ratio test for the Gaussian observable]. Put $\Omega = \mathbf{R}$, $\mathcal{A} = C_0(\Omega)$, $\mathbf{O}_{\sigma^2} \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}^{\rm bd}, F_{(\cdot)}^{\sigma^2})$ in $C_0(\Omega)$ such that:

$$F_{\Xi}^{\sigma^2}(\omega) = \frac{1}{\sqrt{2\pi}\sigma} \int_{\Xi} \exp[-\frac{(x-\omega)^2}{2\sigma^2}] du \qquad (\forall \Xi \in \mathcal{B}_{\mathbf{R}}^{\rm bd}, \quad \forall \omega \in \Omega = \mathbf{R}). \tag{5.61}$$

And thus. consider the product observable $\mathbf{O}_\sigma^2 \equiv (\mathbf{R}^2, \mathcal{B}_{\mathbf{R}^2}^{\rm bd}, F_{(\cdot)}^{\sigma^2}) \times F_{(\cdot)}^{\sigma^2})$ in $C_0(\Omega)$. That is,

$$(F_{\Xi_1}^{\sigma^2} \times F_{\Xi_2}^{\sigma^2})(\omega) = \frac{1}{(\sqrt{2\pi}\sigma)^2} \iint_{\Xi_1 \times \Xi_2} \exp[-\frac{(x_1-\omega)^2 + (x_2-\omega)^2}{2\sigma^2}] dx_1 dx_2$$
$$(\forall \Xi_k \in \mathcal{B}_{\mathbf{R}}^{\rm bd} (k=1,2), \quad \forall \omega \in \Omega = \mathbf{R}). \tag{5.62}$$

[Case(I): Two sided test, i.e., $C_H = \{\omega_0\}$]. Assume that $C_H = \{\omega_0\}$, $\omega_0 \in \Omega = \mathbf{R}$. Then,

$$\Lambda_{\{\omega_0\}}(x_1, x_2) = \lim_{\Xi_1 \times \Xi_2 \to \{(x_1, x_2)\}} \frac{\sup_{\omega \in \{\omega_0\}} P_\omega(\Xi_1 \times \Xi_2)}{\sup_{\omega \in \Omega} P_\omega(\Xi_1 \times \Xi_2)}$$
$$= \frac{\exp[-\frac{(x_1-\omega_0)^2 + (x_2-\omega_0)^2}{2\sigma^2}]}{\exp[-\frac{(x_1-(x_1+x_2)/2)^2 + (x_2-(x_1+x_2)/2)^2}{2\sigma^2}]}$$
$$= \exp[-\frac{[(x_1+x_2)-2\omega_0]^2}{4\sigma^2}] = \exp[-\frac{[(x_1+x_2)/2-\omega_0]^2}{2(\sigma/\sqrt{2})^2}]$$
$$(\forall (x_1, x_2) \in \mathbf{R}^2). \tag{5.63}$$

Also, for any $\epsilon(>0)$, define $[D]_{\{\omega_0\}}^{\epsilon}$ $(\in \mathcal{F})$ such that:

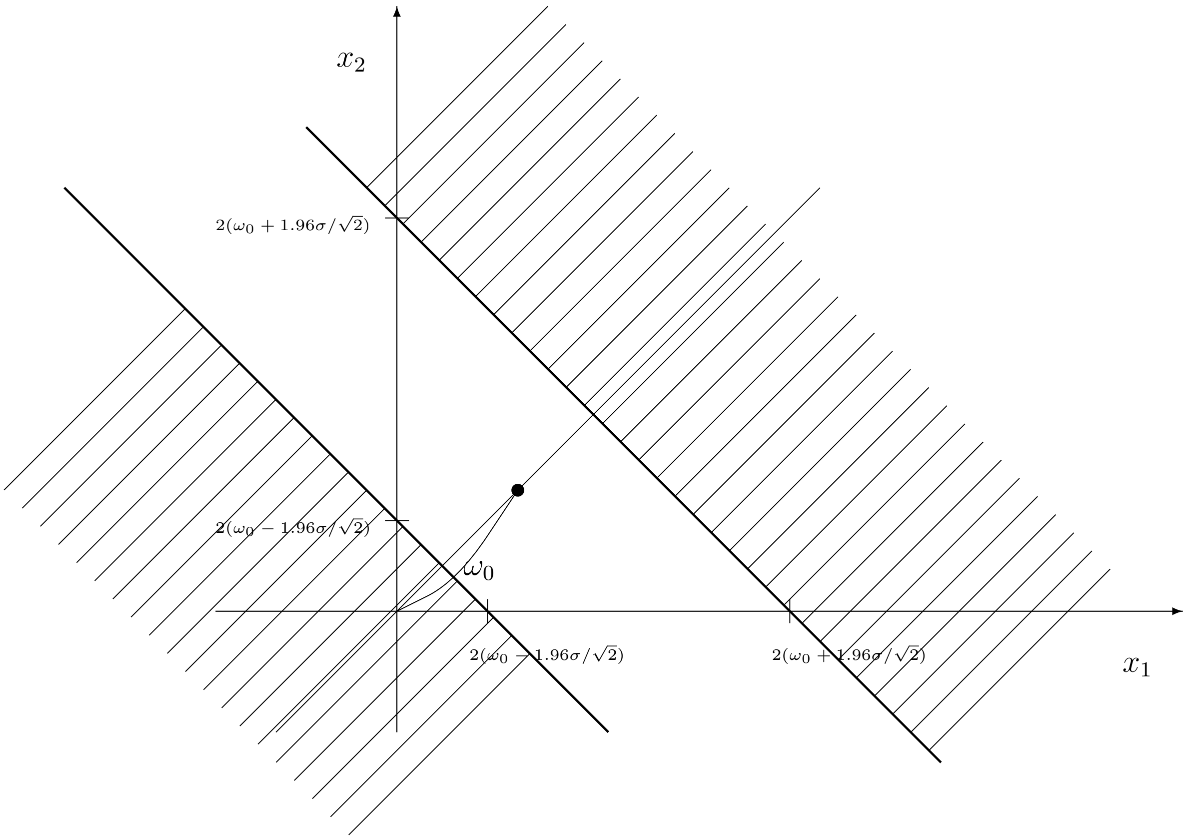
$$[D]_{\{\omega_0\}}^{\epsilon} = \{(x_1, x_2) \in \mathbf{R}^2 \mid \Lambda_{\{\omega_0\}}(x_1, x_2) < \epsilon\}. \tag{5.64}$$

Thus we can define $\epsilon_{\max}^{0.05}$ such that:

$$\epsilon_{\max}^{0.05} = \sup\{\epsilon \mid \sup_{\omega \in \{\omega_0\}} P_\omega([D]_{\{\omega_0\}}^{\epsilon}) \le 0.05\}. \tag{5.65}$$

Now we can conclude that

$$[D]_{\{\omega_0\}}^{\epsilon_{\max}^{0.05}}$$
$$= \{(x_1, x_2) \in \mathbf{R}^2 \mid (x_1+x_2)/2 \le \omega_0 - 1.96\sigma/\sqrt{2}\}$$
$$\bigcup \{(x_1, x_2) \in \mathbf{R}^2 \mid (x_1+x_2)/2 \ge \omega_0 + 1.96\sigma/\sqrt{2}\}$$
$$= \text{"Slash part in the following figure"}$$



[Case(II): One sided test, i.e., $C_H = [\omega_0, \infty)$]. Assume that $C_H = [\omega_0, \infty)$, $\omega_0 \in \Omega = \mathbf{R}$. Then,

$$\Lambda_{[0,\infty)}(x_1, x_2) = \lim_{\Xi_1 \times \Xi_2 \to \{(x_1,x_2)\}} \frac{\sup_{\omega \in [\omega_0,\infty)} P_\omega(\Xi_1 \times \Xi_2)}{\sup_{\omega \in \Omega} P_\omega(\Xi_1 \times \Xi_2)}$$
$$= \begin{cases} \exp[-\frac{[(x_1+x_2)-2\omega_0]^2}{4\sigma^2}] & (\frac{x_1+x_2}{2} < \omega_0) \\ 1 & (\text{ otherwise }) \end{cases} \tag{5.66}$$

Also, for any $\epsilon(> 0)$, define $[D]^\epsilon_{[\omega_0,\infty)}$ $(\in \mathfrak{F})$ such that:

$$[D]^\epsilon_{[0,\infty)} = \{(x_1, x_2) \in \mathbf{R}^2 \mid \Lambda_{[0,\infty)}(x_1, x_2) \le \epsilon\}$$
$$= \{(x_1, x_2) \in \mathbf{R}^2 \mid \frac{x_1 + x_2}{2} - \omega_0 < \sqrt{4\sigma^2 \log \epsilon}\}. \tag{5.67}$$

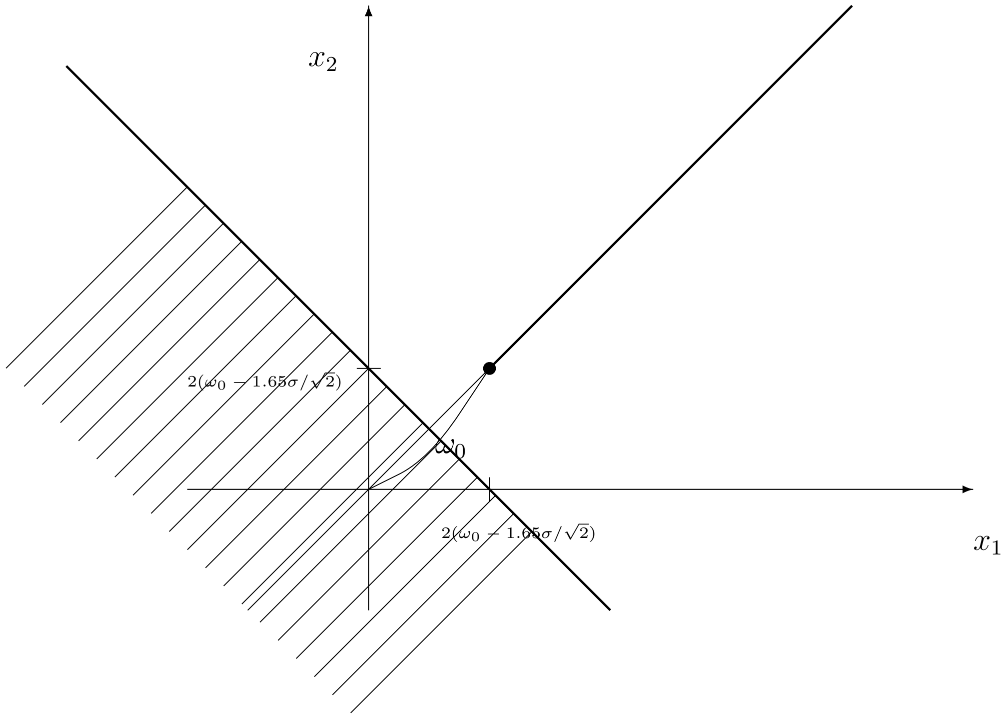
Thus we can define $\epsilon^{0.05}_{\max}$ such that:

$$\epsilon^{0.05}_{\max} = \sup\{\epsilon \mid \sup_{\omega_0 \in [0,\infty)} P_{\omega_0}([D]^\epsilon_{[0,\infty)}) \le 0.05\}. \tag{5.68}$$

Therefore, we can conclude that

$$[D]^{\epsilon^{0.05}_{\max}}_{[0,\infty)}$$
$$=\{(x_1, x_2) \in \mathbf{R}^2 \mid (x_1 + x_2)/2 \le \omega_0 - 1.65\sigma/\sqrt{2}\}. \qquad (cf.\ (2.58)).$$

=“Slash part in the following figure”



■

Example 5.20. [The test using Neyman-Pearson theorem for the Gaussian observable]. Put $\Omega = \mathbf{R}$, $\mathcal{A} = C_0(\Omega)$, $\mathbf{O}_{\sigma^2} \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}^{\rm bd}, F_{(\cdot)}^{\sigma^2})$ and $\mathbf{O}_{\sigma^2}^2 \equiv (\mathbf{R}^2, \mathcal{B}_{\mathbf{R}^2}^{\rm bd}, F_{(\cdot)}^{\sigma^2} \times F_{(\cdot)}^{\sigma^2})$ in $C_0(\Omega)$ are as in the above.

[Case(I): Two sided test, i.e., $C_H = \{\omega_0\}$]. Assume that $C_H = \{\omega_0\}$, $\omega_0 \in \Omega = \mathbf{R}$. Then,

$$\nu_1^{\omega_0}(\Xi_1 \times \Xi_2) = P_{\omega_0}(\Xi_1 \times \Xi_2) = [F(\Xi_1 \times \Xi_2)](\omega_0) \qquad (\forall \Xi_1 \times \Xi_2 \in \mathcal{B}_{\mathbf{R}^2}^{\rm bd})$$

and

$$\nu_2^{E(x_0)} = P_{E(x_0)}(\Xi_1 \times \Xi_2) = [F(\Xi_1 \times \Xi_2)](E(x_0)) \qquad (\forall \Xi_1 \times \Xi_2 \in \mathcal{B}_{\mathbf{R}^2}^{\rm bd}). \quad (5.69)$$

Thus, we have "the 0.05-influential domain of ν_1 for ν_2" such that:

$$[D](\breve{\epsilon}_{\max}^{0.05}, \phi_{\nu_1^{\omega_0}/\nu_2^{E((x_1,x_2))}}) = \begin{cases} \{(x_1,x_2) \mid (x_1+x_2)/2 \le \omega_0 - 1.65\sigma/\sqrt{2}\} & (E(x_0) < \omega_0) \\ \{(x_1,x_2) \mid (x_1+x_2)/2 \ge \omega_0 + 1.65\sigma/\sqrt{2}\} & (E(x_0) > \omega_0). \end{cases}$$

Put

$$[D]_{\{\omega_0\},(x_1,x_2)}^{\breve{\epsilon}_{\max}^{0.05}} = \cap_{\omega_0 \in \{\omega_0\}} [D](\breve{\epsilon}_{\max}^{0.05}, \phi_{\nu_1^{\omega_0}/\nu_2^{E((x_1,x_2))}}) = [D](\breve{\epsilon}_{\max}^{0.05}, \phi_{\nu_1^{\omega_0}/\nu_2^{E((x_1,x_2))}}) \quad (\forall (x_1,x_2) \in \mathbf{R}^2). \tag{5.70}$$

Therefore, we can conclude that

$$[D]^{\check{\epsilon}^{0.05}_{\max}}_{\{\omega_0\}} = \{(x_1,x_2) \in \mathbf{R}^2 \mid (x_1,x_2) \in [D]^{\check{\epsilon}^{0.05}_{\max}}_{\{\omega_0\},(x_1,x_2)}\}$$
$$= \{(x_1,x_2) \mid (x_1+x_2)/2 \le \omega_0 - 1.65\sigma/\sqrt{2}\}$$
$$\bigcup \{(x_1,x_2) \mid (x_1+x_2)/2 \ge \omega_0 + 1.65\sigma/\sqrt{2}\}.$$

[Case(II): One sided test, i.e., $C_H = [\omega_0, \infty)$]. Assume that $C_H = [\omega_0, \infty)$, $\omega_0 \in \Omega = \mathbf{R}$. Then,

$$\nu_1^{\omega_0}(\Xi_1 \times \Xi_2) = P_{\omega_0}(\Xi_1 \times \Xi_2) = [F(\Xi_1 \times \Xi_2)](\omega_0) \qquad (\forall \Xi_1 \times \Xi_2 \in \mathcal{B}^{\rm bd}_{\mathbf{R}^2})$$

and

$$\nu_2^{E(x_0)}(\Xi_1 \times \Xi_2) = P_{E(x_0)}(\Xi_1 \times \Xi_2) = [F(\Xi_1 \times \Xi_2)](E(x_0)) \qquad (\forall \Xi_1 \times \Xi_2 \in \mathcal{B}^{\rm bd}_{\mathbf{R}^2}). \tag{5.71}$$

Thus, we have "the 0.05-influential domain of ν_1 for ν_2" such that:

$$[D](\check{\epsilon}^{0.05}_{\max}, \phi_{\nu_1^{\omega_0}/\nu_2^{E((x_1,x_2))}}) = \begin{cases} \{(x_1,x_2) \mid (x_1+x_2)/2 \le \omega_0 - 1.65\sigma/\sqrt{2}\} & (E(x_0) < \omega_0) \\ \{(x_1,x_2) \mid (x_1+x_2)/2 \ge \omega_0 + 1.65\sigma/\sqrt{2}\} & (E(x_0) > \omega_0). \end{cases}$$

Put

$$[D]^{\check{\epsilon}^{0.05}_{\max}}_{[0,\infty),(x_1,x_2)} = \cap_{\omega_0 \in [0,\infty)} [D](\check{\epsilon}^{0.05}_{\max}, \phi_{\nu_1^{\omega_0}/\nu_2^{E(x)}}) \qquad (\forall (x_1,x_2) \in \mathbf{R}^2). \tag{5.72}$$

Therefore, we can conclude that

$$[D]^{\check{\epsilon}^{0.05}_{\max}}_{[0,\infty)} = \{(x_1,x_2) \in \mathbf{R}^2 \mid (x_1,x_2) \in [D]^{\check{\epsilon}^{0.05}_{\max}}_{[0,\infty),(x_1,x_2)}\}$$
$$= \{(x_1,x_2) \mid (x_1+x_2)/2 \le \omega_0 - 1.65\sigma/\sqrt{2}\}.$$

■

5.5 Measurement error model in PMT

Although we have several kinds of measurement error models in statistics (*cf.* Fuller [25], Cheng, etc. [16]), the following may be the simplest one (i.e., with normal distributions (= Gaussian distributions)):

$$\begin{cases} \widetilde{y}_n = \theta_0 + \theta_1 x_n + e_n, \\ \widetilde{x}_n = x_n + u_n \\ (e_n, u_n) \sim \text{NI}[\text{average}(0,0), \text{variance}(\sigma^2_{ee}, \sigma^2_{uu})], ^9 \\ \qquad (n = 1, 2, ..., N), \end{cases} \tag{5.73}$$

which, of course, corresponds to the conventional statistics (i.e., the measurement equation in the dynamical system theory (1.2)). The first equation is a classical regression specification, but the true explanatory variable x_n is not observed directly. The observed measure of x_n, denoted by $\widetilde{x}_n$, may be obtained by a certain measurement. Our present concern is how to infer the unknown parameters θ_0 and θ_1 from the measured value $\{(\widetilde{x}_n, \widetilde{y}_n)\}_{n=1}^N$. Precisely speaking, the purpose of this section is to study this problem in general situations (i.e., without the assumption of normal distributions).

Put $\mathcal{A}_0 \equiv C(\Omega_0)$ and $\mathcal{A}_1 \equiv C(\Omega_1)$. Let Θ be a compact space, which may be called an *index state space* (or *parameter space*). Consider a parameterized continuous map $\psi^\theta : \Omega_0 \to \Omega_1$, $\theta \in \Theta$, which induces the parameterized homomorphism $\Psi^\theta : C(\Omega_1) \to C(\Omega_0)$ such that (*cf.* (3.14))

$$(\Psi^\theta f_1)(\omega) = f_1(\psi^\theta(\omega)) \quad (\forall f_1 \in C(\Omega_1), \forall \omega \in \Omega_0).$$

Consider observables $\mathbf{O}_0 \equiv (X, \mathcal{F}, F)$ in $C(\Omega_0)$ and $\mathbf{O}_1 \equiv (Y, \mathcal{G}, G)$ in $C(\Omega_1)$. And recall that $\Psi^\theta \mathbf{O}_1$ can be identified with the observable in $C(\Omega_0)$ (*cf.* Remark 3.6 (i)). Thus, we can consider the product observable $\widetilde{\mathbf{O}}^\theta = (X \times Y, \mathcal{F} \times \mathcal{G}, F \times \Psi^\theta G)$ in $C(\Omega_0)$. Thus, we get the measurement $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}^\theta, S_{[\delta_\omega]})$, $(\omega \in \Omega_0)$. Consider the N times repeated measurement of $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}^\theta, S_{[\delta_\omega]})$, which is represented by $\mathbf{M}_{C(\Omega_0^N)}(\bigotimes_{n=1}^N \widetilde{\mathbf{O}}^\theta, S_{[\otimes_{n=1}^N \delta_{\omega_n}]})$. Here, $\otimes_{n=1}^N \delta_{\omega_n} = \delta_{(\omega_1, \omega_2, \dots, \omega_N)}$ $(\in \mathcal{M}_{+1}^p(\Omega_0^N))$ and $\bigotimes_{n=1}^N \widetilde{\mathbf{O}}^\theta = (X^N \times Y^N, \mathcal{F}^N \times \mathcal{G}^N, \bigotimes_{n=1}^N (F \times \Psi^\theta G))$ in $\bigotimes_{n=1}^N C(\Omega_0) \equiv C(\Omega_0^N)$, that is,

$$\begin{aligned} &[(\bigotimes_{n=1}^N (F \times \Psi^\theta G))(\Xi_1 \times \cdots \times \Xi_N \times \Gamma_1 \times \cdots \times \Gamma_N)](\omega_1, \dots, \omega_N) \\ =&[F \times \Psi^\theta G(\Xi_1 \times \Gamma_1)](\omega_1) \cdot [F \times \Psi^\theta G(\Xi_2 \times \Gamma_2)](\omega_2) \cdots [F \times \Psi^\theta G(\Xi_N \times \Gamma_N)](\omega_N) \\ &\qquad (\forall \Xi_n \in \mathcal{F}, \forall \Gamma_n \in \mathcal{G}, \forall (\omega_1, \dots, \omega_N) \in \Omega_0^N). \end{aligned} \tag{5.74}$$

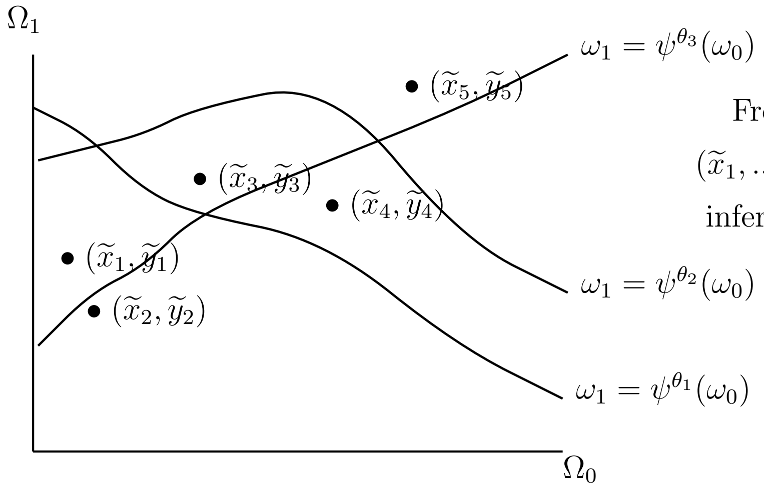
Our present problem is as follows:

(♯) Consider the measurement $\mathbf{M}_{C(\Omega_0^N)}(\bigotimes_{n=1}^N \widetilde{\mathbf{O}}^{\bar{\theta}}, S_{[\otimes_{n=1}^N \delta_{\bar{\omega}_n}]})$ where it is assumed that $\bar{\omega}_1, \bar{\omega}_2, \dots, \bar{\omega}_N$ and $\bar{\theta}$ $(\in \Theta)$ are unknown. Assume that we know that the measured value $(\widetilde{x}_1, \dots, \widetilde{x}_N, \widetilde{y}_1, \dots, \widetilde{y}_N)$ $(\in X^N \times Y^N)$ obtained by the measurement $\mathbf{M}_{C(\Omega_0^N)}$ $(\bigotimes_{n=1}^N \widetilde{\mathbf{O}}^{\bar{\theta}}, S_{[\otimes_{n=1}^N \delta_{\bar{\omega}_n}]})$ belongs to $\prod_{n=1}^N (\Xi_n \times \Gamma_n)$. Then, infer the unknown $\bar{\omega}_1, \bar{\omega}_2, \dots, \bar{\omega}_N$ and $\bar{\theta}$ (particularly, $\bar{\theta}$).

---

[9] Independent random variables with normal distributions

That is, for simplicity under the assumption that $\Omega_0 = X$, $\Omega_1 = Y$, we can illustrate this problem ($\sharp$) as follows:



Question
From the measured data $(\widetilde{x}_1, ..., \widetilde{x}_5, \widetilde{y}_1, ..., \widetilde{y}_5)$ $(\in \Omega_0^5 \times \Omega_1^5)$, infer the reasonable θ.

Answer
"Probably θ_3 !"

This problem is solved as follows: Define the observable $\widehat{\mathbf{O}} \equiv (X^N \times Y^N, \mathcal{F}^N \times \mathcal{G}^N, \widehat{H})$ in $C(\Omega_0^N \times \Theta)$ such that $[\widehat{H}(\Xi_1 \times \cdots \times \Xi_N \times \Gamma_1 \times \cdots \times \Gamma_N)](\omega_1, ..., \omega_N, \theta) = (5.74)$. Note that we have the following identification:

$$\mathbf{M}_{C(\Omega_0^N \times \Theta)}(\widehat{\mathbf{O}}, S_{[(\otimes_{n=1}^N \delta_{\omega_n}) \otimes \delta_\theta]}) = \mathbf{M}_{C(\Omega_0^N)}(\bigotimes_{n=1}^N \widetilde{\mathbf{O}}^\theta, S_{[\otimes_{n=1}^N \delta_{\omega_n}]}).$$

Consider the measurement $\mathbf{M}_{C(\Omega_0^N \times \Theta)}(\widehat{\mathbf{O}}, S_{[(\otimes_{n=1}^N \delta_{\bar{\omega}_n}) \otimes \delta_{\bar{\theta}}]})$ where it is assumed that we do not know $\bar{\omega}_1, \bar{\omega}_2, ..., \bar{\omega}_N, \bar{\theta}$. Then, we can, by Fisher's maximum likelihood method (*cf.* Corollary 5.6), infer the unknown state $(\otimes_{n=1}^N \delta_{\bar{\omega}_n}) \otimes \delta_{\bar{\theta}}$ such that:

$$\begin{aligned} & [\widehat{H}(\Xi_1 \times \cdots \times \Xi_N \times \Gamma_1 \times \cdots \times \Gamma_N)](\bar{\omega}_1, ..., \bar{\omega}_N, \bar{\theta}) \\ = & \max_{(\omega_1, ..., \omega_N, \theta) \in \Omega_0^N \times \Theta} [\widehat{H}(\Xi_1 \times \cdots \times \Xi_N \times \Gamma_1 \times \cdots \times \Gamma_N)](\omega_1, ..., \omega_N, \theta). \end{aligned} \qquad (5.75)$$

This is the answer to the above problem ($\sharp$). It should be noted that the problem ($\sharp$) is stated under the very general situations (i.e., Ω_0, Ω_1, X and Y are not necessarily the real lines $\mathbf{R}$).

In the following example, we apply our result (5.75) to the simple measurement error model (5.73) with normal distributions.

Example 5.21. [The simple example of measurement error model (the case that θ_0, θ_1, ω_1 , ..., ω_N are unknown)]. Let L be a sufficiently large number. Put $\Omega_0 = [-L, L]$, $\Omega_1 = [-L^2 - L, L^2 + L]$, $\Theta = [-L, L]^2$, and define the map $\psi^{(\theta_0, \theta_1)} : \Omega_0 \to \Omega_1$ such that:

$$\psi^{(\theta_1, \theta_2)}(\omega) = \theta_1 \omega + \theta_0 \quad (\forall \omega \in \Omega_0, \forall (\theta_0, \theta_1) \in \Theta.$$

Also, put $(X, \mathcal{F}, F) = (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, G^{\sigma_1})$ in $C(\Omega_0)$ and $(Y, \mathcal{G}, G) = (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, G^{\sigma_2})$ in $C(\Omega_1)$ (*cf.* Example 2.17). Thus, we define the product observable $\widetilde{\mathbf{O}}^{(\theta_0,\theta_1)} = (X \times Y, \mathcal{F} \times \mathcal{G}, H^\theta)$, where $H^\theta \equiv F \times \Psi^\theta G$, in $C(\Omega_0)$ such that:

$$[H^\theta(\Xi \times \Gamma)](\omega) = (\frac{1}{\sqrt{2\pi\sigma_1\sigma_2}})^2 \iint_{\Xi\times\Gamma} \exp[-\frac{(x-\omega)^2}{2\sigma_1^2} - \frac{(y-(\theta_1\omega+\theta_0))^2}{2\sigma_2^2}]dxdy$$
$$(\forall \Xi \in \mathcal{B}_{\mathbf{R}}, \forall \Gamma \in \mathcal{B}_{\mathbf{R}}, \forall \omega \in \Omega_0).$$

Thus, we have the observable $\widehat{\mathbf{O}} = (\mathbf{R}^{2N}, \mathcal{B}_{\mathbf{R}^{2N}}, \widehat{H})$ in $C(\Omega_0^N \times \Theta)$ such that:

$$[\widehat{H}(\Xi_1 \times \cdots \times \Xi_N \times \Gamma_1 \times \cdots \times \Gamma_N)](\omega_1, ..., \omega_N, \theta_0, \theta_1)$$
$$=(\frac{1}{\sqrt{2\pi\sigma_1\sigma_2}})^{2N} \int\cdots\int_{\Pi_{n=1}^N(\Xi_n\times\Gamma_n)} e^{-\frac{\sum_{n=1}^N(x_n-\omega_n)^2}{2\sigma_1^2} - \frac{\sum_{n=1}^N(y_n-(\theta_1\omega_n+\theta_0))^2}{2\sigma_2^2}} dx_1dy_1\cdots dx_Ndy_N. \tag{5.76}$$

Assume the conditions in the problem (♯), and further add that

$$\Xi_n^\epsilon = [\widetilde{x}_n - \epsilon, \widetilde{x}_n + \epsilon], \quad \Gamma_n^\epsilon = [\widetilde{y}_n - \epsilon, \widetilde{y}_n + \epsilon] \quad (\text{for sufficiently small positive } \epsilon).$$

Then, our main result (5.75) says that

$$\max_{(\omega_1,...,\omega_N,\theta_0,\theta_1)\in\Omega_0^N\times\Theta} [\widehat{H}(\Xi_1^\epsilon \times \cdots \times \Xi_N^\epsilon \times \Gamma_1^\epsilon \times \cdots \times \Gamma_N^\epsilon)](\omega_1, ..., \omega_N, \theta)$$
$$\Longleftrightarrow \min_{(\omega_1,...,\omega_N,\theta_0,\theta_1)\in\Omega_0^N\times\Theta} \Big[\sum_{n=1}^N (\frac{\widetilde{x}_n}{\sigma_1} - \frac{\omega_n}{\sigma_1})^2 + \sum_{n=1}^N (\frac{\widetilde{y}_n}{\sigma_2} - (\frac{\theta_1\sigma_1}{\sigma_2}\frac{\omega_n}{\sigma_1} + \frac{\theta_0}{\sigma_2}))^2\Big] \quad (\text{since } \epsilon \text{ is small})$$

(Here, note that the distance between a point $(\frac{\widetilde{x}_n}{\sigma_1}, \frac{\widetilde{y}_n}{\sigma_2})$ and a line $y = \frac{\theta_1\sigma_1}{\sigma_2}x + \frac{\theta_0}{\sigma_2}$ is equal to $\frac{|\widetilde{y}_n - \theta_1\widetilde{x}_n - \theta_0|}{\sqrt{\sigma_2^2+\sigma_1^2\theta_1^2}}$. Then, we see)

$$\Longleftrightarrow \min_{(\theta_0,\theta_1)\in\Theta} \frac{\sum_{n=1}^N(\widetilde{y}_n - \theta_1\widetilde{x}_n - \theta_0)^2}{\sigma_2^2 + \sigma_1^2\theta_1^2} \tag{5.77}$$

$$\Longleftrightarrow \begin{cases} \sum_{n=1}^N(\widetilde{y}_n - \bar{\theta}_1\widetilde{x}_n - \bar{\theta}_0) = 0 & (\leftarrow \frac{\partial}{\partial\theta_0}(5.77) = 0), \\ \sum_{n=1}^N(\bar{\theta}_1\widetilde{y}_n\sigma_1^2 + \widetilde{x}_n\sigma_2^2 - \bar{\theta}_0\bar{\theta}_1\sigma_1^2)(\widetilde{y}_n - \bar{\theta}_1\widetilde{x}_n - \bar{\theta}_0) = 0 & (\leftarrow \frac{\partial}{\partial\theta_1}(5.77) = 0). \end{cases} \tag{5.78}$$

Thus, the unknown parameters $\bar{\theta}_1$ and $\bar{\theta}_2$ are inferred by the solution of this equation (5.78). Note that this is a direct consequence of our main result (5.75).

■

Example 5.22. [The case that θ_0, θ_1, σ_1^2, σ_2^2, ω_1 , ..., ω_N are unknown]. Assume that θ_0, θ_1, σ_1^2, σ_2^2, ω_1 , ..., ω_N are unknown. The log-likelihood is

$$L(\theta_0, \theta_1, \sigma_1^2, \sigma_2^2, \omega_1, ..., \omega_N) = \log[(5.76)]$$
$$= -\frac{N\log\sigma_1^2}{2} - \frac{N\log\sigma_2^2}{2} - \frac{\sum_{n=1}^N(x_n-\omega_n)^2}{2\sigma_1^2} - \frac{\sum_{n=1}^N(y_n-\theta_0-\theta_1\omega_n)^2}{2\sigma_2^2}.$$

Taking partial derivatives with respect to θ_0, θ_1, σ_1^2, σ_2^2 and ω_1, ..., ω_N, and equating the results to zero, gives the likelihood equations,

$$\sum_{n=1}^{N}(y_n-\theta_0-\theta_1\omega_n)=0, \qquad \sum_{n=1}^{N}(y_n-\theta_0-\theta_1\omega_n)\omega_n=0,$$

$$\frac{\sum_{n=1}^{N}(x_n\omega_n)^2}{N}=\sigma_1^2, \qquad \frac{\sum_{n=1}^{N}(y_n-\theta_0-\theta_1\omega_n)^2}{N}=\sigma_2^2,$$

$$\frac{(x_n\omega_n)^2}{2\sigma_1^2}-\frac{(y_n-\theta_0-\theta_1\omega_n)^2}{2\sigma_2^2}=0, \qquad (n=1,2,...,N).$$

Thus we can easily solve it as follows:

$$\theta_1^2=\frac{\sigma_2^2}{\sigma_1^2}=\frac{S_{yy}}{S_{xx}}, \quad 2\sigma_1^2=S_{xx}-\frac{S_{xy}}{\theta_1}, \quad 2\sigma_2^2=S_{yy}-S_{xy}\theta_1,$$

$$\theta_0=\bar{y}-\theta_1\bar{x}, \quad 2\omega_n=x_n+\frac{y_n-\theta_0}{\theta_1}=x_n+\bar{x}+\frac{y_n-\bar{y}_n}{\theta_1},$$

where

$$\bar{x}=\frac{x_1+\cdots+x_N}{N}, \qquad \bar{y}=\frac{y_1+\cdots+y_N}{N},$$

$$S_{xx}=\frac{(x_1-\bar{x})^2+\cdots+(x_N-\bar{x})^2}{N}, \quad S_{yy}=\frac{(y_1-\bar{y})^2+\cdots+(y_N-\bar{y})^2}{N},$$

$$S_{xy}=\frac{(x_1-\bar{x})(y_1-\bar{y})+\cdots+(x_N-\bar{x})(y_N-\bar{y})}{N}.$$

(Cf. Cheng, etc. [16]).

■

5.6 Appendix (Iterative likelihood function method)

In this section we study the "Iterative likelihood function method (*cf.* [47])", which will be related to subjective Bayesian statistics (see §8.6 later).

Consider the "measurement" described in the following "step [1]" and "step [2]",

[1] First we take a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}_1 \equiv (X, 2^X, F), S_{[*]})$, and we know that the measured value is equal to x ($\in X$).

[2] And successively, we take a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}_2 \equiv (Y, 2^Y, G), S_{[*]})$, and we know that the measured value is equal to y ($\in Y$).

Note that "[1]+[2]" ie equal to the following [3][10] :

[3] We take a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}_1 \times \mathbf{O}_2 \equiv \equiv (X \times Y, \mathcal{F} \times \mathcal{G}, H \equiv F \times G)\ S_{[*]})$, and we know that the measured value obtained by $\mathbf{M}_{C(\Omega)}(\mathbf{O}_1 \times \mathbf{O}_2, S_{[*]})$ is equal to $(x, y)\ (\in X \times Y)$.

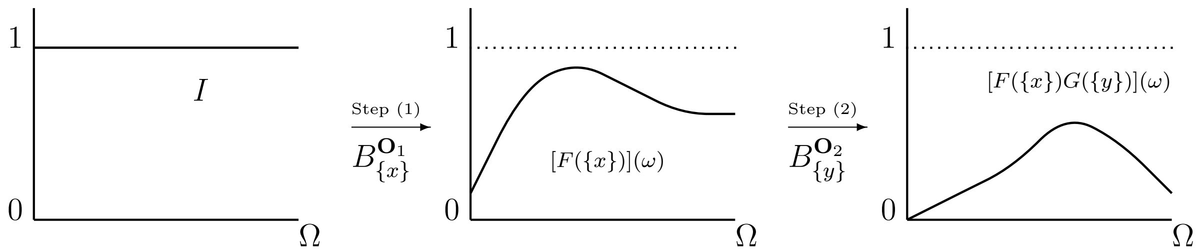
$\big($A non-negative (real-valued) continuous function $F(\Xi)$ in an observable $(X, \mathcal{F}, F)$ is called *a likelihood function*, or, *a likelihood quantity.*$\big)$ Then we can say:

[♭] By Step [1], we get the likelihood function $F(\{x\})$. And further by step [2] (i.e., by "[1]+[2]" (=[3])), we get the new likelihood function $F(\{x\})G(\{y\})\ (\equiv [F \times G](\{x\} \times \{y\}))$.

Using the Bayes operator (*cf.* the formula (5.12)), this statement [♭] can be rewritten as follows:

$$I \xrightarrow[B_{\{x\}}^{\mathbf{O}_1}]{\text{Step (1)}} F(\{x\}) \xrightarrow[B_{\{y\}}^{\mathbf{O}_2}]{\text{Step (2)}} F(\{x\})G(\{y\}) \qquad \text{in } C(\Omega), \tag{5.79}$$

where $I(\in C(\Omega))$ is the identity element, i.e., the constant function such that $I(\omega) = 1 (\forall \omega \in \Omega)$.



It should be noted that:

(F_1) the constant likelihood function "I" (or "$k \times I$" where $k > 0$) is the likelihood function that represents the fact "we have no information about the system $S_{[*]}$".

Now we introduce the following notation. Cf. [47].

Notation 5.23. $[S_{[*]}(\!(G)\!)_{lq}]$. *The system $S_{[*]}$ (formulated in $C(\Omega)$) such that we know it has the likelihood quantity G ($G \in C(\Omega)$, $0 \le G(\omega)$ $(\forall \omega \in \Omega)$) is denoted by $S_{[*]}(\!(G)\!)_{lq}$.*

[10] Recall §2.5 (Remarks(II)), that is, "Only one measurement is permitted to be conducted". Thus, "[1]+[2]" is a methodological explanation.

Thus, the symbol $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\!(kG)\!)_{lq})$ *means "the measurement* $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$ *under the condition that we know the likelihood quantity of the system* $S_{[*]}$ *is equal to* kG*, where* $G \in C(\Omega)$*,* $0 \le G(\omega)$ $(\forall \omega \in \Omega)$*".*

■

Under this notation, the conventional Fisher's maximum likelihood method (i.e., Corollary 5.6) says that:

(F_1') Assume that we first have no information about the system $S_{[*]}$. And we take a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$, i.e., $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\!(kI)\!)_{lq})$. Then, from the fact that the measured value x $(\in X)$ is obtained by the $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\!(kI)\!)_{lq})$, we know that the likelihood quantity of the system $S_{[*]}$ is equal to $k[F(\{x\})](\omega)$. $\Big($Thus, there is a reason to regard the unknown state $[\,*\,]$ as the state $\omega_0 (\in \Omega)$ such that $k[F(\{x\})](\omega_0) = \max_{\omega \in \Omega} k[F(\{x\})](\omega)$.$\Big)$

However, it is usual to assume that we have a little bit of information before a measurement. Thus, let us start from the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (X, 2^X, F), S_{[*]}(\!(G_0)\!)_{lq})$. Here we have the following problem:

($\mathrm{P_G}$) How to infer the new likelihood quantity of the system $S_{[*]}$ from the fact that the measured value x $(\in X)$ is obtained by the $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\!(G_0)\!)_{lq})$.

This is equivalent to the following problem:

($\mathrm{P_G'}$) How to infer the likelihood quantity of the system $S_{[*]}$ from the fact that the measured value (y_0, x) $(\in \{y_0, y_1\} \times X)$ is obtained by the iterated measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}_0 \times \mathbf{O}, S_{[*]}(\!(kI)\!)_{lq})$, where $\mathbf{O}_0 = (\{y_0, y_1\}, 2^{\{y_0, y_1\}}, G)$ and $G(\{y_0\}) = G_0$, $G(\{y_1\}) = I - G_0$.

Thus, from (F_1') and "($\mathrm{P_G}$)$\leftrightarrow$($\mathrm{P_G'}$)", the problem ($\mathrm{P_G}$) is solved as follows:

(F_2) (The answer of the (P_G)): We know that the new likelihood quantity G_{new} of the system $S_{[*]}$ is equal to $B_{\{x\}}^{\mathbf{O}}(G_0)$. Here, Bayes operator $B_{\{x\}}^{\mathbf{O}} : C(\Omega) \to C(\Omega)$ is defined by $B_{\{x\}}^{\mathbf{O}}(G) = F(\{x\})G$ $(\forall G \in C(\Omega))$.

Thus we see:

$$S_{[*]}(\!(I)\!)_{lq} \xrightarrow[x \text{ is obtained}]{\mathbf{M}_{C(\Omega)}(\mathbf{O}_1,\, S_{[*]}(\!(I)\!)_{lq})} S_{[*]}(\!(F(\{x\}))\!)_{lq} \xrightarrow[y \text{ is obtained}]{\mathbf{M}_{C(\Omega)}(\mathbf{O}_2,\, S_{[*]}(\!(F(\{x\}))\!)_{lq})} S_{[*]}(\!(F(\{x\})G(\{y\}))\!)_{lq}$$

where $\mathbf{O}_1 = (X, 2^X, F)$ and $\mathbf{O}_2 = (Y, 2^Y, G)$.

Summing up, we can symbolically describe it as follows:

$$\begin{cases} [F_1] \quad \text{No information quantity} \quad \longleftrightarrow \quad kI(\in C(\Omega)) \\ [F_2] \quad S_{[*]}((G))_{lq} \xrightarrow[x \text{ is obtained}]{\mathbf{M}_{C(\Omega)}(\mathbf{O},\, S_{[*]}((G))_{lq})} S_{[*]}((B^{\mathbf{O}}_{\{x\}}G))_{lq} \Big(= S_{[*]}((F(\{x\})G))_{lq} \Big), \end{cases} \tag{5.80}$$

where $\mathbf{O} = (X, 2^X, F)$.

The following example will promote the understanding of "iterative likelihood function method".

Example 5.24. [The urn problem]. There are two urns ω_1 and ω_2. The urn ω_1 [resp. ω_2] contains 8 white and 2 black balls [resp. 4 white and 6 black balls]. Assume that they can not be distinguished in appearance.

- Choose one urn from the two.

Now you sample, randomly, with replacement after each ball.

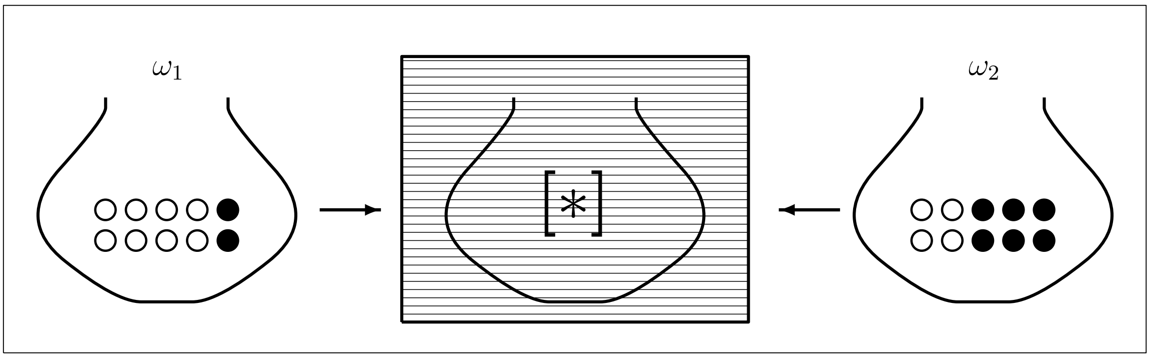
(i). First, you get "white ball".

(Q_1) *Which is the chosen urn, ω_1 or ω_2?*

(ii). Further, assume that you continuously get "black".

(Q_2) *How about the case? Which is the chosen urn, ω_1 or ω_2?*

The illustration of $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$ (or, $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}((kI))_{lq})$)



[*Answers*]. In what follows this problem is studied in the iterative likelihood function method. Put $\Omega = \{\omega_1, \omega_2\}$. $\mathbf{O} = (\{w, b\}, 2^{\{w,b\}}, F)$ where

$$[F(\{w\})](\omega_1) = 0.8,\ [F(\{b\})](\omega_1) = 0.2,\ [F(\{w\})](\omega_2) = 0.4,\ [F(\{b\})](\omega_2) = 0.6. \tag{5.81}$$

The situation of no information in Fisher's method is represented by kI $(k > 0)$. Thus, it suffices to consider the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O},\ S_{[*]}((kI))_{lq})$. Since the measured value "w" was obtained, the new likelihood quantity $G_{\rm new}$ is given as follows:

$$\begin{aligned} G_{\rm new}(\omega_1)\Big(&= kI \cdot [F(\{w\})](\omega_1)\Big) = 0.8k, \\ G_{\rm new}(\omega_2)\Big(&= kI \cdot [F(\{w\})](\omega_2)\Big) = 0.4k. \end{aligned} \tag{5.82}$$

Thus, by Fisher's maximum likelihood method, we see that

(A_1) *there is a reason to infer that* $[\,*\,] = \omega_1$.

For the further case, it suffices to consider the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O},\ S_{[*]}((G_{\rm new}))_{lq})$. Thus we similarly calculate that

$$\begin{aligned} G_{\rm new^2}(\omega_1)\ \Big(&= [G_{\rm new}](\omega_1) \cdot [F(\{b\})](\omega_1)\Big) = 0.16k, \\ G_{\rm new^2}(\omega_2)\ \Big(&= [G_{\rm new}](\omega_2) \cdot [F(\{b\})](\omega_2)\Big) = 0.24k. \end{aligned} \tag{5.83}$$

Thus we, by Fisher's maximum likelihood method, see that

(A_2) *there is a reason to infer that* $[\,*\,] = \omega_2$.

■

Chapter 6

Fisher's statistics II (related to Axioms 1 and 2)

As mentioned in Chapters 2 and 3, measurement theory is formulated as follows:

$$\text{PMT} = \underset{[\text{Axiom 1 (2.37)}]}{\text{measurement}} + \underset{[\text{Axiom 2 (3.26)}]}{\text{the relation among systems}} \quad \text{in } C^*\text{-algebra} \;. \tag{6.1 (=(1.4))}$$

In this chapter we study the relation between Fisher's statistics (mentioned in the previous chapter) and Axiom 2. Particularly we show that regression analysis can be completely understood within the framework of Axioms 1 and 2. We expect that our result will make the readers notice that regression analysis is more profound than they usually think. As mentioned in Chapter 1 (*cf.* Declaration (1.11)), we assert that the results in Chapters 5 and 6 guarantee that "Fisher's statistics is theoretically true (in PMT)".[1]

6.1 Regression analysis I

6.1.1 Introduction

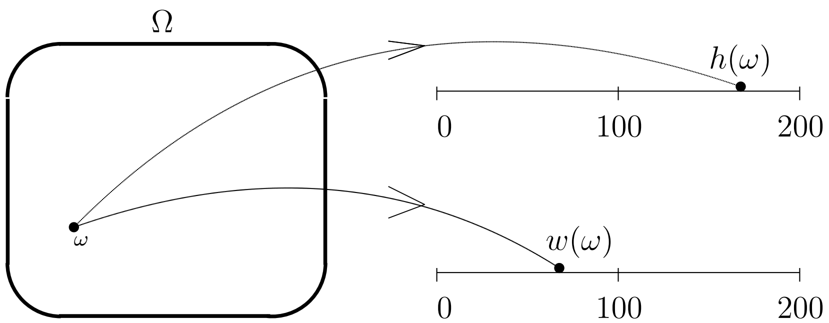
The purpose of this chapter is to study and understand "regression analysis" completely under Axiom 1 and 2 (of measurement theory). The following Example 6.1 is the most typical in all examples of "regression analysis".

Example 6.1. [A typical example of regression analysis]. Let $\Omega \equiv \{\omega_1, \omega_2, ..., \omega_{100}\}$ be a set of all students of a certain high school. Define $h : \Omega \to [0, 200]$ [resp. $w : \Omega \to [0, 200]$]

[1] We believe that only "Fisher's maximum likelihood method" and "regression analysis" are most essential in statistics. Thus we believe that, in order to justify statistics, it suffices to show that the two (i.e., "Fisher's maximum likelihood method" and "regression analysis") are formulated in PMT.

such that:

$$h(\omega_n) = \text{“the height of a student } \omega_n\text{”} \quad (n = 1, 2, ..., 100)$$
$$\Big[\text{ resp. } w(\omega_n) = \text{“the weight of a student } \omega_n\text{”} \quad (n = 1, 2, ..., 100)\Big]$$



(Note that this is a special case of Fig. (3.20).) Assume that:

(1) The principal of this high school knows the both functions h and w. That is, he knows the exact data of the height and weight concerning all students.

Also, assume that:

(2) Some day, a certain student helped a drowned girl. But, he left without reporting the name. Thus, all information that the principal knows is as follows:

(i) he is a student of his high school.

(ii) his height [resp. weight] is about 170 cm [resp. about 80 kg].

Now we have the following question:

- Under the above assumption (1) and (2), how does the principal infer who is he?

This is just what regression analysis says. For the solution, see Regression Analysis I (6.7) later.

■

In order to explain our main assertion, let us begin with the following Example 6.2 (the conventional argument of regression analysis in Fisher's maximum likelihood method), which is easy and well-known.

Example 6.2. [The conventional argument of regression analysis in Fisher's method]. We have a rectangular water tank filled with water. Assume that the height of water at time t is given by the following function $h(t)$:

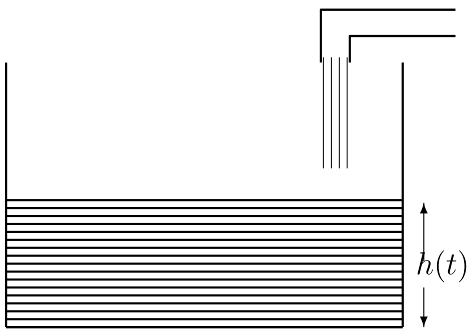
$$h(t) = \alpha_0 + \beta_0 t, \tag{6.2}$$

where α_0 and β_0 are unknown fixed parameters such that α_0 is the height of water filling the tank at the beginning and β_0 is the increasing height of water per unit time. The measured height $h_m(t)$ of water at time t is assumed to be represented by

$$h_m(t) = \alpha_0 + \beta_0 t + e(t), \tag{6.3}$$

where $e(t)$ represents a noise (or more precisely, a measurement error) with some suitable conditions. And assume that we obtained the measured data of the heights of water at $t = 1, 2, 3$ as follows:

$$h_m(1) = 1.9, \quad h_m(2) = 3.0, \quad h_m(3) = 4.7. \tag{6.4}$$



Under this setting, we consider the following problem:

(i) Infer the true value $h(2)$ of the water height at $t = 2$ from the measured data (6.4).

This problem (i) is usually solved as follows: From the theoretical point of view, we can infer, by Fisher's maximum likelihood method and regression analysis, that

$$(\alpha_0, \beta_0) = (0.4, 1.4). \tag{6.5}$$

$\Big($For the derivation of (6.5) from (6.4), see Example 6.4 (6.16) later.$\Big)$ And next, we can infer that

$$h(2) = 3.2, \tag{6.6}$$

by the calculation: $h(2) = 0.4 + 1.4 \times 2 = 3.2$. This is the answer to the problem (i).

■

The above argument in Example 6.2 is, of course, well known and adopted as the usual regression analysis. Thus all statisticians may think that there is no serious problem in regression analysis. However it is not true. For example, we have the basic problem in the argument of Example 6.2 as follows:

(ii) What kinds of axioms are hidden behind the argument in Example 6.2? And moreover, justify the argument in Example 6.2 under the axioms.

It is important. If we have no answer to the question: "What kinds of rules are permitted to be used in statistics?", we can not prove (or, justify) that the argument in Example 6.2 is true (or not). That is because there is no justification without an axiomatic formulation. In this sense, we believe that the above question (ii) is the most important problem in theoretical statistics. Also, if some know the great success of the axiomatic formulation in physics (e.g., the three laws in Newtonian mechanics, or von Neumann's formulation of quantum mechanics, *cf.* [71], [84]), it is a matter of course that they want to understand statistics axiomatically.

Trying to solve the problem (ii), some may consider as follows:

(iii) Firstly, Fisher's maximum likelihood method should be declared as an axiom (*cf.* Corollary 5.6). Also, the derivation of the (6.6) from the (6.5) should be justified under some axioms. That is, it must not be accepted as a common sense.

This opinion (iii) may not be far from our assertion proposed in this chapter. However, in order to describe the above (iii) precisely, we must make vast preparations.

Our standing point of this book is extremely theoretical (and not practical). However we expect that many statisticians will be interested in our proposal. That is because we believe that every statistician may want to know the justification of both the (6.5) and the (6.6) in Example 6.2.

6.1.2 Regression analysis I in measurements

By the results in the previous chapters (i.e., Theorem 3.7 and Corollary 5.6), we can easily propose:

REGRESSION ANALYSIS I [The conventional regression analysis in PMT]. (6.7)
Let $(T \equiv \{0,1,...,N\}, \pi : T \setminus \{0\} \to T)$ *be a tree with root* 0*, and let* $\mathbf{S}_{[*]} \equiv [S_{[*]}; \{\Phi_{\pi(t),t} : C(\Omega_t) \to C(\Omega_{\pi(t)})\}_{t \in T\setminus\{0\}}]$ *be a general system with the initial system* $S_{[*]}$*. And, let an observable* $\mathbf{O}_t \equiv (X_t, 2^{X_t}, F_t)$ *in a* C^**-algebra* $C(\Omega_t)$ *be given for each* $t \in T$*. Let* $\widetilde{\mathbf{O}}_0$ *be the Heisenberg picture representation of the sequential observable* $[\{\mathbf{O}_t\}_{t\in T}, \{\Phi_{t,\pi(t)} : C(\Omega_t) \to C(\Omega_{\pi(t)})\}_{t\in T\setminus\{0\}}]$ *in* $C(\Omega_0)$*. Then, we have a measurement*

$$\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, 2^{\prod_{t\in T} X_t}, \widetilde{F}_0), S_{[*]}). \qquad (\textit{cf. Theorem 3.7}).$$

Assume that the measured value by the measurement $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[*]})$ *belongs to* $\prod_{t\in T} \Xi_t$ $(\in 2^{\prod_{t\in T} X_t})$*. Then, there is a reason to infer that the state* $[*]$ *of the system* S $\Big($*i.e., the state before the measurement* $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0,\ S_{[*]})\Big)$*, the state after the measurement* $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[*]})$ *and the* δ_{ω_0} $(\in \mathcal{M}_{+1}^p(\Omega))$ $\Big($*defined by (6.9)*$\Big)$ *are equal. That is, Corollary 5.6 says that there is a reason to infer that*

$$[\,*\,] = \text{ “the state after the measurement } \mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[*]})\text{”} = \delta_{\omega_0}. \tag{6.8}$$

Here the δ_{ω_0} $(\in \mathcal{M}_{+1}^p(\Omega_0))$ *is defined by*

$$[\widetilde{F}_0(\prod_{t\in T} \Xi_t)](\omega_0) = \max_{\omega\in\Omega_0} [\widetilde{F}_0(\prod_{t\in T} \Xi_t)](\omega). \tag{6.9}$$

■

Remark 6.3. [Regression analysis I]. The above regression analysis is quite applicable. For example, note that the “$\Phi_{\pi(t),t} : C(\Omega_t) \to C(\Omega_{\pi(t)})$” is generally assumed to be Markov operators (and not homomorphisms). In this sense, Regression analysis I may not be “conventional”.

■

Now we shall review Example 6.2 in the light of Regression Analysis I.

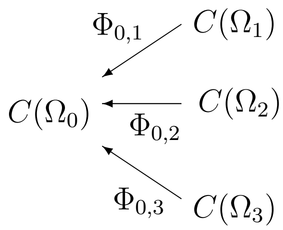
Example 6.4. [Continued from Example 6.2, the conventional argument of regression analysis in Fisher's method]. Put $\Omega_0 = [0.0,\ 1.0] \times [0.0,\ 2.0]$, and put $\Omega_1 = \Omega_2 = \Omega_3 = [0.0,\ 10.0]$. For each t $(\in \{1,2,3\})$, define a continuous map $\phi_{0,t} : \Omega_0 \to \Omega_t$ such that:

$$\Omega_0(\equiv [0.0,\ 1.0] \times [0.0,\ 2.0]) \ni \omega \equiv (\alpha, \beta) \underset{\phi_{0,t}}{\mapsto} \alpha + \beta t \in \Omega_t(\equiv [0.0,\ 10.0]). \tag{6.10}$$

Thus, for each t $(\in \{1,2,3\})$, we have a homomorphism $\Phi_{0,t} : C(\Omega_t) \to C(\Omega_0)$ such that:

$$[\Phi_{0,t} f_t](\omega) = f_t(\phi_{0,t}(\omega)) \qquad (\forall \omega \in \Omega_0, \forall f_t \in C(\Omega_t)). \tag{6.11}$$

It is usual to assume that regression analysis is applied to the system with a parallel structure such as in the figure (6.12). $\Big($From the peculiarity of this problem, we can also assume that this system has a series structure. However, we are not concerned with it.$\Big)$



(6.12)

For each $t \in \{1,2,3\}$, consider the discrete Gaussian observable $\mathbf{O}_{\sigma^2,N} \equiv (X_N, 2^{X_N}, F_{\sigma,N})$ in $C(\Omega_t)$, (*cf.*(2.60) in Example 2.18). That is,

$$\Omega_t = [0.0,\ 10.0], \qquad X_N = \{\frac{k}{N} \mid k = 0, \pm 1, \pm 2, ..., \pm N^2\},$$

and

$$[F_{\sigma,N}(\{k/N\})](\omega) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} \int_{N-\frac{1}{2N}}^{\infty} \exp[-\frac{(x-\omega)^2}{2\sigma^2}]dx & (k = N^2, \forall \omega \in [a,b]), \\ \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\frac{k}{N}-\frac{1}{2N}}^{\frac{k}{N}+\frac{1}{2N}} \exp[-\frac{(x-\omega)^2}{2\sigma^2}]dx & (\forall k = 0, \pm 1, \pm 2, ..., \pm (N^2-1), \quad \forall \omega \in [a,b]), \\ \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{-N+\frac{1}{2N}} \exp[-\frac{(x-\omega)^2}{2\sigma^2}]dx & (k = -N^2, \forall \omega \in [a,b]). \end{cases}$$

(*cf.* (2.aa60) in Example 2.18)

Here, we define the observable $\widetilde{\mathbf{O}}_0 \equiv (X_N^3, 2^{X_N^3}, \widetilde{F}_0)$ in $C(\Omega_0)$ such that:

$$\begin{aligned} &[\widetilde{F}_0(\Xi_1 \times \Xi_2 \times \Xi_3)](\omega) = [\Phi_{0,1} F_{\sigma^2,N}](\omega) \cdot [\Phi_{0,2} F_{\sigma^2,N}](\omega) \cdot [\Phi_{0,3} F_{\sigma^2,N}](\omega) \\ =&[F_{\sigma^2,N}(\Xi_1)](\phi_{0,1}(\omega)) \cdot [F_{\sigma^2,N}(\Xi_2)](\phi_{0,2}(\omega)) \cdot [F_{\sigma^2,N}(\Xi_3)](\phi_{0,3}(\omega)) \\ &(\forall \Xi_1, \Xi_2, \Xi_3 \in 2^{X_N},\ \forall \omega = (\alpha, \beta) \in \Omega_0 = [0.0,\ 1.0] \times [0.0,\ 2.0]). \end{aligned} \tag{6.13}$$

Then, we have the measurement $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[*]})$. The (6.4) says that the measured value obtained by the measurement $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[*]})$ is equal to

$$(1.9,\ 3.0,\ 4.7)\ (\in X_N^3). \tag{6.14}$$

Here, Fisher's method (Corollary 5.6) says that it suffices to solve the problem

$$\text{"Find } (\alpha_0, \beta_0) \text{ such as } \max_{(\alpha,\beta)\in\Omega_0} [\widetilde{F}_0(\{1.9\} \times \{3.0\} \times \{4.7\}(\alpha, \beta)\text{".} \tag{6.15}$$

Putting

$$\Xi_1 = [1.9 - \frac{1}{2N}, 1.9 + \frac{1}{2N}], \Xi_2 = [3.0 - \frac{1}{2N}, 3.0 + \frac{1}{2N}], \Xi_3 = [4.7 - \frac{1}{2N}, 4.7 + \frac{1}{2N}],$$

we see, under the assumption that N is sufficiently large, that

$$\begin{aligned}
&(6.15) \Rightarrow \max_{(\alpha,\beta)\in\Omega_0} \frac{1}{\sqrt{2\pi\sigma^2}^3} \iiint_{\Xi_1\times\Xi_2\times\Xi_3} e^{[-\frac{(x_1-(\alpha+\beta))^2+(x_2-(\alpha+2\beta))^2+(x_3-(\alpha+3\beta))^2}{2\sigma^2}]} dx_1 dx_2 dx_3 \\
&\Rightarrow \max_{(\alpha,\beta)\in\Omega_0} \exp\Big(- [(1.9-(\alpha+\beta))^2 + (3.0-(\alpha+2\beta))^2 + (4.7-(\alpha+3\beta))^2]/(2\sigma^2)\Big) \\
&\Rightarrow \min_{(\alpha,\beta)\in\Omega_0} [(1.9-(\alpha+\beta))^2 + (3.0-(\alpha+2\beta))^2 + (4.7-(\alpha+3\beta))^2] \\
&\qquad\qquad \text{(by the least squares method)} \\
&\Rightarrow \begin{cases} (1.9-(\alpha+\beta)) + (3.0-(\alpha+2\beta)) + (4.7-(\alpha+3\beta)) = 0 \\ (1.9-(\alpha+\beta)) + 2(3.0-(\alpha+2\beta)) + 3(4.7-(\alpha+3\beta)) = 0 \end{cases} \\
&\Rightarrow \quad (\alpha_0, \beta_0) = (0.4, 1.4). \qquad (6.16)
\end{aligned}$$

This is the conclusion of Regression Analysis I (6.7). Also, using the notations in Regression Analysis I, we remark that:

(R) the measurement $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t,\ 2^{\prod_{t\in T} X_t},\ \widetilde{F}_0), S_{[*]})$ is hidden behind the inference (6.16) $\Big(= (6.5)$ in Example 6.2$\Big)$.

This fact will be important in §6.3.

■

The above may be the standard argument of the conventional regression analysis in measurement theory. However, our problem (i) in Example 6.2 is not to infer the (α_0, β_0) but $h(2)$. In this sense the above regression analysis I is not sufficient. As the answer of the problem (i) in Example 6.2, we usually consider that it suffices to calculate $h(2)$ $\Big(\equiv \phi_{0,2}(0.4, 1.4)\Big)$ in the following:

$$h(2) = 0.4 + 1.4 \times 2 = 3.2. \qquad (6.17)$$

However, this is doubtful. $\Big($In fact, this (6.17) is not always true in general situations. (*cf.* Regression analysis II (6.51) later).$\Big)$ We should not rely on "a common sense" but Axioms 1 and 2. That is, we must solve the problem:

- How can the above (6.17) $\Big(= (6.6)$ in Example 6.2$\Big)$ be deduced from Axioms 1 and 2?

In order to do this, we will make some preparations in the next section.

6.2 Bayes operator, Schrödinger picture, and S-states

In order to improve Regression Analysis I (introduced in the previous section), in this section we make some preparations (i.e., Bayes operator, Schrödinger picture, S-state, etc.). Our main assertion (Regression Analysis II) will be mentioned in §6.3. We begin with the following definition, which is a general form of "Bayes operator" in Remark 5.7.

Definition 6.5. [Bayes operator (or precisely, Bayes-Kalman operator)]. *Let* $(T \equiv \{0,1,...,N\}, \pi : T \setminus \{0\} \to T)$ *be a tree with root* 0 *and let* $\mathbf{S}_{[*]} \equiv [S_{[*]}; \{C(\Omega_t) \overset{\Phi_{\pi(t),t}}{\to} C(\Omega_{\pi(t)})\}_{t \in T\setminus\{0\}}]$ *be a general system with the initial system* $S_{[*]}$*. And, let an observable* $\mathbf{O}_t \equiv (X_t, \mathcal{F}_t, F_t)$ *in* $C(\Omega_t)$ *be given for each* $t \in T$*. Let* $\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, \bigotimes_{t\in T} \mathcal{F}_t, \widetilde{F}_0)$ *be as in Theorem 3.7 in the case* $\mathcal{A}_t = C(\Omega_t)$ $(\forall t \in T)$*. That is,* $\widetilde{\mathbf{O}}_0$ *is the Heisenberg picture representation of the sequential observable* $[\{\mathbf{O}_t\}_{t\in T}; \{C(\Omega_t) \overset{\Phi_{\pi(t),t}}{\to} C(\Omega_{\pi(t)})\}_{t\in T\setminus\{0\}}]$*. Let* τ *be any element in* T*. If a positive bounded linear operator* $B^{(0,\tau)}_{\prod_{t\in T}\Xi_t} : C(\Omega_\tau) \to C(\Omega_0)$ *satisfies the following condition* (BO)*, we call* $\{B^{(0,\tau)}_{\prod_{t\in T}\Xi_t} \mid \Xi_t \in \mathcal{F}_t \ (\forall t \in T)\}$ [*resp.* $B^{(0,\tau)}_{\prod_{t\in T}\Xi_t}$] *a family of Bayes operators* [*resp. a Bayes operator*]*:*

(BO) *for any observable* $\mathbf{O}'_\tau \equiv (Y_\tau, \mathcal{G}_\tau, G_\tau)$ *in* $C(\Omega_\tau)$*, there exists an observable* $\widehat{\mathbf{O}}_0 \equiv \Big((\prod_{t\in T} X_t) \times Y_\tau, (\bigotimes_{t\in T}\mathcal{F}_t) \bigotimes \mathcal{G}_\tau), \widehat{F}_0\Big)$ *in* $C(\Omega_0)$ *such that*

- (i) $\widehat{\mathbf{O}}_0$ *is the Heisenberg picture representation of* $[\{\overline{\mathbf{O}}_t\}_{t\in T}; \{C(\Omega_t) \overset{\Phi_{\pi(t),t}}{\to} C(\Omega_{\pi(t)})\}_{t\in T\setminus\{0\}}]$*, where* $\overline{\mathbf{O}}_t = \mathbf{O}_t$ *(if* $t \neq \tau$*),* $= \mathbf{O}_\tau \times \mathbf{O}'_\tau$ *(if* $t = \tau$*),*
- (ii) $\widehat{F}_0((\prod_{t\in T}\Xi_t) \times \Gamma_\tau) = B^{(0,\tau)}_{\prod_{t\in T}\Xi_t}(G_\tau(\Gamma_\tau)) \quad (\forall \Xi_t \in \mathcal{F}_t \ (\forall t \in T), \forall \Gamma_\tau \in \mathcal{G}_\tau)$*,*
- (iii) $\widehat{F}_0((\prod_{t\in T}\Xi_t) \times Y_\tau) = \widetilde{F}_0(\prod_{t\in T}\Xi_t)\Big(\equiv B^{(0,\tau)}_{\prod_{t\in T}\Xi_t}(1_\tau)\Big)$*,* $(\forall \Xi_t \in \mathcal{F}_t \ (\forall t \in T))$*, where* 1_τ *is the identity in* $C(\Omega_\tau)$*.*

Also, define the map $R^{(0,\tau)}_{\prod_{t\in T}\Xi_t} : \mathcal{M}^m_{+1}(\Omega_0) \to \mathcal{M}^m_{+1}(\Omega_\tau)$ *such that:*

$$R^{(0,\tau)}_{\prod_{t\in T}\Xi_t}(\nu) = \frac{(B^{(0,\tau)}_{\prod_{t\in T}\Xi_t})^*(\nu)}{\|(B^{(0,\tau)}_{\prod_{t\in T}\Xi_t})^*(\nu)\|_{\mathcal{M}(\Omega_\tau)}} \qquad (\forall \nu \in \mathcal{M}^m_{+1}(\Omega_0)), \tag{6.18}$$

where $(B^{(0,\tau)}_{\prod_{t\in T}\Xi_t})^* : C(\Omega_0)^* \to C(\Omega_\tau)^*$ *is the adjoint operator of* $B^{(0,\tau)}_{\prod_{t\in T}\Xi_t} : C(\Omega_\tau) \to C(\Omega_0)$*. The map* $R^{(0,\tau)}_{\prod_{t\in T}\Xi_t}$ *is called a "normalized dual Bayes operator". Bayes operator is also called "Bayes-Kalman operator".*

■

We see

$$B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(g_\tau) \leq \Phi_{0,\tau} g_\tau \quad (\forall g_\tau \in C(\Omega_\tau) \text{ such that } g_\tau \geq 0), \tag{6.19}$$

because it holds, for any observable $\mathbf{O}'_\tau \equiv (Y_\tau, \mathcal{G}_\tau, G_\tau)$ in $C(\Omega_\tau)$,

$$\begin{aligned} B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(G_\tau(\Gamma_\tau)) &= \widehat{F}_0((\prod_{t\in T}\Xi_t)\times\Gamma_\tau) \leq \widehat{F}_0((\prod_{t\in T}X_t)\times\Gamma_\tau) \\ &= \Phi_{0,\tau}G_\tau(\Gamma_\tau)\Big(=B^{(0,\tau)}_{\Pi_{t\in T}X_t}(G_\tau(\Gamma_\tau))\Big) \qquad (\forall \Gamma_\tau \in \mathcal{F}_\tau). \end{aligned} \tag{6.20}$$

The following theorem is essential to Regression Analysis II later.

Theorem 6.6. [The existence theorem of the Bayes operator (*cf.* [46, 55])]. *Let* $\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t,\ 2^{\prod_{t\in T} X_t},\ \widetilde{F}_0)$ *be as in Theorem 3.7 in the case* $\mathcal{A}_t = C(\Omega_t)$ $(\forall t \in T)$. *And, for any* s $(\in T)$, *put* $T_s \equiv \{t \in T \mid s \leq t\}$. *Assume that, for each* s $(\in T)$, *there exists an observable* $\widetilde{\mathbf{O}}_s \equiv (\prod_{t\in T_s} X_t, 2^{\prod_{t\in T_s} X_t}, \widetilde{F}_s)$ *in* $C(\Omega_s)$ *such that* $\Phi_{\pi(s),s}\widetilde{F}_s(\prod_{t\in T_s}\Xi_t) = \widetilde{F}_{\pi(s)}\Big((\prod_{t\in T_{\pi(s)}\setminus T_s} X_t)\times(\prod_{t\in T_s}\Xi_t)\Big)$ $(\forall \Xi_t \in 2^{X_t}$ $(\forall t \in T))$, *(cf. Theorem 3.7). Let* τ *be any element in* T. *Then, there exists a family of Bayes operators* $\{B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t} \mid \Xi_t \in 2^{X_t}$ $(\forall t \in T)\}$.

Proof. See [46]. The proof in [46] is essentially true, but it is not complete. That is because the definition of "Bayes operator" (i.e., Definition 6.5) was not mentioned in [46]. Thus, we add the complete proof in what follows. It will be proved by induction. Let $\mathbf{O}'_\tau$ $= (Y_\tau, 2^{Y_\tau}, G_\tau)$ be any observable in $C(\Omega_\tau)$.

[Step 1] First, define the positive bounded linear operator $\widehat{B}^{(\tau,\tau)}_{\Pi_{t\in T_\tau}\Xi_t} : C(\Omega_\tau) \to C(\Omega_\tau)$ such that:

$$\widehat{B}^{(\tau,\tau)}_{\Pi_{t\in T_\tau}\Xi_t}(g_\tau) = \widetilde{F}_\tau(\Pi_{t\in T_\tau}\Xi_t)\times g_\tau \quad (\forall g_\tau \in C(\Omega_\tau)), \tag{6.21}$$

and define the observable $\widehat{\mathbf{O}}_\tau \equiv ((\Pi_{t\in T_\tau} X_t)\times Y_\tau, 2^{X_\tau \times Y_\tau}, \widehat{F}_\tau)$ in $C(\Omega_\tau)$ such that:

$$\widehat{F}_\tau(\Pi_{t\in T_\tau}\Xi_t \times \Gamma_\tau) = \widehat{B}^{(\tau,\tau)}_{\Pi_{t\in T_\tau}\Xi_t}(G_\tau(\Gamma_\tau)) \qquad (\forall \Gamma_\tau \in 2^{Y_\tau}), \tag{6.22}$$

which is clearly the Heisenberg picture representation of the *sequential observable* $[\{\overline{\mathbf{O}}_t\}_{t\in T_\tau}, \{C(\Omega_t) \overset{\Phi_{\pi(t),t}}{\to} C(\Omega_{\pi(t)})\}_{t\in T_\tau\setminus\{\tau\}}]$, where $\overline{\mathbf{O}}_t = \mathbf{O}_t$ (if $t \neq \tau$), $= \mathbf{O}_\tau \times \mathbf{O}'_\tau$ (if $t = \tau$). Thus, the operator $\widehat{B}^{(\tau,\tau)}_{\Pi_{t\in T_\tau}\Xi_t} : C(\Omega_\tau) \to C(\Omega_\tau)$ is the Bayes operator induced from the $\widetilde{\mathbf{O}}_\tau$ $\Big(= (\Pi_{t\in T_\tau} X_t, 2^{\Pi_{t\in T_\tau} X_t}, \widetilde{F}_\tau)\Big)$, which is uniquely determined.

[Step 2] Let s be any element in $T \setminus \{0\}$ such that $s \le \tau$. Here, assume that $\widehat{B}^{(s,\tau)}_{\Pi_{t\in T_s}\Xi_t} : C(\Omega_\tau) \to C(\Omega_s)$ is the Bayes operator induced from the $\widetilde{\mathbf{O}}_s$ $\Big(= (\Pi_{t\in T_s} X_t, 2^{\Pi_{t\in T'_s} X_t}, \widetilde{F}_s)\Big)$. That is, there exists an observable $\widehat{\mathbf{O}}_s \equiv ((\prod_{t\in T_s} X_t) \times Y_\tau, 2^{(\Pi_{t\in T_s} X_t)\times Y_\tau}, \widehat{F}_s)$ in $C(\Omega_s)$ such that

(i) $\widehat{\mathbf{O}}_s$ is the Heisenberg picture representation (*cf.* Theorem 3.7) of the *sequential observable* $[\{\widehat{\mathbf{O}}_t\}_{t\in T_s}, \{C(\Omega_t) \overset{\Phi_{\pi(t),t}}{\to} C(\Omega_{\pi(t)})\}_{t\in T_s\setminus\{s\}}]$, where $\overline{\mathbf{O}}_t = \mathbf{O}_t$ (if $t \ne \tau$), $= \mathbf{O}_\tau \times \mathbf{O}'_\tau$ (if $t = \tau$),

(ii) $\widehat{F}_s((\Pi_{t\in T_s}\Xi_t) \times \Gamma_\tau) = \widehat{B}^{(s,\tau)}_{\Pi_{t\in T_s}\Xi_t}(G_\tau(\Gamma_\tau)) \qquad (\Xi_t \in 2^{X_t}\ (\forall t \in T_s), \forall \Gamma_\tau \in 2^{Y_\tau})$,

(iii) $\widehat{F}_s((\Pi_{t\in T_s}\Xi_t) \times Y_\tau) = \widetilde{F}_s(\prod_{t\in T_s}\Xi_t) \qquad (\Xi_t \in 2^{X_t}\ (\forall t \in T_s))$.

Let $(x_t)_{t\in T_{\pi(s)}}$ be any element in $\Pi_{t\in T_{\pi(s)}} X_t$. Note that $\{(x_t)_{t\in T_{\pi(s)}}\} = \Pi_{t\in T_{\pi(s)}}\{x_t\}$. Define the positive bounded linear operator $\widehat{B}^{(\pi(s),\tau)}_{\Pi_{t\in T_{\pi(s)}}\{x_t\}} : C(\Omega_\tau) \to C(\Omega_{\pi(s)})$ by

$$[\widehat{B}^{(\pi(s),\tau)}_{\Pi_{t\in T_{\pi(s)}}\{x_t\}}(g_\tau)](\omega_{\pi(s)}) = \frac{[\widetilde{F}_{\pi(s)}(\prod_{t\in T_{\pi(s)}}\{x_t\})](\omega_{\pi(s)}) \times [\Phi_{\pi(s),s}\widehat{B}^{(s,\tau)}_{\Pi_{t\in T_s}\{x_t\}}(g_\tau)](\omega_{\pi(s)})}{[\widetilde{F}_{\pi(s)}((\Pi_{t\in T_{\pi(s)}\setminus T_s} X_t) \times \prod_{t\in T_s}\{x_t\})](\omega_{\pi(s)})}$$
$$(\forall g_\tau \in C(\Omega_\tau), \quad \forall \omega_{\pi(s)}\ (\in \Omega_{\pi(s)})). \tag{6.23}$$

Here, the above is assumed to be equal to 0 if the denominator of (6.23) is equal to 0 $\Big($i.e., $[\widetilde{F}_{\pi(s)}((\Pi_{t\in T_{\pi(s)}\setminus T_s} X_t) \times \prod_{t\in T_s}\{x_t\})](\omega_{\pi(s)}) = 0\Big)$. And thus, we can define the positive bounded linear operator $\widehat{B}^{(\pi(s),\tau)}_{\Pi_{t\in T_{\pi(s)}}\Xi_t} : C(\Omega_\tau) \to C(\Omega_{\pi(s)})$ by

$$\widehat{B}^{(\pi(s),\tau)}_{\Pi_{t\in T_{\pi(s)}}\Xi_t} = \sum_{(x_t)_{t\in T_{\pi(s)}} \in \Pi_{t\in T_{\pi(s)}}\Xi_t} \widehat{B}^{(\pi(s),\tau)}_{\{(x_t)_{t\in T_{\pi(s)}}\}}.$$

Define the observable $\widehat{\mathbf{O}}_{\pi(s)} \equiv ((\prod_{t\in T_{\pi(s)}} X_t) \times Y_\tau, 2^{(\Pi_{t\in T_{\pi(s)}} X_t)\times Y_\tau}, \widehat{F}_{\pi(s)})$ in $C(\Omega_{\pi(s)})$ such that:

$$\widehat{F}_{\pi(s)}((\Pi_{t\in T_{\pi(s)}}\Xi_t) \times \Gamma_\tau) = \widehat{B}^{(\pi(s),\tau)}_{\Pi_{t\in T_{\pi(s)}}\Xi_t}(G_\tau(\Gamma_\tau)) \quad (\Xi_t \in 2^{X_t}\ (\forall t \in T_{\pi(s)}), \forall \Gamma_\tau \in 2^{Y_\tau}),$$

which is clearly the Heisenberg picture representation of $[\{\overline{\mathbf{O}}_t\}_{t\in T_{\pi(s)}}, \{C(\Omega_t) \overset{\Phi_{\pi(t),t}}{\to} C(\Omega_{\pi(t)})\}_{t\in T_{\pi(s)}\setminus\{\pi(s)\}}]$, where $\overline{\mathbf{O}}_t = \mathbf{O}_t$ (if $t \ne \tau$), $= \mathbf{O}_\tau \times \mathbf{O}'_\tau$ (if $t = \tau$). Also, it holds that

$$\widehat{F}_{\pi(s)}((\Pi_{t\in T_{\pi(s)}}\Xi_t) \times Y_\tau) = \widetilde{F}_{\pi(s)}(\prod_{t\in T_{\pi(s)}}\Xi_t) \qquad (\Xi_t \in 2^{X_t}\ (\forall t \in T_{\pi(s)})).$$

That is because we see

$$
\begin{aligned}
&\widehat{F}_{\pi(s)}((\Pi_{t\in T_{\pi(s)}}\Xi_t)\times Y_\tau)=\widehat{B}^{(\pi(s),\tau)}_{\Pi_{t\in T_{\pi(s)}}\Xi_t}(1_\tau)=\sum_{(x_t)_{t\in T_{\pi(s)}}\in \Pi_{t\in T_{\pi(s)}}\Xi_t}\widehat{B}^{(\pi(s),\tau)}_{\Pi_{t\in T_{\pi(s)}}\{x_t\}}(1_\tau)\\
&=\sum_{(x_t)_{t\in T_{\pi(s)}}\in \Pi_{t\in T_{\pi(s)}}\Xi_t}\frac{\widetilde{F}_{\pi(s)}(\prod_{t\in T_{\pi(s)}}\{x_t\})\times \Phi_{\pi(s),s}\widehat{B}^{(s,\tau)}_{\Pi_{t\in T_s}\{x_t\}}(1_\tau)}{\widetilde{F}_{\pi(s)}((\Pi_{t\in T_{\pi(s)}\setminus T_s}X_t)\times\prod_{t\in T_s}\{x_t\})}\\
&=\sum_{(x_t)_{t\in T_{\pi(s)}}\in \Pi_{t\in T_{\pi(s)}}\Xi_t}\frac{\widetilde{F}_{\pi(s)}(\prod_{t\in T_{\pi(s)}}\{x_t\})\times \widetilde{F}_{\pi(s)}((\Pi_{t\in T_{\pi(s)}\setminus T_s}X_t)\times\prod_{t\in T_s}\{x_t\})}{\widetilde{F}_{\pi(s)}((\Pi_{t\in T_{\pi(s)}\setminus T_s}X_t)\times\prod_{t\in T_s}\{x_t\})}\\
&=\sum_{(x_t)_{t\in T_{\pi(s)}}\in \Pi_{t\in T_{\pi(s)}}\Xi_t}\widetilde{F}_{\pi(s)}(\prod_{t\in T_{\pi(s)}}\{x_t\})=\widetilde{F}_{\pi(s)}(\prod_{t\in T_{\pi(s)}}\Xi_t). \qquad (6.24)
\end{aligned}
$$

Therefore, we see that $\widehat{B}^{(\pi(s),\tau)}_{\Pi_{t\in T_{\pi(s)}}\Xi_t}:C(\Omega_\tau)\to C(\Omega_{\pi(s)})$ is the Bayes operator induced from the $\widetilde{\mathbf{O}}_{\pi(s)}\Big(=(\Pi_{t\in T_{\pi(s)}}X_t, 2^{\Pi_{t\in T_{\pi(s)}}X_t}, \widetilde{F}_{\pi(s)})\Big)$. Thus, we can, by induction, finish the proof since it suffices to put $B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}=\widehat{B}^{(0,\tau)}_{\Pi_{t\in T_0}\Xi_t}$. □

Let $\widetilde{\mathbf{O}}_0\equiv(\prod_{t\in T}X_t, 2^{\Pi_{t\in T}X_t}, \widetilde{F}_0)$, $\mathbf{O}'_\tau\equiv(Y_\tau, 2^{Y_\tau}, G_\tau)$, $\{B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}\mid \Xi_t\in 2^{X_t}\ (\forall t\in T)\}$, $\widehat{\mathbf{O}}_0\equiv((\prod_{t\in T}X_t)\times Y_\tau, 2^{(\Pi_{t\in T}X_t)\times Y_\tau}, \widehat{F}_0)$ and $\{R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}\mid \Xi_t\in 2^{X_t}\ (\forall t\in T)\}$ be as in Definition 6.5. Assume that

(C_1) we know that the measured value $(x_t)_{t\in T}$ $(\in(\prod_{t\in T}X_t))$ obtained by $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[\delta_{\omega_0}]})$ belongs to $\prod_{t\in T}\Xi_t$.

Note that this (C_1) is the same as the following (C_2).

(C_2) we know that the measured value $((x_t)_{t\in T}, y)$ $(\in(\prod_{t\in T}X_t)\times Y_\tau)$ obtained by $\mathbf{M}_{C(\Omega_0)}$ $(\widehat{\mathbf{O}}_0, S_{[\delta_{\omega_0}]})$ belongs to $(\prod_{t\in T}\Xi_t)\times Y_\tau$.

Thus we see that

(C_3) the probability distribution of unknown y $\Big($under the assumption (C_2) (=(C_1))$\Big)$, i.e., the probability that y $(\in Y_\tau)$ belongs to Γ_τ, is represented by

$$
\frac{{}_{C(\Omega_0)^*}\langle\delta_{\omega_0}, \widehat{F}_0((\prod_{t\in T}\Xi_t)\times\Gamma_\tau)\rangle_{C(\Omega_0)}}{{}_{C(\Omega_0)^*}\langle\delta_{\omega_0}, \widehat{F}_0((\prod_{t\in T}\Xi_t)\times Y_\tau)\rangle_{C(\Omega_0)}}\ \Big(\equiv\frac{{}_{C(\Omega_0)^*}\langle\delta_{\omega_0}, B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(G_\tau(\Gamma_\tau))\rangle_{C(\Omega_0)}}{{}_{C(\Omega_0)^*}\langle\delta_{\omega_0}, B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(1_\tau)\rangle_{C(\Omega_0)}}\Big). \qquad (6.25)
$$

A simple calculation shows:

$$
(6.25)={}_{C(\Omega_\tau)^*}\langle\frac{(B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t})^*(\delta_{\omega_0})}{\|(B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t})^*(\delta_{\omega_0})\|_{\mathcal{M}(\Omega)}}, G_\tau(\Gamma_\tau)\rangle_{C(\Omega_\tau)}={}_{C(\Omega_\tau)^*}\langle R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\delta_{\omega_0}), G_\tau(\Gamma_\tau)\rangle_{C(\Omega_\tau)}.
$$

Therefore, we say that

(C_4) the probability distribution of unknown y $\Big($under (C_2) (=(C_1))$\Big)$ is represented by

$$ {}_{C(\Omega_\tau)^*}\langle R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\delta_{\omega_0}), G_\tau(\Gamma_\tau)\rangle_{C(\Omega_\tau)}. \tag{6.26}$$

Let this (C_4) be, as an abbreviation, denoted (or, called) by

(C_5) the *S-state* $\Big($after the measurement $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[\delta_{\omega_0}]})\Big)$ at τ (in T) is equal to $R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\delta_{\omega_0})$.

For completeness, again note that (C_4) = (C_5), i.e., (C_5) is an abbreviation for (C_4). Note that the concept of "S-state" and that of "state" are completely different. In measurement theory, as seen in Axiom 1, the state always appears as the ρ^p in $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[\rho^p]})$. That is, the state ρ^p is always fixed and never moves. In this sense, the ρ^p may be called a "real state". On the other hand, the "S-state" is used in the abbreviation (C_5) of (C_4).

Summing up the above argument, we have the following definition.

Definition 6.7. [S-state (= Schrödinger picture)]. *Assume the above situation. If the above statement (C_4) holds, then we say "(C_5) holds", i.e., "the S-state (after the measurement* $\mathbf{M}_{C(\Omega_0)}$ $(\widehat{\mathbf{O}}_0, S_{[\delta_{\omega_0}]})$*) at τ ($\in T$) is equal to* $R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\delta_{\omega_0})$*". The representation using "S-state" is called the Schrödinger picture representation. The S-state is also called a Schrödinger state or imaginary state.*

■

As seen in the above argument, we must note that the Bayes operator is always hidden behind the Scrödinger picture representation.

We sum up the above argument $\big($i.e., (C_1)⇒(C_5)$\big)$ as the following lemma.

Lemma 6.8. [S-state]. *Let* $\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, 2^{\prod_{t\in T} X_t}, \widetilde{F}_0)$, $\{B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t} \mid \Xi_t \in 2^{X_t}\ (\forall t\in T)\}$ *and* $\{R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t} \mid \Xi_t \in 2^{X_t}\ (\forall t \in T)\}$ *be as in Definition 6.5. Assume that*

- *we know that the measured value* $(x_t)_{t\in T}$ $(\in \prod_{t\in T} X_t)$ *obtained by* $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[\delta_{\omega_0}]})$ *belongs to* $\prod_{t\in T}\Xi_t$.

Then, we can say

(♯) *the S-state* $\Big($*after the measurement* $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[\delta_{\omega_0}]})\Big)$ *at τ (in T) is equal to* $R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\delta_{\omega_0})$.

■

The following lemma will be used as Theorem 6.13.

Lemma 6.9. [Inference and S-state]. *Let* $\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, 2^{\prod_{t\in T} X_t}, \widetilde{F}_0)$, $\{B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t} \mid \Xi_t \in 2^{X_t} \ (\forall t \in T)\}$ *and* $\{R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t} \mid \Xi_t \in 2^{X_t} \ (\forall t \in T)\}$ *be as in Definition 6.5. Assume that*

($\bullet$) *we know that the measured value* $(x_t)_{t\in T}$ $(\in \prod_{t\in T} X_t)$ *obtained by* $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[*]})$ *belongs to* $\prod_{t\in T}\Xi_t$.

Then, there is a reason to infer that

($\sharp$) *the S-state* $\Big($*after the measurement* $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[*]})\Big)$ *at* τ *(in* T*) is equal to* $R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\delta_{\omega_0})$.

Here the δ_{ω_0} $(\in \mathcal{M}^p_{+1}(\Omega_0))$ *is defined by*

$$[\widetilde{F}_0(\prod_{t\in T}\Xi_t)](\omega_0) = \max_{\omega\in\Omega_0}[\widetilde{F}_0(\prod_{t\in T}\Xi_t)](\omega). \tag{6.27}$$

Proof. The proof is similar to that of Corollary 5.6. Let $(Y_\tau, 2^{Y_\tau}, G_\tau)$ be any observable in $C(\Omega_\tau)$. Note that the above ($\bullet$) is the same as the following:

($\bullet$)' we know the measured value $((x_t)_{t\in T}, y)$ $(\in (\prod_{t\in T} X_t)\times Y_\tau)$ obtained by $\mathbf{M}_{C(\Omega_0)}(\widehat{\mathbf{O}}, S_{[*]})$ belongs to $(\prod_{t\in T}\Xi_t)\times Y_\tau$ (where $\widehat{\mathbf{O}}_0$ is as in Definition 6.5).

Thus we can infer, by Theorem 5.3 (Fisher's method) and the equality $\widetilde{F}_0(\prod_{t\in T}\Xi_t) = \widehat{F}_0((\prod_{t\in T}\Xi_t)\times Y_\tau)$, that the unknown state $[*]$ $\Big($in $\mathbf{M}_{C(\Omega_0)}(\widehat{\mathbf{O}}, S_{[*]})\Big)$ is equal to δ_{ω_0} $\Big($defined by (6.27)$\Big)$. Thus the conditional probability $P_{\Pi_{t\in T}\Xi_t}(\cdot)$ under the condition that we know that $((x_t)_{t\in T}, y) \in (\prod_{t\in T} X_t)\times Y_\tau$ is given by

$$\begin{aligned} P_{\Pi_{t\in T}\Xi_t}(\Gamma_\tau) &= \frac{{}_{C(\Omega_0)^*}\langle \delta_{\omega_0}, \widehat{F}_0((\prod_{t\in T}\Xi_t)\times\Gamma_\tau)\rangle_{C(\Omega_0)}}{{}_{C(\Omega_0)^*}\langle \delta_{\omega_0}, \widehat{F}_0((\prod_{t\in T}\Xi_t)\times Y_\tau)\rangle_{C(\Omega_0)}} = \frac{{}_{C(\Omega_0)^*}\langle \delta_{\omega_0}, B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(G_\tau(\Gamma_\tau))\rangle_{C(\Omega_0)}}{{}_{C(\Omega_0)^*}\langle \delta_{\omega_0}, B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(1_\tau)\rangle_{C(\Omega_0)}} \\ &= {}_{C(\Omega_\tau)^*}\langle R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\delta_{\omega_0}), G_\tau(\Gamma_\tau)\rangle_{C(\Omega_\tau)} \qquad (\forall \Gamma_\tau \in 2^{Y_\tau}). \end{aligned}$$

From the equivalence of (C_4) and (C_5), we can conclude the ($\sharp$). □

Now we consider the simplest case that $T \equiv \{0,\tau\}$ and $\mathbf{S}_{[\delta_{\omega_0}]} \equiv [S_{[\delta_{\omega_0}]}; C(\Omega_\tau) \overset{\Phi_{0,\tau}}{\rightarrow} C(\Omega_0)]$. For each $k = 0, \tau$, consider the null observable $\mathbf{O}^{(\mathrm{nl})}_k \equiv (\{0,1\}, 2^{\{0,1\}}, F^{(\mathrm{nl})}_k)$ in $C(\Omega_k)$ (*cf.* Example 2.21). Then, we have the measurement

$$\mathbf{M}_{C(\Omega_0)}\Big(\widetilde{\mathbf{O}}_0 \equiv (\{0,1\}^2, 2^{\{0,1\}^2}, F^{(\mathrm{nl})}_0 \times \Phi_{0,\tau}F^{(\mathrm{nl})}_\tau), S_{[\delta_{\omega_0}]}\Big). \tag{6.28}$$

Note that:

(i) the probability that the measured value (by $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[\delta_{\omega_0}]})$) is equal to $(1,1)$ is given by 1. That is, the measured value is always (or surely) equal to $(1,1)$.

Thus,

(ii) the measured value obtained by $\mathbf{M}_{C(\Omega_0)}(\widehat{\mathbf{O}}_0, S_{[\delta_{\omega_0}]})$ has always the form $((1,1), y)$ $(\in \{0,1\}^2 \times Y_\tau)$. Here $\widehat{\mathbf{O}}_0$ is defined by

$$(\{0,1\}^2 \times Y_\tau, 2^{\{0,1\}^2 \times Y_\tau}, F_0^{(\mathrm{nl})} \times \Phi_{0,\tau} F_\tau^{(\mathrm{nl})} \times \Phi_{0,\tau} G_\tau) \tag{6.29}$$

for any any observable $(Y_\tau, 2^{Y_\tau}, G_\tau)$ in $C(\Omega_\tau)$.

Note that $\mathbf{M}_{C(\Omega_0)}(\widehat{\mathbf{O}}_0, S_{[\delta_{\omega_0}]})$ and $\mathbf{M}_{C(\Omega_0)}((Y_\tau, 2^{Y_\tau}, \Phi_{0,\tau} G_\tau), S_{[\delta_{\omega_0}]})$ are essentially the same. That is because "to take $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[\delta_{\omega 0}]})$" is essentially the same as "to take no measurement" (*cf.* Example 2.21). Thus, the above (ii) implies that

(iii) the probability distribution of unknown y $\Big($under (ii) $\big(=$ (i)$\big)\Big)$, i.e., the probability that $y \in \Gamma_\tau$, is represented by

$$_{C(\Omega_\tau)^*}\langle \Phi_{0,\tau}^*(\delta_{\omega_0}), G_\tau(\Gamma_\tau) \rangle_{C(\Omega_\tau)}$$

for any $(Y_\tau, 2^{Y_\tau}, G_\tau)$ in $C(\Omega_\tau)$ and any Γ_τ $(\in 2^{Y_\tau})$.

That is because it holds that

$$\frac{_{C(\Omega_0)^*}\langle \delta_{\omega_0}, (F_0^{(\mathrm{nl})} \times \Phi_{0,\tau} F_\tau^{(\mathrm{nl})} \times \Phi_{0,\tau} G_\tau)(\{(1,1)\} \times \Gamma_\tau) \rangle_{C(\Omega_0)}}{_{C(\Omega_0)^*}\langle \delta_{\omega_0}, (F_0^{(\mathrm{nl})} \times \Phi_{0,\tau} F_\tau^{(\mathrm{nl})} \times \Phi_{0,\tau} G_\tau)(\{(1,1)\} \times Y_\tau) \rangle_{C(\Omega_0)}}$$
$$= {}_{C(\Omega_\tau)^*}\langle \Phi_{0,\tau}^*(\delta_{\omega_0}), G_\tau(\Gamma_\tau) \rangle_{C(\Omega_\tau)}.$$

Thus,we get the following (iv), which is short for (iii).

(iv) the S-state at τ $(\in T \equiv \{0, \tau\})$ is equal to $\Phi_{0,\tau}^*(\delta_{\omega_0})$.

Thus we conclude that (i) $\Rightarrow$ (iv). However, note that (i) always holds. Therefore, we think that (iv) always holds.

From the above argument, we have the following lemma. This will be used in the statement (6.33).

Lemma 6.10. [The Schrödinger picture representation]. *Put* $T = \{0, \tau\}$. *Let* $\mathbf{S}_{[\delta_{\omega_0}]} \equiv [S_{[\delta_{\omega_0}]}; \{C(\Omega_\tau) \overset{\Phi_{0,\tau}}{\to} C(\Omega_0)\}]$ *be a general system with an initial state* $S_{[\delta_{\omega_0}]}$. *Then we see that*

($\sharp$) *the S-state at* τ ($\in T \equiv \{0,\tau\}$) *is* $\Phi^*_{0,\tau}(\delta_{\omega_0})$.

Here it should be noted that the measurement $\mathbf{M}_{C(\Omega_0)}((Y_\tau, 2^{Y_\tau}, \Phi_{0,\tau}G_\tau), S_{[\delta_{\omega_0}]})$ $\Big($*or,* $\mathbf{M}_{C(\Omega_0)}(\widehat{\mathbf{O}}_0, S_{[\delta_{\omega_0}]})$, $\Big)$ *is hidden behind the assertion* ($\sharp$).

■

Also, the following lemma is the formal representation of Corollary 5.6 (ii). $\Big($Cf. Remark 6.12.$\Big)$

Lemma 6.11. [Inference and the Schrödinger picture representation]. *Put* $T = \{0,\tau\}$. *Let* $\mathbf{S}_{[*]} \equiv [S_{[*]}; \{\Phi_{0,\tau} : C(\Omega_\tau) \to C(\Omega_0)\}]$ *be a general system with an initial state* $S_{[*]}$. *Let* $\mathbf{O}_0 = (X_0, 2^{X_0}, F_0)$ *be an observable in* $C(\Omega_0)$. *And, let* $\mathbf{O}^{(\mathrm{nl})}_\tau = (\{0,1\}, 2^{\{0,1\}}, F^{(\mathrm{nl})}_\tau)$ *be the null observable in* $C(\Omega_\tau)$ *(cf. Example 2.21). Consider a measurement* $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0 (\equiv \mathbf{O}_0 \times \Phi_{0,\tau}\mathbf{O}^{(\mathrm{nl})}_\tau), S_{[*]})$, *which is essentially the same as* $\mathbf{M}_{C(\Omega_0)}(\mathbf{O}_0, S_{[*]})$. *Assume that*

- *we know that the measured value obtained by* $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0 \equiv \mathbf{O}_0 \times \Phi_{0,\tau}\mathbf{O}^{(\mathrm{nl})}_\tau, S_{[*]})$ *belongs to* $\Xi_0 \times \{1\}$ ($\in 2^{X_0 \times \{0,1\}}$).

Then we see that

($\sharp$) *there is a reason to infer that the S-state (after the measurement* $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[*]})$) *at* τ ($\in T \equiv \{0,\tau\}$) *is* $\Phi^*_{0,\tau}(\delta_{\omega_0})$,

where δ_{ω_0} ($\in \mathcal{M}^p_{+1}(\Omega_0)$) *is defined by*

$$[F_0(\Xi_0)](\omega_0) = \max_{\omega \in \Omega_0} [F_0(\Xi_0)](\omega). \tag{6.30}$$

Proof. Let $B^{(0,\tau)}_{\Xi_0 \times \{1\}} : C(\Omega_\tau) \to C(\Omega_0)$ and $R^{(0,\tau)}_{\Xi_0} : \mathcal{M}^m_{+1}(\Omega_0) \to \mathcal{M}^m_{+1}(\Omega_\tau)$ be as in Definition 6.5. Here, note that, from the property of null observable, it holds that $F_0(\Xi_0) \times \Phi_{0,\tau}F^{(\mathrm{nl})}_\tau(\{1\}) = F_0(\Xi_0)$. Thus we see that $B^{(0,\tau)}_{\Xi_0 \times \{1\}}(g_\tau) = F_0(\Xi_0) \times \Phi_{0,\tau}g_\tau$ for any g_τ ($\in C(\Omega_\tau)$). By Lemma 6.9, it suffices to prove $R^{(0,\tau)}_{\Xi_0}(\delta_{\omega_0}) = \Phi^*_{0,\tau}(\delta_{\omega_0})$. This is shown as follows:

$$
\begin{aligned}
&{}_{C(\Omega_\tau)^*}\langle R^{(0,\tau)}_{\Xi_0 \times \{1\}}(\delta_{\omega_0}), g_\tau \rangle_{C(\Omega_\tau)} = {}_{C(\Omega_\tau)^*}\langle \frac{(B^{(0,\tau)}_{\Xi_0 \times \{1\}})^*(\delta_{\omega_0})}{\|(B^{(0,\tau)}_{\Xi_0 \times \{1\}})^*(\delta_{\omega_0})\|_{\mathcal{M}(\Omega_\tau)}}, g_\tau \rangle_{C(\Omega_\tau)} \\
&= \frac{1}{\|(B^{(0,\tau)}_{\Xi_0 \times \{1\}})^*(\delta_{\omega_0})\|_{\mathcal{M}(\Omega_\tau)}} {}_{C(\Omega_0)^*}\langle \delta_{\omega_0}, B^{(0,\tau)}_{\Xi_0 \times \{1\}}(g_\tau) \rangle_{C(\Omega_0)} = \frac{[F_0(\Xi_0)](\omega_0) \times [\Phi_{0,\tau}g_\tau](\omega_0)}{[F_0(\Xi_0)](\omega_0)} \\
&= {}_{C(\Omega_\tau)^*}\langle \Phi^*_{0,\tau}(\delta_{\omega_0}), g_\tau \rangle_{C(\Omega_\tau)} \qquad (\forall g_\tau \in C(\Omega_\tau)).
\end{aligned} \tag{6.31}
$$

Then, we see that $R^{(0,\tau)}_{\Xi_0 \times \{1\}}(\delta_{\omega_0}) = \Phi^*_{0,\tau}(\delta_{\omega_0})$. This completes the proof. □

The following remark shows that Corollary 5.6 (ii) is a direct consequence of Lemma 6.11.

Remark 6.12. [Continued from Corollary 5.6 (Fisher's maximum likelihood method in classical measurements)]. As mentioned before, the proof of Corollary 5.6 is temporary. Corollary 5.6 should be understood as a corollary of Lemma 6.11 as follows: In Lemma 6.11, put $\Omega_0 = \Omega_\tau = \Omega_{+0}$. And let $\Phi_{0,\tau} : C(\Omega_{+0}) \to C(\Omega_0)$ be the identity map. Since "the S-state (after the measurement $\mathbf{M}_{C(\Omega_0)}(\mathbf{O}_0, S_{[*]})$) at $\tau(=+0)$" $= \Phi_{0,\tau}(\delta_{\omega_0}) = \delta_{\omega_0}$, we easily see that Corollary 5.6 is a consequence of Lemma 6.11. This should be regarded as the formal proof of Corollary 5.6.

■

6.3 Regression analysis II in measurements

Now let us explain the reason why we consider:

(♯) it is worthwhile doubting the derivation of (6.6) $\big(= (6.17)\big)$ from (6.5) $\big(= (6.16)\big)$, i.e., the formula $h(2) = 0.4 + 1.4 \times 2 = 3.2$.

Using the notations in Regression Analysis I (6.7), we recall the statement (R) of Example 6.4 as follows:

(R) the measurement $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t,\ 2^{\prod_{t\in T} X_t},\ \widetilde{F}_0), S_{[*]})$ is hidden behind the inference (6.5) $\big(=(6.16)\big)$.

And we conclude, by Corollary 5.6 (or Remark 6.12), that

$$\begin{aligned} [\,*\,] = &\ \text{"the S-state after the measurement } \mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[*]})\text{"} \\ = &\ \delta_{\omega_0}. \end{aligned} \tag{6.32}$$

Here the δ_{ω_0} $(\in \mathcal{M}^p_{+1}(\Omega_0))$ is defined by $[\widetilde{F}_0(\prod_{t\in T} \Xi_t)](\omega_0) = \max_{\omega\in\Omega_0}[\widetilde{F}_0(\prod_{t\in T} \Xi_t)](\omega)$. On the other hand,

- the map "$\delta_{\omega_0} \mapsto \Phi^*_{0,\tau}(\delta_{\omega_0})$" $\Big($i.e., the derivation of (6.6) $\big(= (6.17)\big)$ from (6.5) $\big(= (6.16)\big)\Big)$ is due to the Schrödinger picture, behind which the measurement $\mathbf{M}_{C(\Omega_0)}(\Phi_{0,\tau}\mathbf{O}'_\tau \equiv (Y_\tau, 2^{Y_\tau}, \Phi_{0,\tau}G_\tau), S_{[\delta_{\omega_0}]})$ is hidden. Cf. Lemma 6.10. (6.33)

Thus, in order to conclude the assertion (6.6) $\big(= (6.17)\big)$, we need the above "two measurements", that is,

$$\text{“}\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0 \equiv (\textstyle\prod_{t\in T} X_t, 2^{\prod_{t\in T} X_t}, \widetilde{F}_0), S_{[*]})\text{” and “}\mathbf{M}_{C(\Omega_0)}(\Phi_{0,\tau}\mathbf{O}'_\tau \equiv (Y_\tau, 2^{Y_\tau}, \Phi_{0,\tau}G_\tau), S_{[\delta_{\omega_0}]})\text{”}. \tag{6.34}$$

However, note that it is forbidden to conduct "two measurements" (*cf.* §2.5(II)). This is the reason that we think that it is worthwhile doubting (6.6) $\big(= (6.17)\big)$. In order to avoid this confusion, it suffices to consider the "simultaneous" measurement:

$$\mathbf{M}_{C(\Omega_0)}(\widehat{\mathbf{O}}_0 \equiv ((\prod_{t\in T} X_t) \times Y_\tau, 2^{(\prod_{t\in T} X_t)\times Y_\tau}, \widehat{F}_0), S_{[*]}), \quad (\text{where } \widehat{\mathbf{O}}_0 \text{ is as in Definition 6.5}), \tag{6.35}$$

instead of (6.34).

Then, we rewrite Lemma 6.9 as an main theorem as follows:

Theorem 6.13. [= Lemma 6.9, Inference in Markov relation]. *Let* $\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, 2^{\prod_{t\in T} X_t}, \widetilde{F}_0)$ *be as in Theorem 3.7 in the case* $\mathcal{A}_t = C(\Omega_t)$ $(\forall t \in T)$*. And consider a measurement* $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[*]})$*. Let* τ *be any element in* T*. Let* $\{R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t} \mid \Xi_t \in 2^{X_t}\ (\forall t \in T)\}$ *be as in Definition 6.5. Assume that we know that the measured value (obtained by* $\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[*]})$*) belongs to* $\prod_{t\in T}\Xi_t$*. Then, there is a reason to infer that*

$$(\sharp) \qquad \text{“the } S\text{-state at } \tau\ (\in T) \text{ after } \mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[*]})\text{”} = R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\delta_{\omega_0}). \tag{6.36}$$

Here δ_{ω_0} $(\in \mathcal{M}^p_{+1}(\Omega))$ *is defined by*

$$[\widetilde{F}_0(\prod_{t\in T}\Xi_t)](\omega_0) = \max_{\omega\in\Omega_0}[\widetilde{F}_0(\prod_{t\in T}\Xi_t)](\omega). \tag{6.37}$$

■

Lastly, we prove the following lemma, which justifies the inference (6.6).

Lemma 6.14. [Some property of homomorphic relation]. *Let* $\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, 2^{\prod_{t\in T} X_t}, \widetilde{F}_0)$ *be as in Theorem 3.7 in the case* $\mathcal{A}_t = C(\Omega_t)$ $(\forall t \in T)$*. Consider the family of Bayes operators* $\{B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t} \mid \Xi_t \in 2^{X_t}\ (t \in T)\}$ *such as in Definition 6.5. Let* τ *be any element in* T*. Assume that* $\Phi_{\pi(t),t} : C(\Omega_t) \to C(\Omega_{\pi(t)})$ $(\forall t \in T$ *such that* $0 < t \le \tau)$ *is homomorphic. Then, it holds that:*

$$B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(G_\tau(\Gamma_\tau)) = \widetilde{F}_0(\prod_{t\in T}\Xi_t) \times \Phi_{0,\tau}G_\tau(\Gamma_\tau) \quad (\forall \Xi_t \in 2^{X_t}\ (\forall t \in T), \forall \Gamma_\tau \in 2^{Y_\tau}), \tag{6.38}$$

for any observable $(Y_\tau, 2^{Y_\tau}, G_\tau)$ *in* $C(\Omega_\tau)$*. That is, we see that the Bayes operator* $B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}: C(\Omega_\tau) \to C(\Omega_0)$ *is determined uniquely under the homomorphic condition.*

Proof. The proof is shown in the following three steps.

[Step 1]. Let ω_0 be any element in Ω_0. And let g_τ and h_τ be in $C(\Omega_\tau)$ such that:

$$0 \le g_\tau \le 1,\ g_\tau(\phi_{0,\tau}(\omega_0)) = 0,\ 0 \le h_\tau \le 1, \text{ and } h_\tau(\phi_{0,\tau}(\omega_0)) = 1. \tag{6.39}$$

where $\phi_{0,\tau} : \Omega_0 \to \Omega_\tau$ is defined by (3.14). Then we see, by (6.19), that

$$0 \le [B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(g_\tau)](\omega) \le (\Phi_{0,\tau} g_\tau)(\omega) = g_\tau(\phi_{0,\tau}(\omega)) \quad (\forall \omega \in \Omega_0). \tag{6.40}$$

Putting $\omega = \omega_0$ in (6.40), we get, by (6.39), that

$$[B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(g_\tau)](\omega_0) = 0. \tag{6.41}$$

Also, from the linearity of Bayes operator and the condition (iii) of Definition 6.5, we get

$$\begin{aligned}[B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(1_\tau - h_\tau)](\omega) &= [B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(1_\tau)](\omega) - [B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(h_\tau)](\omega) \\ &= [\widetilde{F}_0(\prod_{t\in T}\Xi_t)](\omega) - [B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(h_\tau)](\omega) \quad (\forall \omega \in \Omega_0).\end{aligned} \tag{6.42}$$

Thus, putting $\omega = \omega_0$ in (6.42), we get, by (6.39), that

$$\begin{aligned}0 &\le [B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(1_\tau - h_\tau)](\omega_0) \\ &\le [(\Phi_{0,\tau}(1_\tau - h_\tau))](\omega_0) = 1_\tau(\phi_{0,\tau}(\omega_0)) - h_\tau(\phi_{0,\tau}(\omega_0)) = 1 - 1 = 0.\end{aligned} \tag{6.43}$$

Then, we obtain

$$[B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(h_\tau)](\omega_0) = [\widetilde{F}_0(\prod_{t\in T}\Xi_t)](\omega_0). \tag{6.44}$$

[Step 2]. Let ω_0 be any fixed element in Ω_0. Fix any f $(\in C(\Omega_\tau))$ such that $0 \le f \le 1$. Define g_τ, h_τ $(\in C(\Omega_\tau))$ such that:

$$\begin{aligned} g_\tau(\omega_\tau) &= \max\{0, f(\omega_\tau) - f(\phi_{0,\tau}(\omega_0))\} \qquad (\forall \omega_\tau \in \Omega_\tau), \\ h_\tau(\omega_\tau) &= \min\{\frac{f(\omega_\tau)}{f(\phi_{0,\tau}(\omega_0))}, 1\} \qquad (\forall \omega_\tau \in \Omega_\tau).\end{aligned} \tag{6.45}$$

The g_τ and the h_τ clearly satisfy (6.39). And moreover, we see, for any $\omega_\tau \in \Omega_\tau$, that

$$\begin{aligned} & g_\tau(\omega_\tau) + f(\phi_{0,\tau}(\omega_0)) h_\tau(\omega_\tau) \\ &= \max\{0, f(\omega_\tau) - f(\phi_{0,\tau}(\omega_0))\} + \min\{f(\omega_\tau), f(\phi_{0,\tau}(\omega_0))\} \\ &= \begin{cases} (f(\omega_\tau) - f(\phi_{0,\tau}(\omega_0)) + f(\phi_{0,\tau}(\omega_0)), & \text{if } f(\omega_\tau) \ge f(\phi_{0,\tau}(\omega_0)) \\ 0 + f(\omega_\tau), & \text{if } f(\omega_\tau) \le f(\phi_{0,\tau}(\omega_0)) \end{cases} \\ &= f(\omega_\tau). \end{aligned} \tag{6.46}$$

[Step 3]. Let ω_0 be any element in Ω_0. Let Γ_τ be any element in 2^{Y_τ}. From the [step 2], we see that there exist $\widehat{g}_\tau$ $(\in C(\Omega_\tau))$ and $\widehat{h}_\tau$ $(\in C(\Omega_\tau))$ such that $G_\tau(\Gamma_\tau) = \widehat{g}_\tau + [G_\tau(\Gamma_\tau)](\phi_{0,\tau}(\omega_0))\widehat{h}_\tau$, $\widehat{g}_\tau(\phi_{0,\tau}(\omega_0)) = 0$, $\widehat{h}_\tau(\phi_{0,\tau}(\omega_0)) = 1$. Then we see

$$
\begin{aligned}
&[B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(G_\tau(\Gamma_\tau))](\omega) = \Big[B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}\Big(\widehat{g}_\tau + [G_\tau(\Gamma_\tau)](\phi_{0,\tau}(\omega_0))\widehat{h}_\tau\Big)\Big](\omega) \\
=&[B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\widehat{g}_\tau)](\omega) + [G_\tau(\Gamma_\tau)](\phi_{0,\tau}(\omega_0)) \times [B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\widehat{h}_\tau)](\omega) \quad (\forall \omega \in \Omega_0).
\end{aligned} \tag{6.47}
$$

Putting $\omega = \omega_0$, we see, by (6.41) and (6.44), that $[B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\widehat{g}_\tau)](\omega_0) = 0$ and $[B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\widehat{h}_\tau)](\omega_0)$ $= [\widetilde{F}_0(\prod_{t\in T} \Xi_t)](\omega_0)$. And, we see, by (6.47), that

$$
\begin{aligned}
[B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(G_\tau(\Gamma_\tau))](\omega_0) &= [G_\tau(\Gamma_\tau)](\phi_{0,\tau}(\omega_0)) \times [\widetilde{F}_0(\prod_{t\in T} \Xi_t)](\omega_0) \\
&= [\Phi_{0,\tau}G_\tau(\Gamma_\tau)](\omega_0) \times [\widetilde{F}_0(\prod_{t\in T} \Xi_t)](\omega_0).
\end{aligned}
$$

Since ω_0 $(\in \Omega_0)$ is arbitrary, we obtain (6.38). This completes the proof. □

Now we can propose our main assertion as follows:

REGRESSION ANALYSIS II [The new proposal of regression analysis, *cf.*[55]].

(6.48)

Let $(T \equiv \{0, 1, ..., N\}, \pi : T\backslash\{0\} \to T)$ *be a tree with root* 0, *and let* $\mathbf{S}_{[*]} \equiv [S_{[*]}; \{C(\Omega_t) \overset{\Phi_{\pi(t),t}}{\to} C(\Omega_{\pi(t)})\}_{t\in T\backslash\{0\}}]$ *be a general system with the initial system* $S_{[*]}$. *And, let an observable* $\mathbf{O}_t \equiv (X_t, 2^{X_t}, F_t)$ *in a* C^**-algebra* $C(\Omega_t)$ *be given for each* $t \in T$. *Then, we have a measurement*

$$
\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, 2^{\prod_{t\in T} X_t}, \widetilde{F}_0), S_{[*]}) \qquad (\textit{cf. Theorem 3.7}). \tag{6.49}
$$

Assume that the measured value by the measurement $\mathbf{M}_{C(\Omega)}(\widetilde{\mathbf{O}}_0, S_{[*]})$ *belongs to* $\prod_{t\in T} \Xi_t$ $(\in 2^{\prod_{t\in T} X_t})$. *Also define* δ_{ω_0} $(\in \mathcal{M}^p_{+1}(\Omega_0))$ *such that:*

$$
[\widetilde{F}_0(\prod_{t\in T} \Xi_t)](\omega_0) = \max_{\omega\in\Omega_0}[\widetilde{F}_0(\prod_{t\in T} \Xi_t)](\omega). \tag{6.50}
$$

Let τ *be any element in* T. *Let* $\{R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t} \mid \Xi_t \in 2^{X_t}\ (\forall t \in T)\}$ *be as in Definition 6.5.* $\Big($*The existence of* $\{R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t} \mid \Xi_t \in 2^{X_t}\ (\forall t \in T)\}$ *is assumed by Theorem 6.6.*$\Big)$ *Then, we see:*

(i). *[The S-state at τ ($\in T$)]. There is a reason to infer that*

$$(\sharp) \quad \text{“The S-state at } \tau \ (\in T) \text{ after } \mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[*]})\text{”} \ = R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\delta_{\omega_0}). \tag{6.51}$$

Also

(ii). *[The S-state at τ ($\in T$) for homomorphism $\Phi_{0,\tau}$]. Assume that $\Phi_{0,\tau} : C(\Omega_\tau) \to C(\Omega_0)$ is homomorphic $\Big($i.e., $\Phi_{\pi(t),t} : C(\Omega_t) \to C(\Omega_{\pi(t)})$ ($\forall t \in T$ such that $0 < t \leq \tau$) is homomorphic$\Big)$. Then there is a reason to infer that*

$$\text{“the S-state at } \tau \ (\in T) \text{ after } \mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S_{[*]})\text{”} \ = \Phi^*_{0,\tau}(\delta_{\omega_0}). \tag{6.52}$$

Here note that $\Phi^*_{0,\tau}(\delta_{\omega_0}) = \delta_{\phi_{0,\tau}(\omega_0)}$ where $\phi_{0,\tau} : \Omega_0 \to \Omega_\tau$ is defined by (3.14).

Proof. (i). See Theorem 6.13 (= Lemma 6.9).

(ii). We see, by Lemma 6.14, that

$$\begin{aligned}
&{}_{C(\Omega_\tau)^*}\langle R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\delta_{\omega_0}), g_\tau\rangle_{C(\Omega_\tau)} = {}_{C(\Omega_\tau)^*}\langle \frac{(B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t})^*(\delta_{\omega_0})}{(B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t})^*(\delta_{\omega_0})}, g_\tau\rangle_{C(\Omega_\tau)} \\
=&\frac{1}{\|(B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t})^*(\delta_{\omega_0})\|_{\mathcal{M}(\Omega_\tau)}}{}_{C(\Omega_0)^*}\langle \delta_{\omega_0}, B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(g_\tau)\rangle_{C(\Omega_0)} \\
=&\frac{1}{[\widetilde{F}_0(\prod_{t\in T}\Xi_t)](\omega_0)}{}_{C(\Omega_0)^*}\langle \delta_{\omega_0}, \widetilde{F}_0(\prod_{t\in T}\Xi_t) \times \Phi_{0,\tau} g_\tau\rangle_{C(\Omega_0)} \quad \text{(by Lemma 6.14)} \\
=&{}_{C(\Omega_\tau)^*}\langle \Phi^*_{0,\tau}(\delta_{\omega_0}), g_\tau\rangle_{C(\Omega_\tau)} \qquad (\forall g_\tau \in C(\Omega_\tau)).
\end{aligned}$$

Then, we see that $R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\delta_{\omega_0}) = \Phi^*_{0,\tau}(\delta_{\omega_0})$.

□

Remark 6.15. [(i) Continued from Example 6.2]. Note that our problem (i) in Example 6.2 was to infer the $h(2)$ and not (α_0, β_0). Regression analysis II (6.52) is applicable to our problem, that is, the above (6.52) says that there is a reason to calculate $h(2)$ in the following:

$$h(2) = \phi_{0,2}(0.4, 1.4) = 0.4 + 1.4 \times 2 = 3.2. \tag{6.53}$$

[(ii) Interesting logic]. It should be noted that, when $\tau = 0$, the Regression Analysis II is the same as the Regression Analysis I. Thus, we also conclude (6.5), i.e., $(\alpha_0, \beta_0) = (0.4, 1.4)$. After all, the Regression Analysis II says that

(M_1) as the result in the case that $\tau = 0$, the conclusion (6.5) in Example 6.2 is reasonable, or

(M_2) as the result in the case that $\tau \neq 0$, the conclusion (6.6) in Example 6.2 is reasonable.

However, it should be noted that the Regression Analysis II does not guarantee that

(M_3) both (6.5) and (6.6) in Example 6.2 are (simultaneously) reasonable.

That is because two measurements (i.e., the measurement $\mathbf{M}_1$ behind (M_1) and the measurement $\mathbf{M}_2$ behind (M_2)) are included in (M_1) and (M_2). If we want to conclude this (M_3), we must consider the simultaneous measurement of "measurement $\mathbf{M}_1$" and "measurement $\mathbf{M}_2$", that is, we must generalize Definition 6.5 (Bayes operator) such as $B^{(0,(0,\tau))}_{\Pi_{t \in T} \Xi_t} : C(\Omega_0) \times C(\Omega_\tau) \to C(\Omega_0)$ satisfying similar conditions since only one measurement is permitted (*cf.* §2.5(II)). This is, of course, interesting, though it is not discussed in this book.

■

6.4 Conclusions

In this chapter we show that regression analysis can be completely understood in PMT as follows (*cf.* [55]):

$$
\text{measurement theory} \\
\Longrightarrow \left\{ \begin{array}{l}
\text{Axiom 1} \Rightarrow \begin{array}{c}\text{Theorem 5.3} \\ \text{(Fisher's method)}\end{array} \Rightarrow \left\{ \begin{array}{l} \text{Corollary 5.5 (conditional probability)} \\ \text{Corollary 5.6 (classical Fisher's method)} \end{array} \right. \\
\\
\text{Axiom 2} \Rightarrow \left\{ \begin{array}{l} \text{Theorem 3.7 (measurability)} \\ \text{Theorem 6.6 (the existence of Bayes operator)} \\ \text{Lemma 6.14 (some property of homomorphic relation).} \end{array} \right.
\end{array} \right.
$$

And, using these results, we derive "regression analysis" as follows:

(i) : "Corollary 5.6" + "Theorem 3.7" ⇒ "Regression Analysis I ",

$$
\text{(ii)} : \quad \text{"Corollary 5.5"} + \text{"Theorem 6.6"} \Rightarrow \left. \begin{array}{c} \text{Theorem 3.7} \\ \\ \text{"Theorem 6.13"} \\ \text{(Markov inference)} \\ \\ \text{"Lemma 6.14"} \end{array} \right\} \Rightarrow \text{"Regression Analysis II".}
$$

We believe that Regression Analysis II is the best (i.e., precise, wide, deep etc.) in all conventional proposals of regression analysis (though it should be generalized as mentioned in Remark 6.15.). It is surprising that both statistics and quantum mechanics can be understood in the same theory, i.e., measurement theory (6.1) (=(1.4)).

We believe that every statistician may want to know the justification of (6.5) and (6.6) in Example 6.2. Thus we expect that many statisticians will be interested in our axiomatic approach. That is because there is no justification without axioms.

We think that the results in Chapters 5 and 6 guarantee that "Fisher's statistics is theoretically true", (*cf.* Declaration (1.11)).

Chapter 7

Practical logic

It is certain that pure logic (*cf.* [89]) is merely a kind of rule in mathematics (or meta-mathematics). However, if it is so, the logic is not guaranteed to be applicable to our world. For instance, (pure) logic does not assure the following famous statement:

[♯] *Since Socrates is a man and all men are mortal, it follows that Socrates is mortal.*

That is, we think that the problem: "Is the [♯] true or not?" should be answered. Thus, the purpose of this chapter is to prove the statement [♯], or more generally, to propose "practical logic", i.e., "logic with an interpretation", [1] which is formulated in the framework of the measurement theory:

$$\text{PMT} = \underset{\text{[Axiom 1 (2.37)]}}{\text{measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} \quad \text{in } C^*\text{-algebra} \quad . \tag{7.1}$$
(=(1.4))

Firstly, the symbol "$A \Rightarrow B$" (i.e., "implication") is defined in terms of measurements. And we prove the standard syllogism for classical systems:

$$\text{“}A \Rightarrow B,\ B \Rightarrow C\text{” implies “}A \Rightarrow C\text{”} \; ^2 \tag{7.2}$$

(This is not trivial, because the (7.2) does not necessarily hold in quantum systems.) We can assert, by "Declaration (1.11)" in §1.4, that this theorem (7.2) guarantees that the above (7.2) (or, the statement [♯]) is "theoretical true". Several variants may be interesting. For example, under the condition that "$A \Rightarrow B$, $B \Rightarrow C$", we can assert a kind of conclusion such as "$C \Rightarrow A$". For completeness, "pure logic" and "practical logic" must not be confused. The former is a basic rule on which mathematics is founded. On the other hand, the latter is a collection of theorems (whose forms are similar to that of "pure logic") in MT. All results in this chapter are due to [41]. Also, this chapter can be skipped if readers want to study statistics in the framework of SMT firstly (*cf.* Chapters 8).

[1] We have no confidence for the naming "practical logic". We may choose the other namings: "empirical logic", "applied logic", "usual logic". etc.

[2] It is said that the syllogism is said to be, for the first time, introduced by Aristotle (B.C.384-B.C.322)

7.1 Measurement, inference, control and practical logic

The PMT has various aspects. For example, we believe that three concepts: "measurement", "inference", and "control" are different aspects of the same thing. Let us explain it as follows: Let $\mathbf{M}_{\mathcal{A}}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[\rho^p]})$ be a measurement formulated in a C^*-algebra $\mathcal{A}$. Note that Axiom 1 can be regarded as the answer to the following problem:

(M) What kind of measured value is obtained by a measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[\rho^p]})$?

As mentioned in Chapter 5, the measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[\rho^p]})$ is often denoted by $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[*]})$, if we want to stress that we do not know the state ρ^p. Using this notation, we can respectively characterize "inference (I)" and "control (C)" as follows:

(I) Assume that we get a measured value $x(\in X)$ by a measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[*]})$. Then, infer the state $[\,*\,]$,

and

(C) Assume that we want to get a measured value $x(\in X)$ by a measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[*]})$. Then, settle the state $[\,*\,]$.

Of course, Fisher's maximum likelihood method is one of the answers of the above problems (I) and (C).

Also, we think that

(L) "Practical logic" is characterized as "a qualitative theory concerning conditional probability (*cf.* §2.5 (IV)) in PMT".

Thus "practical logic" is also one of the aspects of Axiom 1. Also, since "(practical) logic" is a qualitative aspect of "inference", we can say that "(practical) logic" [resp. "inference"] is used in rough [resp. precise] arguments. For completeness, "pure logic" and "practical logic" must not be confused. The former is a basic rule on which mathematics is built. And thus it is not related to our world. On the other hand, the latter is a collection of theorems (whose forms are similar to that of "pure logic") in PMT. Since practical logic is regarded as a theorem in PMT, it automatically possesses the measurement theoretical interpretation. That is, we think that

"practical logic" = "theorems (whose forms are similar to (pure) logic) in MT".

Recall, throughout this book, that the *measured value set* (or, *label set*) X is assumed to be finite if we write $(X, 2^X, F)$ (or, $(X, \mathcal{P}(X), F)$ and not $(X, \mathcal{F}, F)$. In this chapter we always assume that X is finite.

7.2 Quasi-product observables with dependence

We begin with the following definition.

Definition 7.1. [Marginal observable, quasi-product observable, consistency. (*cf.* Definition 2.10.)]. *Let $\mathcal{A}$ be a C^*-algebras. Let $K = \{1, 2, ..., |K|\}$.*

(i). *Consider an observable* $\mathbf{O} \equiv (\times_{k \in K} X_k,\ 2^{\times_{k\in K} X_k},\ F)$ (*with a label set* $\times_{k \in K} X_k$) *in $\mathcal{A}$. Let D be $D \subseteq K$. An observable* $\mathbf{O}_D \equiv (\times_{k\in D} X_k,\ 2^{\times_{k\in D} X_k},\ F_D)$ *in $\mathcal{A}$ is called a D-marginal observable of* $\mathbf{O}$ *if it satisfies that*

$$F_D(\times_{k\in D} \Xi_k) = F\Big((\times_{k\in D} \Xi_k) \times (\times_{k\in K\setminus D} X_k) \Big),$$

for all $\Xi_k \in 2^{X_k}$, $k \in D$. *Also this* $\mathbf{O}_D$ *is denoted by* $\mathbf{O}\big|_D$. *Here note that the marginal observable* $\mathbf{O}\big|_D$ *is equal to the image observable* $\mathbf{O}_{[g_D]}$ *where* $\times_{k\in K} X_k \ni (x_k)_{k\in K} \overset{g_D}{\longmapsto} (x_k)_{k\in D} \in \times_{k\in D} X_k$.

(ii). *For each* $k \in K$, *consider an observable* $\mathbf{O}_k \equiv (X_k, 2^{X_k}, F_k)$ *in $\mathcal{A}$. If there exists an observable* $\mathbf{O}_K \equiv (\times_{k\in K} X_k, 2^{\times_{k\in K} X_k}, F)$ *in $\mathcal{A}$ such that* $\mathbf{O}_K\big|_{\{k\}} = \mathbf{O}_k$ *for all* $k \in K$, *then* $[\mathbf{O}_k : k \in K]$ *is called consistent. Also, this* $\mathbf{O}_K$ *is called a quasi-product observable of* $[\mathbf{O}_k : k \in K]$, *and is sometimes denoted by* $(\times_{k\in K} X_k, 2^{\times_{k\in K} X_k}, \times^{\mathbf{O}_K}_{k\in K} F_k)$, *or* $\times^{\mathbf{O}_K}_{k\in K} \mathbf{O}_k$ $\Big($*or,* $(\times_{k\in K} X_k, 2^{\times_{k\in K} X_k}, \overset{\text{qp}}{\times}_{k\in K} F_k)$, *or* $\overset{\text{qp}}{\times}_{k\in K} \mathbf{O}_k\Big)$.

■

Note that the consistency of observables $[(X_k, 2^{X_k}, F_k) : k \in K]$ in $\mathcal{A}$ is not guaranteed in general. If the commutativity condition:

$$F_{k_1}(\Xi_{k_1}) F_{k_2}(\Xi_{k_2}) = F_{k_2}(\Xi_{k_2}) F_{k_1}(\Xi_{k_1}) \quad (\forall \Xi_{k_1} \in 2^{X_{k_1}},\ \forall \Xi_{k_2} \in 2^{X_{k_2}},\ k_1 \neq k_2)$$

holds, then we can construct a quasi-product observable $\mathbf{O} \equiv (\times_{k\in K} X_k, 2^{\times_{k\in K} X_k}, F \equiv \times^{\mathbf{O}}_{k\in K} F_k)$ such that:

$$F(\Xi_1 \times \Xi_2 \times \cdots \times \Xi_{|K|}) = F_1(\Xi_1)F_2(\Xi_2)\cdots F_{|K|}(\Xi_{|K|}).$$

It is, of course, the case that the uniqueness is not guaranteed even under the above commutativity condition.

Remark 7.2. [Only one measurement is permitted (*cf.* §2.5. Remarks (II))]. If we want the data concerning both $\mathbf{O}_1$ and $\mathbf{O}_2$ for the system $S_{[\rho^p]}$, we must take a simultaneous measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}_{12} \equiv \mathbf{O}_1 \times^{\mathbf{O}_{12}} \mathbf{O}_2, S_{[\rho^p]})$. Therefore, if a quasi-product observable $\mathbf{O}_{12}$ does not exist (i.e., $[\mathbf{O}_1, \mathbf{O}_2]$ is not consistent), the concept of "the data concerning $\mathbf{O}_1$ and $\mathbf{O}_2$ for the system $S_{[\rho^p]}$" is nonsense, i.e., it has no reality. This is a prevalent notion in quantum theory as in the case that the concept "the momentum and position of a particle" or "the trajectory of a particle" is meaningless in quantum theory. (For the recent results, see [37, 40].) It should be emphasized that the importance of this spirit (i.e., "the consistency of $[\mathbf{O}_1, \mathbf{O}_2]$" ⇔ "the reality of data") is essential.

■

As the classical PMT is rather easy, people tend to overlook important facts in classical systems. Since quantum theory is moderately difficult, it is rather handy compared to classical theory.

Let $X = \{x^1, x^2, ..., x^J\}$. Let $\mathbf{O} \equiv (X, 2^X, F)$ be an observable in a commutative C^*-algebra $\mathcal{A}$ (hence by Gelfand theorem, we can assume that $\mathcal{A} = C(\Omega)$). We can consider the following identification:

$$(X, 2^X, F) \longleftrightarrow \Big[\, [F(\{x^j\})](\omega) \ : \ j = 1, 2, ..., J \, \Big]$$

(where $F(\{x^j\}) \equiv [F(\{x^j\})] \in C(\Omega)$), and therefore denote

$$\mathrm{Rep}[\mathbf{O}] = \mathrm{Rep}[(X, 2^X, F)] = \Big[[F(\{x^j\})](\omega) \ : \ j = 1, 2, ..., J \Big].$$

It is clear that

$$0 \le [F(\{x^j\})](\omega) \le 1 \quad \text{and} \quad \sum_{j=1}^{J} [F(\{x^j\})](\omega) = 1 \qquad (\forall \omega \in \Omega),$$

that is, $\mathrm{Rep}[(X, 2^X, F)]$ is considered to be the resolution of the identity (*cf.* §2.3).

Consider two observables $\mathbf{O}_1 \equiv (X_1, 2^{X_1}, F_1)$ and $\mathbf{O}_2 \equiv (X_2, 2^{X_2}, F_2)$ in $C(\Omega)$ such that:

$$X_1 = \{x_1^1, x_1^2, ..., x_1^{J_1}\} \quad \text{and} \quad X_2 = \{x_2^1, x_2^2, ..., x_2^{J_2}\}.$$

Let $\mathbf{O}_{12} \equiv (X_1 \times X_2, 2^{X_1 \times X_2}, F \equiv F_1 \times^{\mathbf{O}_{12}} F_2)$ be a quasi-product observable with the marginal observables $\mathbf{O}_1$ and $\mathbf{O}_2$. (The existence of $\mathbf{O}_{12}$ is guaranteed by Theorem 2.11 since $C(\Omega)$ is commutative.) Put

$$\text{Rep}[\mathbf{O}_{12}] = \begin{bmatrix} [F(\{(x_1^1, x_2^1)\})](\omega) & [F(\{(x_1^1, x_2^2)\})](\omega) & \dots & [F(\{(x_1^1, x_2^{J_2})\})](\omega) \\ [F(\{(x_1^2, x_2^1)\})](\omega) & [F(\{(x_1^2, x_2^2)\})](\omega) & \dots & [F(\{(x_1^2, x_2^{J_2})\})](\omega) \\ \vdots & \vdots & \ddots & \vdots \\ [F(\{(x_1^{J_1}, x_2^1)\})](\omega) & [F(\{(x_1^{J_1}, x_2^2)\})](\omega) & \dots & [F(\{(x_1^{J_1}, x_2^{J_2})\})](\omega) \end{bmatrix}.$$

Let $X = \{x^1, x^2, ..., x^J\}$. Let $\mathbf{O} \equiv (X, 2^X, F)$ be an observable in a C^*-algebra $\mathcal{A}$. Put $X = \Xi_y \bigcup \Xi_n$ (where $\Xi_y \bigcap \Xi_n = \emptyset$). Define the map $g : X \to X_{(2)} \equiv \{y, n\}$ such that $g(x) = y$ (if $x \in \Xi_y$), $= n$ (if $x \in \Xi_n$). Here we can define the two-valued observable $(X_{(2)} \equiv \{y, n\}, 2^{X_{(2)}}, F_{(2)})$ in $\mathcal{A}$ as the image observable $\mathbf{O}_{[g]}$. This two-valued observable is also called *yes-no observable* or $1 - 0$ *observable.* The following lemma says about the conditions that a quasi-product observable of yes-no observables should satisfy.

Lemma 7.3. [The existence condition of yes-no quasi-product observable]. *Consider yes-no observables* $\mathbf{O}_1 \equiv (X_1, 2^{X_1}, F_1)$ *and* $\mathbf{O}_2 \equiv (X_2,\ 2^{X_2},\ F_2)$ *in a commutative* C^**-algebra* $C(\Omega)$ *such that:*

$$X_1 = \{y_1, n_1\} \quad \textit{and} \quad X_2 = \{y_2, n_2\}.$$

Let $\mathbf{O}_{12} \equiv (X_1 \times X_2, 2^{X_1 \times X_2}, F \equiv F_1 \times^{\mathbf{O}_{12}} F_2)$ *be a quasi-product observable with the marginal observables* $\mathbf{O}_1$ *and* $\mathbf{O}_2$.

Put

$$\text{Rep}[\mathbf{O}_{12}] = \begin{bmatrix} [F(\{(y_1, y_2)\})](\omega) & [F(\{(y_1, n_2)\})](\omega) \\ [F(\{(n_1, y_2)\})](\omega) & [F(\{(n_1, n_2)\})](\omega) \end{bmatrix}$$

$$= \begin{bmatrix} \alpha(\omega) & [F_1(\{y_1\})](\omega) - \alpha(\omega) \\ [F_2(\{y_2\})](\omega) - \alpha(\omega) & 1 + \alpha(\omega) - [F_1(\{y_1\})](\omega) - [F_2(\{y_2\})](\omega) \end{bmatrix}, \tag{7.3}$$

where $\alpha \in C(\Omega)$. $\Big($*Note that* $[F(\{(y_1, y_2)\})](\omega) + [F(\{(y_1, n_2)\})](\omega) = [F_1(\{y_1\})](\omega)$ *and* $[F(\{(y_1, y_2)\})](\omega) + [F(\{(n_1, y_2)\})](\omega) = [F_2(\{y_2\})](\omega)\Big)$.

That is,

	$[F_2(\{y_2\})](\omega)$	$[F_2(\{n_2\})](\omega)$
$[F_1(\{y_1\})](\omega)$	$\alpha(\omega)$	$[F_1(\{y_1\})](\omega)-\alpha(\omega)$
$[F_1(\{n_1\})](\omega)$	$[F_2(\{y_2\})](\omega)-\alpha(\omega)$	$1+\alpha(\omega)-[F_1(\{y_1\})](\omega)-[F_2(\{y_2\})](\omega)$

Then, it holds that

$$\max\{0,[F_1(\{y_1\})](\omega)+[F_2(\{y_2\})](\omega)-1\}\le\alpha(\omega)\le\min\{[F_1(\{y_1\})](\omega),\ [F_2(\{y_2\})](\omega)\}\quad(\forall\omega\in\Omega). \tag{7.4}$$

Conversely, for any α $(\in C(\Omega))$ *that satisfies (7.4), the observable* $\mathbf{O}_{12}$ *defined by (7.3) is a quasi-product observable with the marginal observables* $\mathbf{O}_1$ *and* $\mathbf{O}_2$*. Also, note that*

$$[F(\{(y_1,n_2)\})](\omega)=0\ \Leftrightarrow\ \alpha(\omega)=[F_1(\{y_1\})](\omega)\ \Rightarrow\ [F_1(\{y_1\})](\omega)\le[F_2(\{y_2\})](\omega). \tag{7.5}$$

Proof. Though this lemma is easy, we add a brief proof for completeness. Since $0\le[F(\{(x_1^1,x_2^2)\})]\ (\omega)\le 1$, $(\forall x^1,x^2\in\{y,n\})$, we see, by (7.3), that

$$\begin{aligned}&0\le\alpha(\omega)\le 1,\quad 0\le[F_1(\{y_1\})](\omega)-\alpha(\omega)\le 1,\quad 0\le[F_2(\{y_2\})](\omega)-\alpha(\omega)\le 1,\\&0\le 1+\alpha(\omega)-[F_1(\{y_1\})](\omega)-[F_2(\{y_2\})](\omega)\le 1,\end{aligned} \tag{7.6}$$

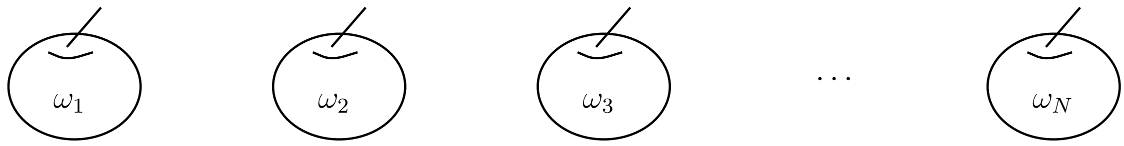
which clearly implies (7.4). Conversely. if α satisfies (7.4), then we easily see (7.6). Also, (7.5) is obvious. This completes the proof. □

Next we provide several examples, which will promote a understanding of our theory.

Example 7.4. [Tomatoes' example]. Let $\Omega=\{\omega_1,\omega_2,....,\omega_N\}$ be a set of tomatoes, which is regarded as a compact Hausdorff space with the discrete topology. Consider yes-no observables $\mathbf{O}_{\rm RD}\equiv(X_{\rm RD},2^{X_{\rm RD}},F_{\rm RD})$ and $\mathbf{O}_{\rm SW}\equiv(X_{\rm SW},2^{X_{\rm SW}},F_{\rm SW})$ in $C(\Omega)$ such that:

$$X_{\rm RD}=\{y_{\rm RD},n_{\rm RD}\}\text{ and }X_{\rm SW}=\{y_{\rm SW},n_{\rm SW}\},$$

where we consider that "$y_{\rm RD}$" and "$n_{\rm RD}$" respectively mean "RED" and "NOT RED". Similarly, "$y_{\rm SW}$" and "$n_{\rm SW}$" respectively mean "SWEET" and "NOT SWEET".
For example, the ω_1 is red and not sweet, the ω_2 is red and sweet, etc. as follows.



$y_{\rm RD}$	$y_{\rm RD}$	$n_{\rm RD}$	$\cdots$	$n_{\rm RD}$
$n_{\rm SW}$	$y_{\rm SW}$	$y_{\rm SW}$	$\cdots$	$n_{\rm SW}$

We see that

(∗) the probability that $x_{\rm RD}$ $(\in X_{\rm RD} \equiv \{y_{\rm RD}, n_{\rm RD}\})$, the measured value by the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}_{\rm RD}, S_{[\delta_{\omega_n}]})$, belongs to $\Xi_{\rm RD}$ $(\subseteq X_{\rm RD} \equiv \{y_{\rm RD}, n_{\rm RD}\})$ is given by

$$\delta_{\omega_n}(F_{\rm RD}(\Xi_{\rm RD})) \quad (= [F_{\rm RD}(\Xi_{\rm RD})](\omega_n)\ ).$$

That is, the probability that the tomato ω_n is observed as "RED" $\big[$ resp. "NOT RED"$\big]$ is given by $[F_{\rm RD}(\{y_{\rm RD}\})]\ (\omega_n)$ $\big[$ resp. $[F_{\rm RD}(\{n_{\rm RD}\})]\ (\omega_n)$ $\big]$. (*Continued to Example 7.5*). ■

Example 7.5. [Tomatoes' example; continued from Example 7.4]. Consider the quasi-product observable as follows:

$$\mathbf{O} = (X_{\rm RD} \times X_{\rm SW}, 2^{X_{\rm RD} \times X_{\rm SW}}, F \equiv F_{\rm RD} \overset{\mathbf{qp}}{\times} F_{\rm SW}),$$

that is,

$$\begin{aligned} \mathrm{Rep}[\mathbf{O}] &= \begin{bmatrix} [F(\{(y_{\rm RD}, y_{\rm SW})\})](\omega) & [F(\{(y_{\rm RD}, n_{\rm SW})\})](\omega) \\ [F(\{(n_{\rm RD}, y_{\rm SW})\})](\omega) & [F(\{(n_{\rm RD}, n_{\rm SW})\})](\omega) \end{bmatrix} \\ &= \begin{bmatrix} \alpha(\omega) & [F_{\rm RD}(\{y_{\rm RD}\})](\omega) - \alpha(\omega) \\ [F_{\rm SW}(\{y_{\rm SW}\})](\omega) - \alpha(\omega) & 1 + \alpha(\omega) - [F_{\rm RD}(\{y_{\rm RD}\})](\omega) - [F_{\rm SW}(\{y_{\rm SW}\})](\omega) \end{bmatrix} \end{aligned}$$

where $\alpha(\omega)$ satisfies (7.4). Hence by Axiom 1, when we observe that the tomato ω_n is "RED", we can see that the probability that the tomato ω_n is "SWEET" is given by

$$\frac{[F(\{(y_{\rm RD}, y_{\rm SW})\})](\omega_n)}{[F(\{(y_{\rm RD}, y_{\rm SW})\})](\omega_n) + [F(\{(y_{\rm RD}, n_{\rm SW})\})](\omega_n)}. \tag{7.7}$$

$\Big($For the conditional probability, see §2.5(IV).$\Big)$ Here note that (7.7) implies ;

$$\text{“}[F(\{(y_{\rm RD}, n_{\rm SW})\})](\omega_n) = 0\text{”} \quad \text{if and only if} \quad \text{“RED”} \Rightarrow \text{“SWEET”}\ , \tag{7.8}$$

which is also clearly equivalent to "NOT SWEET" ⇒ "NOT RED".

■

Being motivated by the above (7.8), we introduce the following definition of "implication" as a general form which is applicable to classical and quantum systems.

Definition 7.6. [Implication]. *Let* $\mathbf{O}_1 \equiv (X_1, 2^{X_1}, F_1)$ *and* $\mathbf{O}_2 \equiv (X_2, 2^{X_2}, F_2)$ *be observables (not necessarily two-valued observables) in a* C^**-algebra* $\mathcal{A}$*. Let* $\mathbf{O}_{12} = (X_1 \times X_2, 2^{X_1} \times X_2, F_1 \overset{\mathbf{O}_{12}}{\times} F_2)$ *be a quasi-product observable of* $\mathbf{O}_1$ *and* $\mathbf{O}_2$*. Let* $\rho^p \in \mathfrak{S}^p(\mathcal{A}^*)$*. Let* $\Xi_1 \in \mathcal{P}(X_1)$ *and* $\Xi_2 \in \mathcal{P}(X_2)$*. Then, the condition*

$$\rho^p\Big((F_1 \overset{\mathbf{O}_{12}}{\times} F_2)(\Xi_1 \times (X_2 \setminus \Xi_2))\Big) = 0 \tag{7.9}$$

is denoted by

$$\mathbf{O}_1^{\Xi_1} \underset{\mathbf{M}_{\mathcal{A}}(\mathbf{O}_{12}, S_{[\rho^p]})}{\Longrightarrow} \mathbf{O}_2^{\Xi_2}. \tag{7.10}$$

■

Remark 7.7. [Contraposition]. Assume that we get a measured value (x_1, x_2) ($\in X_1 \times X_2$) by the measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}_{12}, S_{[\rho^p]})$. And assume the condition (7.10). If we know that $x_1 \in \Xi_1$, then we can assure that $x_2 \in \Xi_2$. Also, (7.9) is of course also equal to $\mathbf{O}_1^{X_1 \setminus \Xi_1} \underset{\mathbf{M}_{\mathcal{A}}(\mathbf{O}_{12}, S_{[\rho^p]})}{\Longleftarrow} \mathbf{O}_2^{X_2 \setminus \Xi_2}$ since $\mathbf{O}_{12} = \mathbf{O}_{\{1,2\}} = \mathbf{O}_{21}$ (i.e., $K = \{1, 2\}$ is not regarded as an ordered set). That is, "$\mathbf{O}_1^{X_1 \setminus \Xi_1} \underset{\mathbf{M}_{\mathcal{A}}(\mathbf{O}_{12}, S_{[\rho^p]})}{\Longleftarrow} \mathbf{O}_2^{X_2 \setminus \Xi_2}$" is the contraposition of (7.10).

■

7.3 Consistency and syllogism

In this section we study the consistent condition for observables. We show several theorems of practical syllogisms (i.e., theorems concerning "implication" in Definition 7.6).

7.3.1 Consistent condition

Though we are not concerned with quantum theory in this chapter, our investigations for classical systems become clearer in comparison with quantum theory. Therefore, the following definitions (Definitions 7.8 and 7.9) are common in both classical and quantum theory.

Definition 7.8. [Covering family]. *Let* $\mathcal{A}$ *be a* C^**-algebra. For each* $k \in K \equiv \{1, 2, ..., |K|-1, |K|\}$*, consider a label set* X_k*. Consider* $\mathcal{D}$ $(\subseteq 2^K)$ *such that* $\bigcup_{D \in \mathcal{D}} D = K$*. Then,* $\mathcal{G} \equiv [\ \mathbf{O}_D \equiv (\ \times_{k \in D} X_k,\ 2^{\times_{k \in D} X_k},\ F_D)\ : D \in \mathcal{D}\ ]$ *is called a covering family of observables in* $\mathcal{A}$*, if it satisfies the following condition:*

$$\mathbf{O}_{D_1}\big|_{D_1 \cap D_2} = \mathbf{O}_{D_2}\big|_{D_1 \cap D_2} \qquad (\forall D_1, \forall D_2 \in \mathcal{D} \text{ such that } D_1 \cap D_2 \neq \emptyset).$$

Note that, if $\mathcal{G}$ *is a covering family, it holds that* $\mathbf{O}_{D_1}\big|_{\{k\}} = \mathbf{O}_{D_2}\big|_{\{k\}}$ *for any* $k \in K$ *and any* $D_1, D_2 \in \mathcal{D}$ *such that* $k \in D_1 \bigcap D_2$*. Thus, a covering family of observables* $\mathcal{G}$ *determines a unique* $\{k\}$*-marginal observable* $\mathbf{O}_k \equiv (X_k, 2^{X_k}, F_k)$ *for each* $k \in K$*.*

■

The following definition is a generalization of Definition 7.1 (i.e., the case that $\mathcal{D} = \{\{1\}, \{2\}, ..., \{|K|\}\}$).

Definition 7.9. [Consistent condition]. *Let* $\mathcal{A}$ *be a* C^**-algebra. A covering family of observable* $\mathcal{G} \equiv [\ \mathbf{O}_D \equiv (\ \times_{k \in D} X_k,\ 2^{\times_{k \in D} X_k},\ F_D)\ : D \in \mathcal{D}\ (\ \subseteq 2^{K)}\ ]$ *in* $\mathcal{A}$ *is called consistent, if there exists an observable* $\mathbf{O}_K \equiv (\ \times_{k \in K} X_k,\ 2^{\times_{k \in K} X_k},\ F)$ *in* $\mathcal{A}$ *such that:*

$$\mathbf{O}_K\big|_D = \mathbf{O}_D \qquad (\forall D \in \mathcal{D}). \tag{7.11}$$

Also, the above relation (7.11) is denoted by

$$[\mathbf{O}_D : D \in \mathcal{D}] \sqsubset \mathbf{O}_K. \tag{7.12}$$

■

Remark 7.10. [Consistent condition]. Under the condition (7.12), the data concerning $\mathcal{G} \equiv [\mathbf{O}_D : D \in \mathcal{D}]$ for the system $S_{[\rho^p]}$ is obtained by the simultaneous measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}_K, S_{[\rho^p]})$. Thus, a covering family $\mathcal{G}$ has no reality, if it is not consistent. Recall the arguments in Remark 7.2, which correspond to the above definition for the case that $\mathcal{D} = \{\{1\}, \{2\}\}$.

■

Lemma 7.11. [Consistent condition]. *Let* $\mathcal{A}$ *be a* C^**-algebra. Let* $\mathcal{G}_1 \equiv [\ \mathbf{O}^1_{D_1}\ : D_1 \in \mathcal{D}_1\ (\ \subseteq 2^K)\ ]$ *be a covering family of observables in* $\mathcal{A}$*. And let* $\mathcal{G}_2 \equiv [\ \mathbf{O}^2_{D_2}\ : D_2 \in \mathcal{D}_2\ (\ \subseteq 2^{K)}\ ]$ *be a consistent covering family of observables in* $\mathcal{A}$*. Assume that for any* $D_1 \in \mathcal{D}_1$ *there* *exists an* D_2 $(\in \mathcal{D}_2)$ *such that:*

$$D_1 \subseteq D_2 \quad \text{and} \quad \mathbf{O}^1_{D_1} = \mathbf{O}^2_{D_2}\big|_{D_1}. \tag{7.13}$$

Then, $\mathcal{G}_1$ *is consistent.*

Proof. Since a covering family $\mathcal{G}_2$ is consistent, there exists an observable $\mathbf{O}_K \equiv (\times_{k\in K} X_k,\ 2^{\times_{k\in K} X_k},\ F_K)$ in $\mathcal{A}$ such that $\mathbf{O}^2_{D_2} = \mathbf{O}_K\big|_{D_2}$ $(\forall D_2 \in \mathcal{D}_2)$. Let D_1 be any element in $\mathcal{D}_1$. Then, by choosing $D_2(\in \mathcal{D}_2)$ satisfying (7.13), we see that $\mathbf{O}^1_{D_1} = \mathbf{O}^2_{D_2}\big|_{D_1}$ $= (\mathbf{O}_K\big|_{D_2})\big|_{D_1} = \mathbf{O}_K\big|_{D_1}$. This completes the proof. □

Lemma 7.12. [Consistent condition and quasi-product observables]. *Let* $\mathcal{A}$ *be a commutative* C^**-algebra (i.e.,* $\mathcal{A} = C(\Omega)$*). Let* D_{12} *and* D_{23} *be subsets of* K*. Put* $D_{123} \equiv D_{12} \bigcup D_{23} \equiv (D_{12} \setminus D_{23}) \bigcap (D_{12} \cap D_{23}) \bigcap (D_{23} \setminus D_{12}) \equiv D_1 \bigcup D_2 \bigcup D_3$*. Consider the following observables in* $C(\Omega)$ *:*

$$\mathbf{O}_{D_{12}} \equiv (\mathop{\times}_{k\in D_{12}} X_k, \mathcal{P}(\mathop{\times}_{k\in D_{12}} X_k), F_{D_{12}}) \quad \text{and} \quad \mathbf{O}_{D_{23}} \equiv (\mathop{\times}_{k\in D_{23}} X_k, \mathcal{P}(\mathop{\times}_{k\in D_{23}} X_k), F_{D_{23}})$$

such that $\mathbf{O}_{D_{12}}\big|_{D_2} = \mathbf{O}_{D_{23}}\big|_{D_2}$*. Then, there exists an observable* $\mathbf{O}_{D_{123}} \equiv (\times_{k\in D_{123}} X_k,$ $\mathcal{P}(\times_{k\in D_{123}} X_k),\ F_{D_{123}})$ *such that* $\mathbf{O}_{D_{123}}\big|_{D_{12}} = \mathbf{O}_{D_{12}}$ *and* $\mathbf{O}_{D_{123}}\big|_{D_{23}} = \mathbf{O}_{D_{23}}$.

Proof. Assume that $D_{12} \bigcap D_{23} \neq \emptyset$. (If $D_{12} \bigcap D_{23} = \emptyset$, this lemma is trivial. Put $Y_m = \times_{k\in D_m} X_k = \{y^1_m, y^2_m, ..., y^{j_m}_m, ..., y^{M_m}_m\}$, $m = 1, 2, 3$. (So, $M_m = \prod_{k\in D_m} |X_k|$.) Thus, we can put, by $Y_1 \times Y_2 = \times_{k\in D_{12}} X_k$ and $Y_2 \times Y_3 = \times_{k\in D_{23}} X_k$, that

$$\mathbf{O}_{D_{12}} = (Y_1 \times Y_2, \mathcal{P}(Y_1 \times Y_2), F_{12} \equiv F_{D_{12}})$$

and

$$\mathbf{O}_{D_{23}} = (Y_2 \times Y_3, \mathcal{P}(Y_2 \times Y_3), F_{23} \equiv F_{D_{23}}).$$

Define the observable $\mathbf{O}_{D_{123}} \equiv (\times^3_{m=1} Y_m, \mathcal{P}(\times^3_{m=1} Y_m), F_{123})$ in $C(\Omega)$ such that:

$$[F_{123}(\{(y^{j_1}_1, y^{j_2}_2, y^{j_3}_3)\})](\omega) = \begin{cases} \frac{[F_{12}(\{(y^{j_1}_1, y^{j_2}_2)\})](\omega)\cdot[F_{23}(\{(y^{j_2}_2, y^{j_3}_3)\})](\omega)}{[F_2(\{y^{j_2}_2\})](\omega)} & \text{if } [F_2(\{y^{j_2}_2\})](\omega) \neq 0 \\ 0 & \text{if } [F_2(\{y^{j_2}_2\})](\omega) = 0 \end{cases}$$

for $1 \leq \forall j_1 \leq M_1$, $1 \leq \forall j_2 \leq M_2$, $1 \leq \forall j_3 \leq M_3$. Therefore, it is clear that this lemma holds. For example, $\mathbf{O}_{D_{123}}\big|_{D_{23}} = \mathbf{O}_{D_{23}}$ is easily seen as follows:

$$[F_{123}(Y_1 \times \{(y_2^{j_2}, y_3^{j_3})\})](\omega) = \sum_{y_1^{j_1} \in Y_1} [F_{123}(\{(y_1^{j_1}, y_2^{j_2}, y_3^{j_3})\})](\omega)$$

$$= \sum_{y_1^{j_1} \in Y_1} \frac{[F_{12}(\{(y_1^{j_1}, y_2^{j_2})\})](\omega) \cdot [F_{23}(\{(y_2^{j_2}, y_3^{j_3})\})](\omega)}{[F_2(\{y_2^{j_2}\})](\omega)}$$

$$= \frac{[F_{12}(Y_1 \times \{y_2^{j_2}\})](\omega) \cdot [F_{23}(\{(y_2^{j_2}, y_3^{j_3})\})](\omega)}{[F_2(\{y_2^{j_2}\})](\omega)} = \frac{[F_2(\{y_2^{j_2}\})](\omega) \cdot [F_{23}(\{(y_2^{j_2}, y_3^{j_3})\})](\omega)}{[F_2(\{y_2^{j_2}\})](\omega)}$$

$$= [F_{23}(\{(y_2^{j_2}, y_3^{j_3})\})](\omega) \qquad (\forall \omega \in \Omega,\ 1 \leq \forall j_2 \leq M_2,\ 1 \leq \forall j_3 \leq M_3).$$

This completes the proof. □

The following theorem is a kind of generalization of Theorem 2.11 $\big($which essentially corresponds to the result for $\mathcal{D} = \{\{1\}, \{2\}, ..., \{|K|\}\}$ in the following theorem$\big)$. Here note that a covering family $[\mathbf{O}_D : D \in \mathcal{D}]$ is equivalent to $[\mathbf{O}_{D'} : D' \in \{D' : D' \subseteq D$ for some $D \in \mathcal{D}\}]$ where $\mathbf{O}_{D'} = \mathbf{O}_D\big|_{D'}$ for any D' such that $D' \subseteq D$.

Theorem 7.13. [Consistent condition and quasi-product observables]. *Let* $\mathcal{D} = \{\{1, 2\}, \{2, 3\}, ..., \{|K|-1, |K|\}\}$ $(\subseteq 2^K)$. *Let* $\mathcal{G} = [\mathbf{O}_D = (\times_{k \in D} X_k, 2^{\times_{k\in D} X_k}, F_D) : D \in \mathcal{D}]$ *be a covering family of observables in a commutative* C^**-algebra* $C(\Omega)$. $\big($*Here we can put* $\mathcal{G} = [\ \mathbf{O}_{k,k+1} \equiv (X_k \times X_{k+1}, \mathcal{P}(X_k \times X_{k+1}), F_{k,k+1} \equiv F_k \overset{\mathbf{O}_{k,k+1}}{\times} F_{k+1}) : k = 1, 2, ..., |K|-1\ ]$.$\big)$ *Then,* $\mathcal{G} = [\ \mathbf{O}_{k,k+1} : k = 1, 2, ..., |K|-1\ ]$ *is consistent.*

Proof. Put $D_{12} = \{1, 2\}$ and $D_{23} = \{2, 3\}$. By Lemma 7.12, we get $\mathbf{O}_{123}$ $(= \mathbf{O}_{D_{123}})$ such that $\mathcal{G}_3 = [\mathbf{O}_{123}, \mathbf{O}_{34}, \mathbf{O}_{45}, ..., \mathbf{O}_{|K|-1,|K|}]$ is a covering family in $C(\Omega)$ where $\mathbf{O}_{12} = \mathbf{O}_{123}\big|_{\{1,2\}}$ and $\mathbf{O}_{23} = \mathbf{O}_{123}\big|_{\{2,3\}}$. Iteratively, we get $\mathcal{G}_{|K|-1} = [\mathbf{O}_{123\cdots|K|-1}, \mathbf{O}_{|K|-1,|K|}]$ and $\mathcal{G}_{|K|} = [\mathbf{O}_{123\cdots|K|-1,|K|}] \equiv [\mathbf{O}_K]$, which is clearly consistent. So, by Lemma 7.11, we see that $\mathcal{G}_{|K|-1} \sqsubset \mathbf{O}_K$. Therefore, we iteratively get $\mathcal{G} \sqsubset \mathbf{O}_K$. This completes the proof. □

Remark 7.14. [Quantum PMT]. This theorem is due to the commutativity of a C^*-algebra $C(\Omega)$. In general (particularly in quantum systems, i.e., $\mathcal{A} = \mathcal{C}(V)$), there exists no $\mathbf{O}_{123}$ such that $[\mathbf{O}_{12}, \mathbf{O}_{23}] \sqsubset \mathbf{O}_{123}$ (i.e., $[\mathbf{O}_{12}, \mathbf{O}_{23}]$ is not consistent in general). Thus, we have no simultaneous measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}_{123}, S_{[\rho^p]})$. Therefore, in general, we can not get information (i.e., data) concerning the covering family $[\mathbf{O}_{12}, \mathbf{O}_{23}]$ for the quantum system $S_{[\rho^p]}$. That is, in general, the covering family $[\mathbf{O}_{12}, \mathbf{O}_{23}]$ has no reality in quantum mechanics.

■

The following notation is the preparation for Theorems 7.19 and 7.23.

Notation 7.15. [Preparation for Theorems 7.19 and 7.23]. *Let* $\mathcal{G} = [\mathbf{O}_{12}, \mathbf{O}_{23}, ...\mathbf{O}_{|K|-1,|K|}] \equiv [\ (X_k \times X_{k+1},\ \mathcal{P}(X_k \times X_{k+1}),\ F_{k,k+1} \equiv F_k \times^{\mathbf{O}_{k,k+1}} F_{k+1}) : k = 1, 2, ..., |K| - 1\ ]$ *be a covering family of observables in a commutative* C^**-algebra* $C(\Omega)$*. (So,* $\mathcal{G}$ *is consistent as in Theorem 7.13). Suppose that* $X_k = \{y_k, n_k\}$ *for each* $k \in K$*. As in Definition 7.8, put*

$$\mathrm{Rep}[\mathbf{O}_\mathrm{k}] = \mathrm{Rep}[(\mathrm{X_k}, 2^{\mathrm{X_k}}, \mathrm{F_k})] = \Big[[\mathrm{F_k}(\{\mathrm{y_k}\})](\omega), [\mathrm{F_k}(\{\mathrm{n_k}\})](\omega)\Big] \equiv \Big[\mathrm{p}^1_\mathrm{k}(\omega), \mathrm{p}^0_\mathrm{k}(\omega)\Big]$$

for all $k = 1, 2, 3, ..., |K|$*. And put*

$$\begin{aligned}
&\mathrm{Rep}[\mathbf{O}_{\mathrm{k,k+1}}] = \mathrm{Rep}[(\mathrm{X_k} \times \mathrm{X_{k+1}}, 2^{\mathrm{X_k} \times \mathrm{X_{k+1}}}, \mathrm{F_{k,k+1}})] \\
&= \begin{bmatrix} [F_{k,k+1}(\{y_k\} \times \{y_{k+1}\})](\omega) & [F_{k,k+1}(\{y_k\} \times \{n_{k+1}\})](\omega) \\ [F_{k,k+1}(\{n_k\} \times \{y_{k+1}\})](\omega) & [F_{k,k+1}(\{n_k\} \times \{n_{k+1}\})](\omega) \end{bmatrix} \\
&\equiv \begin{bmatrix} p^{11}_{k,k+1}(\omega) & p^{10}_{k,k+1}(\omega) \\ p^{01}_{k,k+1}(\omega) & p^{00}_{k,k+1}(\omega) \end{bmatrix} \\
&\equiv \begin{bmatrix} p^{11}_{k,k+1}(\omega) & p^1_k(\omega) - p^{11}_{k,k+1}(\omega) \\ p^1_{k+1}(\omega) - p^{11}_{k,k+1}(\omega) & 1 + p^{11}_{k,k+1}(\omega) - p^1_k(\omega) - p^1_{k+1}(\omega) \end{bmatrix}
\end{aligned} \tag{7.14}$$

for all $k = 1, 2, ..., |K| - 1$*, where* $p^{11}_{k,k+1}(\omega)$ *satisfies (7.4). Let* $\mathbf{O}_K \equiv (\times_{k \in K} X_k,\ \mathcal{P}(\times_{k \in K} X_k),\ F_K)$ *be any observable in* $C(\Omega)$ *such that:*

$$[\mathbf{O}_{12}, \mathbf{O}_{23}, ...\mathbf{O}_{|K|-1,|K|}] \sqsubset \mathbf{O}_K. \tag{7.15}$$

(The existence of $\mathbf{O}_K$ *is guaranteed by Theorem 7.13.) Put*

$$\Big[p^{j_1,j_2,...,j_{|K|}}_{1,2,...,|K|}(\omega)\ :\ j_1, j_2, ..., j_{|K|} = 1, 0\Big] \equiv \Big[[F_K(\times_{k=1}^{|K|} \{x^{j_k}_k\})](\omega)\ :\ j_1, j_2, ..., j_{|K|} = 1, 0\Big], \tag{7.16}$$

where $x^{j_k}_k = y_k$ *(if* $j_k = 1$ *) and* $x^{j_k}_k = n_k$ *(if* $j_k = 0$ *). Define* $\mathbf{O}_{1,|K|} \equiv (X_1 \times X_{|K|},\ \mathcal{P}(X_1 \times X_{|K|}),\ F_{1,|K|})$ *such that* $\mathbf{O}_{1,|K|} = \mathbf{O}_K\big|_{\{1,|K|\}}$*. Put*

$$\begin{aligned}
&\mathrm{Rep}[\mathbf{O}_{1,|\mathrm{K}|}] = \mathrm{Rep}[(\mathrm{X_1} \times \mathrm{X_{|K|}}, 2^{\mathrm{X_1} \times \mathrm{X_{|K|}}}, \mathrm{F_{1,|K|}})] \\
&= \begin{bmatrix} [F_{1,|K|}(\{y_1\} \times \{y_{|K|}\})](\omega) & [F_{1,|K|}(\{y_1\} \times \{n_{|K|}\})](\omega) \\ [F_{1,|K|}(\{n_1\} \times \{y_{|K|}\})](\omega) & [F_{1,|K|}(\{n_1\} \times \{n_{|K|}\})](\omega) \end{bmatrix} \\
&\equiv \begin{bmatrix} p^{11}_{1,|K|}(\omega) & p^{10}_{1,|K|}(\omega) \\ p^{01}_{1,|K|}(\omega) & p^{00}_{1,|K|}(\omega) \end{bmatrix} \equiv \begin{bmatrix} p^{11}_{1,|K|}(\omega) & p^1_1(\omega) - p^{11}_{1,|K|}(\omega) \\ p^1_{|K|}(\omega) - p^{11}_{1,|K|}(\omega) & 1 + p^{11}_{1,|K|}(\omega) - p^1_1(\omega) - p^1_{|K|}(\omega) \end{bmatrix}.
\end{aligned} \tag{7.17}$$

(*Continued to Lemmas 7.16 and 7.17 and Theorem 7.19 for* $K = \{1, 2, 3\}$, *and to Theorem 7.23 for general case*).

■

Lemma 7.16. [Continued from Notation 7.15]. *Under Notation 7.15 for* $K = \{1, 2, 3\}$, *we see, (putting* $p_{123}^{j_1 j_2 j_3} = p_{123}^{j_1 j_2 j_3}(\omega)$ *in (7.16),* $p_{123}^{111} = A$ *and* $p_{123}^{101} = B$*),*

$$
\begin{aligned}
&p_{123}^{111} = A(\omega), && p_{123}^{011} = p_{23}^{11} - A(\omega),\\
&p_{123}^{110} = p_{12}^{11} - A(\omega), && p_{123}^{010} = p_{2}^{1} - p_{12}^{11} - p_{23}^{11} + A(\omega),\\
&p_{123}^{101} = B(\omega), && p_{123}^{001} = p_{3}^{1} - p_{23}^{11} - B(\omega),\\
&p_{123}^{100} = p_{1}^{1} - p_{12}^{11} - B(\omega), && p_{123}^{000} = 1 - p_{1}^{1} - p_{2}^{1} - p_{3}^{1} + p_{12}^{11} + p_{23}^{11} + B(\omega),
\end{aligned} \tag{7.18}
$$

where

$$
\max\{0, -p_2^1(\omega) + p_{12}^{11}(\omega) + p_{23}^{11}(\omega)\} \le A(\omega) \le \min\{p_{12}^{11}(\omega), p_{23}^{11}(\omega)\} \tag{7.19}
$$

and

$$
\begin{aligned}
&\max\{0, p_1^1(\omega) + p_2^1(\omega) + p_3^1(\omega) - p_{12}^{11}(\omega) - p_{23}^{11}(\omega) - 1\}\\
&\le B(\omega) \le \min\{p_1^1(\omega) - p_{12}^{11}(\omega), p_3^1(\omega) - p_{23}^{11}(\omega)\}.
\end{aligned} \tag{7.20}
$$

Proof. From (7.16), (7.15) and (7.14) for $K = \{1, 2, 3\}$, we see

$$
\begin{aligned}
&p_{123}^{111} + p_{123}^{110} = p_{12}^{11}, && p_{123}^{101} + p_{123}^{100} = p_{12}^{10} = p_1^1 - p_{12}^{11},\\
&p_{123}^{011} + p_{123}^{010} = p_{12}^{01} = p_2^1 - p_{12}^{11}, && p_{123}^{001} + p_{123}^{000} = p_{12}^{00} = 1 + p_{12}^{11} - p_1^1 - p_2^1,\\
&p_{123}^{111} + p_{123}^{011} = p_{23}^{11}, && p_{123}^{110} + p_{123}^{010} = p_{23}^{10} = p_2^1 - p_{23}^{11},\\
&p_{123}^{101} + p_{123}^{001} = p_{23}^{01} = p_3^1 - p_{23}^{11}, && p_{123}^{100} + p_{123}^{000} = p_{23}^{00} = 1 - p_{23}^{11} - p_2^1 - p_3^1.
\end{aligned}
$$

After a small computation, we get (7.18). Since $0 \le p_{123}^{j_1 j_2 j_3}(\omega) \le 1$, we see, from (7.18), that

$$
\begin{aligned}
&0 \le A \le 1, \quad p_{23}^{11} - 1 \le A \le p_{23}^{11}, \quad p_{12}^{11} - 1 \le A \le p_{12}^{11},\\
&-p_2^1 + p_{12}^{11} + p_{23}^{11} \le A \le 1 - p_2^1 + p_{12}^{11} + p_{23}^{11},\\
&0 \le B \le 1, \quad p_3^1 - p_{23}^{11} - 1 \le B \le p_3^1 - p_{23}^{11}, \quad p_1^1 - p_{12}^{11} - 1 \le B \le p_1^1 - p_{12}^{11},\\
&p_1^1 + p_2^1 + p_3^1 - p_{12}^{11} - p_{23}^{11} - 1 \le B \le p_1^1 + p_2^1 + p_3^1 - p_{12}^{11} - p_{23}^{11}.
\end{aligned}
$$

This implies (7.19) and (7.20). This completes the proof. □

Lemma 7.17. [Continued from Notation 7.15]. *Under Notation 7.15 for* $K = \{1, 2, 3\}$, *we see*

$$\begin{aligned}
&\max\{0, -p_2^1(\omega) + p_{12}^{11}(\omega) + p_{23}^{11}(\omega)\} \\
&\qquad\qquad + \max\{0, p_1^1(\omega) + p_2^1(\omega) + p_3^1(\omega) - p_{12}^{11}(\omega) - p_{23}^{11}(\omega) - 1\} \\
&\leq p_{13}^{11}(\omega) \qquad (7.21)\\
&\leq \min\{p_{12}^{11}(\omega), p_{23}^{11}(\omega)\} + \min\{p_1^1(\omega) - p_{12}^{11}(\omega), p_3^1(\omega) - p_{23}^{11}(\omega)\}. \qquad (7.22)
\end{aligned}$$

Proof. Since $p_{13}^{11}(\omega) = p_{123}^{111}(\omega) + p_{123}^{101}(\omega) = A(\omega) + B(\omega)$ in Lemma 7.16, by (7.19) and (7.20) we can easily get (7.21) and (7.22). This completes the proof. □

Remark 7.18. [Comparison]. Let us compare the result in Lemma 7.17 with the result (7.4) in Lemma 7.3 (i.e., the result without consistent condition). Note that (7.4) implies

$$C_1 \equiv \max\{0, p_1^1(\omega) + p_3^1(\omega) - 1\} \leq p_{13}^{11}(\omega) \leq \min\{p_1^1(\omega), p_3^1(\omega)\} \equiv C_2.$$

Here we can easily see that $C_1 \leq$ (7.21) and (7.22) $\leq C_2$ from the following trivial inequalities:

$$\max\{0, \alpha_1 + \alpha_2\} \leq \max\{0, \max\{0, \alpha_1\} + \max\{0, \alpha_2\}\} = \max\{0, \alpha_1\} + \max\{0, \alpha_2\}$$

and

$$\begin{aligned}
&\min\{\alpha_1, \alpha_2\} + \min\{\alpha_3, \alpha_4\} = \min\{\ \alpha_1 + \alpha_3,\ \alpha_1 + \alpha_4,\ \alpha_2 + \alpha_3,\ \alpha_2 + \alpha_4\ \} \\
&\leq \min\{\ \alpha_1 + \alpha_3,\ \alpha_2 + \alpha_4\ \}\ .
\end{aligned}$$

Therefore, we see in Lemma 7.17 that the value $p_{13}^{11}(\omega)$ is restricted under the consistent condition of $[\mathbf{O}_{12}, \mathbf{O}_{23}]$.

■

7.3.2 Practical syllogism

Now we show several theorems of practical syllogisms (i.e., theorems concerning "implication" in Definition 7.6) as the consequences of our arguments.

Theorem 7.19. [Practical syllogism, [41]]. *Assume Notation 7.15 for* $K = \{1, 2, 3\}$. *That is,* $[\mathbf{O}_{12}, \mathbf{O}_{23}]$ *is a covering family of observables in a commutative* C^**-algebra* $C(\Omega)$.

Let $\delta_{\omega_0} \in \mathcal{M}^p_{+1}(\Omega)$ *for any fixed* $\omega_0 \in \Omega$*. Let* $\mathbf{O}_{123}$ $(= \mathbf{O}_K)$ *be any observable such that* $[\mathbf{O}_{12}, \mathbf{O}_{23}] \sqsubset \mathbf{O}_{123}$ *and let* $\mathbf{O}_{13} = \mathbf{O}_{123}\big|_{\{1,3\}}$*. (The existence of* $\mathbf{O}_{123}$ *is guaranteed by Theorem 7.13.) Then we have the following statements* [1] $\sim$ [3]:

[1]. *Assume that*

$$\mathbf{O}_1^{\{y_1\}} \underset{\mathbf{M}_{C(\Omega)}(\mathbf{O}_{12}, S_{[\delta_{[\omega_0]}]})}{\Longrightarrow} \mathbf{O}_2^{\{y_2\}}, \qquad \mathbf{O}_2^{\{y_2\}} \underset{\mathbf{M}_{C(\Omega)}(\mathbf{O}_{23}, S_{[\delta_{\omega_0}]})}{\Longrightarrow} \mathbf{O}_3^{\{y_3\}}. \tag{7.23}$$

Then, we see that

$$\begin{bmatrix} p_{13}^{11}(\omega_0) & p_{13}^{10}(\omega_0) \\ p_{13}^{01}(\omega_0) & p_{13}^{00}(\omega_0) \end{bmatrix} = \begin{bmatrix} p_1^1(\omega_0) & 0 \\ p_3^1(\omega_0) - p_1^1(\omega_0) & 1 - p_3^1(\omega_0) \end{bmatrix}, \tag{7.24}$$

hence, we see that

$$\mathbf{O}_1^{\{y_1\}} \underset{\mathbf{M}_{C(\Omega)}(\mathbf{O}_{13}, S_{[\delta_{\omega_0}]})}{\Longrightarrow} \mathbf{O}_3^{\{y_3\}}. \tag{7.25}$$

[2]. *Assume that*

$$\mathbf{O}_1^{\{y_1\}} \underset{\mathbf{M}_{C(\Omega)}(\mathbf{O}_{12}, S_{[\delta_{\omega_0}]})}{\Longleftarrow} \mathbf{O}_2^{\{y_2\}}, \qquad \mathbf{O}_2^{\{y_2\}} \underset{\mathbf{M}_{C(\Omega)}(\mathbf{O}_{23}, S_{[\delta_{\omega_0}]})}{\Longrightarrow} \mathbf{O}_3^{\{y_3\}}. \tag{7.26}$$

Then, we see that

$$\begin{bmatrix} p_{13}^{11}(\omega_0) & p_{13}^{10}(\omega_0) \\ p_{13}^{01}(\omega_0) & p_{13}^{00}(\omega_0) \end{bmatrix} = \begin{bmatrix} \alpha(\omega_0) & p_1^1(\omega_0) - \alpha(\omega_0) \\ p_3^1(\omega_0) - \alpha(\omega_0) & 1 + \alpha(\omega_0) - p_1^1(\omega_0) - p_3^1(\omega_0) \end{bmatrix}$$

where

$$\max\{p_2^1(\omega_0), p_1^1(\omega_0) + p_3^1(\omega_0) - 1\} \le \alpha(\omega_0) \le \min\{p_1^1(\omega_0), p_3^1(\omega_0)\}. \tag{7.27}$$

Also (7.26) is equivalent to

$$\mathbf{O}_2^{\{y_2\}} \underset{\mathbf{M}_{C(\Omega)}(\mathbf{O}_{123}, S_{[\delta_{\omega_0}]})}{\Longrightarrow} \mathbf{O}_{13}^{\{(y_1,y_3)\}}. \tag{7.28}$$

[3]. *Assume that*

$$\mathbf{O}_1^{\{y_1\}} \underset{\mathbf{M}_{C(\Omega)}(\mathbf{O}_{12}, S_{[\delta_{\omega_0}]})}{\Longrightarrow} \mathbf{O}_2^{\{y_2\}}, \qquad \mathbf{O}_2^{\{y_2\}} \underset{\mathbf{M}_{C(\Omega)}(\mathbf{O}_{23}, S_{[\delta_{\omega_0}]})}{\Longleftarrow} \mathbf{O}_3^{\{y_3\}}. \tag{7.29}$$

Then, we see that

$$\begin{bmatrix} p_{13}^{11}(\omega_0) & p_{13}^{10}(\omega_0) \\ p_{13}^{01}(\omega_0) & p_{13}^{00}(\omega_0) \end{bmatrix} = \begin{bmatrix} \alpha(\omega_0) & p_1^1(\omega_0) - \alpha(\omega_0) \\ p_3^1(\omega_0) - \alpha(\omega_0) & 1 + \alpha(\omega_0) - p_1^1(\omega_0) - p_3^1(\omega_0) \end{bmatrix}$$

where

$$\max\{0, p_1^1(\omega_0) + p_3^1(\omega_0) - p_2^1(\omega_0)\} \le \alpha(\omega_0) \le \min\{p_1^1(\omega_0), p_3^1(\omega_0)\}. \tag{7.30}$$

Also (7.29) is equivalent to

$$\mathbf{O}_{13}^{\{(y_1,y_3),(y_1,n_3),(n_1,y_3)\}} \underset{\mathbf{M}_{C(\Omega)}(\mathbf{O}_{123}, S_{[\delta_{\omega_0}]})}{\Longrightarrow} \mathbf{O}_2^{\{y_2\}}. \tag{7.31}$$

Proof. [1]. By (7.23) and (7.5), we see that $p_{12}^{10} = p_{23}^{10} = 0$, so, $p_{12}^{11} = p_1^1 \le p_2^1 = p_{23}^{11} \le p_3^1$. Therefore, we see that (7.21) $= p_{12}^{11} + \max\{0, p_3^1 - 1\} = p_1^1$. And (7.22) $= p_1^1 + 0 = p_1^1$. This implies that $p_{13}^{11} = p_1^1$, i.e., (7.24). Also, (7.25) follows from $p_{13}^{10} = 0$.
[2]. By (7.26) and (7.5), we see that $p_{12}^{01} = p_{23}^{10} = 0$, so, $p_{12}^{11} = p_2^1 \le p_1^1$ and $p_{23}^{11} = p_2^1 \le p_3^1$. Therefore, we see that (7.21) $= p_{23}^{11} + \max\{0, p_1^1 - p_2^1 + p_3^1 - 1\} = \max\{p_2^1, p_1^1 + p_3^1 - 1\}$. And (7.22) $= \min\{p_2^1, p_2^1\} + \min\{p_1^1 - p_2^1, p_3^1 - p_2^1\} = \min\{p_1^1, p_3^1\}$. This implies (7.27). Also, we see that (7.26) $\Leftrightarrow p_{12}^{01} = p_{23}^{10} = 0 \Leftrightarrow p_{123}^{010} = p_{123}^{011} = p_{123}^{110} = 0 \Leftrightarrow$ (7.28).
[3]. By (7.29) and (7.5), we see that $p_{12}^{10} = p_{23}^{01} = 0$, so, $p_{12}^{11} = p_1^1 \le p_2^1$ and $p_{23}^{11} = p_3^1 \le p_2^1$. Therefore, we see that (7.21) $= \max\{0, p_1^1 - p_2^1 + p_3^1\} + \max\{0, p_2^1 - 1\} = \max\{0, p_1^1 - p_2^1 + p_3^1\}$. And (7.22) $= \min\{p_1^1, p_3^1\}$. This implies (7.30). Also, (7.29) $\Leftrightarrow p_{12}^{10} = p_{23}^{01} = 0 \Leftrightarrow p_{123}^{101} = p_{123}^{100} = p_{123}^{001} = 0 \Leftrightarrow$ (7.31). This completes the proof. □

Remark 7.20. [Practical logic and pure logic]. The reader must not confuse the result (for example, (7.23)⇒(7.25)) in Theorem 7.19 with pure logic (i.e., mathematical logic). Theorem 7.19 is a consequence of Axiom 1. Note that Theorem 7.19 is due to Theorem 7.13, i.e., the commutativity of C^*-algebra $C(\Omega)$. That means the results in Theorem 7.19 can not be expected in quantum systems. In comparison with quantum theory, Theorem 7.19 becomes clearer. For example, in general, the syllogism is meaningless in quantum systems. This is easily shown as follows. Put $V = \mathbf{C}^5$, and $\mathcal{A} = B(\mathbf{C}^5)$. And

$$\vec{e}_1 = \begin{bmatrix}1\\0\\0\\0\\0\end{bmatrix}, \quad \vec{e}_2 = \begin{bmatrix}0\\1\\0\\0\\0\end{bmatrix}, \quad \vec{e}_3 = \begin{bmatrix}0\\0\\1\\0\\0\end{bmatrix}, \quad \vec{e}_4 = \begin{bmatrix}0\\0\\0\\1\\0\end{bmatrix}, \quad \vec{e}_5 = \begin{bmatrix}0\\0\\0\\0\\1\end{bmatrix},$$

and put $\vec{f}_4 = \frac{\vec{e}_4}{\sqrt{2}} + \frac{\vec{e}_5}{\sqrt{2}}$, $\vec{f}_5 = \frac{\vec{e}_4}{\sqrt{2}} - \frac{\vec{e}_5}{\sqrt{2}}$. Define the three observables $\mathbf{O}_1 \equiv (X_1 \equiv \{a_1, b_1, c_1\}, 2^{X_1}, F_1)$, $\mathbf{O}_2 \equiv (X_2 \equiv \{a_2, b_2, c_2\}, 2^{X_2}, F_2)$ and $\mathbf{O}_3 \equiv (X_3 \equiv \{a_3, b_3, c_3\}, 2^{X_3}, F_3)$ such that

$$F_1(\{a_1\}) = |\vec{e}_1\rangle\langle\vec{e}_1|, \quad F_1(\{b_1\}) = |\vec{e}_2\rangle\langle\vec{e}_2| + |\vec{e}_3\rangle\langle\vec{e}_3| + |\vec{e}_4\rangle\langle\vec{e}_4|, \quad F_1(\{c_1\}) = |\vec{e}_5\rangle\langle\vec{e}_5|,$$

$$F_2(\{a_2\}) = |\vec{e}_1\rangle\langle\vec{e}_1| + |\vec{e}_2\rangle\langle\vec{e}_2|, \quad F_2(\{b_2\}) = |\vec{e}_3\rangle\langle\vec{e}_3|, \quad F_2(\{c_2\}) = |\vec{e}_4\rangle\langle\vec{e}_4| + |\vec{e}_5\rangle\langle\vec{e}_5|,$$

$$F_3(\{a_3\}) = |\vec{e}_1\rangle\langle\vec{e}_1| + |\vec{e}_2\rangle\langle\vec{e}_2| + |\vec{e}_3\rangle\langle\vec{e}_3|, \quad F_3(\{b_3\}) = |\vec{f}_4\rangle\langle\vec{f}_4|, \quad F_3(\{c_3\}) = |\vec{f}_5\rangle\langle\vec{f}_5|.$$

Note that $\mathbf{O}_1$ and $\mathbf{O}_2$ [resp. $\mathbf{O}_2$ and $\mathbf{O}_3$] commute. Let $\mathbf{O}_{12} = (X_1 \times X_2, 2^{X_1} \times {}^{X_2}, F_1 \times F_2)$ be the product observable of $\mathbf{O}_1$ and $\mathbf{O}_2$. And let $\mathbf{O}_{23} = (X_2 \times X_3, 2^{X_2} \times {}^{X_3}, F_2 \times F_3)$ be the product observable of $\mathbf{O}_2$ and $\mathbf{O}_3$. Let ρ^p be any pure state (i.e., $\rho^p \in \mathfrak{S}^p(B(\mathbf{C}^5)^*)$). Then, we have

$$\mathbf{O}_1^{\{a_1\}} \underset{\mathbf{M}_{\mathcal{A}}(\mathbf{O}_{12}, S_{[\rho^p]})}{\Longrightarrow} \mathbf{O}_2^{\{a_2\}}, \quad \mathbf{O}_2^{\{a_2\}} \underset{\mathbf{M}_{\mathcal{A}}(\mathbf{O}_{23}, S_{[\rho^p]})}{\Longrightarrow} \mathbf{O}_3^{\{a_3\}}.$$

since we see

$$\rho^p\Big((F_1 \times F_2)(\{a_1\} \times (\{b_2, c_2\}))\Big) = 0, \quad \rho^p\Big((F_1 \times F_2)(\{a_2\} \times (\{b_3, c_3\}))\Big) = 0.$$

However, it should be noted that we have no product observable of $\mathbf{O}_1$, $\mathbf{O}_2$ and $\mathbf{O}_3$. Thus, the implication:

$$\mathbf{O}_1^{\{a_1\}} \underset{\mathbf{M}_{\mathcal{A}}(\mathbf{O}_{13}, S_{[\rho^p]})}{\Longrightarrow} \mathbf{O}_3^{\{a_3\}}$$

is nonsense since $\mathbf{O}_{13}$ can not be defined.

■

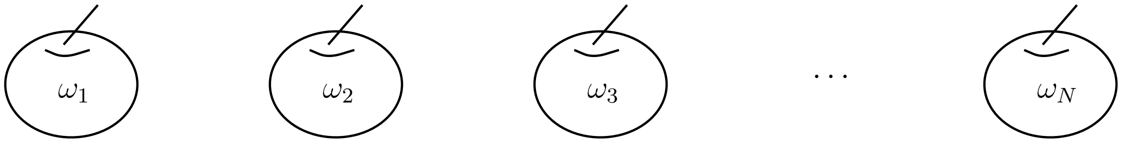
Example 7.21. [Continued from Example 7.4, [41]]. Let Ω, $C(\Omega)$, $\mathbf{O}_1 \equiv \mathbf{O}_{\rm SW} \equiv (X_{\rm SW}, 2^{X_{\rm SW}}, F_{\rm SW})$ and $\mathbf{O}_3 \equiv \mathbf{O}_{\rm RD} \equiv (X_{\rm RD}, 2^{X_{\rm RD}}, F_{\rm RD})$ be as in Example 7.4. Let $\mathbf{O}_2 \equiv \mathbf{O}_{\rm RP} \equiv (X_{\rm RP}, 2^{X_{\rm RP}}, F_{\rm RP})$ be an observable in $C(\Omega)$ such that:

$$X_{\rm RP} = \{y_{\rm RP}, n_{\rm RP}\},$$

where "$y_{\rm RP}$" and "$n_{\rm RP}$" respectively mean "RIPE" and "NOT RIPE". Put

$$\begin{aligned}
{\rm Rep}[\mathbf{O}_1] &= \Big[[F_{\rm SW}(\{y_{\rm SW}\})](\omega), [F_{\rm SW}(\{n_{\rm SW}\})](\omega)\Big], \\
{\rm Rep}[\mathbf{O}_2] &= \Big[[F_{\rm RP}(\{y_{\rm RP}\})](\omega), [F_{\rm RP}(\{n_{\rm RP}\})](\omega)\Big], \\
{\rm Rep}[\mathbf{O}_3] &= \Big[[F_{\rm RD}(\{y_{\rm RD}\})](\omega), [F_{\rm RD}(\{n_{\rm RD}\})](\omega)\Big].
\end{aligned}$$

For example,



$n_{\rm SW}$	$y_{\rm SW}$	$y_{\rm SW}$	$\cdots$	$n_{\rm SW}$
$y_{\rm RP}$	$n_{\rm RP}$	$y_{\rm RP}$	$\cdots$	$n_{\rm RP}$
$y_{\rm RD}$	$y_{\rm RD}$	$n_{\rm RD}$	$\cdots$	$n_{\rm RD}$

Consider the following quasi-product observables:

$$\mathbf{O}_{12} = (X_{\rm SW} \times X_{\rm RP}, 2^{X_{\rm SW} \times X_{\rm RP}}, F_{12} \equiv F_{\rm SW} \overset{\mathbf{O}_{12}}{\times} F_{\rm RP})$$

and

$$\mathbf{O}_{23} = (X_{\rm RP} \times X_{\rm RD}, 2^{X_{\rm RP} \times X_{\rm RD}}, F_{23} \equiv F_{\rm RP} \overset{\mathbf{O}_{23}}{\times} F_{\rm RD}).$$

Let $\delta_{\omega_n} \in \mathcal{M}^p_{+1}(\Omega)$ for any fixed $\omega_n \in \Omega$. Assume that

$$\mathbf{O}_1^{\{y_1\}} \underset{\mathbf{M}_{C(\Omega)}(\mathbf{O}_{12}, S_{[\delta_{\omega_0}]})}{\Longrightarrow} \mathbf{O}_2^{\{y_2\}}, \qquad \mathbf{O}_2^{\{y_2\}} \underset{\mathbf{M}_{C(\Omega)}(\mathbf{O}_{23}, S_{[\delta_{\omega_0}]})}{\Longrightarrow} \mathbf{O}_3^{\{y_3\}}. \tag{7.32}$$

Then, we see, by Theorem 7.19 [1], that

$$\mathrm{Rep}[\mathbf{O}_{13}] = \begin{bmatrix} [F_{13}(\{y_{\rm SW}\} \times \{y_{\rm RD}\})](\omega_n) & [F_{13}(\{y_{\rm SW}\} \times \{n_{\rm RD}\})](\omega_n) \\ [F_{13}(\{n_{\rm SW}\} \times \{y_{\rm RD}\})](\omega_n) & [F_{13}(\{n_{\rm SW}\} \times \{n_{\rm RD}\})](\omega_n) \end{bmatrix} \tag{7.33}$$

$$= \begin{bmatrix} [F_{\rm SW}(\{y_{\rm SW}\})](\omega_n) & 0 \\ [F_{\rm RD}(\{y_{\rm RD}\})](\omega_n) - [F_{\rm SW}(\{y_{\rm SW}\})](\omega_n) & 1 - [F_{\rm RD}(\{y_{\rm RD}\})](\omega_n) \end{bmatrix}.$$

So, when we observe that the tomato ω_n is "RED", we can infer, by the fuzzy inference $\mathbf{M}_{C(\Omega)}(\mathbf{O}_{13}, S_{[\delta_{\omega_n}]})$ (equivalently, $\mathbf{M}_{C(\Omega)}(\mathbf{O}_{31}, S_{[\delta_{\omega_n}]})$), the probability that the tomato ω_n is "SWEET" is given by

$$\frac{[F_{13}(\{y_{\rm SW}\} \times \{y_{\rm RD}\})](\omega_n)}{[F_{13}(\{y_{\rm SW}\} \times \{y_{\rm RD}\})](\omega_n) + [F_{13}(\{n_{\rm SW}\} \times \{y_{\rm RD}\})](\omega_n)} = \frac{[F_{\rm SW}(\{y_{\rm SW}\})](\omega_n)}{[F_{\rm RD}(\{y_{\rm RD}\})](\omega_n)}. \tag{7.34}$$

Also, when we observe that the tomato ω_n is "SWEET", we can infer, by the fuzzy inference $\mathbf{M}_{C(\Omega)}(\mathbf{O}_{13}, S_{[\delta_{\omega_n}]})$, the probability that the tomato ω_n is "RED" is given by

$$\frac{[F_{13}(\{y_{\rm SW}\} \times \{y_{\rm RD}\})](\omega_n)}{[F_{13}(\{y_{\rm SW}\} \times \{y_{\rm RD}\})](\omega_n) + [F_{13}(\{y_{\rm SW}\} \times \{n_{\rm RD}\})](\omega_n)} = \frac{[F_{\rm RD}(\{y_{\rm RD}\})](\omega_n)}{[F_{\rm RD}(\{y_{\rm RD}\})](\omega_n)} = 1. \tag{7.35}$$

Note that (7.32) implies (and is implied by)

$$\text{"SWEET"} \Longrightarrow \text{"RIPE"} \qquad \text{and} \qquad \text{"RIPE"} \Longrightarrow \text{"RED"} . \tag{7.36}$$

(Recall (7.8)). So, it is "reasonable" to reach the conclusion:

$$\text{"SWEET"} \Longrightarrow \text{"RED"} , \tag{7.37}$$

which is implied by the above (7.35). $\Big($Here we are afraid that the most important fact may be overlooked. For completeness, note that the conclusion "(7.36) $\Rightarrow$ (7.37)" is a consequence of Theorem 7.19 (and therefore, our axiom).$\Big)$ However, the result (7.34) is due to the peculiarity of fuzzy inferences. That is, in spite of the fact (7.36), we get the conclusion (7.34) that is somewhat like

$$\text{"RED"} \Longrightarrow \text{"SWEET"} . \tag{7.38}$$

Note that the conclusion (7.37) is not valuable in the market. What we want in the market is the conclusion such as (7.38) (or (7.34)).

■

Example 7.22. [Continued from Example 7.21, [41]]. Instead of (7.32), assume that

$$\mathbf{O}_1^{\{y_1\}} \underset{\mathbf{M}_{C(\Omega)}(\mathbf{O}_{12}, S_{[\delta_{\omega_n}]})}{\Longleftarrow} \mathbf{O}_2^{\{y_2\}}, \qquad \mathbf{O}_2^{\{y_2\}} \underset{\mathbf{M}_{C(\Omega)}(\mathbf{O}_{23}, S_{[\delta_{\omega_n}]})}{\Longrightarrow} \mathbf{O}_3^{\{y_3\}}. \tag{7.39}$$

Assume the notation (7.33). When we observe that the tomato ω_n is "RED", we can infer, by the fuzzy inference $\mathbf{M}_{C(\Omega)}(\mathbf{O}_{13}, S_{[\delta_{\omega_n}]})$, the probability that the tomato ω_n is "SWEET" is given by

$$Q = \frac{[F_{13}(\{y_{\rm SW}\} \times \{y_{\rm RD}\})](\omega_n)}{[F_{13}(\{y_{\rm SW}\} \times \{y_{\rm RD}\})](\omega_n) + [F_{13}(\{n_{\rm SW}\} \times \{y_{\rm RD}\})](\omega_n)} \tag{7.40}$$

which is, by (7.27), estimated as follows:

$$\begin{aligned} &\max\left\{\frac{[F_{\rm RP}(\{y_{\rm RP}\})](\omega_n)}{[F_{\rm RD}(\{y_{\rm RD}\})](\omega_n)}, \frac{[F_{\rm SW}(\{y_{\rm SW}\})] + [F_{\rm RD}(\{y_{\rm RD}\})] - 1}{[F_{\rm RD}(\{y_{\rm RD}\})](\omega_n)}\right\} \\ &\le Q \le \min\{\frac{[F_{\rm SW}(\{y_{\rm SW}\})](\omega_n)}{[F_{\rm RD}(\{y_{\rm RD}\})](\omega_n)}, 1\}. \end{aligned} \tag{7.41}$$

Note that (7.39) implies (and is implied by)

$$\text{"RIPE"} \Longrightarrow \text{"SWEET"} \qquad \text{and} \qquad \text{"RIPE"} \Longrightarrow \text{"RED"} . \tag{7.42}$$

And note that the conclusion (7.41) is somewhat like

$$\text{"RED"} \Longrightarrow \text{"SWEET"} . \tag{7.43}$$

Therefore, this conclusion is peculiar to "fuzziness".

■

The following theorem is a generalization of the first part of Theorem 7.19.

Theorem 7.23. [Standard syllogism, *cf.* [41]]. *Assume Notation 7.15. Let* $\delta_{\omega_0} \in \mathcal{M}_{+1}^p(\Omega)$. *Assume that*

$$\mathbf{O}_k^{\{y_k\}} \underset{\mathbf{M}_{C(\Omega)}(\mathbf{O}_{k,k+1}, S_{[\delta_{\omega_0}]})}{\Longrightarrow} \mathbf{O}_{k+1}^{\{y_{k+1}\}} \qquad (\forall k = 1, 2, ..., |K| - 1), \tag{7.44}$$

Let $\mathbf{O}_K$ *be any observable as in Notation 7.15, i.e.,* $\mathcal{G} = [\mathbf{O}_{12}, \mathbf{O}_{23}, \mathbf{O}_{34}, ..., \mathbf{O}_{|K|-1,|K|}] \sqsubset \mathbf{O}_K$. *Put* $\mathbf{O}_{1,|K|} = \mathbf{O}_K\big|_{\{1,|K|\}}$. *Then, we see that*

$$\mathrm{Rep}[\mathbf{O}_{1,|K|}]at\,\omega_0 = \begin{bmatrix} p_{1,|K|}^{11}(\omega_0) & p_{1,|K|}^{10}(\omega_0) \\ p_{1,|K|}^{01}(\omega_0) & p_{1,|K|}^{00}(\omega_0) \end{bmatrix} = \begin{bmatrix} p_1^1(\omega_0) & 0 \\ p_{|K|}^1(\omega_0) - p_1^1(\omega_0) & 1 - p_{|K|}^1(\omega_0) \end{bmatrix}, \tag{7.45}$$

hence, we see that

$$\mathbf{O}_1^{\{y_1\}} \underset{\mathbf{M}_{C(\Omega)}(\mathbf{O}_{1,|K|}, S_{[\delta_{\omega_0}]})}{\Longrightarrow} \mathbf{O}_{|K|}^{\{y_{|K|}\}}. \tag{7.46}$$

Proof. Let $\mathbf{O}_K$ be any observable such that $\mathcal{G} = [\mathbf{O}_{12}, \mathbf{O}_{23}, \mathbf{O}_{34}, ..., \mathbf{O}_{|K|-1,|K|}] \sqsubset \mathbf{O}_K$. Thus, we see that $[\mathbf{O}_K\big|_{\{1,3\}}, \mathbf{O}_{34}, ..., \mathbf{O}_{|K|-1,|K|}] \sqsubset \mathbf{O}_K\big|_{K\setminus\{2\}}$. Note that $(\mathbf{O}_K\big|_{\{1,3\}})\big|_{\{m\}} = \mathbf{O}_m$, $m = 1, 3$. Also note, by (7.24), that

$$\mathrm{Rep}[\mathbf{O}_K\big|_{\{1,3\}}]\mathrm{at}\,\omega_0 = \begin{bmatrix} p_1^1(\omega_0) & 0 \\ p_3^1(\omega_0) - p_1^1(\omega_0) & 1 - p_3^1(\omega_0) \end{bmatrix},$$

and therefore $\mathbf{O}_1^{\{y_1\}} \underset{\mathbf{M}_{C(\Omega)}(\mathbf{O}_K|_{\{1,3\}}, S_{[\delta_{\omega_0}]})}{\Longrightarrow} \mathbf{O}_3^{\{y_3\}}$. Hence, by induction, we see that $\mathrm{Rep}[\mathbf{O}_{1,|K|}] \equiv \mathrm{Rep}[\mathbf{O}_K\big|_{\{1,|K|\}}] =$ (7.45) at $\omega = \omega_0$. This completes the proof. □

7.4 Conclusion

It is certain that (pure) logic is merely a kind of rule in mathematics. However, if it is so, the logic is not guaranteed to be applicable to our world. For instance, (pure) logic does not assure the truth of the following famous statement:

[♯] *Since Socrates is a man and all men are mortal, it follows that Socrates is mortal.*

That is, we think that the problem: "Is this [♯] (theoretical) true or not?" is unsolved. Thus, the purpose of this chapter was to prove the [♯], or more generally, to propose "practical logic", i.e., a collection of theorems (whose forms are similar to that of "pure logic") in PMT.

Firstly, the symbol "$A \Rightarrow B$" (i.e., "implication") is defined in terms of measurements (*cf.* Definition 7.6). And we prove the standard syllogism for classical systems:

$$\text{"}A \Rightarrow B,\ B \Rightarrow C\text{" implies "}A \Rightarrow C\text{"}, \tag{7.47}$$

which is the same as the above (♯). (This (7.47) is not trivial since it does not necessarily hold in quantum systems.) We can assert, by "Declaration (1.11)" in §1.4, that PMT guarantees that the above statement [♯] is true.

Several variants may be interesting. For example, under the condition that "$A \Rightarrow B$, $B \Rightarrow C$", we can assert a kind of conclusion such as "$C \Rightarrow A$". That is,

$$\text{"}A \Rightarrow B,\ B \Rightarrow C\text{" implies "}C \Rightarrow A\text{"} \qquad \text{in some sense.} \tag{7.48}$$

For completeness, "pure logic" and "practical logic" must not be confused. The former is a basic rule on which mathematics is founded. On the other hand, the latter is a collection of theorems (whose forms are similar to that of "pure logic") in PMT.

7.5 Appendix (Zadeh's fuzzy sets theory)

7.5.1 What is Zadeh's fuzzy sets theory?

As mentioned in Chapter 1 (i.e., the footnote below Problem 1.2), one of motivations of our research is motivated by Zadeh's fuzzy sets theory. In 1965, L.A. Zadeh proposed a certain system theory, in which *a membership function* $f : \Omega \to [0,1]$, which is asserted to represent "fuzziness", plays an important role. The membership function is considered as a kind of generalization of a characteristic function. Here, the characteristic function χ_D of D ($\subseteq \Omega$) is defined by $\chi_D : \Omega \to \{0,1\}$ such that:

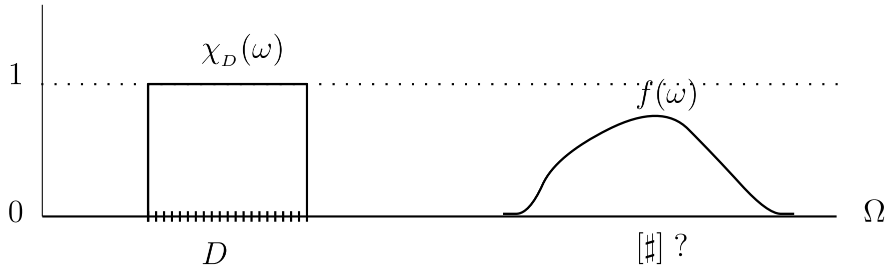
$$\chi_D(\omega) = \begin{cases} 1 & (\omega \in D) \\ 0 & (\omega \notin D). \end{cases}$$

Consider the identification:

$$\text{“characteristic function } \chi_D\text{”} \quad \longleftrightarrow \quad \text{“set } D\text{”},$$

which gives us the question "What is the following [♯]?"

$$\text{“membership function } f\text{”} \quad \longleftrightarrow [\sharp].$$



The [♯] is called a *fuzzy set* by Zadeh. Thus we think that Zadeh's fuzzy sets theory has two aspects $[A_1]$ and $[A_2]$ as follows:

$$\text{Zadeh's fuzzy sets theory} \begin{cases} [A_1] : \text{membership functions (analytic aspect)}, \\ \\ [A_2] : \text{fuzzy sets (logical aspect)}. \end{cases} \tag{7.49}$$

Zadeh's fuzzy sets theory acquired a lot of believers. In fact, his paper [93] is one of the most cited papers in all fields of 20th century science. However, his theory seems "fuzzy" rather than "difficult". Thus, it is natural that the following problem arises:

$[\sharp_1]$ Is Zadeh's fuzzy sets theory true or not?

When we examine the problem, we are immediately confronted with the following problem:

$[\sharp_2]$ What is "true or not"? Or, if we want to assert "Zadeh's fuzzy sets theory is true [or not]", what do we say?

And when we study the problems $[\sharp_1]$ and $[\sharp_2]$, we immediately notice the fact that we have not yet the clear answer to even the question: "Is Fisher's statistics true or not?". [3] As mentioned in Chapter 1, our research starts from the above questions $[\sharp_1]$ and $[\sharp_2]$. And we conclude "Declaration (1.11)" in §1.4 as follows:

- MT is entitled to check all theories in theoretical informatics. In other words, we can, by using MT, introduce the criterion: "true or not" into theoretical informatics. That is, MT can be regarded as "the Construction of theoretical informatics".

[3] In Chapters 5 and 6, it is proved that Fisher's statistics is theoretically true.

Now, consider an observable $(X, 2^X, F)$ in $C(\Omega)$. Note that, for any Ξ ($\subseteq X$), $F(\Xi)$ is a membership function on Ω. Since $F(\Xi) \in C(\Omega)$, the $F(\Xi)$, of course, has various analytic aspects. Also, in this chapter we see that the membership function $F(\Xi)$ has various logical aspects. Thus, someone may conclude that Zadeh's fuzzy sets theory (i.e., the analytic aspect $[A_1]$ and the logical aspect $[A_2]$ in (7.49)) is understood in the framework of measurement theory, that is, Zadeh's fuzzy sets theory is true (*cf.* "Declaration (1.11)" in §1.4). We may agree with this opinion. In fact, these kinds of aspects $[A_1]$ and $[A_2]$ can not be found in the conventional formulation of system theory (*cf.* (1.2)) such as

$$\boxed{\text{"dyn. syst. theor."}} = \begin{cases} \frac{dx(t)}{dt} = f(x(t), u_1(t), t), \; x(0) = x_0 & \cdots \text{(state equation)}, \\ y(t) = g(x(t), u_2(t), t) & \text{(measurement equation)}. \end{cases} \tag{7.50}$$
(=(1.2))

That is because the conventional formulation (7.50) does not possess the concept of "observable in the sense of Definition 2.7".

The believers of Zadeh's fuzzy sets theory say too much (*cf.* [64]). And thus, we have no firm answer to the question: "What is the essence of Zadeh's theory?". If we can assume that:

($\sharp$) Zadeh wanted to assert that *DST* (7.50) *and "logic" are closely connected* $\Big($*or precisely, "logic" is one of the aspects of DST* (7.50)$\Big)$ *though the two are, in appearance, independent,*

then we can understand his assertion. That is because in this section we study "logic" in measurement theory, which is a kind of generalization of the system theory (7.50). This is our opinion for Zadeh's theory. Of course, there may be another opinion, that is, someone may assert that Zadeh said something much more than the ($\sharp$). If it is so, we may not understand his theory in the framework of measurement theory.

Recall the arguments in Chapter 1 (particularly, "Declaration (1.11)" in §1.4, tables (1.7) and (1.8)). Now, we have only two options, i.e.,

(i) Zadeh's fuzzy sets theory is characterized as the theory concerning membership functions in measurement theory.

(ii) Zadeh's fuzzy sets theory is not characterized in measurement theory. Thus another fundamental theory (*cf.* $\boxed{\text{The third mathematical scientific theory}}$ in (1.7)) should be proposed.

Although there is a possibility that (ii) is reasonable, that is, Zadeh's fuzzy sets theory may be understood in another fundamental theory (*cf.* The third mathematical scientific theory in (1.7)), we should note that the proposal of another fundamental theory is much more remarkable than the justification of Zadeh's fuzzy sets theory. Thus we choose the (i) even if the essential part of Zadeh's assertion (e.g., the scientific part asserted in [64]) can not be characterized in MT. Thus we conclude that Zadeh's assertion can not be completely understood in measurement theory, i.e.,

- Zadeh's assertion is not completely "theoretical true" (*cf.* Declaration 1.11), though practical logic somewhat has the property like "fuzzy set".

This is our present opinion.

7.5.2 Why is Zadeh's paper cited frequently?

Although we believe that the above argument in §7.5.1 is proper, it does not explain the reason why Zadeh's paper is cited frequently. As mentioned before, Zadeh's paper [93] is one of the most cited papers of all scientific papers. This is an established fact. This fact may imply that there is something interesting behind Zadeh's assertion. Thus, we think that the question "Why is Zadeh's paper cited frequently?" is more important than the question "What is Zadeh's fuzzy sets theory?". Thus we shall consider the question:

- Why does the term "fuzzy" look attractive?

We think that the reason is that Zadeh's spirit is regarded as *the antithesis of the myth: "Science must be exact, clear, strict, etc".* This myth seems to be due to Newtonian mechanics (and moreover, theoretical physics), which has been located in the center of all science. That is, we think that

- many people want another science, which is fuzzy, rough, vague, etc.

If it is so, we should recall Table 1.8 (in Chapter 1), which asserts mathematical science is classified as follows:

$$\begin{cases} \text{theoretical physics ('TOE')} & \text{..... exact mathematical science} \\ \text{theoretical informatics (measurement theory)} & \text{..... fuzzy mathematical science.} \end{cases} \tag{7.51}$$

If it is true, we can understand the reason why the term "fuzzy" was accepted widely. Thus we do not deny the following opinion:

(♯) "measurement theory" = "fuzzy theory". (Cf. [42].) Or, the attractive parts of Zadeh's assertions are mostly included in measurement theory.

That is because we believe that

(♭) Measurement theory is the very theory that represents the anti-spirit against the myth: "Science must be exact, clear, strict, etc".

In fact, the terms

- *fuzzy statement (cf. the footnote below Example 2.16), ready-made, useful or not, subjective, popularity, likes or dislikes, (in "Theoretical informatics of Table (1.8)")*

seem to belong to the category of "fuzziness". On the other hand, the terms

- *precise statement (cf. the footnote below Example 2.16), made to order, empirical true or not, objective, truth, (in "Theoretical informatics of Table (1.8)")*

obviously belong to the category of "exactness".

Chapter 8

Statistical measurements in C^*-algebraic formulation

As mentioned in the beginning of Chapter 2, measurement theory (MT) can be classified into two subjects, i.e., "(pure) measurement theory (PMT)" and "statistical measurement theory (SMT)". That is,

$$\text{MT (="measurement theory")} \begin{cases} \text{PMT (="(pure) measurement theory") in Chapters } 2 \sim 7 \\ \\ \text{SMT (="statistical measurement theory") in Chapters } 8 \sim \end{cases} \tag{8.1}$$

PMT is essential, and it is formulated as follows:

$$\text{PMT} = \underset{\text{[Axiom 1 (2.37)]}}{\text{measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} \quad \text{in } C^*\text{-algebra} . \qquad \begin{matrix}(8.2)\\ (=(1.4))\end{matrix}$$

Here it should be noted that the state ρ^p is always assumed to be pure, i.e., $\rho^p \in \mathfrak{S}^p(\mathcal{A}^*)$. In this chapter we study the statistical measurement for a *statistical state,* i.e., the measurement in the case that the state is distributed. The distribution (i.e., a statistical state) is represented by a mixed state ρ^m ($\in \mathfrak{S}^m(\mathcal{A}^*)$). The Statistical MT (i.e., SMT) is formulated as follows:

$$\text{SMT} = \underset{\text{[Proclaim 1 (8.10)]}}{\text{statistical measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} \quad \text{in } C^*\text{-algebra} , \tag{8.3}$$

where Proclaim 1 is characterized as follows:

$$\text{"Proclaim 1"} = \text{"Axiom 1"} + \underset{\text{(the probabilistic interpretation of mixed state)}}{\text{"statistical state"}} \tag{8.4}$$

Thus, the (8.3) is also rewritten such as

$$\text{SMT} = \underset{\text{(Axioms 1 and 2)}}{\text{PMT}} + \underset{\text{(the probabilistic interpretation of mixed state)}}{\text{"statistical state"}} \quad \text{in } C^*\text{-algebra} . \tag{8.5}$$

Therefore it should be noted that *there is no SMT without PMT.* Also, we add "belief measurement theory" in §8.6 and "principal components analysis" in §8.7.

8.1 Statistical measurements (C^*-algebraic formulation)

8.1.1 General theory of statistical measurements

Axiom 1 (proposed in §2.4) says that the measurement of an observable $\mathbf{O}\big(\equiv (X, \mathcal{F}, F)\big)$ for the system with the state ρ^p $(\in \mathfrak{S}^p(\mathcal{A}^*))$ induces the sample space $(X, \mathcal{F}, P(\cdot) \equiv \rho^p(F(\cdot)))$. That is, Axiom 1 says symbolically that:

$$\boxed{\begin{array}{c}\text{"observable"}\\ (X,\mathcal{F},F) \text{ in } \mathcal{A}\end{array}} \text{ and } \boxed{\begin{array}{c}\text{"state"}\\ \rho^p\in\mathfrak{S}^p(\mathcal{A}^*)\end{array}} \underset{\text{measurement}}{\Longrightarrow} \boxed{\begin{array}{c}\text{"sample space"}\\ (X,\mathcal{F},P(\cdot)\equiv\rho^p(F(\cdot)))\end{array}}.$$

Here it should be noted that the state must be always pure, i.e., $\rho^p \in \mathfrak{S}^p(\mathcal{A}^*)$ in Axiom 1. However we sometimes want to generalize the concept of "state", i.e., to introduce "statistical state", which is represented by a mixed state ρ^m $(\in \mathfrak{S}^m(\mathcal{A}^*))$. That is, we assert (in Proclaim 1 later) that

$$[\sharp] \quad \text{"statistical state"} = \underset{\text{(mathematics)}}{\text{"mixed state"}} + \text{"probabilistic interpretation".}$$

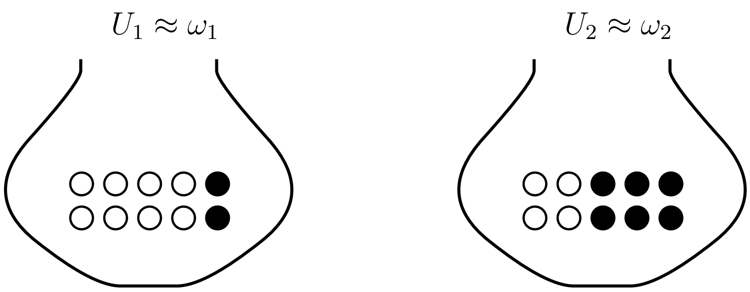
Also, it should be noted that we have already studied "S-states" in Chapter 6, which is one of the aspects of the statistical state. Although the statistical state has various aspects, we begin with the following example, which will promote a better understanding of the concept of "statistical state".

Example 8.1. [Coin-tossing and urn problem]. There are two urns U_1 and U_2. The urn U_1 [resp. U_2] contains 8 white and 2 black balls [resp. 4 white and 6 black balls]. Under the following identification (*cf.* (5.16) in Example 5.8):

$$U_1 \approx \omega_1, \qquad U_2 \approx \omega_2,$$

we regard Ω $\big(\equiv \{\omega_1, \omega_2\}\big)$ as the state space. And consider the observable $\mathbf{O}\big(\equiv (X \equiv \{w, b\}, 2^{\{w,b\}}, F)\big)$ in $C(\Omega)$ where

$$[F(\{w\})](\omega_1) = 0.8, \qquad [F(\{b\})](\omega_1) = 0.2,$$
$$[F(\{w\})](\omega_2) = 0.4, \qquad [F(\{b\})](\omega_2) = 0.6. \tag{8.6}$$



Here consider the following procedures (P_1) and (P_2).

(P_1) One of the two (i.e., ω_1 or ω_2) is chosen by an unfair tossed-coin ($C_{p,1-p}$), i.e.,

$$\text{Head } (100p\%) \to \omega_1, \ \text{Tail } (100(1-p)\%) \to \omega_2 \qquad (0 \le p \le 1). \tag{8.7}$$

The chosen urn is denoted by $[*](\in \{\omega_1, \omega_2\})$. Here define the mixed state $\nu_0(\in \mathcal{M}_{+1}^m(\Omega))$ such that $\nu_0 = p\delta_{\omega_1} + (1-p)\delta_{\omega_2}$ (i.e., $\nu_0(\{\omega_1\}) = p$, $\nu_0(\{\omega_2\}) = 1-p$), which is considered to be *"the distribution of* $[*]$*."* Thus we call the ν_0 *a statistical state.*

(P_2) Take one ball, at random, out of the urn chosen by the procedure (P_1). That is, we take the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$.

Then we have the following question:

(Q) Calculate the probability that a measured value "w" [resp. "b"] is obtained by the above measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$.

[*Answer*]. The "measurement" defined in the above (P_1) and (P_2) is denoted by

$$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}([\delta_{\omega_1}; p] \oplus [\delta_{\omega_2}; 1-p])). \tag{8.8}$$

This may be called a "probabilistic measurement", and the symbol $[\delta_{\omega_1}; p] \oplus [\delta_{\omega_2}; 1-p]$ may be called a "probabilistic state." Note that:

(i) the probability that $[\,*\,] = \delta_{\omega_1}$ [resp. $[\,*\,] = \delta_{\omega_2}$] is given by p [resp. $1-p$].

(ii) If $[\,*\,] = \delta_{\omega_1}$ [resp. if $[\,*\,] = \delta_{\omega_2}$], the probability that the measured value obtained by $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$ is equal to x ($\in \{w, b\}$) is, by Axiom 1, given by

$$\begin{aligned} &{}_{\mathcal{M}(\Omega)}\langle \delta_{\omega_1}, F(\{x\}) \rangle_{C(\Omega)} = 0.8 \ \ (\text{ if } x = w), \quad = 0.2 \ \ (\text{ if } x = b), \\ &\Big[\text{resp. } {}_{\mathcal{M}(\Omega)}\langle \delta_{\omega_2}, F(\{x\}) \rangle_{C(\Omega)} = 0.4 \ \ (\text{ if } x = w), \quad = 0.6 \ \ (\text{ if } x = b) \quad \Big]. \end{aligned}$$

Thus, under the condition (P_1), the probability that the measured value obtained by the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$ is equal to x ($\in \{w, b\}$) is given by

$$\begin{aligned} P(\{x\}) &= \int_\Omega {}_{\mathcal{M}(\Omega)}\langle \delta_\omega, F(\{x\}) \rangle_{C(\Omega)} \nu_0(d\omega) = {}_{\mathcal{M}(\Omega)}\langle \nu_0, F(\{x\}) \rangle_{C(\Omega)} \\ &= \begin{cases} 0.8p + 0.4(1-p) & (\text{ if } x = w), \\ 0.2p + 0.6(1-p)) & (\text{ if } x = b). \end{cases} \end{aligned}$$

This is the answer to the above question (Q). Summing up, we see:

($\sharp$) There is a reason that the "measurement" $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}([\delta_{\omega_1}; p] \oplus [\delta_{\omega_2}; 1-p]))$ is one of interpretations of the *"statistical measurement"* $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0))$, *(cf. Proclaim 1 (8.10) later).* Here the mixed state $\nu_0 (\in \mathcal{M}_{+1}^m(\Omega))$ is called a *"statistical state"*, which represents the distribution of $[\,*\,]$. And, the probability that the measured value x ($\in \{w, b\}$) is obtained by the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0))$, is given by

$$_{C(\Omega)^*}\Big\langle \nu_0, F(\{x\}) \Big\rangle_{C(\Omega)} \Big(\equiv \int_\Omega {}_{C(\Omega)^*}\Big\langle \delta_\omega, F(\{x\}) \Big\rangle_{C(\Omega)} \nu_0(d\omega) \Big).$$

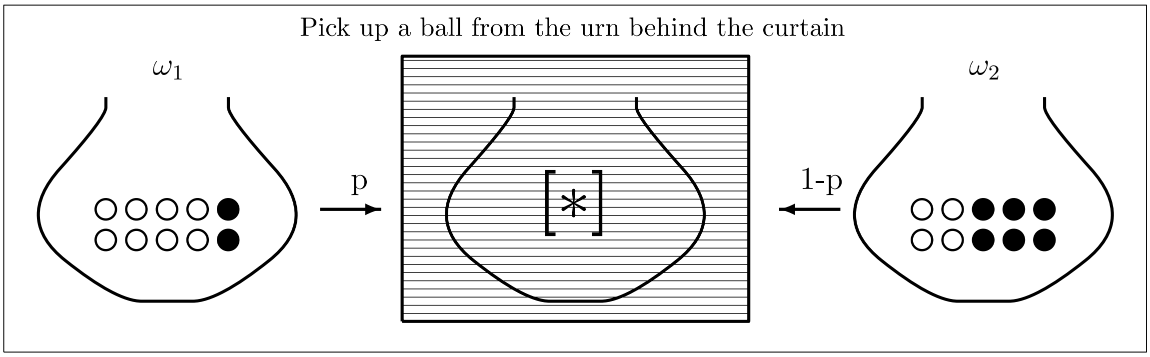
Thus we consider that

$$S_{[*]}([\delta_{\omega_1}; p] \oplus [\delta_{\omega_2}; 1-p])) \xrightleftharpoons[\text{statistical form}]{\text{probabilistic form}} S_{[*]}(\nu_0) \tag{8.9}$$

That is, the statistical state ν_0 is *the mixed state with probabilistic interpretation,* or, *the mixed state generated by coin-tossing.*

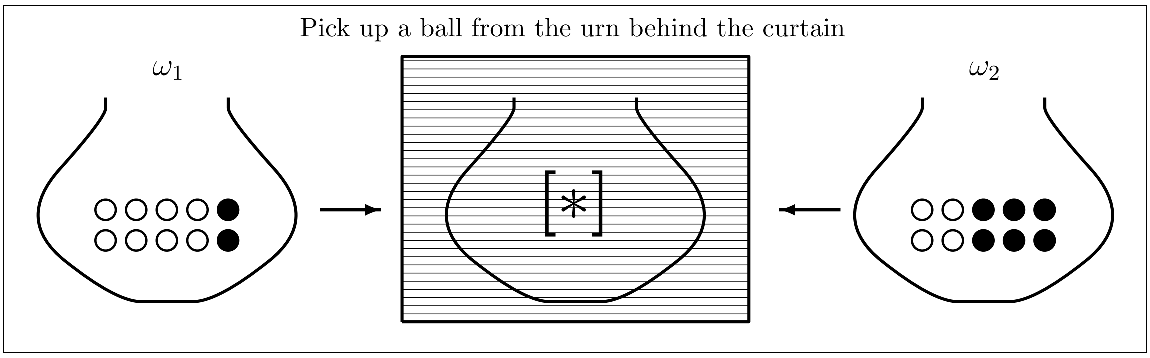
Thus, we see

The typical example of $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0))$



On the other hand, we recall that

The typical example of $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$



■

Now, we introduce "statistical measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[*]}(\rho^m)\ )$". The mixed state ρ^m (with the probabilistic interpretation) is called an *statistical state.* We propose the following "Proclaim 1", which should be read by the hint of the statement (♯) in Example 8.1.

PROCLAIM 1. [The probabilistic interpretation of mixed states, *cf.* [44]]. *Consider a statistical measurement* $\mathbf{M}_{\mathcal{A}}\big(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[*]}(\rho^m)\big)$ *formulated in a* C^**-algebra* $\mathcal{A}$*. Then, the probability that* x $(\in X)$*, the measured value obtained by the statistical measurement* $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[*]}(\rho^m))$*, belongs to a set* Ξ $(\in \mathcal{F})$ *is given by*

$$\rho^m(F(\Xi)) \Big(\equiv {}_{\mathcal{A}^*}\Big\langle \rho^m, F(\Xi) \Big\rangle_{\mathcal{A}} \Big).$$

The statistical measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[*]}(\rho^m))$ *is sometimes denoted by* $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S(\rho^m))$. (8.10)

That is, Proclaim 1[1] asserts that

$$[\sharp] \quad \text{"statistical state"} = \underset{\text{(mathematics)}}{\text{"mixed state"}} + \underset{\text{(such as coin-tossing)}}{\text{"probabilistic interpretation"}}\,^{2} \tag{8.11}$$

Note that the above "Proclaim 1" should be understood as

$$\text{"Proclaim 1"} = \text{"Axiom 1"} + \underset{\text{(the probabilistic interpretation of mixed state)}}{\text{"statistical state"}}$$

Therefore, the Statistical MT (i.e., SMT) is formulated as follows:

$$\begin{aligned} \text{SMT} &= \underset{\text{[Proclaim 1 (8.10)]}}{\text{statistical measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} \\ &= \underset{\text{(Axioms 1 and 2)}}{\text{PMT}} + \underset{\text{(the probabilistic interpretation of mixed state)}}{\text{"statistical state"}} \quad \text{in } C^*\text{-algebra.} \end{aligned}$$

Therefore, we stress:

- there is no SMT without PMT. (8.12)

Also, for the relation between PMT and SMT, see Remark 8.3 [hybrid measurement theory] later.

The following definition is the same as Definition 3.1. Here, it should be noted that "Markov relation among systems (i.e., $\{\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2)\in T^2_{\leqq}}$)" and "sequential

[1] Proclaim 1 is somewhat methodological. Thus, in [44], "Proclaim 1" was called "Method 1".

[2] As seen later (i.e., §8.7), Bertrand's paradox is due to the confusion between mixed states (mathematical concept) and statistical states (measurement theoretical concept). In order to avoid this confusion, it may be recommended to remember that there is always "coin-tossing" behind "statistical state".

observable (i.e., $[\{\mathbf{O}\}_{t \in T}, \{\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2)\in T^2_{\leq}}]$)" are common to both PMT and SMT. This implies that Axiom 2 is common to PMT and SMT.

Definition 8.2. [General systems in statistical measurements, *cf.* Definition 3.1]. *The pair* $\mathbf{S}_{[*]}(\rho^m_{t_0}) \equiv [S(\rho^m_{t_0}), \{\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2)\in T^2_{\leq}}]$ *is called a general system with an initial state* $S(\rho^m_{t_0})$ *if it satisfies the following conditions* (i)∼(iii).

(i) *With each* t $(\in T)$, *a* C^**-algebra* $\mathcal{A}_t$ *is associated.*

(ii) *Let* t_0 $(\in T)$ *be the root of* T. *And, assume that a system* S *has the state* $\rho^m_{t_0}$ $(\in \mathfrak{S}^m(\mathcal{A}^*_{t_0}))$ *at* t_0, *that is, the initial state is equal to* $\rho^p_{t_0}$.

(iii) *For every* $(t_1, t_2) \in T^2_{\leq}$, *Markov operator* $\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}$ *is defined such that* $\Phi_{t_1,t_2}\Phi_{t_2,t_3} = \Phi_{t_1,t_3}$ *holds for all* $(t_1,t_2), (t_2,t_3) \in T^2_{\leq}$.

The family $\{\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2)\in T^2_{\leq}}$ *is also called a "Markov relation among systems". Let an observable* $\mathbf{O}_t \equiv (X_t, \mathcal{F}_t, F_t)$ *in a* C^**-algebra* $\mathcal{A}_t$ *be given for each* $t \in T$. *The pair* $[\{\mathbf{O}\}_{t\in T}, \{\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2)\in T^2_{\leq}}]$ *is called a "sequential observable".*

■

Again note that Axiom 2 is common to PMT and SMT. Thus we see,

	measurements	relation among systems
PMT	Axiom 1 (2.37)	Axiom 2 (3.26)
SMT	Proclaim 1 (8.10)	Axiom 2 (3.26)

In what follows, we introduce some examples, which promote a better understanding of Proclaim 1. That is, readers will see that statistical states are not only generated by "coin-tossing" but also by several causes, for example, "Schrödinger picture", "Bayes theorem", etc.

Remark 8.3. [(i) Axiom 1 and Proclaim 1, hybrid measurement theory (= "HMT")]. For example, consider a pure state class $\mathfrak{S}^p(C(\Omega_1)^*))$ ($\equiv \mathcal{M}^p_{+1}(\Omega_1)$) in Axiom 1 and a mixed state class $\mathfrak{S}^m(C(\Omega_2)^*))$ ($\equiv \mathcal{M}^m_{+1}(\Omega_2)$) in Proclaim 1. Then we sometimes consider the tensor state class $\mathfrak{S}^p(C(\Omega_1)^*)) \otimes \mathfrak{S}^m(C(\Omega_2)^*))$, which is defined by

$$\Big\{\delta_{\omega_1} \otimes \rho^m_1 \in \mathcal{M}^m_{+1}(\Omega_1 \times \Omega_2) \;\Big|\; \omega_1 \in \Omega_1, \rho^m_2 \in \mathcal{M}^m_{+1}(\Omega_2)\Big\}.$$

This is called a *"hybrid state class".* In applications, we often devote ourselves to the *hybrid measurement theory* (= HMT).

[(ii) Axiom 1 and Proclaim 1, hybrid measurement theory]. For each $\mu(\in \mathbf{R})$, consider a mixed state $\rho_\mu^m(\in \mathcal{M}_{+1}^m(\mathbf{R}))$ such that

$$\rho_\mu^m(D) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_D \exp[-\frac{(\omega-\mu)^2}{2\sigma^2}]d\omega \quad (\forall D \in \mathcal{B}_{\mathbf{R}}, \text{ Borel field}),$$

where σ is a fixed positive number. Let $\mathbf{O} \equiv (X, \mathcal{F}, F)$ be an observable in $C_0(\mathbf{R})$. Then, we have the (statistical) measurement $\mathbf{M}_{C_0(\mathbf{R})}(\mathbf{O}, S(\rho_\mu^m))$. On the other hand, define the observable $\widehat{\mathbf{O}} = (X, \mathcal{F}, \widehat{F})$ in $C_0(\mathbf{R})$ such that:

$$[\widehat{F}(\Xi)](\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbf{R}} [F(\Xi)](\omega) \exp[-\frac{(\omega-\mu)^2}{2\sigma^2}]d\omega \quad (\forall \mu \in \mathbf{R}, \Xi \in \mathcal{F}).$$

Also note that

$$\begin{aligned} {}_{C_0(\mathbf{R})^*}\Big\langle \rho_\mu^m, F(\Xi) \Big\rangle_{C_0(\mathbf{R})} &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbf{R}} [F(\Xi)](\omega) \exp[-\frac{(\omega-\mu)^2}{2\sigma^2}]d\omega \\ &= {}_{C_0(\mathbf{R})^*}\Big\langle \delta_\mu, \widehat{F}(\Xi) \Big\rangle_{C_0(\mathbf{R})}, \end{aligned}$$

which urges us to consider the following identification:

$$\underset{\text{(statistical measurement)}}{\mathbf{M}_{C_0(\mathbf{R})}(\mathbf{O}, S(\rho_\mu^m))} \longleftrightarrow \underset{\text{(pure measurement)}}{\mathbf{M}_{C_0(\mathbf{R})}(\widehat{\mathbf{O}}, S_{[\delta_\mu]})}\ .$$

[(iii): Axiom 1 and Proclaim 1, hybrid measurement theory]. Let Λ_1 and Λ_2 be compact spaces (or compact index sets). For each $\lambda_1(\in \Lambda_1)$, consider a (parameterized) mixed state $\rho_{\lambda_1}^m(\in \mathcal{M}_{+1}^m(\Omega))$. And further, for each $\lambda_2(\in \Lambda_2)$, consider a parameterized observable $\mathbf{O}_{\lambda_2} \equiv (X, \mathcal{F}, F_{\lambda_2})$ in $C(\Omega)$. Then, we have the (statistical) measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}_{\lambda_2}, S(\rho_{\lambda_1}^m))$ in $C(\Omega)$. Define the observable $\widehat{\mathbf{O}} = (X, \mathcal{F}, \widehat{F})$ in $C(\Lambda_1 \times \Lambda_2)$ such that:

$$[\widehat{F}(\Xi)](\lambda_1, \lambda_2) = {}_{C(\Omega)^*}\Big\langle \rho_{\lambda_1}^m, F_{\lambda_2}(\Xi) \Big\rangle_{C(\Omega)} \quad (\forall(\lambda_1, \lambda_2) \in \Lambda_1 \times \Lambda_2, \Xi \in \mathcal{F}).$$

That is, we see

$${}_{C(\Omega)^*}\Big\langle \rho_{\lambda_1}^m, F_{\lambda_2}(\Xi) \Big\rangle_{C(\Omega)} = {}_{C(\Lambda_1\times\Lambda_2)^*}\Big\langle \delta_{(\lambda_1,\lambda_2)}, \widehat{F}(\Xi) \Big\rangle_{C(\Lambda_1\times\Lambda_2)},$$

which urges us to consider the following identification:

$$\underset{\text{(statistical measurement)}}{\mathbf{M}_{C(\Omega)}(\mathbf{O}_{\lambda_2}, S(\rho_{\lambda_1}^m))} \longleftrightarrow \underset{\text{(pure measurement)}}{\mathbf{M}_{C(\Lambda_1\times\Lambda_2)}(\widehat{\mathbf{O}}, S_{[\delta_{(\lambda_1,\lambda_2)}]})}\ . \tag{8.13}$$

Such an identification is often used in measurement theory. In this sense, the classification (8.1) should be considered to be flexible. ■

Remark 8.4. [Natural mixed state[3] and statistical state, Bertrand's paradox]. For example, consider the square $[0,1]\times[0,1]$ ($\subset \mathbf{R}^2$). This square has a natural measure m (which is usually called the Lebesgue measure) such that $m([a,b]\times[c.d]) = |b-a|\cdot|d-c|$ ($0 \le a \le b \le 1$ and $0 \le c \le d \le 1$). Here, it should be noted that m is a mixed state (i.e., $m \in \mathcal{M}_{+1}^m([0,1]\times[0,1])$), however, it is not a statistical state. That is, the natural mixed state is not always a statistical state. We should recall that there is no statistical state without the probabilistic interpretation (such as coin-tossing). This is just what Bertrand's paradox (*cf.* [35], also see §8.7 Appendix (Bertrand'd paradox)) teaches us. That is because Bertrand's paradox says that, if "the natural mixed state" is unreasonably regarded as "statistical state", we encounter a serious paradox (since a natural mixed state is not always unique). Also, recall Chapter 4 (Boltzmann's statistical mechanics), in which the normalized invariant measure is not regarded as "probability"[4] but "normalized staying time". (Continued to §8.7 Appendix (Bertrand's paradox))

■

8.1.2 Examples of statistical measurements

In Example 8.1, we showed "$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}([\delta_{\omega_1};p]\oplus[\delta_{\omega_2};1-p]))$" as the typical example of statistical measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0))$. In this section, we study the other typical examples.

The following example (Schrödinger picture) was already studied more precisely in Chapter 6.

Example 8.5. [(i): Schrödinger picture I]. Let $\Psi_{0,1} : \mathcal{A}_1 \to \mathcal{A}_0$ be a Markov operator. Let $\rho_0^p \in \mathfrak{S}^p(\mathcal{A}_0^*)$. That is, we consider the following general system:

$$\underset{\text{(pure) state}\rho_0^p}{[\mathcal{A}_0]} \xleftarrow{\Psi_{0,1}} [\mathcal{A}_1]. \tag{8.14}$$

Also, consider any observable $\mathbf{O}_1 \equiv (X_1, \mathcal{F}_1, F_1)$ in a C^*-algebra $\mathcal{A}_1$. And put $\widetilde{\mathbf{O}}_0 =$

[3]The "natural mixed state ρ" usually means the "invariant mixed state ρ" for some "natural"homomorphism $\Phi : \mathcal{A} \to \mathcal{A}$. That is, it holds that $\Phi^*(\rho) = \rho$.

[4]Such probability may be called "a priori probability". Thus we consider that the concept of "a priori probability" is nonsense.

$(X_1, \mathcal{F}_1, \Psi_{0,1}F_1)$. Thus we have the measurement

$$\mathbf{M}_{\mathcal{A}_0}(\widetilde{\mathbf{O}}_0, S_{[\rho_0^p]}).$$

Axiom 1 says that the measurement $\mathbf{M}_{\mathcal{A}_0}(\widetilde{\mathbf{O}}_0, S_{[\rho_0^p]})$ generates the sample space $(X_1, \mathcal{F}_1, P)$ such that:

$$P(\Xi_1) = {}_{\mathcal{A}_0^*}\Big\langle \rho_0^p, \Psi_{0,1}F_1(\Xi_1)\Big\rangle_{\mathcal{A}_0} \tag{8.15}$$

$$= {}_{\mathcal{A}_1^*}\Big\langle \Psi_{0,1}^*\rho_0^p, F_1(\Xi_1)\Big\rangle_{\mathcal{A}_1} \qquad (\forall \Xi_1 \in \mathcal{F}). \tag{8.16}$$

This implies that the measurement $\mathbf{M}_{\mathcal{A}_0}(\widetilde{\mathbf{O}}_0, S_{[\rho_0^p]})$ can be considered to be equal to the statistical measurement $\mathbf{M}_{\mathcal{A}_1}(\mathbf{O}_1, S(\Psi_{0,1}^*\rho_0^p))$. That is, $\mathbf{M}_{\mathcal{A}_0}(\widetilde{\mathbf{O}}_0, S_{[\rho_0^p]})$ is the representation due to the Heisenberg picture, and $\mathbf{M}_{\mathcal{A}_1}(\mathbf{O}_1, S_{[*]}(\Psi_{0,1}^*\rho_0^p))$ is the representation due to the Schrödinger picture. Summing up, we have the identification:

$$\underset{\text{(meaningful in the sense of Axiom 1)}}{\overset{\text{[the representation by Heisenberg picture]}}{\mathbf{M}_{\mathcal{A}_0}(\Psi_{0,1}\mathbf{O}_1, S_{[\rho_0^p]})}} \quad \overset{\text{identification}}{\longleftrightarrow} \quad \underset{\text{(meaningful in the sense of Proclaim 1)}}{\overset{\text{[the representation by Schödinger picture]}}{\mathbf{M}_{\mathcal{A}_1}(\mathbf{O}_1, S(\Psi_{0,1}^*\rho_0^p))}} \tag{8.17}$$

in which the left-hand side is understood in Axiom 1 and the right-hand side is understood in Proclaim 1. For completeness, we explain the meaning of the identification (8.17) as follows: The left-hand side of (8.17) means that

($\bullet_1$) Taking a measurement $\mathbf{M}_{\mathcal{A}_0}(\Psi_{0,1}\mathbf{O}_1, S_{[\rho_0^p]})$ N-times $\Big($that is, taking a measurement $\mathbf{M}_{\mathcal{A}_0}(\Psi_{0,1}\mathbf{O}_1, S_{[\rho_0^p]})$, and taking a measurement $\mathbf{M}_{\mathcal{A}_0}(\Psi_{0,1}\mathbf{O}_1, S_{[\rho_0^p]})$,..., and taking a measurement $\mathbf{M}_{\mathcal{A}_0}(\Psi_{0,1}\mathbf{O}_1, S_{[\rho_0^p]})\Big)$, we obtain measured values x_1, x_2,...,x_N. And thus we have the sample space $(X, \mathcal{F}, \rho_0^p(\Psi_{0,1}F(\,\cdot\,)))$ $(= (8.15))$.

The right-hand side of (8.17) means that

($\bullet_2$) Taking a statistical measurement $\mathbf{M}_{\mathcal{A}_1}(\mathbf{O}_1, S(\Psi_{0,1}^*\rho_0^p))$ N-times $\Big($that is, taking a measurement $\mathbf{M}_{\mathcal{A}_1}(\mathbf{O}_1, S_{[*1]}(\Psi_{0,1}^*\rho_0^p))$, and taking a measurement $\mathbf{M}_{\mathcal{A}_1}(\mathbf{O}_1, S_{[*2]}(\Psi_{0,1}^*\rho_0^p))$,..., and taking a measurement $\mathbf{M}_{\mathcal{A}_1}(\mathbf{O}_1, S_{[*N]}(\Psi_{0,1}^*\rho_0^p))\Big)$, we obtain measured values x'_1, x'_2,...,x'_N. And thus we have the sample space $(X, \mathcal{F}, (\Psi_{0,1}^*\rho_0^p)(F(\,\cdot\,)))$ $(= (8.16))$.

Since (8.15) = (8.16), we identify ($\bullet_1$) with ($\bullet_2$).[5]

[(ii): Schrödinger picture II]. Let $\Psi_{1,2} : \mathcal{A}_2 \to \mathcal{A}_1$ be a Markov operator. Let ρ_1^m $(\in \mathfrak{S}^m(\mathcal{A}_1^*))$ be a statistical state. That is, we consider the following general system:

[5]Strictly speaking. we must say "we regard ($\bullet_2$) as ($\bullet_1$)". That is because Axiom 2 says that Heisenberg picture representation is more fundamental than Schrödinger picture representation.

$$\underset{\text{statistical state}\rho_1^m}{[\mathcal{A}_1]} \xleftarrow{\Psi_{0,1}} [\mathcal{A}_2]. \tag{8.18}$$

Here, let $\mathbf{O}_2 \equiv (X_2, \mathcal{F}_2, F_2)$ be an observable in a C^*-algebra $\mathcal{A}_2$. And put $\widetilde{\mathbf{O}}_1 = (X_2, \mathcal{F}_2, \Psi_{1,2}F_2)$. Since ρ_1^m $(\in \mathfrak{S}^m(\mathcal{A}_1^*))$ is a statistical state (i.e., the probabilistic interpretation is added), we have the statistical measurement

$$\mathbf{M}_{\mathcal{A}_1}(\widetilde{\mathbf{O}}_1 \equiv (X_2, \mathcal{F}_2, \Psi_{1,2}F_2), S(\rho_1^m)), \tag{8.19}$$

which generates the sample space $(X_2, \mathcal{F}_2, P)$ such that:

$$P(\Xi_2) = {}_{\mathcal{A}_1^*}\Big\langle \rho_1^p, \Psi_{0,1}F_2(\Xi_2)\Big\rangle_{\mathcal{A}_1}. \tag{8.20}$$

This is equal to

$${}_{\mathcal{A}_2^*}\Big\langle \Psi_{0,1}^*\rho_1^m, F_2(\Xi_2)\Big\rangle_{\mathcal{A}_2}, \tag{8.21}$$

which implies that the statistical measurement $\mathbf{M}_{\mathcal{A}_1}(\widetilde{\mathbf{O}}_1, S(\rho_1^m))$ can be considered to be equal to the statistical measurement $\mathbf{M}_{\mathcal{A}_2}(\mathbf{O}_2, S(\Psi_{1,2}^*\rho_1^m))$. That is, $\mathbf{M}_{\mathcal{A}_1}(\widetilde{\mathbf{O}}_1, S(\rho_1^m))$ is the representation due to Heisenberg picture, and $\mathbf{M}_{\mathcal{A}_2}(\mathbf{O}_2, S(\Psi_{1,2}^*\rho_1^m))$ is the representation due to Schrödinger picture. Summing up, we have the identification:[6]

$$\underset{\text{(meaningful in the sense of Proclaim 1)}}{\overset{\text{[the representation by Heisenberg picture]}}{\mathbf{M}_{\mathcal{A}_1}(\Psi_{1,2}\mathbf{O}_2, S(\rho_1^m))}} \xleftrightarrow{\text{identification}} \underset{\text{(meaningful in the sense of Proclaim 1)}}{\overset{\text{[the representation by Schödinger picture]}}{\mathbf{M}_{\mathcal{A}_2}(\mathbf{O}_2, S(\Psi_{1,2}^*\rho_1^m))}} \tag{8.22}$$

in which the both sides are understood in Proclaim 1.

■

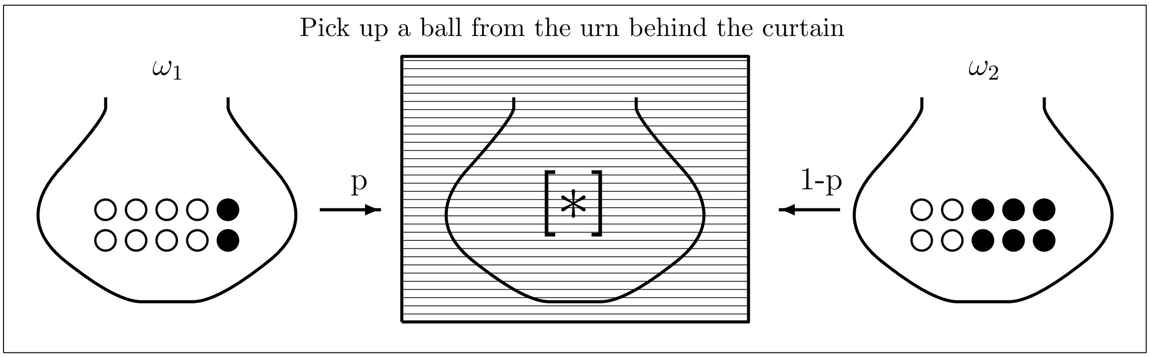
The statistical state also appears in Bayes theorem, which was already studied in Chapter 6.

Example 8.6. [A statistical state in Bayes theorem]. (*continued from Example 8.1*) Assume the situation $(\mathrm{P}_1) \sim (\mathrm{P}_2)$ in Example 8.1 (Coin-tossing). That is, consider the following statistical measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0))$:

---

[6] Recall Axiom 2, which says that $\mathbf{M}_{\mathcal{A}_1}(\Psi_{1,2}\mathbf{O}_2, S(\rho_1^m))$ is more fundamental than $\mathbf{M}_{\mathcal{A}_2}(\mathbf{O}_2, S(\Psi_{1,2}^*\rho_1^m))$.

The picture of $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0))$



Next, consider the following procedure.

(P$_3$) We find that the ball sampled in (P$_2$) is a white one. That is, by the statistical measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\nu_0))$ in (P$_2$), we obtain the measured value $w (\in \{w, b\})$.

(P$_4$) After the above (P$_3$), we further take a "measurement" of an observable $\mathbf{O}_1 \equiv (Y, \mathcal{G}, G)$. And, we know that the measured value belongs to Γ $(\in \mathcal{G})$.

In what follows we study the above (P$_3$) and (P$_4$). The procedures (P$_1$) $\sim$ (P$_4$) can be characterized as the statistical measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O} \times \mathbf{O}_1, S(\nu_0))$. The probability that the measured value $(w, y) (\in \{w, b\} \times \Gamma)$ obtained by $\mathbf{M}_{C(\Omega)}(\mathbf{O} \times \mathbf{O}_1, S(\nu_0))$ belongs to Γ is given by

$$\langle \nu_0, F(\{w\}) \times G(\Gamma) \rangle.$$

Then, under the condition that we know (P$_3$), the probability that the measured value y $(\in Y)$ is obtained in (P$_4$) is given by the conditional probability

$$\frac{{}_{\mathcal{M}(\Omega)}\langle \nu_0, F(\{w\}) \times G(\Gamma) \rangle_{C(\Omega)}}{{}_{\mathcal{M}(\Omega)}\langle \nu_0, F(\{w\}) \rangle_{C(\Omega)}} \Big(= {}_{\mathcal{M}(\Omega)}\Big\langle \frac{F(\{w\}) \times \nu_0}{{}_{\mathcal{M}(\Omega)}\langle \nu_0, F(\{w\}) \rangle_{C(\Omega)}}, G(\Gamma) \Big\rangle_{C(\Omega)} \Big). \tag{8.23}$$

Since $\mathbf{O}_1 (\equiv (Y, \mathcal{G}, G))$ is arbitrary observable in $C(\Omega)$, this implies the following state-reduction:

$$\underset{\text{before "white" is obtained in } (P_2)}{\text{pretest state "}\nu_0\text{"}} \longrightarrow \underset{\text{after "white" is obtained in } (P_2)}{\text{posttest state "}\nu_1\text{"}} \Big(= \frac{F(\{w\}) \times \nu_0}{\langle \nu_0, F(\{w\}) \rangle} \Big). \tag{8.24}$$

That is because the probability that the measured value obtained by $\mathbf{M}_{C(\Omega)}(\mathbf{O}_1, S(\nu_1))$ belongs to Γ is given by

$${}_{\mathcal{M}(\Omega)}\Big\langle \nu_1, G(\Gamma) \Big\rangle_{C(\Omega)} \tag{8.25}$$

and it must hold that (8.23)=(8.25). Here, note that this new mixed state $\nu_1 (\in \mathcal{M}_{+1}^m(\Omega))$ satisfies

$$\nu_1(\{\omega\}) = \frac{\nu_0(\{\omega\}) \times [F(\{w\})](\omega)}{\nu_0(\omega_1) \times [F(\{w\})](\omega_1) + \nu_0(\omega_2) \times [F(\{w\})](\omega_2)} \qquad (\forall \omega \in \Omega \equiv \{\omega_1, \omega_2\}). \tag{8.26}$$

Then, it holds that

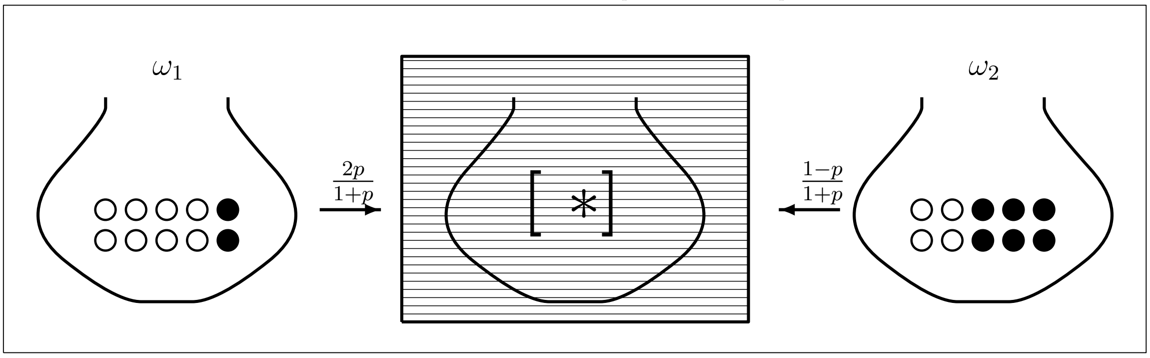
$$\begin{aligned} \nu_1(\{\omega_1\}) &= \frac{0.8p}{0.8p + 0.4(1-p)} = \frac{2p}{1+p}, \\ \nu_1(\{\omega_2\}) &= \frac{0.4(1-p)}{0.8p + 0.4(1-p)} = \frac{1-p}{1+p}. \end{aligned} \tag{8.27}$$

Since

[$\bullet$] the ν_1 is the statistical state after the (P_3),

the "measurement" in (P_4) is represented by the statistical measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}_2, S(\nu_1))$, that is,

The picture of $S([\delta_{\omega_1}; \frac{2p}{1+p}] \oplus [\delta_{\omega_2}; \frac{1-p}{1+p}])$ $(\approx S(\nu_1))$



■

Example 8.7. [(i): A statistical state in the repeated measurement]. Let $\rho^m \in \mathfrak{S}^m(\mathcal{A}^*)$. By the Krein-Milman theorem (*cf.* [92]), we can choose a sequence $\{\rho_k^p\}_{k=1}^N$ in $\mathfrak{S}^p(\mathcal{A}^*)$ such that:

$$\frac{1}{N} \sum_{k=1}^N \rho_k^p \approx \rho^m \qquad \text{(in the sense of the weak}^*\text{-topology of } \mathfrak{S}^m(\mathcal{A}^*)). \tag{8.28}$$

for a sufficiently large natural number N. Consider an observable $\mathbf{O} \equiv (X, \mathcal{F}, F)$ in $\mathcal{A}$. And consider the measurement $\mathbf{M}_{\otimes \mathcal{A}}(\otimes_{k=1}^N \mathbf{O} \equiv (X^N, \mathcal{F}^N, \bigotimes_{k=1}^N F), S_{[\otimes_{k=1}^N \rho_k^p]})$ formulated in the tensor C^*-algebra $\bigotimes_{k=1}^N \mathcal{A}$, where $(\otimes_{k=1}^N F)(X^{m-1} \times \Xi_m \times X^{N-m}) = (\otimes_{k=1}^{m-1} I) \otimes F(\Xi_m) \otimes (\otimes_{k=m+1}^N I)$ $(\forall \Xi_m \in \mathcal{F}, 1 \le \forall m \le N)$. For completeness, note the measurement $\mathbf{M}_{\otimes\mathcal{A}}\big(\otimes_{k=1}^N \mathbf{O},\ S_{[\otimes_{k=1}^N \rho_k^p]}\big)$ is meaningful in the sense of Axiom 1. Let $(x_1, x_2, ..., x_N)$ be a measured value obtained by the measurement $\mathbf{M}_{\otimes\mathcal{A}}\big(\otimes_{k=1}^N \mathbf{O},\ S_{[\otimes_{k=1}^N \rho_k^p]}\big)$. Thus, by Axiom 1, we can "almost surely" expect that

$$\rho^m(F(\Xi)) \approx \frac{\sharp[\{k : x_k \in \Xi\}]}{N} \qquad (\forall \Xi \in \mathcal{F}) \tag{8.29}$$

holds for a sufficiently large N, where $\sharp[B]$ is the number of the elements of a set B. That is because the probability that a measured value obtained by $\mathbf{M}_{\mathcal{A}}\big(\mathbf{O},\ S_{[\rho_k^p]}\big)$ belongs to Ξ $(\in \mathcal{F})$ is given by $\rho^p(F(\Xi))$. In the above sense (8.29), the mathematical symbol $\mathbf{M}_{\mathcal{A}}\big(\mathbf{O},\ S(\rho^m)\big)$ (or, $\mathbf{M}_{\mathcal{A}}\big(\mathbf{O},\ S(\frac{1}{N}\sum_{k=1}^N \rho_k^p)\big)$) can be considered as the statistical measurement, which may be called a "repeated measurement".

[(ii)]. Let Ω be a finite set, i.e., $\Omega \equiv \{\omega_1, \omega_2, ..., \omega_M\}$. Let $\mathbf{O} \equiv (X, \mathcal{F}, F)$ be an observable in $C(\Omega)$. Consider the repeated measurement $\mathbf{M}_{\otimes_{n=1}^{NM} C(\Omega)}\big(\otimes_{n=1}^{NM} \mathbf{O},\ S_{[\otimes_{n=1}^{NM} \delta_{\omega_{\mathrm{mod}_M[n]}}]}\big)$ (which may be called a cyclic measurement), where $\mathrm{mod}_M[n]$ is the integer such that $n = M\dot{j} + \mathrm{mod}_M[n]$ and $0 \le \mathrm{mod}_M[n] \le M-1$. Let $(x_1, x_2, ..., x_{NM})$ be a measured value obtained by the cyclic measurement $\mathbf{M}_{\otimes_{n=1}^{NM} C(\Omega)}\big(\otimes_{n=1}^{NM} \mathbf{O},\ S_{[\otimes_{n=1}^{NM} \delta_{\omega_{\mathrm{mod}_M[n]}}]}\big)$ $\Big(= \otimes_{n=1}^{NM} \mathbf{M}_{C(\Omega)}(\mathbf{O},\ S_{[\delta_{\omega_{\mathrm{mod}_M[n]}}]})\Big)$. Thus, by Axiom 1, we can "almost surely" expect that

$$ {}_{C(\Omega)^*}\Big\langle \frac{\delta_{\omega_1} + \delta_{\omega_2} + \cdots + \delta_{\omega_M}}{M}, F(\Xi) \Big\rangle_{C(\Omega)} \approx \frac{\sharp[\{k : x_k \in \Xi\}]}{NM} \qquad (\forall \Xi \in \mathcal{F}) \tag{8.30}$$

holds for a sufficiently large N. In this sense,

- we often use the repeated statistical measurement $\otimes_{n=1}^N \mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\frac{\delta_{\omega_1}+\delta_{\omega_2}+\cdots+\delta_{\omega_M}}{M}))$ $\Big($or more precisely, the repeated probabilistic measurement $\otimes_{n=1}^N \mathbf{M}_{C(\Omega)}$ $(\mathbf{O}, S_{[*]}(\oplus_{m=1}^M$ $[\delta_{\omega_m}\ ; 1/M]))$, *cf.* (8.8)$\Big)$ as a substitute for $\mathbf{M}_{\otimes_{n=1}^{NM} C(\Omega)}\big(\otimes_{n=1}^{NM} \mathbf{O},\ S_{[\otimes_{n=1}^{NM} \delta_{\omega_{\mathrm{mod}_M[n]}}]}\big)$.

That is, in the following table (in the case that $\Omega = \{\omega_1, \omega_2\}$), the measured data $(x_1, x_2, ..., x_{2N})$ and the measured data $(y_1, y_2, ..., y_{2N})$ have the same statistical properties (e.g., average, variance, etc.).

measurement	$\cdots\cdots$	measured value	measurement	$\cdots\cdots$	measured value
$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_1}]})$	$\cdots\cdots$	x_1	$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\frac{\delta_{\omega_1}+\delta_{\omega_2}}{2}))$	$\cdots\cdots$	y_1
$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_2}]})$	$\cdots\cdots$	x_2	$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\frac{\delta_{\omega_1}+\delta_{\omega_2}}{2}))$	$\cdots\cdots$	y_2
$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_1}]})$	$\cdots\cdots$	x_3	$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\frac{\delta_{\omega_1}+\delta_{\omega_2}}{2}))$	$\cdots\cdots$	y_3
$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_2}]})$	$\cdots\cdots$	x_4	$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\frac{\delta_{\omega_1}+\delta_{\omega_2}}{2}))$	$\cdots\cdots$	y_4
	$\cdots\cdots$	$\cdots$		$\cdots\cdots$	$\cdots$
$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_1}]})$	$\cdots\cdots$	x_{2N-1}	$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\frac{\delta_{\omega_1}+\delta_{\omega_2}}{2}))$	$\cdots\cdots$	y_{2N-1}
$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_1}]})$	$\cdots\cdots$	x_{2N}	$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\frac{\delta_{\omega_1}+\delta_{\omega_2}}{2}))$	$\cdots\cdots$	y_{2N}

■

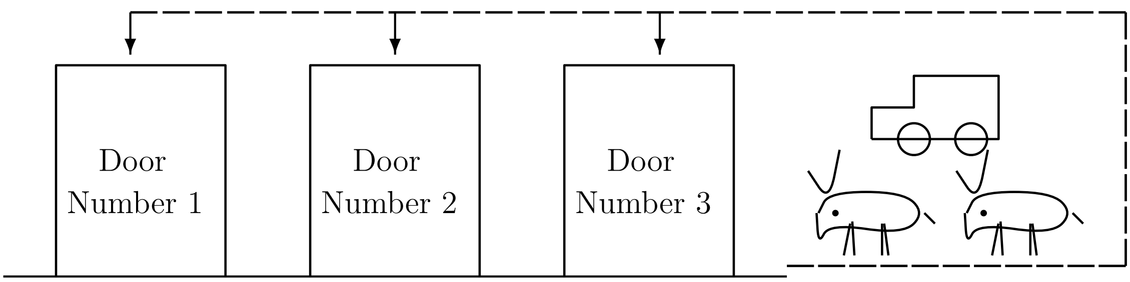
8.1.3 Problems (statistical measurements)

Problem 8.8. [Monty Hall problem, *cf.*[33]]. The Monty Hall problem is as follows (*cf.* Problem 5.12, Remark 5.13 and Problem 11.13) :

(P) Suppose you are on a game show, and you are given the choice of three doors (i.e., "number 1", "number 2", "number 3"). Behind one door is a car, behind the others, goats.

(C) You know that the probability that behind the k-th door (i.e., "number k") is a car is given by p_k $(k = 1, 2, 3)$. $\Big($For example, consider the two cases that $p_1 = p_2 = p_3 = 1/3$, and $p_1 = 3/7$, $p_2 = 1/7$, $p_3 = 3/7$.$\Big)$

You pick a door, say number 1, and the host, who knows what's behind the doors, opens another door, say "number 3", which has a goat. He says to you, "Do you want to pick door number 2?" Is it to your advantage to switch your choice of doors?



[*Answer*]. Put $\Omega = \{\omega_1, \omega_2, \omega_3\}$, where

$\omega_1 \cdots\cdots$ the state that the car is behind the door number 1

$\omega_2 \cdots\cdots$ the state that the car is behind the door number 2

$\omega_3 \cdots\cdots$ the state that the car is behind the door number 3.

Define the observable $\mathbf{O} \equiv (\{1,2,3\}, 2^{\{1,2,3\}}, F)$ in $C(\Omega)$ such that

$$[F(\{1\})](\omega_1) = 0.0, \quad [F(\{2\})](\omega_1) = 0.5, \quad [F(\{3\})](\omega_1) = 0.5,$$ [7]

$$[F(\{1\})](\omega_2) = 0.0, \quad [F(\{2\})](\omega_2) = 0.0, \quad [F(\{3\})](\omega_2) = 1.0,$$

$$[F(\{1\})](\omega_3) = 0.0, \quad [F(\{2\})](\omega_3) = 1.0, \quad [F(\{3\})](\omega_3) = 0.0. \tag{8.31}$$

Define the statistical state ν_0 $(\in \mathcal{M}_{+1}^m(\Omega))$ such that:

$$\nu_0(\{\omega_1\}) = p_1, \quad \nu_0(\{\omega_2\}) = p_2, \quad \nu_0(\{\omega_3\}) = p_3 \tag{8.32}$$

where $p_1 + p_2 + p_3 = 1$, $0 \le p_1, p_2, p_3 \le 1$. Thus we have a statistical measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0))$. Note that

(1) : "measured value 1 is obtained" $\Longleftrightarrow$ the host says "Door (number 1) has a goat" (probability $\longleftrightarrow$ 0)

(2) : "measured value 2 is obtained" $\Longleftrightarrow$ the host says "Door (number 2) has a goat" (probability $\longleftrightarrow 0.5p_1 + 1.0p_3$)

(3) : "measured value 3 is obtained" $\Longleftrightarrow$ the host says "Door (number 3) has a goat" (probability $\longleftrightarrow 0.5p_1 + 1.0p_2$)

Here, assume that

- By the statistical measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0))$, you obtain a measured value 3.

which corresponds to the fact that the host said "Door (number 3) has a goat". Then, the posttest state ν_{post} $(\in \mathcal{M}_{+1}^m(\Omega))$ is given by

$$\nu_{\text{post}} = \frac{F(\{3\}) \times \nu_0}{\langle \nu_0, F(\{3\}) \rangle}. \tag{8.33}$$

That is,

$$\nu_{\text{post}}(\{\omega_1\}) = \frac{\frac{p_1}{2}}{\frac{p_1}{2} + p_2}, \quad \nu_{\text{post}}(\{\omega_2\}) = \frac{p_2}{\frac{p_1}{2} + p_2}, \quad \nu_{\text{post}}(\{\omega_3\}) = 0. \tag{8.34}$$

Thus,

[7] Strictly speaking, $F(\{1\})(\omega_1) = 0.5$ and $F(\{2\})(\omega_1) = 0.5$ should be assumed in the problem (P).

- if $p_1 = p_2 = p_3 = 1/3$, then it holds that $\nu_{\rm post}(\{\omega_1\}) = 1/3$, $\nu_{\rm post}(\{\omega_2\}) = 2/3$, $\nu_{\rm post}(\{\omega_3\}) = 0$, *and thus, you should pick Door (number 2).*
- *if* $p_1 = 3/7$, $p_2 = 1/7$ *and* $p_3 = 3/7$, *then it holds that* $\nu_{\rm post}(\{\omega_1\}) = 3/5$, $\nu_{\rm post}(\{\omega_2\}) = 2/5$, $\nu_{\rm post}(\{\omega_3\}) = 0$, *and thus, you should not pick Door (number 2).*

Also, more generally, we can say that

$$\begin{cases} \text{if } \nu_{\rm post}(\{\omega_1\}) \le \nu_{\rm post}(\{\omega_2\}) (\text{i.e.,} p_1 \le 2p_2), \text{ then, you should pick Door (number 2)} \\ \text{if } \nu_{\rm post}(\{\omega_1\}) \ge \nu_{\rm post}(\{\omega_2\}) (\text{i.e.,} p_1 \ge 2p_2), \text{ then, you should not pick Door (number 2).} \end{cases}$$

■

Remark 8.9. [P. Erdös]. I learnt the Monty Hall problem in the book [33] ("The Man Who Loved Only Numbers, The story of Paul Erdös and the search for mathematical truth"). This problem is famous as the problem in which even P. Erdös made a mistake. I think that this problem is too profound to understand without measurement theory. In fact, everyone may confuse the above Problem (P) for $p_1 = p_2 = p_3 = 1/3$ with Problem 5.12 (i.e., the above problem (P) without the condition (C)). In fact, in [33] (page 234), it is written as follows:

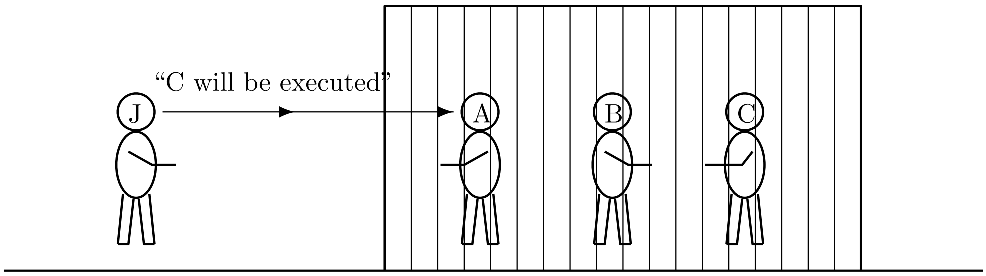
(Q) *You're on a game show and you're given the choice of three doors. Behind one door is a car, and behind the other two are goats. You choose, say, door 1, and the host, who knows where the car is, opens another door, behind which is a goat. He now gives you the choice of sticking with door 1 or switching to the other door? What should you do?*

If you read this description of the Monty Hall problem (in [33]), you may think that the correct answer should be due to Fisher's likelihood method, i.e, the answer presented in Problem 5.12. However, Problem 5.12, Remark 5.13 and Problem 8.8 are not all of the Monty Hall problem. See Problem 11.13 later (which may be my final answer to the Monty Hall problem).

■

Problem 8.10. [The problem of three prisoners].
Consider the following problem:

(P) Three men, A, B, and C were in jail. A knew that one of them was to be set free and the other two were to be executed. But he did not know who was the one to be spared. (He knew that the probability that A [resp. B, C] will be set free is equal to 1/3 [resp. 1/3, 1/3], or more generally, p_a^f [resp. p_b^f, p_c^f].) To the jailer who did know, A said, "Since two out of the three will be executed, it is certain that either B or C will be, at least. You will give me no information about my own chances if you give me the name of one man, B or C, who is going to be executed." Accepting this argument after some thinking, the jailer said, "C will be executed." Thereupon A felt happier because now either he or C would go free, so his chance had increased from 1/3 to 1/2. This prisoner's happiness may or may not be reasonable. What do you think?



[*Answer*]. Put $\Omega = \{\omega_a, \omega_b, \omega_c\}$, where

$$\omega_a \cdots\cdots \text{ the state that A will be set free}$$
$$\omega_b \cdots\cdots \text{ the state that B will be set free}$$
$$\omega_c \cdots\cdots \text{ the state that B will be set free .}$$

Define the observable $\mathbf{O} \equiv (\{x_A, x_B, x_C\}, 2^{\{x_A,x_B,x_C\}}, F)$ in $C(\Omega)$ such that

$$[F(\{x_A\})](\omega_a) = 0.0, \quad [F(\{x_B\})](\omega_a) = 0.5, \quad [F(\{x_C\})](\omega_a) = 0.5,^{8}$$
$$[F(\{x_A\})](\omega_b) = 0.0, \quad [F(\{x_B\})](\omega_b) = 0.0, \quad [F(\{x_C\})](\omega_b) = 1.0,$$
$$[F(\{x_A\})](\omega_c) = 0.0, \quad [F(\{x_B\})](\omega_c) = 1.0, \quad [F(\{x_C\})](\omega_c) = 0.0. \tag{8.35}$$

Define the statistical state ν_0 ($\in \mathcal{M}_{+1}^m(\Omega)$) such that:

$$\nu_0(\{\omega_a\}) = p_a^f, \quad \nu_0(\{\omega_b\}) = p_b^f, \quad \nu_0(\{\omega_c\}) = p_c^f \tag{8.36}$$

where $p_a^f + p_b^f + p_c^f = 1$, $0 \le p_a^f, p_b^f, p_c^f \le 1$, though it may suffice to assume that $p_a^f = p_b^f = p_c^f = 1/3$. Here, note that the following (i) and (ii) are equivalent:

[8] Strictly speaking, $[F(\{x_B\})](\omega_a) = 0.5$ and $[F(\{x_C\})](\omega_a) = 0.5$ should be assumed in the problem (P)

(i) The jailer said to A "C will be executed".

(ii) By the statistical measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0))$, A obtains a measured value x_C

Thus, the posttest state $\nu_{\rm post}$ ($\in \mathcal{M}_{+1}^m(\Omega)$) is given by

$$\nu_{\rm post} = \frac{F(\{x_C\}) \times \nu_0}{\langle \nu_0, F(\{x_C\}) \rangle}. \tag{8.37}$$

That is,

$$\nu_{\rm post}(\{\omega_a\}) = \frac{\frac{p_a^f}{2}}{\frac{p_a^f}{2} + p_b^f}, \quad \nu_{\rm post}(\{\omega_b\}) = \frac{p_b^f}{\frac{p_a^f}{2} + p_b^f}, \quad \nu_{\rm post}(\{\omega_c\}) = 0. \tag{8.38}$$

Thus,

- if $p_a^f = p_b^f = p_c^f = 1/3$, it holds that $\nu_{\rm post}(\{\omega_a\}) = 1/3$, $\nu_{\rm post}(\{\omega_b\}) = 2/3$, $\nu_{\rm post}(\{\omega_c\}) = 0$, *and thus, the prisoner's happiness is not reasonable.* That is because $p_a^f = 1/3 = \nu_{\rm post}(\{\omega_a\})$.

- if $p_a^f = 3/7$, $p_b^f = 1/7$, $p_c^f = 3/7$, it holds that $\nu_{\rm post}(\{\omega_a\}) = 3/5$, $\nu_{\rm post}(\{\omega_b\}) = 2/5$, $\nu_{\rm post}(\{\omega_c\}) = 0$, *and thus, the prisoner's happiness is reasonable.* That is because $p_a^f = 3/7 < 3/5 = \nu_{\rm post}(\{\omega_a\})$.

- if $p_a^f = 1/4$, $p_b^f = 1/2$, $p_c^f = 1/4$, it holds that $\nu_{\rm post}(\{\omega_a\}) = 1/5$, $\nu_{\rm post}(\{\omega_b\}) = 4/5$, $\nu_{\rm post}(\{\omega_c\}) = 0$, *and thus, the prisoner's unhappiness is reasonable.* That is because $p_a^f = 1/3 > 1/5 = \nu_{\rm post}(\{\omega_a\})$.

Also, more generally, we can say that

$$\begin{cases} \text{if } p_a^f \le \nu_{\rm post}(\{\omega_a\})(\text{i.e.,} p_a^f + 2p_b^f \ge 1), & \text{the prisoner's happiness is reasonable} \\ \text{if } p_a^f \ge \nu_{\rm post}(\{\omega_a\})(\text{i.e.,} p_a^f + 2p_b^f \le 1), & \text{the prisoner's unhappiness is reasonable.} \end{cases}$$

■

Remark 8.11. [(i).The problem of three prisoners in PMT]. Recall that the Monty Hall problem is also studied in PMT, that is, Problem 5.12 (Fisher's method) and Remark 5.13 (The moment method). On the other hand, it should be noted that the problem of three prisoners can not be solved in PMT.
[(ii): The relation between the Monty Hall problem and the problem of three prisoners]. Since the Monty Hall problem and the problem of three prisoners are similar, we add something concerning the relation between the two. Consider the (P) (in Problem 8.8) and the (Q) mentioned below.

(Q) (Continued from the (P) in Problem 8.10). There is a woman, who was proposed to by the three prisoners A, B and C. She listened to the conversation between A and the jailer. Thus, assume that she has the same information as A has. Then, we have the following problem:

($\sharp$) Whose proposal should she accept?

[*Answer*]. For simplicity, consider the case that $p_a^f = p_b^f = p_c^f = 1/3$. Then we see that

$$\nu_{\rm post}(\{\omega_a\}) = 1/3, \quad \nu_{\rm post}(\{\omega_b\}) = 2/3, \quad \nu_{\rm post}(\{\omega_c\}) = 0. \tag{8.39}$$

Thus, *she should choose the prisoner B.* Here it should be noted that the problem ($\sharp$) is the same as the Monty Hall problem. That is, the problem:

"(P) in Problem 8.10" + "(Q) in the above"

includes both the Monty Hall problem and the problem of three prisoners.

■

8.2 General statistical system (Example)

As mentioned in the previous section, the Statistical MT (i.e., SMT) is formulated as follows:

$$\text{PMT} = \underset{\text{[Axiom 1 (2.37)]}}{\text{measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} \quad \text{in } C^*\text{-algebra},$$

and

$$\text{SMT} = \underset{\text{[Proclaim 1 (8.10)]}}{\text{statistical measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} \quad \text{in } C^*\text{-algebra},$$

where it should be noted that

$$\text{“Proclaim 1”} = \text{“Axiom 1”} + \underset{\text{(the probabilistic interpretation of mixed state)}}{\text{“statistical state”}}. \tag{8.40}$$

Thus we see

$$\begin{aligned}\text{SMT} &= \underset{\text{[Proclaim 1 (8.10)]}}{\text{statistical measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} \\ &= \underset{\text{(Axioms 1 and 2)}}{\text{PMT}} + \underset{\text{(the probabilistic interpretation of mixed state)}}{\text{“statistical state”}} \quad \text{in } C^*\text{-algebra}.\end{aligned}$$

That is, Axiom 2 is common to PMT and SMT. This will be explicitly seen in the following example (= Example 8.12), which should be compared with Example 3.4. Also recalling Remark 8.3 [hybrid measurement theory (= HMT)], we say that

$$\text{HMT} = \underset{\text{[Axiom 1 (2.37) and Proclaim 1 (8.10)]}}{\text{hybrid measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} \text{ in } C^*\text{-algebra} . \tag{8.41}$$

Here note that PMT and SMT are respectively regarded as one of the aspects of HMT.

Since Axiom 2 is common to PMT and SMT, it is a matter of course that Example 3.4 (in PMT) and Example 8.12 (in SMT) are almost similar.

Example 8.12. [(Continued from Example 3.4) A simple general statistical system, Heisenberg picture]. Suppose that a tree ($T \equiv \{0, 1, ..., 6, 7\}, \pi$) has an ordered structure such that $\pi(1) = \pi(6) = \pi(7) = 0$, $\pi(2) = \pi(5) = 1$, $\pi(3) = \pi(4) = 2$. (See the figure (8.42).) Consider a general system $\mathbf{S}(\rho_0^m) \equiv [S(\rho_0^m), \{\mathcal{A}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{A}_{\pi(t)}\}_{t \in T \setminus \{0\}}]$ with the initial system $S(\rho_0^m)$.

$$\begin{array}{ccccccc}
 & & & & & \Phi_{2,3} & \mathcal{A}_3 \\
 & & & \Phi_{1,2} & \mathcal{A}_2 & \swarrow & \\
 & & & \swarrow & & \nwarrow & \\
 & \Phi_{0,1} & \mathcal{A}_1 & & & \Phi_{2,4} & \mathcal{A}_4 \\
 & \swarrow & & \nwarrow & \Phi_{1,5} & & \\
\mathcal{A}_0 & \leftarrow & \mathcal{A}_6 & & \mathcal{A}_5 & & \\
 & \Phi_{0,6} & & & & & \\
 & \nwarrow & & & & & \\
 & \Phi_{0,7} & \mathcal{A}_7 & & & &
\end{array} \tag{8.42}$$

Also, for each $t \in \{0, 1, ..., 6, 7\}$, consider an observable $\mathbf{O}_t \equiv (X_t, 2^{X_t}, F_t)$ in a C^*-algebra $\mathcal{A}_t$. Thus, we have a sequential observable $[\{\mathbf{O}_t\}_{t \in T}, \{\Phi_{t,\pi(t)} : \mathcal{A}_t \to \mathcal{A}_{\pi(t)}\}_{t \in T \setminus \{0\}}]$. Now we want to consider the following "measurement",

(♯) for a statistical system $S(\rho_0^m)$, take a measurement of *"a sequential observable* $[\{\mathbf{O}_t\}_{t \in T}, \{\mathcal{A}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{A}_{\pi(t)}\}_{t \in T \setminus \{0\}}]$," i.e., take a measurement of an observable $\mathbf{O}_0$ at $0(\in T)$, and next, take a measurement of an observable $\mathbf{O}_1$ at $1(\in T)$, $\cdots\cdots$, and finally take a measurement of an observable $\mathbf{O}_7$ at $7(\in T)$,

which is symbolized by $\mathfrak{M}(\{\mathbf{O}_t\}_{t \in T}, S(\rho_0^m))$. Note that the $\mathfrak{M}(\{\mathbf{O}_t\}_{t \in T}, \mathbf{S}(\rho_0^m))$ is merely a symbol since only one measurement is permitted (*cf.* §2.5 Remark(II)). In what follows let us describe the above ($\sharp$) ($= \mathfrak{M}(\{\mathbf{O}_t\}_{t\in T}, \mathbf{S}(\rho_0^m))$) precisely. Put

$$\widetilde{\mathbf{O}}_t = \mathbf{O}_t \quad \text{and thus} \quad \widetilde{F}_t = F_t \quad (t = 3, 4, 5, 6, 7).$$

First we construct the quasi-product observable $\widetilde{\mathbf{O}}_2$ in $\mathcal{A}_2$ such as

$$\widetilde{\mathbf{O}}_2 = (X_2 \times X_3 \times X_4, 2^{X_2\times X_3\times X_4}, \widetilde{F}_2) \quad \text{where } \widetilde{F}_2 = F_2 \overset{\text{qp}}{\times} (\overset{\text{qp}}{\times}_{t=3,4} \Phi_{2,t}\widetilde{F}_t),$$

if it exists. Iteratively, we construct the following:

$$\begin{array}{ccccc}
\mathcal{A}_0 & \xleftarrow{\Phi_{0,1}} & \mathcal{A}_1 & \xleftarrow{\Phi_{1,2}} & \mathcal{A}_2 \\
F_0 \overset{\text{qp}}{\times} \Phi_{0,6}\widetilde{F}_6 \overset{\text{qp}}{\times} \Phi_{0,7}\widetilde{F}_7 & & F_1 \overset{\text{qp}}{\times} \Phi_{1,5}\widetilde{F}_5 & & \\
\downarrow & & \downarrow & & \\
\underset{(F_0 \overset{\text{qp}}{\times} \Phi_{0,6}\widetilde{F}_6 \overset{\text{qp}}{\times} \Phi_{0,7}\widetilde{F}_7 \overset{\text{qp}}{\times} \Phi_{0,1}\widetilde{F}_1)}{\widetilde{F}_0} & \xleftarrow{\Phi_{0,1}} & \underset{(F_1 \overset{\text{qp}}{\times} \Phi_{1,5}\widetilde{F}_5 \overset{\text{qp}}{\times} \Phi_{1,2}\widetilde{F}_2)}{\widetilde{F}_1} & \xleftarrow{\Phi_{1,2}} & \underset{(F_2 \overset{\text{qp}}{\times} \Phi_{2,3}\widetilde{F}_3 \overset{\text{qp}}{\times} \Phi_{2,4}\widetilde{F}_4)}{\widetilde{F}_2}
\end{array}.$$

That is, we get the quasi-product observable $\widetilde{\mathbf{O}}_1 \equiv (\prod_{t=1}^5 X_t, 2^{\prod_{t=1}^5 X_t}, \widetilde{F}_1)$ of $\mathbf{O}_1$, $\Phi_{1,2}\widetilde{\mathbf{O}}_2$ and $\Phi_{1,5}\widetilde{\mathbf{O}}_5$, and finally, the quasi-product observable $\widetilde{\mathbf{O}}_0 \equiv (\prod_{t=0}^7 X_t, 2^{\prod_{t=0}^7 X_t}, \widetilde{F}_0)$ of $\mathbf{O}_0$, $\Phi_{0,1}\widetilde{\mathbf{O}}_1$, $\Phi_{0,6}\widetilde{\mathbf{O}}_6$ and $\Phi_{0,7}\widetilde{\mathbf{O}}_7$, if it exists. Here, $\widetilde{\mathbf{O}}_0$ is called *the realization* (*or, the Heisenberg picture representation*) of *a sequential observable* $[\{\mathbf{O}_t\}_{t\in T}, \{\mathcal{A}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{A}_{\pi(t)}\}_{t\in T\setminus\{0\}}]$. Then, we have the measurement

$$\mathbf{M}_{\mathcal{A}_0}(\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, 2^{\prod_{t\in T} X_t}, \widetilde{F}_0), S(\rho_0^m)),$$

which is called *the realization* (*or, the Heisenberg picture representation*) *of the symbol* $\mathfrak{M}(\{\mathbf{O}_t\}_{t\in T}, S(\rho_0^m))$.

■

8.3 Bayes theorem in statistical MT

Now let us review "Bayes operator" (Definition 6.5 in §6.2), which plays an important role in SMT as well as PMT. Or, we may say that Bayes operator is more natural in STM than in PMT.

Let $(T \equiv \{0, 1, ..., N\}, \pi : T \setminus \{0\} \to T)$ be a tree with root 0 and let $\mathbf{S}_{[*]} \equiv [S_{[*]}, C(\Omega_t) \overset{\Phi_{\pi(t),t}}{\to} C(\Omega_{\pi(t)}) \ (t \in T \setminus \{0\})]$ be a general system with the initial system $S_{[*]}$. And, let an observable $\mathbf{O}_t \equiv (X_t, \mathcal{F}_t, F_t)$ in a commutative C^*-algebra $C(\Omega_t)$ be given for each $t \in T$. Let $\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, \bigotimes_{t\in T} \mathcal{F}_t, \widetilde{F}_0)$ be as in Theorem 3.7 in the case $\mathcal{A}_t = C(\Omega_t)$ $(\forall t \in T)$. That is, $\widetilde{\mathbf{O}}_0$ is the Heisenberg picture representation of the sequential observable $[\{\mathbf{O}_t\}_{t\in T}, \{C(\Omega_t) \overset{\Phi_{\pi(t),t}}{\to} C(\Omega_{\pi(t)})\}_{t\in T\setminus\{0\}}]$. Let τ be any element in T. If a positive bounded linear operator $B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t} : C(\Omega_\tau) \to C(\Omega_0)$ satisfies the following condition (BO), we call $\{B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t} : \Xi_t \in X_t \ (\forall t \in T)\}$ [resp. $B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}$] a *family of Bayes operators* [resp. a *Bayes operator*]:

(BO) for any observable $\mathbf{O}'_\tau \equiv (Y_\tau, \mathcal{G}_\tau, G_\tau)$ in $C(\Omega_\tau)$, there exists an observable $\widehat{\mathbf{O}}_0 \equiv ((\prod_{t\in T} X_t) \times Y, (\bigotimes_{t\in T} \mathcal{F}_t) \bigotimes \mathcal{G}_\tau, \widehat{F}_0)$ in $C(\Omega_0)$ such that

- (i) $\widehat{\mathbf{O}}_0$ is the Heisenberg picture representation (*cf.* Theorem 3.7) of $[\{\overline{\mathbf{O}}_t\}_{t\in T}; C(\Omega_t) \overset{\Phi_{\pi(t),t}}{\to} C(\Omega_{\pi(t)})$ $(t \in T \setminus \{0\})]$, where $\overline{\mathbf{O}}_t = \mathbf{O}_t$ (if $t \neq \tau$), $= \mathbf{O}_\tau \times \mathbf{O}'_\tau$ (if $t = \tau$),
- (ii) $\widehat{F}_0((\Pi_{t\in T}\Xi_t) \times \Gamma_\tau) = B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(G_\tau(\Gamma_\tau)) \qquad (\Xi_t \in \mathcal{F}_t \ (\forall t \in T), \forall \Gamma_\tau \in \mathcal{G}_\tau)$,
- (iii) $\widehat{F}_0((\Pi_{t\in T}\Xi_t) \times Y_\tau) = \widetilde{F}_0(\prod_{t\in T} \Xi_t) = B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(1_\tau)$, $(\Xi_t \in \mathcal{F}_t \ (\forall t \in T))$, where 1_τ is the identity in $C(\Omega_\tau)$.

Also, define $R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t} : \mathcal{M}^m_{+1}(\Omega_0) \to \mathcal{M}^m_{+1}(\Omega_\tau)$ such that:

$$R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\nu) = \frac{[B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}]^*(\nu)}{\|[B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}]^*(\nu)\|_{\mathcal{M}(\Omega_0)}} \qquad (\forall \nu \in \mathcal{M}^m_{+1}(\Omega_0)),$$

which is called *"a normalized dual Bayes operator"*.

■

It is quite important to see that the Bayes operator $B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t} : C(\Omega_\tau) \to C(\Omega_0)$ is described in terms of the Heisenberg picture. This implies that the Bayes operator $B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t} : C(\Omega_\tau) \to C(\Omega_0)$ is common to PMT and SMT. That is, the dual form $R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t} : \mathcal{M}^m_{+1}(\Omega_0) \to \mathcal{M}^m_{+1}(\Omega_\tau)$ can be applicable to both PMT and SMT and PMT_{PEW} (i.e., subjective Bayesian PMT) mentioned later (in §6.4).

The following theorem is an analogy of Theorem 6.13. This theorem (= Theorem 8.13, Remark 8.14) is also called "Bayes' method".

Theorem 8.13. [Generalized Bayes theorem, Bayes' method, *cf.* [46]]. *Let* $(T \equiv \{0, 1, ..., N\}$, $\pi : T \setminus \{0\} \to T)$ *be a tree with the root* 0 *and let* $\mathbf{S}(\nu_0) \equiv [S(\nu_0), C(\Omega_t) \overset{\Phi_{\pi(t),t}}{\to} C(\Omega_{\pi(t)})$ $(t \in T \setminus \{0\})]$ *be a general system with the initial system* $S(\nu_0)$. *And, let an* *observable* $\mathbf{O}_t \equiv (X_t, \mathcal{F}_t, F_t)$ *in a* C^**-algebra* $C(\Omega_t)$ *be given for each* $t \in T$*. Then, we have a statistical measurement*

$$\mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0 \equiv (\prod_{t \in T} X_t, \bigotimes_{t \in T} \mathcal{F}_t, \widetilde{F}_0), S(\nu_0)). \qquad (cf.\ \text{Theorem 3.7}).$$

Assume that the measured value by the statistical measurement $\mathbf{M}_{C(\Omega)}(\widetilde{\mathbf{O}}_0, S(\nu_0))$ *belongs to* $\prod_{t \in T} \Xi_t$ $(\in \bigotimes_{t \in T} \mathcal{F}_t)$*. Let* τ *be any element in* T*. Then, we see*

$$(a) \qquad \text{"the (statistical) S-state at } \tau(\in T) \text{ after } \mathbf{M}_{C(\Omega_0)}(\widetilde{\mathbf{O}}_0, S(\nu_0))\text{"} = R^{(0,\tau)}_{\prod_{t \in T} \Xi_t}(\nu_0). \tag{8.43}$$

Proof. Since the sequential observable $[\{\mathbf{O}_t\}_{t \in T}, \{C(\Omega_t) \overset{\Phi_{\pi(t),t}}{\to} C(\Omega_{\pi(t)})\}_{t \in T \setminus \{0\}}]$ is common to PMT and SMT, Theorem 3.7 is applicable. Also, by the same argument in Theorem 6.13, the (8.43) immediately follows. □

Remark 8.14. [(i): Bayes operator in Remark 5.7, Bayes' method]. Let $\mathbf{O} \equiv (X, \mathcal{F}, F)$ be an observable in $C(\Omega)$. For each Ξ $(\in \mathcal{F})$, define the continuous linear operator $B^{(0,0)}_{\Xi}$ (or, $B^{\mathbf{O}}_{\Xi}$, $B^{\mathbf{O},(0,0)}_{\Xi}$) $: C(\Omega) \to C(\Omega)$ such that:

$$B^{(0,0)}_{\Xi}(g) = F(\Xi) \cdot g \qquad (\forall g \in C(\Omega)),$$

which is called the *Bayes operator* (or, *simplest Bayes operator*). Define the map $R^{(0,0)}_{\Xi}$: $\mathcal{M}^m_{+1}(\Omega) \to \mathcal{M}^m_{+1}(\Omega)$ (called *"normalized Bayes dual operator"*) such that:

$$(B_1) \qquad R^{(0,0)}_{\Xi}(\nu) = \frac{[B^{(0,0)}_{\Xi}]^*(\nu)}{\|[B^{(0,0)}_{\Xi}]^*(\nu)\|_{\mathcal{M}(\Omega)}} \qquad (\forall \nu \in \mathcal{M}^m_{+1}(\Omega)),$$

that is,

$$[R^{(0,0)}_{\Xi}(\nu)](D_0) = \frac{\int_{D_0}[F(\Xi)](\omega)\nu(d\omega)}{\int_{\Omega}[F(\Xi)](\omega)\nu(d\omega)} \qquad (\forall D_0 \in \mathcal{B}_{\Omega}).$$

Thus, we can describe the well known Bayes theorem (*cf.* [86]) such as

$$\mathcal{M}^m_{+1}(\Omega) \ni \nu \ (= \text{pretest state}) \mapsto \ (\text{posttest state} =) R^{(0,0)}_{\Xi}(\nu) \in \mathcal{M}^m_{+1}(\Omega). \tag{8.44}$$

As a particular case of the above, assume that $\nu = \delta_{\omega_0}$ $(\in \mathcal{M}^p_{+1}(\Omega))$. Then we see that

$$\mathcal{M}^p_{+1}(\Omega) \ni \delta_{\omega_0} \ (= \text{pretest state}) \mapsto \ (\text{posttest state} =) R^{(0,0)}_{\Xi}(\delta_{\omega_0}) = \delta_{\omega_0} \in \mathcal{M}^p_{+1}(\Omega).$$

That is, a pure state δ_{ω_0} is invariant.

[(ii): The conventional Bayes theorem in mathematics]. The above theorem should be compared with the following conventional Bayes theorem (B_2).

(B_2) Let $(\mathcal{S}, \mathcal{B}_{\mathcal{S}}, P)$ be a probability space. Let $\{E_1, E_2, ..., E_n\}$ be a (measurable) decomposition of $\mathcal{S}$, $\Big($i.e., $E_k \in \mathcal{B}_{\mathcal{S}}, \cup_{k=1}^n E_k = \mathcal{S}$, $E_i \cap E_k = \emptyset$(if $i \neq k$)$\Big)$. Let $E \in \mathcal{B}_{\mathcal{S}}$. Then

$$P_E(E_k) = \frac{P(E_k)P_{E_k}(E)}{P(E_1)P_{E_1}(E) + ... + P(E_n)P_{E_n}(E)},$$

where $P_E(E_k) = \frac{P(E \cap E_k)}{P(E)}$, $P_{E_k}(E) = \frac{P(E \cap E_k)}{P(E_k)}$.

The (B_2) is, of course, a mathematical theorem. Thus, when we use the (B_2), we must add a certain interpretation to the (B_2). In measurement theory, this is automatically done as follows:

$$(B_1) \quad = \quad (B_2) \quad + \quad \text{“measurement theoretical interpretation”}.$$

[(iii): The collapse (reduction) of wave packet in quantum mechanics]. The reduction such as (8.44) may happen even in quantum mechanics. In fact, it is called *“the collapse (reduction) of wave packet in quantum mechanics”.* Assume that a measured value obtained by a measurement $\mathbf{M}_{\mathcal{C}(V)}((X, \mathcal{F}, F), S(\rho))$ belongs to Ξ ($\in \mathcal{F}$). Then, we may see the following reduction (i.e., the collapse of wave packet):

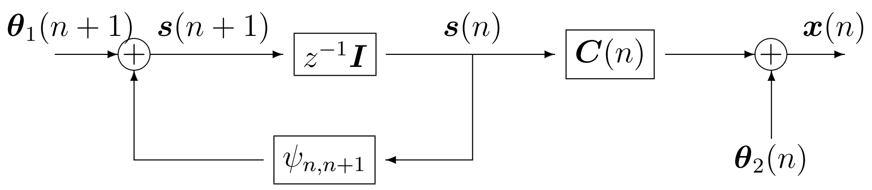
$$Tr_{+1}^m(V) \ni \rho \ (= \text{pretest state}) \mapsto \ (\text{posttest state} =) \frac{F(\Xi)\rho F(\Xi)}{\|F(\Xi)\rho F(\Xi)\|_{Tr(V)}} \in Tr_{+1}^m(V).$$

Note that, even in the case that $\rho = |u\rangle\langle u| \in Tr_{+1}^p(V)$, the above reduction happens (i.e., not invariant). However, I believe that the collapse of wave packet is due to a non-standard argument in quantum mechanics, though the collapse may be indispensable for the intuitive understanding of “quantum Zeno effect (*cf.* [65])”, etc. That is, I have an opinion that *from the pure theoretical point of view quantum mechanics says nothing after a measurement.* That is because, from the theoretical point of view, we always devote ourselves to the Heisenberg picture representation and not the Schrödinger picture representation. And further, it should be noted that the collapse of wave packet in quantum mechanics is not a direct consequence of MT (i.e., Axioms 1 and 2, Proclaim 1) $\Big($though the (8.44) (i.e., the classical reduction) is a consequence of Theorem 8.13 in MT$\Big)$. Thus, in this book we are not concerned with the collapse of wave packet in quantum mechanics.

■

8.4 Kalman filter in noise

As a consequence of Theorem 8.13 (and Theorem 6.13), in this section we reconsider Kalman filter [51], and formulate "Kalman filter" in SMT, which is proposed in [55]. Consider the conventional Kalman filter in the following system:



(Figure (8.45))

where $\boldsymbol{s}(n)$: L-dimensional state vector at time $n(=0,1,...,N)$, $\boldsymbol{x}(n)$: M-dimensional measured data vector, $(\omega \in \Omega)$. In the framework of dynamical system theory (2.1), $\boldsymbol{s}(n)$ and $\boldsymbol{x}(n)$ are described by the following equations: for each $\omega \in \Omega$ where $(\Omega, \mathcal{B}_\Omega, P)$ is a probability space,

$$\begin{cases} \boldsymbol{s}(n+1,\omega) = \psi_{n,n+1}(\boldsymbol{s}(n,\omega)) + \boldsymbol{\theta}_1(n+1,\omega) & \text{: stochastic difference state equation} \\ & \qquad (n=0,1,...,N-1). \\ \boldsymbol{x}(n,\omega) = \boldsymbol{C}(n)\boldsymbol{s}(n,\omega) + \boldsymbol{\theta}_2(n,\omega) & \text{: measurement equation} \end{cases} \tag{8.46}$$

Here, it is assumed that $\psi_{n,n+1}$, $\boldsymbol{C}(n)$, $\boldsymbol{\theta}_1(n,\cdot)$ (and its initial distribution) and $\boldsymbol{\theta}_2(n,\cdot)$ are known where $\psi_{n,n+1}$: $K \times K$-dimensional transition matrix, $\boldsymbol{\theta}_1(n,\cdot)$: L-dimensional input vector which represents a white noise, $\boldsymbol{C}(n)$: $L \times K$-dimensional measurement matrix, $\boldsymbol{\theta}_2(n,\cdot)$: L-dimensional vector which represents a measurement error. Here, our problem is as follows:

($\sharp$) Let τ be any integer such that $0 \le \tau \le N$. Let $\Xi_k \in \mathcal{B}_{\mathbf{R}}$ $(k=0,1,2,...,N)$. Then infer the state vector $\boldsymbol{s}(\tau,\omega)$ at time τ from the fact that

$$(\boldsymbol{x}(0,\omega), \boldsymbol{x}(1,\omega), \boldsymbol{x}(2,\omega), ..., \boldsymbol{x}(N,\omega)) \in \Xi_0 \times \Xi_1 \times \Xi_2 \times \cdots \times \Xi_N.$$

Also, note the original equation of the stochastic difference equation (8.46) is the following equation:

$$\bar{\boldsymbol{s}}(n+1) = \psi_{n,n+1}(\bar{\boldsymbol{s}}(n)) \quad (n=0,1,...,N-1). \tag{8.47}$$

The problem ($\sharp$) was firstly answered in the framework of dynamical system theory (8.46). Now, we consider the ($\sharp$) in the framework of SMT (8.3).

8.4.1 The measurement theoretical formulation of Figure (8.45)

Firstly, we formulate the (8.45) in SMT, (or HMT in Remark 8.3). Assume, for simplicity, that T ($\equiv \{0, 1, ..., N\}$) is a tree with a series structure (though this assumption is not needed). For each t ($\in T$), consider compact Hausdorff spaces $\mathcal{S}_t$ and Θ_t. Although it is natural to assume that $\mathcal{S}_0 = \mathcal{S}_1 = \cdots = \mathcal{S}_N$ and $\Theta_0 = \Theta_1 = \cdots = \Theta_N$, we can do well without this assumption. Now, consider the following two Markov relations among systems: $[\{\Psi_{t_1,t_2} : C(\mathcal{S}_{t_2}) \to C(\mathcal{S}_{t_1})\}_{(t_1,t_2)\in T^2_{\leqq}}]$ and $[\{\Upsilon_{t_1,t_2} : C(\Theta_{t_2}) \to C(\Theta_{t_1})\}_{(t_1,t_2)\in T^2_{\leqq}}]$ such as

$$[C(\mathcal{S}_0)] \xleftarrow{\Psi_{0,1}} [C(\mathcal{S}_1)] \xleftarrow{\Psi_{1,2}} \cdots \xleftarrow{\Psi_{N-2,N-1}} [C(\mathcal{S}_{N-1})] \xleftarrow{\Psi_{N-1,N}} [C(\mathcal{S}_N)] \tag{8.48}$$

where the initial state δ_{s_0} ($\in \mathcal{M}^p_{+1}(\mathcal{S}_0)$) is assumed to be unknown, and

$$[C(\Theta_0)] \xleftarrow{\Upsilon_{0,1}} [C(\Theta_1)] \xleftarrow{\Upsilon_{1,2}} \cdots \xleftarrow{\Upsilon_{N-2,N-1}} [C(\Theta_{N-1})] \xleftarrow{\Upsilon_{N-1,N}} [C(\Theta_N)]$$
$$\big(\text{with the known initial state } \nu_0^\Theta \ (\in \mathcal{M}^m_{+1}(\Theta_0))\big). \tag{8.49}$$

Here, it should be noted that the above (8.48) [resp. (8.49)] is the measurement theoretical formulation of (8.47) [resp. the $\boldsymbol{\theta}_1$ in (8.45)]. Also, note that the (8.48) is equivalent to

$$[\mathcal{M}^m_{+1}(\mathcal{S}_0)] \xrightarrow{\Psi^*_{0,1}} [\mathcal{M}^m_{+1}(\mathcal{S}_1)] \xrightarrow{\Psi^*_{1,2}} \cdots \xrightarrow{\Psi^*_{N-2,N-1}} [\mathcal{M}^m_{+1}(\mathcal{S}_{N-1})] \xrightarrow{\Psi^*_{N-1,N}} [\mathcal{M}^m_{+1}(\mathcal{S}_N)]$$

where $\Psi^*_{n,n+1} : \mathcal{M}^m_{+1}(\mathcal{S}_n) \to \mathcal{M}^m_{+1}(\mathcal{S}_{n+1})]$ is the dual operator of $\Psi_{n,n+1} : C(\mathcal{S}_{n+1}) \to C(\mathcal{S}_n)$. Since the (8.48) corresponds to the conventional (8.47), it is natural to assume that the (8.48) is deterministic, i.e., $\Psi_{n,n+1}$ is homomorphic. Thus, for each $n = 0, 1, ..., N-1$, there exists a continuous map $\psi_{n,n+1} : \mathcal{S}_n \to \mathcal{S}_{n+1}$, i.e.,

$$[\mathcal{S}_0] \xrightarrow{\psi_{0,1}} [\mathcal{S}_1] \xrightarrow{\psi_{1,2}} \cdots \xrightarrow{\psi_{N-2,N-1}} [\mathcal{S}_{N-1}] \xrightarrow{\psi_{N-1,N}} [\mathcal{S}_N]$$

where

$$f_{n+1}(\psi_{n,n+1}(s_n)) = (\Psi_{n,n+1}(f_{n+1}))(s_n) \quad (\forall f_{n+1} \in C(\mathcal{S}_{n+1}), \forall s_n \in \mathcal{S}_n).$$

Next, consider a continuous map $\lambda_n : \mathcal{S}_n \times \Theta_n \to \mathcal{S}_n$, that is,

$$\mathcal{S}_n \times \Theta_n \ni (s_n, \theta_n) \mapsto \lambda_n(s_n, \theta_n) \in \mathcal{S}_n \quad (n = 0, 1, ..., N) \tag{8.50}$$

which should be regarded as the corresponding thing of the left $\oplus$ in (8.45). The continuous map $\lambda_n : \mathcal{S}_n \times \Theta_n \to \mathcal{S}_n$ induces the continuous map $\Lambda_n : \mathcal{M}^m_{+1}(\mathcal{S}_n \times \Theta_n) \to \mathcal{M}^m_{+1}(\mathcal{S}_n)$

such that:

$$(\Lambda_n(\nu_n^{\mathcal{S}} \otimes \nu_n^{\Theta}))(B_n) = (\nu_n^{\mathcal{S}} \otimes \nu_n^{\Theta})(\lambda_n^{-1}(B_n))$$
$$(\forall (\nu_n^{\mathcal{S}} \otimes \nu_n^{\Theta}) \in \mathcal{M}_{+1}^m(\mathcal{S}_n \times \Theta_n), \forall B_n \subseteq \mathcal{S}_n : \text{open}). \tag{8.51}$$

Further, define the continuous map $\widehat{\Phi}_{n,n+1}^* : \mathcal{M}_{+1}^m(\mathcal{S}_n \times \Theta_n) \to \mathcal{M}_{+1}^m(\mathcal{S}_{n+1} \times \Theta_{n+1})$, such that

$$\mathcal{M}_{+1}^m(\mathcal{S}_n \times \Theta_n) \ni \nu_n^{\mathcal{S}} \otimes \nu_n^{\Theta} \mapsto \widehat{\Phi}_{n,n+1}^*(\nu_n^{\mathcal{S}} \otimes \nu_n^{\Theta})$$
$$\equiv [\Lambda_{n+1}(\Psi_{n,n+1}^* \nu_n^{\mathcal{S}} \otimes \Upsilon_{n,n+1}^* \nu_n^{\Theta})] \otimes \Upsilon_{n,n+1}^* \nu_n^{\Theta} \in \mathcal{M}_{+1}^m(\mathcal{S}_{n+1} \times \Theta_{n+1})$$

where $\Upsilon_{n,n+1}^* : \mathcal{M}_{+1}^m(\Theta_n) \to \mathcal{M}_{+1}^m(\Theta_{n+1})$ is a dual operator of $\Upsilon_{n,n+1} : C(\Theta_{n+1}) \to C(\Theta_n)$. That is,

$$\nu_{n+1}^{\mathcal{S}} \otimes \nu_{n+1}^{\Theta} \big(\equiv \widehat{\Phi}_{n,n+1}^*(\nu_n^{\mathcal{S}} \otimes \nu_n^{\Theta}) \big)$$
$$= [\Lambda_{n+1}(\Psi_{n,n+1}^* \nu_n^{\mathcal{S}} \otimes \Upsilon_{n,n+1}^* \nu_n^{\Theta})] \otimes \Upsilon_{n,n+1}^* \nu_n^{\Theta} \quad (n = 0, 1, ..., N-1) \tag{8.52}$$

which (or, the following (8.53)) corresponds to the state equation (8.46). Thus, we have the Markov relation $[\{\widehat{\Phi}_{n,n+1} : C(\mathcal{S}_{n+1} \times \Theta_{n+1}) \to C(\mathcal{S}_n \times \Theta_n)\}_{n=0}^{N-1}]$:

$$[C(\mathcal{S}_0 \times \Theta_0)] \xleftarrow{\widehat{\Phi}_{0,1}} [C(\mathcal{S}_1 \times \Theta_1)] \xleftarrow{\widehat{\Phi}_{1,2}} \cdots \xleftarrow{\widehat{\Phi}_{N-2,N-1}} [C(\mathcal{S}_{N-1} \times \Theta_{N-1})] \xleftarrow{\widehat{\Phi}_{N-1,N}} [C(\mathcal{S}_N \times \Theta_N)] \tag{8.53}$$

where $\widehat{\Phi}_{n,n+1}$ is the pre-dual operator of $\widehat{\Phi}_{n,n+1}^*$ (i.e., $(\widehat{\Phi}_{n,n+1})^* = \widehat{\Phi}_{n,n+1}^*$). That is, the (8.53) is equivalent to

$$[\mathcal{M}_{+1}^m(\mathcal{S}_0 \times \Theta_0)] \xrightarrow{\widehat{\Phi}_{0,1}^*} [\mathcal{M}_{+1}^m(\mathcal{S}_1 \times \Theta_1)] \xrightarrow{\widehat{\Phi}_{1,2}^*} \cdots [\mathcal{M}_{+1}^m(\mathcal{S}_{N-1} \times \Theta_{N-1})] \xrightarrow{\widehat{\Phi}_{N-1,N}^*} [\mathcal{M}_{+1}^m(\mathcal{S}_N \times \Theta_N)] \tag{8.53$'$}$$

Next, we consider the measurement theoretical characterization of the measurement equation (8.46). That is, consider the following Markov relation:

$$[C(\Theta_0')] \xleftarrow{\Upsilon_{0,1}'} [C(\Theta_1')] \xleftarrow{\Upsilon_{1,2}'} \cdots \xleftarrow{\Upsilon_{N-2,N-1}'} [C(\Theta_{N-1}')] \xleftarrow{\Upsilon_{N-1,N}'} [C(\Theta_N')]$$
$$\text{(with the initial state } \nu_0^{\Theta'} \ (\in \mathcal{M}_{+1}^m(\Theta_0'))),$$

which corresponds to the $\boldsymbol{\theta_2}$ in (8.46). Also, for each n ($\in T$), consider an observable $\mathbf{O}_n = (X_n, 2^{X_n}, F_n)$ in $C(\mathcal{S}_n \times \Theta_n')$, which corresponds to the measurement equation (8.46). Note that the observable $\mathbf{O}_n = (X_n, 2^{X_n}, F_n)$ in $C(\mathcal{S}_n \times \Theta_n')$ can be also regarded as an observable in $C(\mathcal{S}_n \times \Theta_n \times \Theta_n')$. Thus, we see that the (8.46) corresponds to the following:

$$\begin{matrix} [C(\mathcal{S}_0 \times \Theta_0 \times \Theta_0')] & \xleftarrow{\widehat{\widehat{\Phi}}_{0,1}} & [C(\mathcal{S}_1 \times \Theta_1 \times \Theta_1')] & \xleftarrow{\widehat{\widehat{\Phi}}_{1,2}} & \cdots & \xleftarrow{\widehat{\widehat{\Phi}}_{N-1,N}} & [C(\mathcal{S}_N \times \Theta_N \times \Theta_N')] \\ (X_0, 2^{X_0}, F_0) & & (X_1, 2^{X_1}, F_1) & & \cdots & & (X_N, 2^{X_N}, F_N) \end{matrix} \tag{8.54}$$

with the initial state $\delta_{s_0} \otimes \nu_0^{\Theta} \otimes \nu_0^{\Theta'}$ where $\widehat{\widehat{\Phi}}_{n,n+1} \equiv \widehat{\Phi}_{n,n+1} \otimes \Upsilon'_{n,n+1})$. Here, note that ν_0^{Θ} $(\in \mathcal{M}_{+1}^m(\Theta_0))$ and $\nu_0^{\Theta'}$ $(\in \mathcal{M}_{+1}^m(\Theta'_0))$ are known, but δ_{s_0} $(\in \mathcal{M}_{+1}^p(\mathcal{S}_0))$ is unknown. Therefore, we have the correspondence:

$$(8.46) \text{ in DST} \leftrightarrow (8.54) \text{ in SMT (or precisely, HMT, } cf. \text{ Remark 8.3).}$$

Thus, we can skip to the next section §8.4.2. However, in what follows we add the concrete form of the family $\{\mathbf{O}_n = (X_n, 2^{X_n}, F_n)\}_{n=0}^N$ (in (8.54)), which corresponds to the measurement equation (8.46) in detail.

Let $\mathcal{S}'_n$ and $\mathcal{S}''_n$ be compact spaces. Let $C : \mathcal{S}_n \to \mathcal{S}''_n$ be a continuous map, which induces the continuous map $\Lambda_n^C : \mathcal{M}_{+1}^m(\mathcal{S}_n) \to \mathcal{M}_{+1}^m(\mathcal{S}''_n)$ such that:

$$(\Lambda_n^C(\nu_n^{\mathcal{S}}))(A'_n) = \nu_n^{\mathcal{S}}((\lambda_n^C)^{-1}(A'_n)) \qquad (\forall \nu_n^{\mathcal{S}} \in \mathcal{M}_{+1}^m(\mathcal{S}_n), \forall A'_n \subseteq \mathcal{S}''_n : \text{open}).$$

And consider a continuous map $\lambda'_n : \mathcal{S}''_n \times \Theta'_n \to \mathcal{S}'_n$, which induces the continuous map $\Lambda'_n : \mathcal{M}_{+1}^m(\mathcal{S}''_n \times \Theta'_n) \to \mathcal{M}_{+1}^m(\mathcal{S}'_n)$ such that:

$$(\Lambda'_n(\nu_n^{\mathcal{S}''} \otimes \nu_n^{\Theta'}))(B'_n) = (\nu_n^{\mathcal{S}''} \otimes \nu_n^{\Theta'})((\lambda'_n)^{-1}(B'_n))$$
$$(\forall(\nu_n^{\mathcal{S}''} \otimes \nu_n^{\Theta'}) \in \mathcal{M}_{+1}^m(\mathcal{S}''_n \times \Theta'_n), \quad \forall B'_n \subseteq \mathcal{S}'_n : \text{open}).$$

For each n $(= 0, 1, ..., N)$, consider an observable $\mathbf{O}'_n = (X_n, 2^{X_n}, F'_n)$ in $C(\mathcal{S}'_n)$, which may be an (approximate) exact observable (*cf.* Example 2.20). Thus, for each n $(\in T)$, we can define the observable $\mathbf{O}_n = (X_n, 2^{X_n}, F_n)$ (in (8.54)) in $C(\mathcal{S}_n \times \Theta'_n)$ such that:

$${}_{C(\mathcal{S}_n \times \Theta'_n)^*}\langle \nu_n^{\mathcal{S}} \otimes \nu_n^{\Theta'}, F_n(\Xi_n) \rangle_{C(\mathcal{S}_n \times \Theta'_n)} = {}_{C(\mathcal{S}'_n)^*}\langle \Lambda'_n(\Lambda_n^C(\nu_n^{\mathcal{S}}) \otimes \nu_n^{\Theta'}), F'_n(\Xi_n) \rangle_{C(\mathcal{S}'_n)}$$
$$(\forall(\nu_n^{\mathcal{S}} \otimes \nu_n^{\Theta'}) \in \mathcal{M}_{+1}^m(\mathcal{S}_n \times \Theta'_n)).$$

8.4.2 Kalman filter in Noise

For simplicity, put $\widehat{\Theta}_n = \Theta_n \times \Theta'_n$ and $\nu_0^{\widehat{\Theta}} = \nu_0^{\Theta} \otimes \nu_0^{\Theta'}$. And, we rewrite the (8.54) as follows:

$$\begin{matrix} [C(\mathcal{S}_0 \times \widehat{\Theta}_0)] & \xleftarrow{\widehat{\widehat{\Phi}}_{0,1}} & [C(\mathcal{S}_1 \times \widehat{\Theta}_1)] & \xleftarrow{\widehat{\widehat{\Phi}}_{1,2}} & \cdots \xleftarrow{\widehat{\widehat{\Phi}}_{N-2,N-1}} & [C(\mathcal{S}_{N-1} \times \widehat{\Theta}_{N-1})] & \xleftarrow{\widehat{\widehat{\Phi}}_{N-1,N}} & [C(\mathcal{S}_N \times \widehat{\Theta}_N)] \\ (X_0, 2^{X_0}, F_0) & & (X_1, 2^{X_1}, F_1) & & \cdots & (X_{N-1}, 2^{X_{N-1}}, F_{N-1}) & & (X_N, 2^{X_N}, F_N) \end{matrix}$$

with the initial state $\delta_{s_0} \otimes \nu_0^{\widehat{\Theta}}$, where $\nu_0^{\widehat{\Theta}}$ $(\in \mathcal{M}_{+1}^m(\widehat{\Theta}_0))$ is known $\big($that is, ν_0^{Θ} $(\in \mathcal{M}_{+1}^m(\Theta_0))$ and $\nu_0^{\Theta'}$ $(\in \mathcal{M}_{+1}^m(\Theta'_0))$ are known$\big)$, but δ_{s_0} $(\in \mathcal{M}_{+1}^p(\mathcal{S}_0))$ is unknown.

Now, we get the sequential observable $[\mathbf{O}_T] \equiv [\{\mathbf{O}_t\}_{t\in T}; \{\widehat{\widehat{\Phi}}_{t_1,t_2} : C(\mathcal{S}_{t_2} \times \widehat{\Theta}_{t_2}) \to C(\mathcal{S}_{t_1} \times \widehat{\Theta}_{t_1})\}_{(t_1,t_2)\in T^2_{\leqq}}]$. Then, we can construct the observable $\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, 2^{\prod_{t\in T} X_t}, \widetilde{F}_0)$ in $C(\mathcal{S}_0 \times \widehat{\Theta}_0)$, which is the realization of the sequential observable $[\mathbf{O}_T]$, such as

$$\begin{array}{ccccccccc}
[C(\mathcal{S}_0 \times \widehat{\Theta}_0)] & \xleftarrow{\widehat{\widehat{\Phi}}_{0,1}} & [C(\mathcal{S}_1 \times \widehat{\Theta}_1)] & \xleftarrow{\widehat{\widehat{\Phi}}_{1,2}} & \cdots \xleftarrow{\widehat{\widehat{\Phi}}_{N-2,N-1}} & [C(\mathcal{S}_{N-1} \times \widehat{\Theta}_{N-1})] & \xleftarrow{\widehat{\widehat{\Phi}}_{N-1,N}} & [C(\mathcal{S}_N \times \widehat{\Theta}_N)] \\
F_0 & & F_1 & & \cdots & F_{N-1} & & F_N \\
\downarrow & & \downarrow & & & \downarrow & & \downarrow \\
\boxed{\begin{matrix}(F_0 \overset{\text{qp}}{\times} \widehat{\widehat{\Phi}} \widetilde{F}_1) \\ = \widetilde{F}_0\end{matrix}} & \xleftarrow{\widehat{\widehat{\Phi}}_{0,1}} & \begin{matrix}(F_1 \overset{\text{qp}}{\times} \widehat{\widehat{\Phi}} \widetilde{F}_2) \\ = \widetilde{F}_1\end{matrix} & \xleftarrow{\widehat{\widehat{\Phi}}_{1,2}} & \cdots \xleftarrow{\widehat{\widehat{\Phi}}_{N-2,N-1}} & \begin{matrix}(F_{N-1} \overset{\text{qp}}{\times} \widehat{\widehat{\Phi}} \widetilde{F}_N) \\ = \widetilde{F}_{N-1}\end{matrix} & \xleftarrow{\widehat{\widehat{\Phi}}_{N-1,N}} & \begin{matrix}(F_N) \\ = \widetilde{F}_N\end{matrix}
\end{array} \tag{8.55}$$

(The existence of the $\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, 2^{\prod_{t\in T} X_t}, \widetilde{F}_0)$ is assured by Theorem 3.7.) Thus, we can represent the "measurement" $\mathfrak{M}(\{\mathbf{O}_t\}_{t\in T}, \mathbf{S}(\delta_{s_0} \otimes \nu_0^{\widehat{\Theta}}))$ such as

$$\mathfrak{M}(\{\mathbf{O}_t\}_{t\in T}, \mathbf{S}(\delta_{s_0} \otimes \nu_0^{\widehat{\Theta}})) = \mathbf{M}_{C(\mathcal{S}_0 \times \widehat{\Theta}_0)}(\widetilde{\mathbf{O}}_0, S(\delta_{s_0} \otimes \nu_0^{\widehat{\Theta}})).$$

Here, assume that

($\sharp$) we know that the measured value $(x_t)_{t\in T}$ $(\in \prod_{t\in T} X_t)$, obtained by the measurement $\mathbf{M}_{C(\mathcal{S}_0 \times \widehat{\Theta}_0)}(\widetilde{\mathbf{O}}_0, S(\delta_{s_0} \otimes \nu_0^{\widehat{\Theta}}))$, belongs to $\prod_{t\in T} \Xi_t$.

Fisher's maximum likelihood method (*cf.* Theorem 5.3, Corollary 5.6) says that there is a reason to infer that the unknown s_0 $(\in \mathcal{S}_0)$ is determined by

$${}_{C(\mathcal{S}_0\times\widehat{\Theta}_0)^*}\langle \delta_{s_0} \otimes \nu_0^{\widehat{\Theta}}, \widetilde{F}_0(\prod_{t\in T} \Xi_t)\rangle_{C(\mathcal{S}_0\times\widehat{\Theta}_0)} = \max_{s\in\mathcal{S}_0} {}_{C(\mathcal{S}_0\times\widehat{\Theta}_0)^*}\langle \delta_s \otimes \nu_0^{\widehat{\Theta}}, \widetilde{F}_0(\prod_{t\in T} \Xi_t)\rangle_{C(\mathcal{S}_0\times\widehat{\Theta}_0)}.$$

Let $\tau \in T$, and let $\{B^{(0,\tau)}_{\prod_{t\in T}\Xi_t} \mid \prod_{t\in T} \Xi_t \in 2^{\prod_{t\in T} X_t}\}$ be a family of Bayes operators. (The existence is assured by Theorem 6.6.) Then, we see, by Lemma 8.9, that the new S-state $\nu^{\mathcal{S}\times\widehat{\Theta}_\tau}_{\tau,\text{new}}$ $(\in \mathcal{M}^m_{+1}(\mathcal{S}_\tau \times \widehat{\Theta}_\tau))$ is defined by

$$\nu^{\mathcal{S}\times\widehat{\Theta}_\tau}_{\tau,\text{new}} = R^{(0,\tau)}_{\prod_{t\in T}\Xi_t}(\delta_{s_0} \otimes \nu_0^{\widehat{\Theta}})$$

where $R^{(0,\tau)}_{\prod_{t\in T}\Xi_t} : \mathcal{M}^m_{+1}(\mathcal{S}_0 \times \widehat{\Theta}_0) \to \mathcal{M}^m_{+1}(\mathcal{S}_\tau \times \widehat{\Theta}_\tau)$ is a normalized dual Bayes operator, i.e., $R^{(0,\tau)}_{\prod_{t\in T}\Xi_t}(\nu) = \frac{(B^{(0,\tau)}_{\prod_{t\in T}\Xi_t})^*(\nu)}{\|(B^{(0,\tau)}_{\prod_{t\in T}\Xi_t})^*(\nu)\|}$ $(\forall \nu \in \mathcal{M}^m_{+1}(\mathcal{S}_0 \times \widehat{\Theta}_0))$. Thus there is a reason to think that the new S-state (in $\mathcal{M}^m_{+1}(\mathcal{S}_\tau)$) is equal to $\nu^{\mathcal{S}}_{\tau,\text{new}}$ such that:

$$\nu^{\mathcal{S}_\tau}_{\tau,\text{new}}(D_\tau) \equiv \nu^{\mathcal{S}_\tau\times\widehat{\Theta}_\tau}_{\tau,\text{new}}(D_\tau \times \widehat{\Theta}_\tau) \quad (\forall D_\tau (\subseteq \mathcal{S}_\tau) : \text{open set}).$$

Remark 8.15. [Stochastic differential equation] It is important to generalize the stochastic difference state equation in (8.46) to the stochastic differential equation (1.2a). In order to do it in SMT, we must prepare the W^*-algebraic formulation of SMT (in Chapter 9). Thus we do not touch this problem in this book.

8.5 Information and entropy

As one of applications (of Bayes theorem), we now study the "entropy" of the measurement. Here we have the following definition.

Definition 8.16. [Information quantity, the entropy of measurement (= fuzzy entropy), *cf.* [42]]. *Consider a statistical measurement* $\mathbf{M}_{C(\Omega)}\big(\mathbf{O} \equiv (X, 2^X, F),\ S(\rho_0)\big)$ *in a commutative* C^**-algebra* $C(\Omega)$*, where the label set* X *is assumed to be at most countable, i.e.,* $X = \{x_1, x_2, ..., x_n, ...\}$*. Then, the* $H(\mathbf{M})$*, the (fuzzy) entropy of* $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\rho_0))$*, is defined by*

$$
\begin{aligned}
&H\Big(\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\rho_0))\Big) \\
&= \sum_{n=1}^{\infty} \Big(\int_\Omega [F(\{x_n\})](\omega)\rho_0(d\omega) \int_\Omega \frac{[F(\{x_n\})](\omega)}{\int_\Omega [F(\{x_n\})](\omega)\rho_0(d\omega)} \log \frac{[F(\{x_n\})](\omega)}{\int_\Omega [F(\{x_n\})](\omega)\rho_0(d\omega)} \rho_0(d\omega)\Big) \qquad (8.56) \\
&= \sum_{n=1}^{\infty} \cdot I(\{x_n\})
\end{aligned}
$$

$$
\text{where, } P(\{x_n\}) = \int_\Omega [F(\{x_n\})](\omega)\rho_0(d\omega)
$$

$$
\Big(= \text{ the probability that a measured value } x_n \text{ is obtained}\Big)
$$

$$
\begin{aligned}
I(\{x_n\}) &= \int_\Omega \frac{[F(\{x_n\})](\omega)}{\int_\Omega [F(\{x_n\})](\omega)\rho_0(d\omega)} \log \frac{[F(\{x_n\})](\omega)}{\int_\Omega [F(\{x_n\})](\omega)\rho_0(d\omega)} \rho_0(d\omega) \\
&= \frac{1}{P(\{x_n\})} \int_\Omega [F(\{x_n\})](\omega) \log [F(\{x_n\})](\omega)\rho_0(d\omega) - \log P(\{x_n\}) \\
&\Big(= \text{ the information quantity when a measured value } x_n \text{ is obtained}\Big)
\end{aligned} \qquad (8.57)
$$

$\mathbf{M}_{C(\Omega)}\big(\mathbf{O}, S(\rho_0)\big)$ *is the normalized* W^**-algebraic representation of a* C^**-measurement* $\mathbf{M}_{C_0(\Omega)}\big(\mathbf{O} \equiv (X, \mathcal{P}_0(X), F),\ S(\rho_0)\big)$*, the entropy* $H\Big(\mathbf{M}_{C_0(\Omega)}(\mathbf{O}, S(\rho_0))\Big)$ *is also defined by* $H\Big(\mathbf{M}_{C(\Omega)}\ (\mathbf{O}, S(\rho_0))\Big)$.

■

The definition is derived from the following consideration. Assume that we get the measured value x ($\in X$) by the statistical measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\rho_0))$. Note that its probability $P(\{x\})$ is given by $P(\{x\}) = {}_{C(\Omega)^*}\Big\langle \rho_0, F(\{x\}) \Big\rangle_{C(\Omega)} = \int_\Omega [F(\{x\})](\omega)\ \rho_0(d\omega)$. Also, we consider, by (8.44) (or, (5.13)), that the new statistical state $\overline{\rho}_x$ $(\in \mathcal{M}^m_{+1}(\Omega))$ is given by

$$\overline{\rho}_x(D) = \frac{\int_D [F(\{x\})](\omega)\rho_0(d\omega)}{\int_\Omega [F(\{x\})](\omega)\rho_0(d\omega)} \qquad (\forall D \in \mathcal{B}_\Omega),$$

whose information quantity $I(x)$ is of course determined by $I(\{x\}) = \int_\Omega \frac{d\overline{\rho}_x}{d\rho_0}(\omega) \log \frac{d\overline{\rho}_x}{d\rho_0}\rho_0(d\omega)$, where the Radon-Nikodým derivative $\frac{d\overline{\rho}_x}{d\rho_0}(\omega)$ is defined by $\frac{[F(\{x\})](\omega)}{\int_\Omega [F(\{x\})](\omega)\rho_0(d\omega)}$. Thus, the average information quantity, i.e., entropy, is given by

$$H\Big(\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\rho_0))\Big) = \sum_{n=1}^{\infty} P(\{x_n\}) \cdot I(\{x_n\}),$$

which is equal to (8.56). Also it should be noted that the formula (8.56) can easily calculated as follows:

$$H(\mathbf{M}) = \sum_{n=1}^{\infty} \int_\Omega [F(\{x_n\})](\omega) \log [F(\{x_n\})](\omega) \rho_0(d\omega) - \sum_{n=1}^{\infty} P(\{x_n\}) \log P(\{x_n\}). \quad (8.58)$$

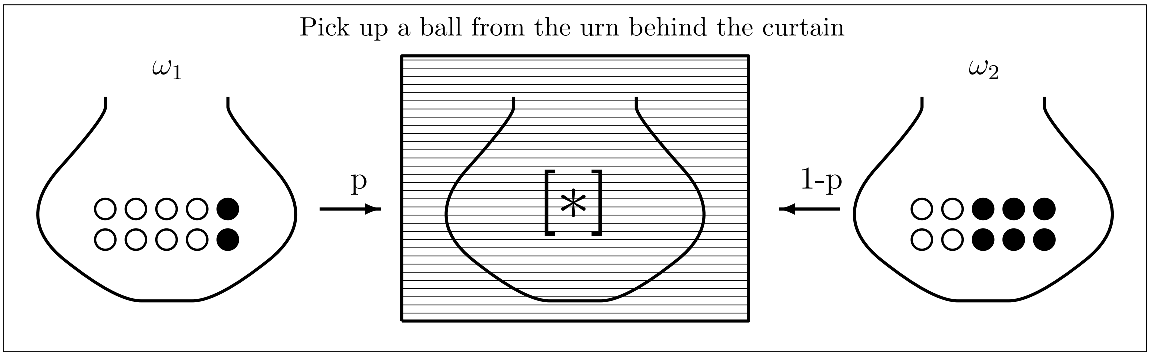
Also, if $\mathbf{O}$ is crisp, we see that $H(\mathbf{M}) = -\sum_{n=1}^{\infty} P(\{x_n\}) \log P(\{x_n\})$.

Example 8.17. [Urn problem (in Example 8.1)]. There are two urns ω_1 and ω_2. The urn ω_1 [resp. ω_2] contains $8N$ white and $2N$ black balls [resp. $4N$ white and $6N$ black balls], where N is a sufficiently large number. We regard Ω $\big(\equiv \{\omega_1, \omega_2\}\big)$ as the state space. And consider the observable $\mathbf{O}\big(\equiv (X \equiv \{w, b\}, 2^{\{w,b\}}, F)\big)$ in $C(\Omega)$ where

$$[F(\{w\})](\omega_1) = 0.8, \qquad [F(\{b\})](\omega_1) = 0.2,$$
$$[F(\{w\})](\omega_2) = 0.4, \qquad [F(\{b\})](\omega_2) = 0.6.$$

Here define the statistical state $\nu_0 (\in \mathcal{M}^m_{+1}(\Omega))$ such that $\nu_0(\{\omega_1\}) = p$, $\nu_0(\{\omega_2\}) = 1-p$. And consider a statistical measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0))$.

The illustration of $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0))$



Put

$$
\begin{aligned}
P(\{x\}) &: \text{ the probability that a measured value } x \ (\in \{w,b\}) \text{ is obtained}\\
I(\{x\}) &: \text{ the information quantity that is acquired when we know that}\\
&\quad\ \text{a measured value } x \ (\in \{w,b\}) \text{ is obtained}\\
\nu_1^x &: \text{ the posttest state after a measured value } x \ (\in \{w,b\}) \text{ is obtained}
\end{aligned}
$$

Then,

$$
P(\{w\}) = 0.8p + 0.4(1-p), \qquad P(\{b\}) = 0.2p + 0.6(1-p),
$$

$$
\begin{aligned}
I(\{w\}) &= \frac{0.8p \log 0.8 + 0.4(1-p)\log 0.4}{0.8p + 0.4(1-p)} - \log(0.8p + 0.4(1-p)),\\
I(\{b\}) &= \frac{0.2p \log 0.2 + 0.6(1-p)\log 0.6}{0.2p + 0.6(1-p)} - \log(0.2p + 0.6(1-p)),
\end{aligned}
$$

$$
\begin{aligned}
\nu_1^w(\{\omega_1\}) &= \frac{0.8p}{0.8p+0.4(1-p)}, \qquad \nu_1^w(\{\omega_2\}) = \frac{0.4(1-p)}{0.8p+0.4(1-p)},\\
\nu_1^b(\{\omega_1\}) &= \frac{0.2p}{0.2p+0.6(1-p)}, \qquad \nu_1^b(\{\omega_2\}) = \frac{0.6(1-p)}{0.2p+0.6(1-p)}.
\end{aligned}
$$

Then, we see, by (8.58), that

$$
\begin{aligned}
&H(\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0)))\\
=&[F(\{w\})](\omega_1)\log[F(\{w\})](\omega_1)p + [F(\{w\})](\omega_2)\log[F(\{w\})](\omega_2)(1-p)\\
&+ [F(\{b\})](\omega_1)\log[F(\{b\})](\omega_1)p + [F(\{b\})](\omega_2)\log[F(\{b\})](\omega_2)(1-p)\\
&- P(\{w\})\log P(\{w\}) - P(\{b\})\log P(\{b\})\\
=&0.8(\log 0.8)p + 0.4(\log 0.4)(1-p) + 0.2(\log 0.2)p + 0.6(\log 0.6)(1-p)\\
&- (0.8p + 0.4(1-p))\log(0.8p+0.4(1-p)) - (0.2p+0.6(1-p))\log(0.2p+0.6(1-p)).
\end{aligned}
$$

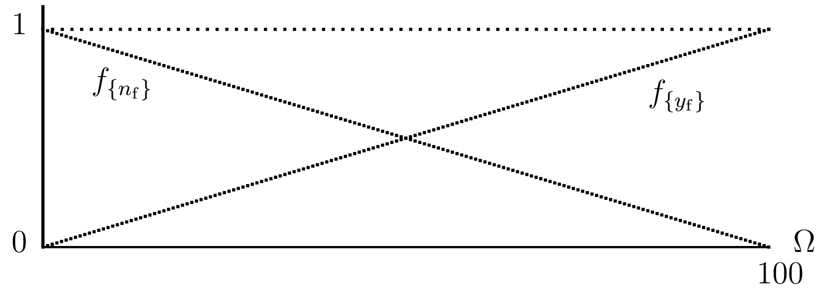
Assume that $p = 1/2$. Then, we see that

$$
H(\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0))) = 0.6 - 0.3\log_2 3 = 0.123\cdots(\text{bit}).
$$

■

Example 8.18. [Fuzzy information (fast or not fast), *cf.* [42]]. Let $\Omega \equiv \{\omega_1, \omega_2, ..., \omega_{100}\}$ be a set of pupils in some school. Let $\mathbf{O}_{\rm b} \equiv (X = \{y_{\rm b}, n_{\rm b}\}, 2^X, b_{(\cdot)})$ be the crisp C^*-observable in the commutative C^*-algebra $C(\Omega)$ such that $b_{\{y_{\rm b}\}}(\omega_n) = 0$ (n is odd), $= 1$

(n is even), and $b_{\{n_{\rm b}\}}(\omega_n) = 1 - b_{\{y_{\rm b}\}}(\omega_n)$. Also, let $\mathbf{O}_{\rm f} \equiv (Y = \{y_{\rm f}, n_{\rm f}\}, 2^Y, f_{(\cdot)})$ be the C^*-observable in C^*-algebra $C(\Omega)$ such that $f_{\{y_{\rm f}\}}(\omega_n) = (n-1)/99$ $(\forall \omega_n \in \Omega)$ and $f_{\{n_{\rm f}\}}(\omega_n) = 1 - f_{\{y_{\rm f}\}}(\omega_n)$. Let $\rho_0 \in \mathcal{M}_{+1}^m(\Omega)$, for example, assume that $\rho_0 = \nu_{\rm u}$, i.e., the equal weight on Ω, namely, $\nu_{\rm u}(\{\omega_n\}) = 1/100$ $(\forall n)$. Thus we have two measurements $\mathbf{M}_{C(\Omega)}(\mathbf{O}_{\rm b}, S(\nu_u))$ and $\mathbf{M}_{C(\Omega)}(\mathbf{O}_{\rm f}, S(\nu_u))$.



Then, we see, by (8.58), that

$$
\begin{aligned}
& H\Big(\mathbf{M}_{C(\Omega)}(\mathbf{O}_{\rm b}, S(\nu_u))\Big) \\
& = -\|b_{\{y_{\rm b}\}}\|_{L^1(\Omega,\nu_u)} \log \|b_{\{y_{\rm b}\}}\|_{L^1(\Omega,\nu_u)} - \|b_{\{n_{\rm b}\}}\|_{L^1(\Omega,\nu_u)} \log \|b_{\{n_{\rm b}\}}\|_{L^1(\Omega,\nu_u)} \\
& = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = \log_2 2 = 1 \ \ (\text{bit}), \qquad (8.59)
\end{aligned}
$$

and

$$
\begin{aligned}
& H\Big(\mathbf{M}_{C(\Omega)}(\mathbf{O}_{\rm f}, S(\nu_u))\Big) \\
& = \int_\Omega f_{\{y_{\rm f}\}}(\omega) \log f_{\{y_{\rm f}\}}(\omega) \nu_u(d\omega) + \int_\Omega f_{\{n_{\rm f}\}}(\omega) \log f_{\{n_{\rm f}\}}(\omega) \nu_u(d\omega) \\
& \quad - \|f_{\{y_{\rm f}\}}\|_{L^1(\Omega,\nu_u)} \log \|f_{\{y_{\rm f}\}}\|_{L^1(\Omega,\nu_u)} - \|f_{\{n_{\rm f}\}}\|_{L^1(\Omega,\nu_u)} \log \|f_{\{n_{\rm f}\}}\|_{L^1(\Omega,\nu_u)} \qquad (8.60) \\
& \approx 2\int_0^1 \lambda \log_2 \lambda d\lambda + 1 = -\frac{1}{2\log_e 2} + 1 = 0.278\cdots \ \ (\text{bit}). \qquad (8.61)
\end{aligned}
$$

For example, assume that the symbol "$y_{\rm b}$" [resp. "$n_{\rm b}$"] in X is interpreted by "boy" [resp. "girl"]. And "$y_{\rm f}$" [resp. "$n_{\rm f}$"] in Y is interpreted by "fast runner" [resp. "not fast runner"]. When we guess the pure state $(*)$ of the system S $(= S_{(*)}(\nu_u))$ in the above situation, the (8.60) and (8.61) say that the crisp information "boy or girl" is more efficient than the fuzzy information "fast or not fast".

■

Remark 8.19. [Fuzzy information theory]. "Shannon's entropy" is usually defined as follows (*cf.* [79]). Let $(\Omega, \mathcal{B}, P)$ be a probability space. Let $\mathbf{D} = \{D_1, D_2, ...\}$ be the countable decomposition of Ω. Then, the entropy $H(\mathbf{D})$ of $\mathbf{D}$ is defined by $H(\mathbf{D}) = -\sum_{n=1}^{\infty} P(D_n) \log P(D_n)$. Note that Definition 8.16 is the natural extension of Shannon's entropy if we regard the observable $\mathbf{O}$ as a "fuzzy decomposition" (*cf.* the formula (2.30)).

■

8.6 Belief measurement theory (=BMT)

In this section we study "belief measurement theory (=BMT)", which is considered to be closely related to "subjective Bayesian statistics".[9]

Firstly let us consider the following problem:

(P) For example, consider a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (X, 2^X, F), S_{[*]})$ formulated in $C(\Omega)$, where $\Omega = \{\omega_1, \omega_2\}$, and further, assume that we have no information about the $[*]$. How do we represent "having no information about the $[*]$" mathematically? Or, how do we infer the statistical state?

We prepare three answers to the problem (P) in this book. That is, we consider three kinds of "having no information about the $[*]$" (or, "having no belief whether $[*] = \omega_1$ or $[*] = \omega_2$) as follows:

(A_1) Iterative likelihood function method in PMT. See $\mathbf{M}_{C(\Omega)}(\mathbf{O},\ S_{[*]}((kI))_{lq})$ in §5.5.

(A_2) The principle of equal probability (= "PEP"). As seen later (i.e., Theorem 11.12), this is essentially equivalent to the hypothesis that the $[\,*\,]$ is chosen by a fair coin-tossing (e.g., $p = 0.5$ in (8.7)). That is, it suffices to consider the statistical measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O},\ S_{[*]}(\nu_u))$, where $\nu_u(\{\omega_1\}) = \nu_u(\{\omega_2\}) = 1/2$.

(A_3) The principle of equal weight(="PEW" =Bayes' postulate). See §8.6.2 later. This method will be called "belief measurement theory" (or, "BMT").

[9] This is not sure since my understanding of the subjective Bayesian statistics (*cf.* [21]) is not sufficient.

Thus we may have the following classification (and correspondence):

$$\text{MT}\left\{\begin{array}{lll} \text{PMT} = \underset{\text{[Axiom 1 (2.37)]}}{\text{measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} & \longleftrightarrow (A_1) \\ \text{SMT} = \underset{\text{(Axioms 1 and 2)}}{\text{PMT}} + \underset{\text{(the probabilistic interpretation of mixed state)}}{\text{“statistical state”}} & \longleftrightarrow (A_2) \\ \text{BMT} = \underset{\text{(Axioms 1 and 2)}}{\text{PMT}} + \underset{\text{(the principle of equal weight)}}{\text{“belief weight”}} & \longleftrightarrow (A_3) \end{array}\right. \tag{8.62}$$

8.6.1 The general argument about BMT

In §8.1∼§8.5, we studied SMT (i.e., Proclaim 1 (= the probabilistic interpretation of “mixed state”) + Axiom 2), in which “mixed state” has the probabilistic interpretation. In this section, we propose another interpretation of “mixed state”, which may be called “belief interpretation”. That is, we want to assert:

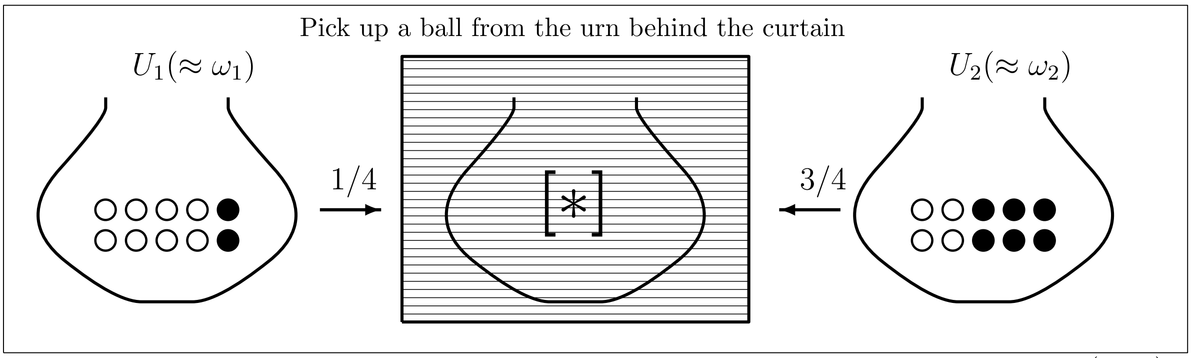
$$\underset{\text{(mixed state)}}{\rho^m \in \mathfrak{S}^m(\mathcal{A}^*)} \cdots \left\{\begin{array}{ll} \underset{\text{[Proclaim 1 (8.10)]}}{\text{“probabilistic interpretation”}} & \longrightarrow \underset{\text{in §8.1∼8.5}}{\text{“SMT”}} \\ \underset{\text{[the principle of equal weight (8.72)]}}{\text{“belief interpretation”}} & \longrightarrow \underset{\text{in this §8.6}}{\text{“BMT”}} \end{array}\right. \tag{8.63}$$

The purpose of this section is, of course, to propose “belief measurement theory” (or, “BMT”).

We begin with a simplest example as follows. Consider the statistical measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (\{w,b\}, 2^{\{w,b\}}, F), S_{[*]}(\nu_0))$. Here

$$[F(\{w\})](\omega_1) = 0.8, \quad [F(\{b\})](\omega_1) = 0.2, \quad [F(\{w\})](\omega_2) = 0.4, \quad [F(\{b\})](\omega_2) = 0.6,$$

and, $\nu_0(\{\omega_1\}) = 1/4$ and $\nu_0(\{\omega_1\}) = 3/4$. Recall that this measurement is symbolically described as follows.

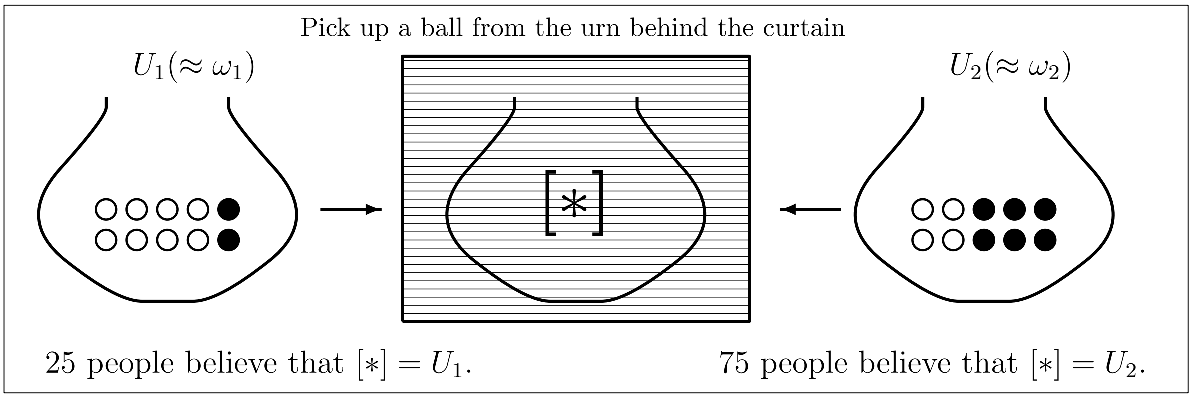


Figure(8.64)

By a hint of the Figure (8.64), we can introduce "BMT" as follows. Assume that there are 100 people. And moreover assume that[10]

$$\begin{cases} \text{25 people (in 100 people) believe that } [*] = U_1 \\ \text{75 people (in 100 people) believe that } [*] = U_2 \end{cases}$$

That is, we have the following picture (instead of Figure (8.64)):



Figure(8.65)

This is just the "belief measurement", which is denoted by $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}((\nu_0))_{bw})$. Also, the ν_0 is called a *belief weight* (or, *approval rate, conviction degree*).[11]

We add the following remark:

(R_1) Note that the $[*]$ (in $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}((\nu_0))_{bw})$) is assumed to be unknown. Thus, the triplet $\left(X, 2^X, {}_{\mathcal{M}(\Omega)}\left\langle \nu, F(\cdot) \right\rangle_{C(\Omega)}\right)$ is a merely mathematical symbol and not a sample space. In other words, it is nonsense to consider the probability that the measured value obtained by $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (X, 2^X, F), S_{[*]}((\nu))_{bw})$ belongs to $\Xi(\in 2^X)$. That is, Proclaim 1(8.10) does not hold for a belief measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (X, 2^X, F), S_{[*]}((\nu))_{bw})$, or equivalently, a belief measurement has no sample space.

This (R_1) is clear. That is because the argument mentioned in Example 8.1 is invalid for a belief measurement, since ν (in $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}((\nu))_{bw})$) is a belief weight and not a statistical state.

However (i.e., in spite of the fact that Proclaim 1(8.10) is invalid), we have the following theorem:

Theorem 8.20. (Bayes theorem for belief measurements). *Assume that we know that a measured value obtained by a belief measurement* $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[*]}((\nu))_{bw})$

[10]Recall "parimutuel betting", which is very applicable. For example, we may consider *the "probability" that life exists on Mars.*

[11]Thus, outsiders may think that $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}((\nu_0))_{bw})$ and $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$ are the same. That is because the number of the believers is not related to the measurement itself.

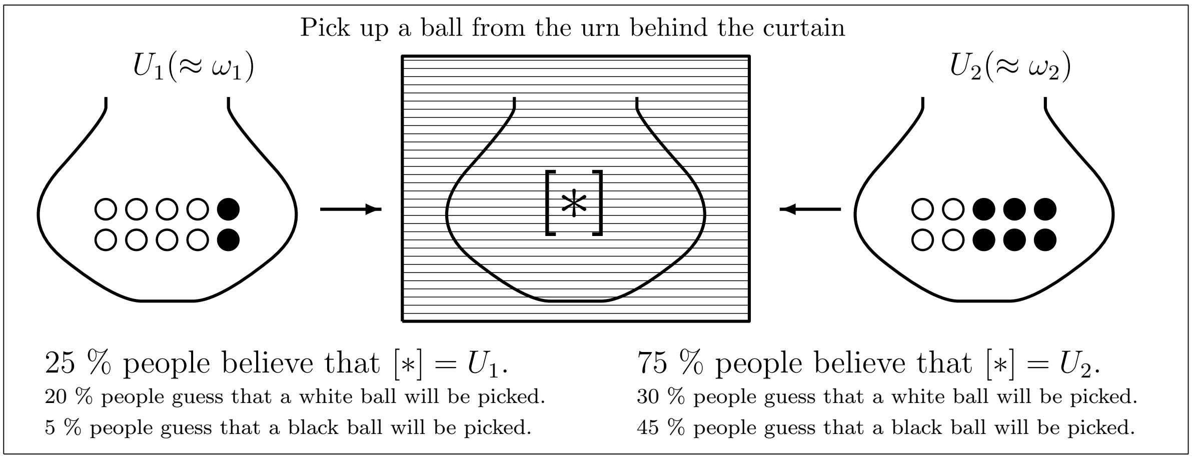
belongs to Ξ $(\in \mathcal{F})$. *Then, we have the "Bayes theorem" such that*

$$\mathcal{M}_{+1}^m(\Omega) \ni \nu (= \text{priori belief weight}) \mapsto (\text{posterior belief weight} =) R_{\Xi}^{(0,0)}(\nu) \in \mathcal{M}_{+1}^m(\Omega). \tag{8.66}$$

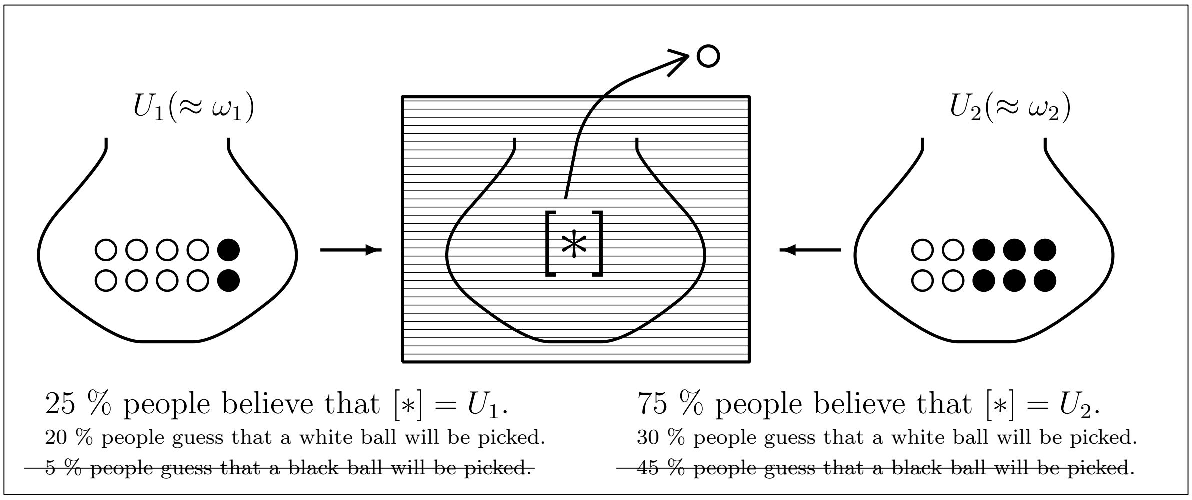
where

$$[R_{\Xi}^{(0,0)}(\nu)](D_0) = \frac{\int_{D_0} [F(\Xi)](\omega)\nu(d\omega)}{\int_{\Omega} [F(\Xi)](\omega)\nu(d\omega)} \qquad (\forall D_0 \subseteq \Omega;\ \text{Borel set}). \tag{8.67}$$

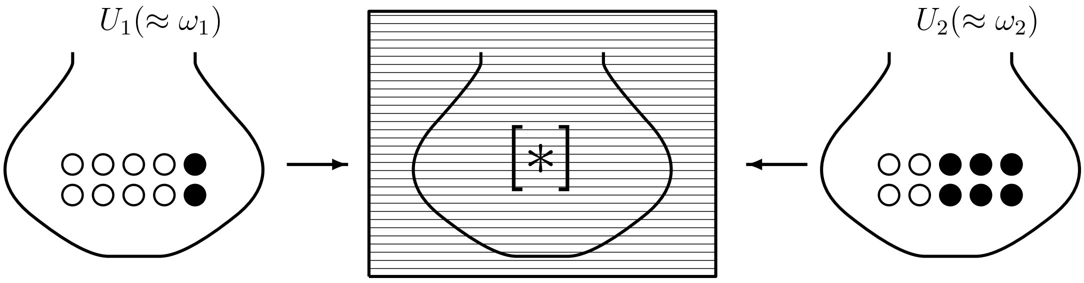
Proof. It suffices to prove a simple case since the proof of the general case is similar. For example, consider the following figure, which is essentially the same as Figure (8.65).



Assume that a "*white ball*" is picked in the above picture. Then, we see:



which is equivalent to the following figure:



40 % people believe that $[*] = U_1$. 60 % people believe that $[*] = U_2$.

Thus we see that Bayes theorem holds for belief measurements. That is because Theorem 8.20 (Bayes theorem for belief measurements) says:

$$\mathcal{M}_{+1}^m(\Omega) \ni \nu_0(= \text{priori belief weight}) \mapsto (\text{posterior belief weight} =) R_{\Xi}^{(0,0)}(\nu_0) \in \mathcal{M}_{+1}^m(\Omega). \quad (8.68)$$

where

$$[R_{\{w\}}^{(0,0)}(\nu_0)](\{\omega\}) = \frac{\int_{\{\omega\}}[F(\{w\})](\omega)\nu_0(d\omega)}{\int_{\Omega}[F(\{w\})](\omega)\nu_0(d\omega)} = \begin{cases} \frac{\frac{8}{10}\times\frac{1}{4}}{\frac{8}{10}\times\frac{1}{4}+\frac{4}{10}\times\frac{3}{4}} = \frac{40}{100} & (\text{if } \omega = \omega_1) \\ \frac{\frac{4}{10}\times\frac{3}{4}}{\frac{8}{10}\times\frac{1}{4}+\frac{4}{10}\times\frac{3}{4}} = \frac{60}{100} & (\text{if } \omega = \omega_2) \end{cases} \quad (8.69)$$

Although this proof is easy, it should be noted that this is different from the proof of Bayes theorem for a statistical measurement. That is because Proclaim 1 (8.20) can not be used in the proof of Theorem 8.26.

□

Remark 8.21. (Extensive interpretation in theoretical informatics). Seeing Figure (8.65), some may think that the belief weight ν (in $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[*]}((\nu))_{bw})$ represents the only "public opinion". However, this is wrong. Recall the spirit of theoretical informatics (in the footnote below the statement (1.12) in Chapter 1), i.e., "extensive interpretation". Thus, we consider that the belief weight ν (in $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[*]}((\nu))_{bw})$ often represents "personal belief".

■

8.6.2 The principle of equal weight

As mentioned in the previous section (i.e., §8.6.1) we have the following notation:

Notation 8.22. $[\mathbf{M}_{C(\Omega)}(\mathbf{O},\ S_{[*]}(\!(\nu)\!)_{bw})\ ]$. *The symbol* $\mathbf{M}_{C(\Omega)}(\mathbf{O},\ S_{[*]}(\!(\nu)\!)_{bw})$, *(*$\nu \in \mathcal{M}_{+1}^m(\Omega)$*), is assumed to represent the measurement* $\mathbf{M}_{C(\Omega)}\ (\mathbf{O} \equiv (X, \mathcal{F}, F),\ S_{[*]})$ *under the hypothesis that the belief weight of the system* $S_{[*]}$ *is* ν*. And it is called a belief measurement.*

■

Now let us explain "Bayes postulate" (= "the principle of equal weight"). Assume that Ω is finite (i.e., $\Omega = \{\omega_1, \omega_2, ..., \omega_N\}$). Then, there is a reason to think that the mixed state ν_u ($\in \mathcal{M}_{+1}^m(\Omega)$) defined by

$$\nu_u(D) = \frac{\sharp[D]}{N} \qquad (\forall D \subseteq \Omega) \tag{8.70}$$

represents "the loosest belief" or "knowing nothing about $S_{[*]}$". (The ν_u is called the *"equal weight."* Cf. Remark 8.23 later). If Ω is infinite, we have no firm opinion.[12] Thus in this section we always assume that Ω is finite.

We add the following remark.

Remark 8.23. [Mathematical properties of equal weight ν_u, [42]]. Let $\Omega \equiv \{\omega_1, \omega_2, ..., \omega_N\}$ be a finite set with the discrete topology. Let ρ_0^m be arbitrary belief weight (i.e., $\rho_0^m \in \mathcal{M}_{+1}^m(\Omega)$). Then, define the entropy $H(\rho_0^m)$ of the ρ_0^m by

$$H\Big(\rho_0^m\Big) = -\sum_{n=1}^{N} \rho_0^m(\{\omega_n\}) \log \rho_0^m(\{\omega_n\}).$$

Here, it is well known that

(i) $\sup\Big\{ H\Big(\rho_0^m\Big) : \rho_0^m \in \mathcal{M}_{+1}^m(\Omega) \Big\} = \log N,$ (8.71)

(ii) "$\rho_0^m(\{\omega_n\}) = 1/N (\forall n)$" $\Longleftrightarrow$ "$H(\rho_0^m) = \log N$".

(iii) Let $T_{av} : C(\Omega) \to \mathbf{C}$ be the *average functional on* $C(\Omega)$*, i.e., a linear positive functional* such that:

(a) $T_{av}(1) = 1$

(b) $T_{av}(f) = T_{av}(f \circ \phi) \qquad (\forall f \in C(\Omega), \forall \text{ bijection } \phi : \Omega \to \Omega)$

where $(f \circ \phi)(\omega) = f(\phi(\omega))$.

---

[12] For example, we may consider as follows: Let Ω be not finite. Let $\mathcal{S}_\Omega$ be a subset of $\{\Phi \mid \Phi : C(\Omega) \to C(\Omega)$ is a Markov operator $\}$. Assume that the $\mathcal{S}_\Omega$ has the unique invariant state ν_u ($\in \mathcal{M}_{+1}^m(\Omega)$), that is, $\Phi^* \nu_u = \nu_u$ ($\forall \Phi \in \mathcal{S}_\Omega$). And further assume that $\nu_u(U) > 0$ ($\forall U (\subseteq \Omega$, open). Then, we may say that the ν_u represents "no belief weight (concerning $\mathcal{S}_\Omega$)" or "completely shuffled weight." Also, see [47].

(iv) T_{av} is uniquely determined such as $T_{av}(f) = \int_\Omega f(\omega)\nu_u(d\omega)$ $\Big(\equiv \frac{\sum_{n=1}^N f(\omega_n)}{N}\Big)$ $(\forall f \in C(\Omega))$.

■

Therefore, we can assert:

The principle of equal weight (= "PEW" = Bayes' postulate). [The belief interpretation of mixed states]. *Consider a system $S_{[*]}$ formulated in $C(\Omega)$ where the state space Ω $(\equiv \{\omega_1, \omega_2, ..., \omega_N\})$ is a finite set. The belief weight is represented by a mixed state ν $(\in \mathcal{M}_{+1}^m(\Omega))$. In particular, the equal weight ν_u $(\equiv \frac{1}{N}\sum_{n=1}^N \delta_{\omega_n} \in \mathcal{M}_{+1}^m(\Omega))$ represents "the loosest belief".* (8.72)

Thus BMT is summarized as follows.

[BMT$_1$] the equal weight ν_u $(\in \mathcal{M}_{+1}^m(\Omega))$ represents "the most loosest belief".

[BMT$_2$] After we get the measured value x by a belief measurement $\mathbf{M}_{C(\Omega)}$ $(\mathbf{O} \equiv (X, 2^X, F)$, $S_{[*]}((\rho_0^m))_{bw})$, the new belief weight of the system $S_{[*]}$ is changed to ρ^m_{new} $(\in \mathcal{M}_{+1}^m(\Omega))$ such that $\rho^m_{\text{new}}(B) = \frac{\int_B [F(\{x\})](\omega)\rho_0^m(d\omega)}{\int_\Omega [F(\{x\})](\omega)\rho_0^m(d\omega)}$ $(\forall B \in \mathcal{B}_\Omega$, Borel field).

Define the map $[R_{\{x\}}^{(0,0)}] : \mathcal{M}_{+1}^m(\Omega) \to \mathcal{M}_{+1}^m(\Omega)$ such that:

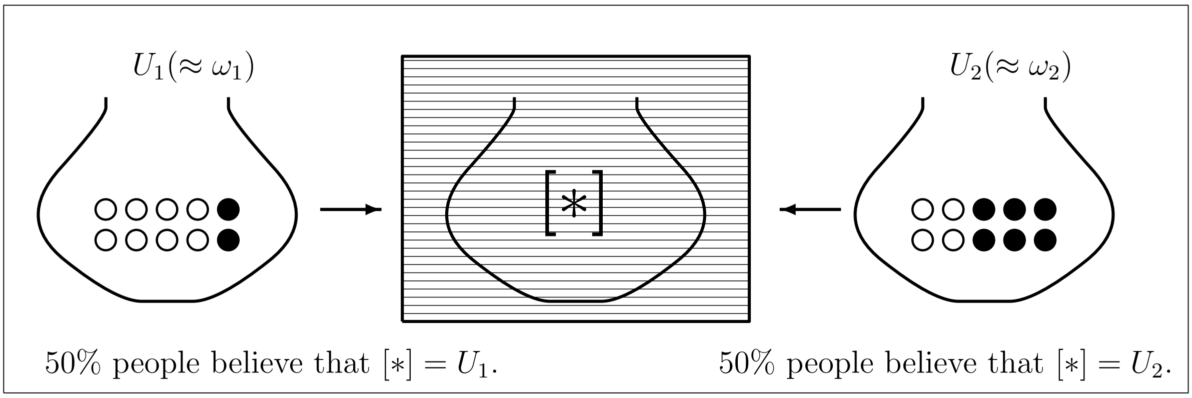
$$[R_{\{x\}}^{(0,0)}](\rho^m) = \frac{\int_{D_0} [F(\{x\})](\omega)\nu(d\omega)}{\int_\Omega [F(\{x\})](\omega)\nu(d\omega)} \qquad (\forall D_0 \subseteq \Omega;\ \text{Borel set }). \tag{8.73}$$

Then, we can symbolically describe it as follows:

$$[\text{BMT}] = \begin{cases} [\text{BMT}_1] & \text{the loosest belief weight} \quad \longleftrightarrow \quad \nu_u(\in \mathcal{M}_{+1}^m(\Omega)) \\ [\text{BMT}_2] & S_{[*]}((\rho))_{bw} \xrightarrow[x \text{ is obtained}]{\mathbf{M}_{C(\Omega)}(\mathbf{O},\, S_{[*]}((\rho))_{bw})} S_{[*]}(([R_{\{x\}}^{(0,0)}](\rho)))_{bw}, \end{cases} \tag{8.74}$$

which should be compared with the characterization (5.80) of "Iterative likelihood function method".

Example 8.24. [= Example 5.24 (the urn problem)]. There are two urns ω_1 and ω_2. The urn ω_1 [resp. ω_2] contains 8 white and 2 black balls [resp. 4 white and 6 black balls].



Figure(8.75)

Assume that they can not be distinguished in appearance.

- Choose one urn from the two. (8.76)

Now you sample, randomly, with replacement after each ball.

(i). First, you get "white ball".

(Q_1) *Do you believe which the chosen urn is,* ω_1 *or* ω_2*?*

(ii). Further, assume that you continuously get "black".

(Q_2) *How about the case? Do you believe which the chosen urn is,* ω_1 *or* ω_2*?*

And further,

(Q_3) *Also, study the case that the urn is chosen by a fair coin-tossing in (8.76).*

[*Answers*]. In what follows this problem is studied in BMT. Put $\Omega = \{\omega_1, \omega_2\}$. $\mathbf{O} = (\{w, b\}, 2^{\{w,b\}}, F)$ where $[F(\{w\})](\omega_1) = 0.8$, $[F(\{b\})](\omega_1) = 0.2$, $[F(\{w\})](\omega_2) = 0.4$, $[F(\{b\})](\omega_2) = 0.6$. The PEW (8.72) says that the loosest belief is represented by ν_u (i.e., $\nu_u(\{\omega_1\}) = \nu_u(\{\omega_2\}) = 1/2$)]. Thus we have the belief measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}((\nu_u))_{bw})$.

(A_1). Thus, consider $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}((\nu_u))_{bw})$. Since the measured value "w" was obtained, the new belief weight ρ^m_{new}

$$\rho^m_{\text{new}}(\{\omega_1\})\Big(= \frac{\int_{\{\omega_1\}}[F(\{w\})](\omega)\nu_u(d\omega)}{\int_{\Omega}[F(\{w\})](\omega)\nu_u(d\omega)}\Big) = \frac{0.8 \times \frac{1}{2}}{0.8 \times \frac{1}{2} + 0.4 \times \frac{1}{2}} = \frac{2}{3}, \quad \rho^m_{\text{new}}(\{\omega_2\}) = \frac{1}{3}.$$

(A_2). Next, consider the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}((\rho^m_{\text{new}}))_{_{bw}})$. Since the measured value "b" was obtained, the new belief weight $\rho^m_{\text{new}^2}$ is represented by

$$\rho^m_{\text{new}^2}(\{\omega_1\})\Big(= \frac{\int_{\{\omega_1\}}[F(\{b\})](\omega)\rho^m_{\text{new}}(d\omega)}{\int_\Omega [F(\{b\})](\omega)\rho^m_{\text{new}}(d\omega)}\Big) = \frac{0.2 \times \frac{2}{3}}{0.2 \times \frac{2}{3} + 0.6 \times \frac{1}{3}} = \frac{2}{5},$$

$$\rho^m_{\text{new}}(\{\omega_2\})\Big(= \frac{\int_{\{\omega_2\}}[F(\{b\})](\omega)\rho^m_{\text{new}}(d\omega)}{\int_\Omega [F(\{b\})](\omega)\rho^m_{\text{new}}(d\omega)}\Big) = \frac{0.6 \times \frac{1}{3}}{0.2 \times \frac{2}{3} + 0.6 \times \frac{1}{3}} = \frac{3}{5}.$$

(A_3) Also, when the urn is chosen by a fair coin-tossing, the above ρ^m_{new} and $\rho^m_{\text{new}^2}$ acquire the probabilistic interpretation. That is, ρ^m_{new} and $\rho^m_{\text{new}^2}$ are regarded as statistical states.

[Remark]. In order to make a belief measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}((\nu_u))_{_{bw}})$ change a statistical measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_u))$, we have two methods. One is the fair coin-tossing method as mentioned in the above (A_3) (and (Q_3)). Another will be proposed as SMT_{PEP} in §11.4, i.e., "the principle of equal probability". Also, note that Theorem 11.12 says that the two methods are equivalent.

■

8.6.3 Is BMT necessary?

Now we have the following classification:

$$\text{MT}\begin{cases} \text{PMT} = \underset{[\text{Axiom 1 (2.37)}]}{\text{measurement}} + \underset{[\text{Axiom 2 (3.26)}]}{\text{the relation among systems}} \\[2ex] \text{SMT} = \underset{(\text{Axioms 1 and 2})}{\text{PMT}} + \underset{(\text{the probabilistic interpretation of mixed state})}{\text{“statistical state”}} \\[2ex] \text{BMT} = \underset{(\text{Axioms 1 and 2})}{\text{PMT}} + \underset{(\text{the principle of equal weight})}{\text{“belief weight”}} \end{cases} \tag{8.77}$$

However, we must consider and answer the following question:

(Q) Is BMT necessary?

In fact, some may think that

(A) BMT is not necessary. It suffices to substitute SMT for BMT *carefully*. In theoretical informatics, the "economical" should come before the "exact".

I may agree with them. However, it should be remarked that

(R) It is clear that we can not use SMT *carefully* without the understanding of the relation between SMT and BMT (i.e., without the understanding of the contents in §8.1 $\sim$ 8.6.2). Especially, note that Proclaim 1 (8.10) is not valid in BMT.

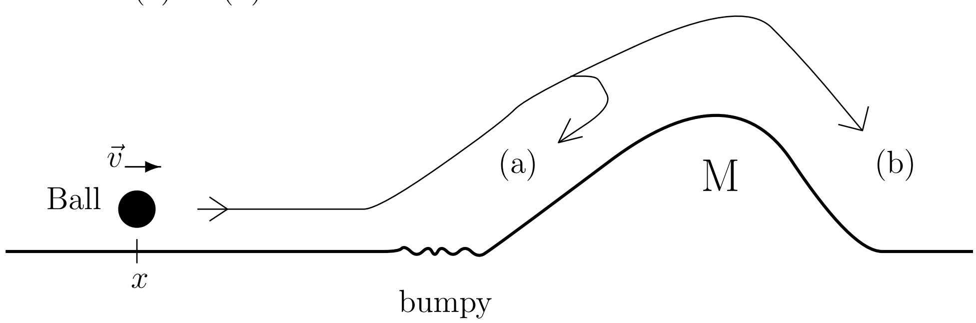
If this (R) is admitted, I agree to the above opinion (A). Thus, I recommend readers to use BMT at least until becoming accustomed to BMT. Also, it should be noted that there is a great confusion in the conventional statistics.

Remark 8.25. [The term: "subjectivity"]. Since the term "subjectivity" is frequently used in statistics, we must be careful for the usage of "subjectivity". For example, consider the following phrase:

- the probability that tomorrow is fine. (8.78)

The above term: "probability" is usually called a "subjective probability". However, the "probability" in (8.78) is the same as the "probability" in the following problem (which is due to Newtonian mechanics, and thus, deterministic). In spite of the deterministic system, we have the following question:

"Calculate the probability that the ball surmounts the mountain M." (8.79)
That is, the case (a) or (b)?



where the initial condition x(position) and $\vec{v}$(velocity) are values with errors, and also, the differential equation is not completely known. However, it should be noted that this problem is usual in engineering. Thus, if this is subjective (or, if a dearth of information implies "subjective"), we consider that almost every problem in engineering is subjective.[13]

[13]Recall the argument in Chapter 1. That is, in theoretical physics we must be in the objective standing point. On the other hand, in theoretical informatics (and its applications) we are, more or less, in the subjective standing point. Recall the engineer's spirit "Use everything available". Thus we may ask the excellent bookmaker about the problem (8.79). However, it should be noted that the bookmaker may calculate the "subjective probability in the sense of BMT (or, parimutuel betting among general people)".

There is a reason to consider that the probaility in the problem (8.79) can be regarded as the "subjective probability in the sense of parimutuel betting among a certain set of specialists". However, it is so, every probability may be regarded as the subjective probability. Thus, in this book, the term: "subjective probability" is used in the case that it is regarded as the probability in the sense of parimutuel betting.

■

Remark 8.26. [Differential geometry and operator algebra, *cf.* Table (1.8a)(4)]. In mathematics, differential geometry is flexible, but the theory of operator algebras (i.e, C^*-algebra and W^*-algebra) somewhat lacks adaptability. Thus, in MT[14], we can not prepare so many ready-made theories. For example, we have two ready-made theories (i.e., BMT and $\mathrm{SMT}_{\mathrm{PEP}}$ (*cf.* §11.4)). This fact (i.e., few ready-made theories can be proposed) is just what we want. That is because to choose one from too many ready-made theories is essentially the same as to create a made-to-order theory. On the other hand, in order to create a made-to-order theory in theoretical physics, the flexibility of differential geometry is essential.

■

8.7 Appendix (Bertrand's paradox)

As mentioned in Remark 8.4, a natural mixed state is not always a ststistical state. In fact we see, in §8.6, that the no informational weight ν_u ($\in \mathcal{M}^m_{+1}(\Omega)$, where Ω is finite) defnied by (8.70) can not be unconditionally regarded as the statistical state.[15] (As seen later (in §11.4), the term "unconditionally" is important.) In this section, we study Bertrand's paradox, which promote our understanding of the relation between a natural mixed state and a ststistical state.

8.7.1 Review (Bertrand's paradox)

Here, let us review the usual argument about Bertrand's paradox (*cf.* [35]). Consider

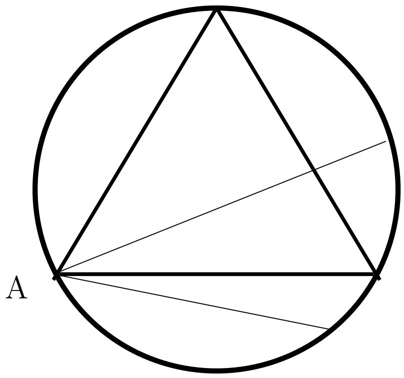
[14] Although Fisher information is closely related to Riemann manifold (in differential geometry, cf [5], [24]), it is not the axiom of MT but a kind of method.
[15] The ν_u is invariant concerning any bijection ϕ on Ω, i.e., $\phi(\nu_u) = \nu_u$. In this sense, it is natural.

the following problem:

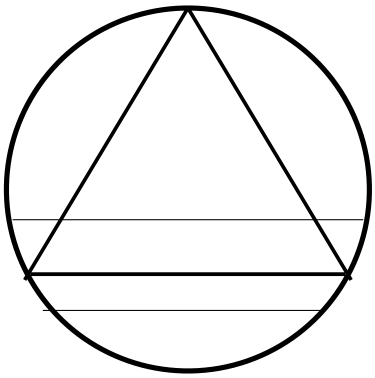
(P_1) Given a circle with the radius 1. Suppose a chord of the circle is chosen at random. What is the probability that the chord is longer than $\sqrt{3}$ (i.e., the side of an inscribed equilateral triangle)?

The problem has apparently several solutions as follows:

(Fig.1)



(Fig.2)



[First Solution (Fig.1)]. The "random endpoints" method: Choose a point A on the circumference and rotate the triangle so that the point is at one vertex. Choose another point on the circle and draw the chord joining it to the first point. For points on the arc between the endpoints of the side opposite the first point, the chord is longer than a side of the triangle. The length of the arc is one third of the circumference of the circle, therefore the probability a random chord is longer than a side of the inscribed triangle is one third.

[Second Solution (Fig.2)]. The "random radius" method: Choose a radius of the circle and rotate the triangle so a side is perpendicular to the radius. Choose a point on the radius and construct the chord whose midpoint is the chosen point. The chord is longer than a side of the triangle if the chosen point is nearer the center of the circle than the point where the side of the triangle intersects the radius. Since the side of the triangle bisects the radius, it is equally probable that the chosen point is nearer or farther. Therefore the probability a random chord is longer than a side of the inscribed triangle is one half.

8.7.2 Bertrand's paradox in measurement theory

We assert that

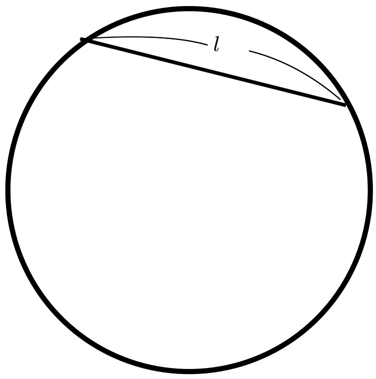
($\sharp$) If Bertrand's paradox is a paradox (i.e., if the argument in §8.7.1 is considered to be strange), it is due to the confusion between statistical states and mixed states (*cf.* (8.11)).

In what follows, we shall explain it. Consider the following problem:

(P_2) Given a circle with the radius 1. Define the state space Ω by the set composed of all chords of this circle. Then, find a natural mixed state ρ ($\in \mathcal{M}_{+1}^m(\Omega)$).

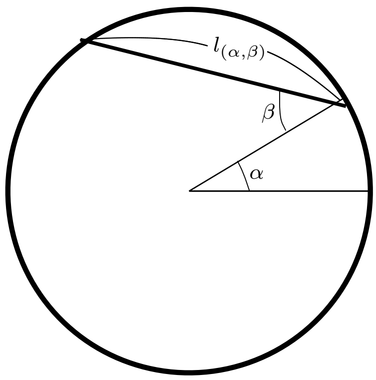
The reader will find that the (P_2) is essentially the same as the problem (P_1) in §8.7.1. Thus, the above problem has also apparently several solutions as follows:

(Fig.0)

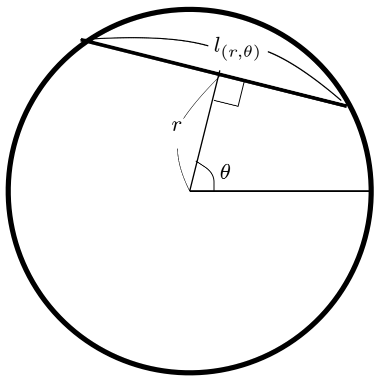


Represent a chord l
by a natural coordinate!

(Fig.1′)



(Fig.2′)



[First Solution (Fig.1′)]. See Fig.0 (Represent a chord by a natural coordinate!). In Fig.1′, we see that the chord l is represented by a point (α, β) in the rectangle $R_1 \equiv \{(\alpha, \beta) \mid 0 \le \alpha \le 2\pi,\ 0 \le \beta \le \pi/2(\text{radian})\}$. That is, we have the following identification:

$$\Omega \ni l_{(\alpha,\beta)} \longleftrightarrow (\alpha, \beta) \in R_1.$$

Under the identification, we get the natural mixed state ρ_1 ($\in \mathcal{M}_{+1}^m(\Omega) \approx \mathcal{M}_{+1}^m(R_1)$) such that $\rho_1(A) = \frac{\text{Area}[A]}{\text{Area}[R_1]} = \frac{\text{Area}[A]}{\pi^2}$ $(\forall A \in \mathcal{B}_{R_1})$, where "Area" = "Lebesgue measure". Therefore, we see

$$\begin{aligned} &\rho_1(\{l_{(\alpha,\beta)} \in \Omega \mid \text{"the length of } l_{(\alpha,\beta)}\text{"} \geq \sqrt{3}\}) \\ =&\frac{\text{Area}[\{(\alpha,\beta) \mid 0 \leq \alpha \leq 2\pi,\ 0 \leq \beta \leq \pi/6\}]}{\text{Area}[\{(\alpha,\beta) \mid 0 \leq \alpha \leq 2\pi,\ 0 \leq \beta \leq \pi/2\}]} \\ =&\frac{2\pi \times (\pi/6)}{2\pi \times (\pi/2)} = \frac{1}{3}. \end{aligned} \tag{8.80}$$

[Second Solution (Fig.2′)]. See Fig.0 (Represent a chord by a natural coordinates). In Fig.2′, we see that the chord l is represented by a point (r,θ) in the rectangle $R_2 \equiv \{(r,\theta) \mid 0 \leq r \leq 1,\ 0 \leq \theta \leq 2\pi\}$. That is, we have the following identification:

$$\Omega \ni l_{(r,\theta)} \longleftrightarrow (r,\theta) \in R_2.$$

Under the identification, we get the natural mixed state ρ_2 $(\in \mathcal{M}_{+1}^m(\Omega) \approx \mathcal{M}_{+1}^m(R_2))$ such that $\rho_2(A) = \frac{\text{Area}[A]}{\text{Area}[R_2]} = \frac{\text{Area}[A]}{2\pi}$ $(\forall A \in \mathcal{B}_{R_2})$. Therefore, we see

$$\begin{aligned} &\rho_2(\{l_{(\alpha,\beta)} \in \Omega \mid \text{"the length of } l_{(r,\theta)}\text{"} \geq \sqrt{3}\}) \\ =&\frac{\text{Area}[\{(r,\theta) \mid 0 \leq r \leq 1/2,\ 0 \leq \theta \leq 2\pi\}]}{\text{Area}[\{(r,\theta) \mid 0 \leq r \leq 1,\ 0 \leq \theta \leq 2\pi\}]} = \frac{1}{2}. \end{aligned} \tag{8.81}$$

Since the above argument is related to "mixed state" and not "statistical state", we have no paradox in the above arguments. That is, if Bertrand's paradox is a paradox (in §8.7.1), it is due to the confusion between mixed states (mathematical concept) and statistical states (measurement theoretical concept).

Some may assert that:

- it suffices to test (8.80) or (8.81) experimentally.

However, it is not true. For completeness, we add the following remark.

Remark 8.27. [Mixed state and statistical state]. In the above arguments, note that ρ_1 $(\in \mathcal{M}_{+1}^m(\Omega))$ and ρ_2 $(\in \mathcal{M}_{+1}^m(\Omega))$ are mixed states and not statistical states. In order to regard a mixed state ρ_1 $(\in \mathcal{M}_{+1}^m(\Omega))$ as a statistical state, we must add the probabilistic interpretation to the mixed state ρ_1. This is, for example, done as follows:

(R_1) Prepare two urns A and B, which respectively contain 100 balls (i.e., "ball 1", "ball 2", ...,"ball 100"). Pick out one ball from the urn A. Assume that the ball is "ball m". Next, pick out one ball from the urn B. Assume that the ball is "ball n". Define (α, β) in the rectangle R_1 such that:

$$\alpha = \frac{2\pi m}{100}, \qquad \beta = \frac{\pi n}{200}.$$

Then, if (α, β) is chosen according to the above rule (R_1), the mixed state ρ_1 ($\in \mathcal{M}^m_{+1}(\Omega)$) acquires the probabilistic interpretation. And thus, it can be regarded as a statistical state. In fact, if we take an exact measurement, we see that the probability that the length of the chord is longer than $\sqrt{3}$ is given by $1/3$. Of course, by a similar way, we can add the probabilistic interpretation to the ρ_2 (in the second solution). That is, it suffices to choose a chord as follows.

(R_2) Prepare two urns A and B, which respectively contain 100 balls (i.e., "ball 1", "ball 2", ...,"ball 100"). Pick out one ball from the urn A. Assume that the ball is "ball m". Next, pick out one ball from the urn B. Assume that the ball is "ball n". Define (r, θ) in the rectangle R_1 such that:

$$r = \frac{m}{100}, \qquad \theta = \frac{2\pi n}{200}.$$

■

Summing up, we conclude as follows. Consider the following problem:

$(P_1)'$ Given a circle with the radius 1. And choose a chord. Find the probability that the chord chosen is longer than $\sqrt{3}$ (i.e., the side of an inscribed equilateral triangle).

Then, we see:

(A_1) If we know that the chord was chosen by the rule (R_1) in Remark 8.27, we can conclude that the probability that the chord chosen be longer than $\sqrt{3}$ is $1/3$.

(A_2) If we know that the chord was chosen by the rule (R_2) in Remark 8.27, we can conclude that the probability that the chord chosen be longer than $\sqrt{3}$ is $1/2$.

(A_3) If we know that the chord was chosen by the physical experiment (conducted in [49]), we may conclude that the probability that the chord chosen be longer than $\sqrt{3}$ is about $1/2$ (*cf.* [49]).

(A_4) etc.

We consider that something like a (physical) coin-tossing (such as Brownian motion, radioactive atom, etc.) is hidden behind the physical experiment (in (A_3)). Thus, we again stress that

- A "coin-tossing" is always hidden behind a statistical state. Or there is no statistical state without a "coin-tossing" (or, "dice-throwing", "urn problem").

Also, it should be noted that we are in theoretical informatics and not in theoretical physics.

Chapter 9

Statistical measurements in W^*-algebraic formulation

The Statistical MT (= SMT) has two kinds of formulations. One is SMT^{C^*} (i.e., the C^*-algebraic formulation of SMT), which was introduced in the previous chapter, that is,

$$\mathrm{SMT}^{C^*} = \underset{\text{[Proclaim 1 (8.10)]}}{\text{statistical measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} \quad \text{in } C^*\text{-algebra.} \tag{9.1 (=(8.2))}$$

In this chapter we introduce another formulation of SMT (i.e., SMT^{W^*}), that is,

$$\mathrm{SMT}^{W^*} = \underset{\text{[Proclaim}^{W^*}\text{1 (9.9)]}}{\text{statistical measurement}} + \underset{\text{[Proclaim}^{W^*}\text{2 (9.23)]}}{\text{the relation among systems}} \quad \text{in } W^*\text{-algebra,} \tag{9.2}$$

which is called the W^*-algebraic formulation of SMT. Of course, "SMT^{C^*}" and "SMT^{W^*}" are essentially the same. The difference between the two is that of the mathematical tools (i.e., C^*-algebra and W^*-algebra). Thus, "SMT^{W^*}" should be understood by an analogy of "SMT^{C^*}". Although the C^*-algebraic formulation is most fundamental, the W^*-algebraic formulation is rather handy from the mathematical point of view.

9.1 Statistical measurements (W^*-algebraic formulation)

The Statistical MT (= SMT) has two kinds of formulations. One is the C^*-algebraic formulation of SMT (= SMT^{C^*}), which was introduced in the previous chapter. In order to develop "Statistical MT", in this chapter we introduce the W^*-algebraic formulation of Statistical MT (= SMT^{W^*}).[1] Here, it should be noted that "SMT^{C^*}" and "SMT^{W^*}" are

[1] Of course, the (pure) measurement theory (= PMT) has also two kinds of formulations, i.e., PMT^{C^*} and PMT^{W^*}. However, the commutative PMT^{W^*} has a demerit such that a pure state can not be represented in the commutative PMT^{W^*} in general. (*cf.* the statement (9.3)). Thus, we usually focus on SMT^{W^*} and not PMT^{W^*}. However, it should be noted that as far as quantum mechanics, PMT^{W^*} is superior to PMT^{C^*}. Cf. §9.3.

essentially the same. The difference between the two is that of the mathematical tools (i.e., C^*-algebra and W^*-algebra).

The C^*-algebraic formulation stated in the previous chapter is, of course, most fundamental. However, from the mathematical (or technical) point of view, the topology of a C^*-algebra $\mathcal{A}$ is somewhat too strong. Note that any C^*-algebra $\mathcal{A}$ can be imbedded into $B(V)$, the algebra composed of all bounded linear operators on a Hilbert space V (*cf.* Theorem 2.4 (the GNS-construction in [50, 76, 82])). Thus, using the imbedding: $\mathcal{A} \subseteq B(V)$, we may start from the weak*-closure $\overline{\mathcal{A}}$ (of $\mathcal{A}$) in $B(V)$. This $\overline{\mathcal{A}}$ is called a *W^*-algebra.* This method (i.e., to formulate measurement theory in terms of W^*-algebras) is called the W^*-algebraic formulation. Though this method is somewhat methodological, it is rather handy from the mathematical point of view. $\Big($For example, this will be seen in Theorem 10.1 in Chapter 10.$\Big)$

Let $\mathcal{N}$ be a W^*-algebra, that is,

[$\sharp_1$] $\mathcal{N}$ is a weak* closed subalgebra of a certain $B(V)$.

It is well known (see, for example, [76]) that this is equivalent to

[$\sharp_2$] $\mathcal{N}$ is a C^*-algebra with the pre-dual Banach space $\mathcal{N}_*$ (i.e., $\mathcal{N} = (\mathcal{N}_*)^*$).

Also, it is well known that the uniqueness of the pre-dual Banach space $\mathcal{N}_*$ is assured. However, we may sometimes call the pair $(\mathcal{N}, \mathcal{N}_*)$ a W^*-algebra.

An element F in $\mathcal{N}$ is called *self-adjoint* if it holds that $F = F^*$. A self-adjoint element F in $\mathcal{N}$ is called *positive* (and denoted by $F \geq 0$) if there exists an element F_0 in $\mathcal{N}$ such that $F = F_0^* F_0$ where F_0^* is the adjoint element of F_0. Also, a positive element F is called a *projection* if $F = F^2$ holds.

Now we can define the *normal state-class* $\mathfrak{S}^n(\mathcal{N}_*)$ such as

$$\mathfrak{S}^n(\mathcal{N}_*) \equiv \{\rho^n \in \mathcal{N}_* \;:\; \|\rho^n\|_{\mathcal{N}_*} = 1 \text{ and } \rho^n \geq 0 \text{ (i.e., } \rho^n(T^*T) \geq 0 \text{ for all } T \in \mathcal{N})\}.$$

The element ρ^n (in $\mathfrak{S}^n(\mathcal{N}_*)$) is called a *normal state* (or, *density state*). The linear functional $\rho^n(T)$ is sometimes denoted by $\Big\langle \rho^n, T \Big\rangle$, or precisely, ${}_{\mathcal{N}_*}\Big\langle \rho^n, T \Big\rangle_{\mathcal{N}}$. Also, note that

- a W^*-algebra $\mathcal{N}$ has a lot of projections,

that is, the set of all finite linear combinations of projections is dense in $\mathcal{N}$ in the weak* topology $\sigma(\mathcal{N}; \mathcal{N}_*)$. Also, note that

- $\mathcal{N}$ has always the identity $I_{\mathcal{N}}$.

Example 9.1. [(i): Commutative W^*-algebras ; $L^\infty(\Omega,\mu)$)]. Let $(\Omega, \mathcal{B}_\Omega, \mu)$ be a measure space. For any $1 \le p \le \infty$, define $L^p(\Omega,\mu)$ $\Big(\equiv L^p(\Omega, \mathcal{B}_\Omega, \mu) \Big) = \{ f : f$ is a complex valued measurable function such that $\|f\|_{L^p} \equiv [\int_\Omega |f(\omega)|^p \mu(d\omega)]^{1/p} < \infty \}$. $\Big($Here, of course, $\|f\|_{L^\infty} = \text{ess.sup}\,\{|f(\omega)| \;:\; \omega \in \Omega\}.\Big)$ Then, the $\mathcal{N} \equiv L^\infty(\Omega,\mu)$ is a commutative W^*-algebra with the pre-dual Banach space $\mathcal{N}_* = L^1(\Omega,\mu)$. We see, of course, that $\mathfrak{S}^n(\mathcal{N}_*) = L^1_{+1}(\Omega,\mu) \equiv \{\rho^n \in L^1(\Omega,\mu) : \rho^n \ge 0, \int_\Omega \rho^n(\omega)\mu(d\omega) = 1$, i.e., ρ^n is a density function on Ω $\}$. Also, it is well known that any commutative W^*-algebra $\mathcal{N}$ is represented by some $L^\infty(\Omega,\mu)$. It should be noted that

- a "pure state" can not be generally represented in terms of the commutative W^*-algebra $L^\infty(\Omega,\mu)$, (9.3)

since we see[2] that δ_{ω_0} (i.e., a point measure at ω_0 $(\in \Omega)$) does not necessarily belong to $L^1(\Omega,\mu)$. Summing up (and recalling Example 2.2), we see,

	commutative C^*-algebra	commutative W^*-algebra
concrete form	$C(\Omega)$	$L^\infty(\Omega;\mu)$
dual space	$\mathcal{M}(\Omega)$ $(= C(\Omega)^*)$	not important
pre-dual space	nothing	$L^1(\Omega;\mu)$ $(= L^\infty(\Omega;\mu)_*)$
pure state	$\delta_{\omega_0} \in \mathcal{M}^p_{+1}(\Omega) \approx \Omega$	(no representation in general)
mixed (normal) state	$\nu \in \mathcal{M}^m_{+1}(\Omega)$	$\overline{\rho} \in L^1_{+1}(\Omega,\mu)$
characteristics[3]	topological approach	measure theoretical approach

[(ii): The case that Ω is countable or finite]. Of course, the above table is in the case that Ω is general. In the case that $\Omega \equiv \{\omega_1, \omega_2, ..., \omega_n\}$ is finite, we can easily see that "commutative C^*-algebra" = "commutative W^*-algebra", that is, we see the following identifications:

$$C(\{\omega_1,\omega_2,...,\omega_n\}) \approx \mathbf{C}^n(cf. \text{ the formula (2.15)}) \approx L^\infty(\{\omega_1,\omega_2,...,\omega_n\},\mu) \tag{9.4}$$

where μ is a measure such that $\mu(\{\omega_k\}) > 0$ $(\forall k = 1,2,...,n)$. Next consider the case that $\Omega \equiv \{\omega_1, \omega_2, ..., \omega_k, ...\}$ is countable infinite. The commutative W^*-algebra $\mathcal{N}$ is defined by

[2]In this sense the W^*-algebraic formulation is fit to SMT rather than PMT. However note our spirit (8.12) : *"There is no SMT without PMT."* Thus we think that PMT (i.e., the concept of "pure state") is not only hidden in the C^*-algebraic formulation of SMT but also in the W^*-algebraic formulation of SMT.

[3]The Ω in $C(\Omega)$ is a topological space. On the other hand, the Ω in $L^\infty(\Omega;\mu)$ is a measure space. Cf. Remark 9.14 later.

$L^\infty(\Omega,\mu)$, where $\mu(\{\omega_k\})>0$ ($\forall k=1,2,...$). In this case, a pure state ρ_{ω_k} ($k=1,2,...$), is defined by $\rho_{\omega_k}(\omega)=\frac{1}{\mu(\{\omega_k\})}$ (if $\omega=\omega_k$), $=0$ (if $\omega\neq\omega_k$).

■

Example 9.2. [Non-commutative W^*-algebras; $B(V)$]. When $\mathcal{N}=B(V)$, we see that $\mathcal{N}_*=Tr(V)$ (*cf.* Example 2.3) and $\mathfrak{S}^n(\mathcal{N}_*)=Tr^m_{+1}(V)\equiv\{\rho^n\in Tr(V):\rho^n\geq 0,\|\rho^n\|_{Tr(V)}=1\}$. Also, note that ${}_{Tr(V)}\Big\langle\rho^n,T\Big\rangle_{B(V)}=tr[\rho^n{\cdot}T]_V$. Here, $tr[A]_V\equiv\sum_{\lambda\in\Lambda}\langle e_\lambda,Ae_\lambda\rangle_V$ where $\{e_\lambda|\lambda\in\Lambda\}$ is a complete orthonormal basis in V. Also, it is well known that the value $tr[A]_V$ is independent of the choice of a complete orthonormal basis $\{e_\lambda|\lambda\in\Lambda\}$ in V. Further, any ρ^n ($\in Tr^m_{+1}(V)$) is represented by $\rho^n=\sum_{\lambda\in\lambda}\alpha_\lambda|e_\lambda\rangle\langle e_\lambda|$ (in the trace norm $\|\cdot\|_{Tr(V)}$) for some complete orthonormal basis $\{e_\lambda|\lambda\in\Lambda\}$ in V and some sequence $\{a_\lambda\}_{\lambda\in\Lambda}$ of non-negative numbers such that $\sum_{\lambda\in\Lambda}a_\lambda=1$. Also it should be noted that any $|v\rangle\langle v|$, ($\|v\|=1$), is just a pure state[4]. Summing up (and recalling Example 2.3), we see,

	non-commutative C^*-algebra	non-commutative W^*-algebra
concrete form	$\mathcal{C}(V)$	$B(V)$
dual space	$Tr(V)$ ($=\mathcal{C}(V)^*$)	not important
pre-dual space	nothing	$Tr(V)$ ($=B(V)_*$)
pure state	$\|v\rangle\langle v\|\in Tr^p_{+1}(V)$	$\|v\rangle\langle v\|\in Tr^p_{+1}(V)$
mixed (normal) state	mixed state: $\rho^m\in Tr^m_{+1}(V)$	normal state: $\rho^n\in Tr^m_{+1}(V)$

(9.5)

■

The following definition is the W^*-algebraic form of Definition 2.7 (C^*-observables).

Definition 9.3. [W^*-observables]. *Let $\mathcal{N}$ be a W^*-algebra. A W^*-observable (or in short, observable, fuzzy observable) $\overline{\mathbf{O}}\equiv(X,\mathcal{F},F)$ in $\mathcal{N}$ is defined such that it satisfies that*

(i) [*σ-field*]. *$(X,\mathcal{F})$ is a measurable space, that is, $\mathcal{F}$ ($\subseteq 2^X$) is a σ-field on X, i.e., it satisfies that*

$$\emptyset\in\mathcal{F},\quad \Xi_k\in\mathcal{F}\ (k=1,2,...)\Longrightarrow\cup_{k=1}^\infty\Xi_k\in\mathcal{F},\quad \Xi\in\mathcal{F}\Longrightarrow\Xi^c\in\mathcal{F},$$

[4]This fact (i.e., a pure state can be represented in terms of W^*-algebra $B(V)$) is remarkable. Thus, The W^*-algebra $B(V)$ has a power to describe quantum PMT as well as quantum SMT. Cf. §9.4.

(ii) *for every* $\Xi \in \mathcal{F}$, $F(\Xi)$ *is a positive element in* $\mathcal{N}$ *(i.e.,* $0 \le F(\Xi) \in \mathcal{N}$*) such that* $F(\emptyset) = 0$ *and* $F(X) = I_{\mathcal{N}}$*, where* 0 *is the* 0*-element and* $I_{\mathcal{N}}$ *is the identity element in* $\mathcal{N}$*, and,*

(iii) [*countably additivity*]. *For any countable decomposition* $\{\Xi_1, \Xi_2, ..., \Xi_j, ...\}$ *of* Ξ, $\Big($*i.e.,* $\Xi, \Xi_j \in \mathcal{F}, \cup_{j=1}^{\infty} \Xi_j = \Xi$, $\Xi_j \cap \Xi_i = \emptyset($ *if* $j \ne i)\Big)$*, it holds that*

$$F(\Xi) = \sum_{j=1}^{\infty} F(\Xi_j)$$

where the series is convergent in the sense of the weak-topology* $\sigma(\mathcal{N}; \mathcal{N}_*)$ *in* $\mathcal{N}$.

If $F(\Xi)$ *is a projection for every* Ξ $(\in \mathcal{F})$*, a* W^**-observable* $(X, \mathcal{F}, F)$ *in* $\mathcal{N}$ *is called a crisp* W^**-observable in* $\mathcal{N}$*. Also, a crisp observable* $\overline{\mathbf{O}} \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, F)$ *(or,* $(\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, F)$*) in* $\mathcal{N}$ *is called a quantity* (or, $\mathbf{R}^n$*-valued quantity*) in W^*-algebra $\mathcal{N}$.

■

Now we show several W^*-observables (in Example 9.4 $\sim$ 9.7).

Example 9.4. [Crisp W^*-observables in $L^\infty(\Omega, \mu)$]. (i). As a typical crisp W^*-observable in $L^\infty(\Omega, \mu)$, the *exact observable* $\overline{\mathbf{O}}_{\text{EXA}} \equiv (\Omega, \mathcal{B}_\Omega, \chi_{(\cdot)})$ is frequently used where χ_Ξ is the characteristic function of Ξ $(\in \mathcal{B}_\Omega)$ $\Big($i.e., $\chi_\Xi(\omega) = 1(\omega \in \Xi), = 0$ (otherwise)$\Big)$. This observable is finest in $L^\infty(\Omega, \mu)$, i.e., it includes all projections.

(ii). Consider the commutative W^*-algebra $L^\infty(\Omega, \mu)$. Let $a : \Omega \to \mathbf{R}$ be a measurable function. Then, we can define the crisp W^*-observable $\overline{\mathbf{O}}_a = (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, F)$ in $L^\infty(\Omega, \mu)$ such that $[F(\Xi)](\omega) = \chi_{a^{-1}(\Xi)}(\omega)$ $(\forall \Xi \in \mathcal{B}_{\mathbf{R}}, \forall \omega \in \Omega)$. Note that we can identify the real-valued measurable function $a(\cdot)$ with the $\overline{\mathbf{O}}_a$. That is, we see

$$\underset{\text{(real valued measurable function on }\Omega)}{a : \Omega \to \mathbf{R}} \qquad \longleftrightarrow \qquad \underset{\text{(crisp observable)}}{(\mathbf{R}, \mathcal{B}_{\mathbf{R}}, F)} \quad \text{in } L^\infty(\Omega, \mu).$$

That is because it holds that $[F((-\infty, \lambda))](\omega) = 0$ (if $\lambda < a(\omega)$), $= 1$ (if $\lambda \ge a(\omega)$), and therefore, the $a(\omega)$ is determined by the equality $a(\omega) = \int_{\mathbf{R}} \lambda \delta_{a(\omega)}(d\lambda) = \int_{\mathbf{R}} \lambda [F(d\lambda)](\omega)$ (a.e. μ). A real-valued measurable function on Ω is called a (*classical*) *quantity* in $L^\infty(\Omega, \mu)$ (though it is not always a bounded function).

■

Example 9.5. [Gaussian W^*-observable]. Define the W^*-observable $\overline{\mathbf{O}} \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, F_{(\cdot)})$ in $\mathcal{N} \equiv L^\infty(\mathbf{R}, d\omega)$ such that:

$$F^{\sigma}_{\Xi}(\omega) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\Xi} e^{-\frac{(u-\omega)^2}{2\sigma^2}} du \qquad (\forall\omega \in \mathbf{R},\ \forall\Xi \in \mathcal{B}_{\mathbf{R}}). \quad (\sigma^2\text{: variance}).$$

This is, of course, the W^*-algebraic form of the Gaussian C^*-observable $\mathbf{O} \equiv (\mathbf{R}, \mathcal{B}^{\rm bd}_{\mathbf{R}}, F_{(\cdot)})$ (*cf.* Example 2.17). Note that the $\mathcal{B}_{\mathbf{R}}$ in $\overline{\mathbf{O}}$ is a σ-field, and the $\mathcal{B}^{\rm bd}_{\mathbf{R}}$ in $\mathbf{O}$ is a σ-ring.

■

Remark 9.6. [The vagueness of a crisp observable]. Let ν be a probability measure on an index set Θ. For each θ ($\in \Theta$), consider a crisp observable $\overline{\mathbf{O}}_\theta \equiv (X, \mathcal{F}, E_\theta)$ in W^*-algebra $\mathcal{N}$. Define the observable $\overline{\mathbf{O}} \equiv (X, \mathcal{F}, F)$ in W^*-algebra $\mathcal{N}$ such that:

$$F(\Xi) = \int_{\Theta} E_\theta(\Xi)\nu(d\theta) \qquad (\forall\Xi \in \mathcal{F})$$

which is not crisp but fuzzy in general. Thus we think that

(F) "fuzzy observable"$\Longleftrightarrow$ "To understand a dearth of information concerning a crisp observable by a fuzzy observable".

This is one of the aspects of "fuzzy observable". When we want to stress this statistical aspect, the "observable" is often called a *"fuzzy observable"* (or, *"random observable"*). This will be again discussed in §11.4.

■

Example 9.7. [(i): Crisp W^*-observables in quantum $B(V)$]. Here, consider the quantum version of the (ii) in Example 9.4. Let A be a self-adjoint operator (not necessarily bounded) on a Hilbert space V. Recall the spectral representation: $A = \int_{\mathbf{R}} \lambda E_A(d\lambda)$. Here, the spectral measure $\overline{\mathbf{O}}_A \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, E_A)$ is of course the crisp W^*-observable in $B(V)$. Conversely, any crisp W^*-observable $(\mathbf{R}, \mathcal{B}_{\mathbf{R}}, F)$ in $B(V)$ determines a unique self-adjoint operator A_F on V such that $A_F = \int_{\mathbf{R}} \lambda F(d\lambda)$. Therefore, we have the identification:

$$\underset{\text{(self-adjoint operator on } V)}{A} \quad \longleftrightarrow \underset{\text{(crisp observable)}}{\overline{\mathbf{O}}_A = (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, F)} \text{ in } B(V) \qquad \Big(\text{i.e., } A = \int_{\mathbf{R}} \lambda F(d\lambda)\Big).$$

A self-adjoint operator A on a Hilbert space V is called a (*unbounded*) *quantity* in $B(V)$ (though A is not always a bounded linear operator).

[(ii): Position quantity, momentum quantity]. Put $V \equiv L^2(\mathbf{R}; dq))$, and define the (unbounded) self-adjoint operator Q [resp. P], which is called the *position quantity* [resp. *momentum quantity*], such that:

$$(Q\psi)(q) = q \cdot \psi(q), \qquad \Big[\text{ resp. } \quad (P\psi)(q) = -i\frac{\hbar d\psi(q)}{dq}\Big].$$

By the following spectral representations,

$$Q = \int_{\mathbf{R}} \lambda E_Q(d\lambda) \qquad \text{and} \qquad P = \int_{\mathbf{R}} \lambda E_P(d\lambda),$$

we see the following identifications:

$$\underset{\text{(self-adjoint operator on } V)}{Q} \longleftrightarrow \underset{\text{(crisp observable)}}{\overline{\mathbf{O}}_Q = (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, E_Q)} \text{ in } B(V)$$

and

$$\underset{\text{(self-adjoint operator on } V)}{P} \longleftrightarrow \underset{\text{(crisp observable)}}{\overline{\mathbf{O}}_P = (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, E_P)} \text{ in } B(V).$$

Here note that

$$[E_Q(\Xi)\psi](q) = \chi_\Xi(q) \cdot \psi(q) \qquad (\forall \psi \in L^2(\mathbf{R}), \forall \Xi \in \mathcal{B}_{\mathbf{R}}, q \in \mathbf{R})$$

and

$$E_P(\Xi)\psi = \mathfrak{F}^*(\chi_\Xi \cdot (\mathfrak{F}\psi)) \qquad (\forall \psi \in L^2(\mathbf{R}),\ \forall \Xi \in \mathcal{B}_{\mathbf{R}}, q \in \mathbf{R}),$$

where the Fourier transform $\mathfrak{F} : L^2(\mathbf{R}, dx) \to L^2(\mathbf{R}, dy)$ is defined by

$$(\mathfrak{F}\psi)(y) = \sqrt{\frac{\hbar}{2\pi}} \int_{\mathbf{R}} \psi(x) e^{-i\hbar xy} dx.$$

Note that both the position observable and momentum observable, which are most important in quantum mechanics, can not be defined in the C^*-algebraic formulation.

[(iii): Glauber-Sudarshan representation]. Consider ψ_0 ($\in V \equiv L^2(\mathbf{R}; dq)$) such that $\|\psi_0\|_{L^2(\mathbf{R})} = 1$ and

$$\langle \psi_0, P\psi_0 \rangle_V = 0, \qquad \langle \psi_0, Q\psi_0 \rangle_V = 0.$$

If we define $\phi_{x,y}(q) = e^{ixy}\psi_0(q - x)$, then an elementary computation shows that

$$\langle P\phi_{x,y}, \phi_{x,y} \rangle_{L^2(\mathbf{R})} = y, \quad \langle Q\phi_{x,y}, \phi_{x,y} \rangle_{L^2(\mathbf{R})} = x. \tag{9.6}$$

Here we can define the W^*-observable $(\mathbf{R}^2, \mathcal{B}_{\mathbf{R}^2}, G)$ in $B(L^2(\mathbf{R}; dq))$ such that:

$$G(\Xi) = \iint_\Xi |\phi_{x,y}\rangle\langle\phi_{x,y}| dxdy \qquad (\forall \Xi \in \mathcal{B}_{\mathbf{R}^2}).$$

This observable is essential in semi-classical mechanics (*cf.* [34]).

■

The following theorem is the W^*-algebraic form of Theorem 2.13. Since W^*-algebra $\mathcal{N}$ has a lot of projections, it is much more useful than Theorem 2.13.

Theorem 9.8. [The W^*-algebraic form of Theorem 2.13, *cf.* [42]]. *Let $\mathcal{N}$ be a W^*-algebra. Let $\overline{\mathbf{O}}_1 \equiv (X_1, \mathcal{F}_1, F_1)$ and $\overline{\mathbf{O}}_2 \equiv (X_2, \mathcal{F}_2, F_2)$ be W^*-observables in $\mathcal{N}$ such that at least one of them is crisp. $\Big($So, without loss of generality, we assume that $\overline{\mathbf{O}}_2$ is crisp$\Big)$. Then, the following statements are equivalent:*

(i) *There exists a quasi-product observable $\overline{\mathbf{O}}_{12} \equiv (X_1 \times X_2, \mathcal{F}_1 \times \mathcal{F}_2, F_1 \overset{\text{qp}}{\times} F_2)$ with marginal observables $\overline{\mathbf{O}}_1$ and $\overline{\mathbf{O}}_2$.*

(ii) *$\overline{\mathbf{O}}_1$ and $\overline{\mathbf{O}}_2$ commute, that is, $F_1(\Xi_1)F_2(\Xi_2) = F_2(\Xi_2)F_1(\Xi_1)$ $(\forall \Xi_1 \in \mathcal{F}_1, \forall \Xi_2 \in \mathcal{F}_2)$.*

Furthermore, if the above statements (i) and (ii) hold, the uniqueness of $\overline{\mathbf{O}}_{12}$ is guaranteed. (So, we can write that $\overline{\mathbf{O}}_{12} = (X_1 \times X_2,\ \mathcal{F}_1 \times \mathcal{F}_2,\ F_1 \times F_2) = \overline{\mathbf{O}}_1 \times \overline{\mathbf{O}}_2$.)

Proof. The proof is essentially the same as that of Theorem 2.13. □

The purpose of this chapter is to propose the W^*-algebraic formulation of SMT, that is,

$$\text{SMT}^{W^*} = \underset{[\text{Proclaim}^{W^*}1\ (9.9)]}{\text{statistical measurement}} + \underset{[\text{Proclaim}^{W^*}2\ (9.23)]}{\text{the relation among systems}} \quad \text{in } W^*\text{-algebra}\ . \tag{9.7 (=(9.2))}$$

In order to do it, we must recall the C^*-algebraic formulation of SMT, that is,

$$\text{SMT}^{C^*} = \underset{[\text{Proclaim 1 (8.10)}]}{\text{statistical measurement}} + \underset{[\text{Axiom 2 (3.26)}]}{\text{the relation among systems}} \quad \text{in } C^*\text{-algebra}\ . \tag{9.8 (=(9.1))}$$

As mentioned before, we want to understand SMT^{W^*} by an analogy of SMT^{C^*}. Here, it should be recalled that

- [Proclaim 1 (8.10), (The probabilistic interpretation of mixed states)]. *Consider a statistical measurement $\mathbf{M}_{\mathcal{A}}\big(\mathbf{O} \equiv (X, \mathcal{F}, F),\ S(\rho^m)\big)$ formulated in a C^*-algebra $\mathcal{A}$. Then, the probability that x $(\in X)$, the measured value obtained by the statistical measurement $\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S(\rho^m)\ )$, belongs to a set Ξ $(\in \mathcal{F})$ is given by*

$$\rho^m(F(\Xi)) \Big(\equiv {}_{\mathcal{A}^*}\Big\langle \rho^m, F(\Xi) \Big\rangle_{\mathcal{A}} \Big).$$

By an analogy of Proclaim 1, we can propose Proclaim$^{W^*}$1 as follows: Cf [44].

PROCLAIM$^{W^*}$1. [Statistical measurements in the W^*-algebraic formulation]. *Consider a statistical measurement* $\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{O}} \equiv (X, \mathcal{F}, F), \overline{S}(\rho^n))$ *formulated in a* W^**-algebra* $\mathcal{N}$. *The probability that* x $(\in X)$, *the measured value obtained by the statistical measurement* $\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{O}} \equiv (X, \mathcal{F}, F), \overline{S}(\rho^n))$, *belongs to* Ξ $(\in \mathcal{F})$ *is given by*

$$\rho^n(F(\Xi)) \Big(\equiv {}_{\mathcal{N}_*}\Big\langle \rho^n, F(\Xi) \Big\rangle_{\mathcal{N}} \Big). \tag{9.9}$$

This will be easily read by the above [Proclaim 1] and the following [TABLE (Statistical measurement theory)].

Statistical measurement theory (9.10)

[C^*-algebraic formulation]	$\longleftrightarrow$	[W^*-algebraic formulation]
Proclaim 1 (8.10)	$\longleftrightarrow$	Proclaim$^{W^*}$1 (9.9)
$\mathfrak{S}^m(\mathcal{A}^*) \ni \rho^m$	$\longleftrightarrow$	$\rho^n \in \mathfrak{S}^n(\mathcal{N}_*)$
C^*-observable $\mathbf{O} \equiv (X, \mathcal{F}, F)$	$\longleftrightarrow$	W^*-observable $\overline{\mathbf{O}} \equiv (X, \mathcal{F}, F)$
$\mathbf{M}_{\mathcal{A}}(\mathbf{O} \equiv (X, \mathcal{F}, F), S(\rho^m))$	$\longleftrightarrow$	$\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{O}} \equiv (X, \mathcal{F}, F), \overline{S}(\rho^n))$

Remark 9.9. [The W^*-algebraic formulation of PMT]. Though the commutative PMT$^{W^*}$ has a demerit such that a pure state can not be represented in the commutative PMT$^{W^*}$ in general (*cf.* the statement (9.3)), a pure state can be represented in the non-commutative PMT$^{W^*}$ (i.e., in $B(V)$, *cf.* Example 9.2). Thus, it is worthwhile mentioning the following Axiom$^{W^*}$1 (9.11). If $\mathcal{N} = B(V)$ or $\mathcal{N} = L^\infty(\Omega, \mu)$ (where Ω is finite or countable infinite), the concept "pure state" is valid (*cf.* (9.4) and (9.5)). Thus, in this case, we can propose "Axiom$^{W^*}$1 (9.11)" (i.e., the W^*-algebraic formulation of Axiom 1) as follows:

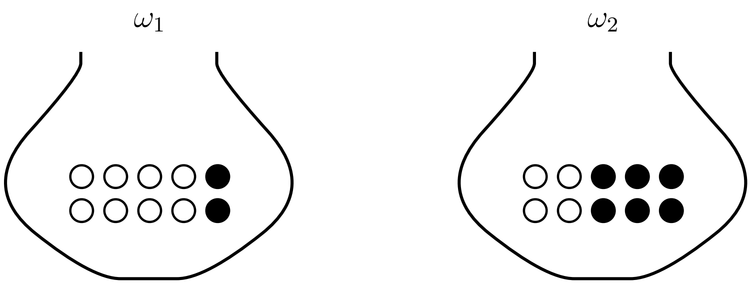
AXIOM$^{W^*}$1. [The W^*-algebraic formulation of Axiom 1]. *Consider a measurement* $\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{O}} \equiv (X, \mathcal{F}, F), \overline{S}_{[\rho^p]})$ *formulated in a* W^**-algebra* $\mathcal{N}$, *where* ρ^p *is a pure state. Assume that the measured value* x $(\in X)$ *is obtained by the measurement* $\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{O}}, \overline{S}_{[\rho^p]})$. *Then, the probability that the* x $(\in X)$ *belongs to a set* Ξ $(\in \mathcal{F})$ *is given by* $\rho^p(F(\Xi))$ $\Big(\equiv {}_{\mathcal{N}_*}\Big\langle \rho^p, F(\Xi) \Big\rangle_{\mathcal{N}} \Big)$. (9.11)

■

In the following example, we see that the C^*-algebraic formulation and the W^*-algebraic formulation are essentially the same.

Example 9.10. [(i): The review of Example 8.1] . There are two urns ω_1 and ω_2. The urn ω_1 [resp. ω_2] contains 8 white and 2 black balls [resp. 4 white and 6 black balls]. We regard Ω $\big(\equiv \{\omega_1, \omega_2\}\big)$ as the state space. And consider the observable $\mathbf{O}\big(\equiv (X \equiv \{w, b\}, 2^{\{w,b\}}, F)\big)$ in $C(\Omega)$ where

$$[F(\{w\})](\omega_1) = 0.8, \qquad [F(\{b\})](\omega_1) = 0.2,$$
$$[F(\{w\})](\omega_2) = 0.4, \qquad [F(\{b\})](\omega_2) = 0.6.$$



Here consider the following procedures (P_1) and (P_2).

(P_1) One of the two (i.e., ω_1 or ω_2) is chosen by an unfair tossed-coin ($C_{p,1-p}$), i.e.,

$$\text{Head } (100p\%) \to \omega_1, \text{ Tail } (100(1-p)\%) \to \omega_2 \qquad (0 \le p \le 1).$$

The chosen urn is denoted by $[*] (\in \{\omega_1, \omega_2\})$. Note, for completeness, that we do not know whether $[*]$ is ω_1 or ω_2. Here define the mixed state $\nu_0 (\in \mathcal{M}^m_{+1}(\Omega))$ such that $\nu_0(\{\omega_1\}) = p$, $\nu_0(\{\omega_2\}) = 1 - p$, which is considered to be *"the distribution of* $[*]$*."*

(P_2) Take one ball, at random, out of the urn chosen by the procedure (P_1). $\Big($That is, we take the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]})$.$\Big)$

[(ii): Continued from the above (i): C^*-algebraic formulation]. As seen in Example 8.1,

- "$(P_1) + (P_2)$" is notated by $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (X, 2^X, F), S(\nu_0))$.

Of course, we see

- the probability that the measured value x ($\in \{w, b\}$) is obtained by the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0))$, is given by

$$
{}_{C(\Omega)^*}\Big\langle \nu_0, F(\{x\}) \Big\rangle_{C(\Omega)} \; \Big(\equiv \int_\Omega {}_{C(\Omega)^*}\Big\langle \delta_\omega, F(\{x\}) \Big\rangle_{C(\Omega)} \nu_0(d\omega) \Big)
$$
$$
= \begin{cases} 0.8p + 0.4(1-p) & (\text{ if } x = w), \\ 0.2p + 0.6(1-p)) & (\text{ if } x = b). \end{cases} \tag{9.12}
$$

[(iii): Continued from the above (i): W^*-algebraic formulation]. Define the measure μ on Ω, for example, such that

$$
\mu(\{\omega_1\}) = \mu(\{\omega_2\}) = 1.
$$

Thus we have the commutative W^*-algebra $L^\infty(\Omega, \mu)$. And consider the observable $\overline{\mathbf{O}}\big(\equiv (X \equiv \{w, b\}, 2^{\{w,b\}}, F)\big)$ in $L^\infty(\Omega, \mu)$ where

$$
[F(\{w\})](\omega_1) = 0.8, \qquad [F(\{b\})](\omega_1) = 0.2,
$$
$$
[F(\{w\})](\omega_2) = 0.4, \qquad [F(\{b\})](\omega_2) = 0.6.
$$

Also define the normal state ρ^n ($\in L^1_{+1}(\Omega, \mu)$) such that[5]

$$
\rho^n(\omega_1) = p, \qquad \rho^n(\omega_2) = 1 - p.
$$

Then, we have the W^*-measurement $\overline{\mathbf{M}}_{L^\infty(\Omega,\mu)}(\overline{\mathbf{O}}, \overline{S}(\rho^n))$ in $L^\infty(\Omega, \mu)$. Of course, we see,

- the probability that the measured value x ($\in \{w, b\}$) is obtained by the measurement $\overline{\mathbf{M}}_{L^\infty(\Omega,\mu)}(\overline{\mathbf{O}}, \overline{S}(\rho^n))$, is given by

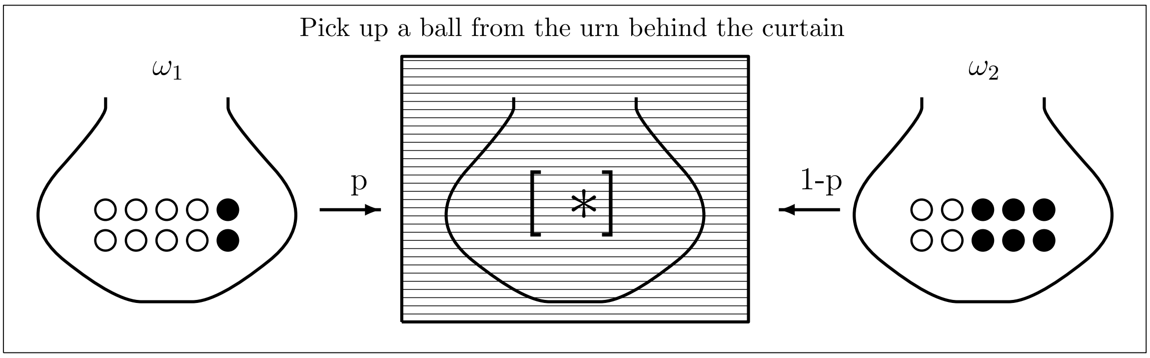
$$
{}_{L^1(\Omega,\mu)}\Big\langle \rho^n, F(\{x\}) \Big\rangle_{L^\infty(\Omega,\mu)} \; \Big(\equiv \int_\Omega [F(\{x\})](\omega) \rho^n(\omega) \mu(d\omega) \Big)
$$
$$
= \begin{cases} 0.8p + 0.4(1-p) & (\text{ if } x = w), \\ 0.2p + 0.6(1-p)) & (\text{ if } x = b). \end{cases} \tag{9.13}
$$

[5]Note that μ is arbitrary (*cf.* the formula (9.4)). If $\mu(\{\omega_1\}) = 1/3$ and $\mu(\{\omega_2\}) = 2$, it suffices to define that $\rho^n(\omega_1) = 3p$ and $\rho^n(\omega_1) = (1-p)/2$.

Thus we see that $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0))$ and $\overline{\mathbf{M}}_{L^\infty(\Omega,\mu)}(\overline{\mathbf{O}}, \overline{S}(\rho^n))$ are essentially the same (*cf.* (9.12) and (9.13)).

Also, we see:

The illustration of $\overline{\mathbf{M}}_{L^\infty(\Omega,\mu)}(\overline{\mathbf{O}}, \overline{S}(\rho^n))$



■

9.2 The relation among systems (Proclaim$^{W^*}$2 in SMT$^{W^*}$)

We mentioned "statistical measurement" [Proclaim$^{W^*}$1 (9.9)] in the previous section. Thus in this section, we devote ourselves to the "relation among systems (i.e., Proclaim$^{W^*}$2)" in the W^*-algebraic formulation of SMT [6]. That is, we want to propose

$$\text{SMT}^{W^*} = \underset{[\text{Proclaim}^{W^*}1\ (9.9)]}{\text{statistical measurement}} + \underset{[\text{Proclaim}^{W^*}2\ (9.23)]}{\text{the relation among systems}} \quad \text{in } W^*\text{-algebra } \mathcal{N}\,. \tag{9.14}$$

(=(9.2))

Let $\mathcal{N}_1$ and $\mathcal{N}_2$ with weak*-topologies $\sigma(\mathcal{N}_1, (\mathcal{N}_1)_*)$ and $\sigma(\mathcal{N}_2, (\mathcal{N}_2)_*)$ respectively. A continuous linear operator $\Psi_{1,2} : \mathcal{N}_2 \to \mathcal{N}_1$ is called a *Markov operator,* if it satisfies that

(i) $\Psi_{1,2}(F_2) \geq 0$ for any positive element F_2 in $\mathcal{N}_2$,

(ii) $\Psi_{1,2}(I_2) = I_1$, where I_k is the identity in $\mathcal{N}_k$ $(k = 1, 2)$.

Here note that, for any observable $(X, \mathcal{F}, F_2)$ in $\mathcal{N}_2$, the $(X, \mathcal{F}, \Psi_{1,2}F_2)$ is an observable in $\mathcal{N}_1$, which is denoted by $\Psi_{12}\mathbf{O}_2$. For example, it is easy to see that, for any countable decomposition $\{\Xi_j\}_{j=1}^\infty$ of Ξ, $(\Xi_j,\ \Xi \in \mathcal{F})$,

[6] If $\mathcal{N} = B(V)$ or $\mathcal{N} = L^\infty(\Omega, \mu)$ (where Ω is finite), the concept "pure state" is valid (*cf.* (9.4) and (9.5)). Thus, in the case, we can propose "PMT$^{W^*}$" (i.e., the W^*-algebraic formulation of PMT) as follows:

$$PMT^{W^*} = \underset{[\text{Axiom}^{W^*}1\ (9.11)]}{\text{statistical measurement}} + \underset{[\text{Proclaim}^{W^*}2\ (9.23)]}{\text{the relation among systems}} \quad \text{in } W^*\text{-algebra } \mathcal{N}\,.$$

$$[\Psi_{1,2}F_2](\Xi) = (w^*)\text{-}\lim_{n\to\infty}\Psi_{1,2}(F_2(\cup_{j=1}^n \Xi_j)) = (w^*)\text{-}\lim_{n\to\infty}\Psi_{1,2}(\sum_{j=1}^n F_2(\Xi_j))$$
$$=(w^*)\text{-}\lim_{n\to\infty}\sum_{j=1}^n[\Psi_{1,2}(F_2)](\Xi_j).$$

Also, a Markov operator $\Psi_{1,2} : \mathcal{N}_2 \to \mathcal{N}_1$ is called a *homomorphism* (or precisely, W^*-*homomorphism*), if it satisfies that

(i) $\Psi_{1,2}(F_2)\Psi_{1,2}(G_2) = \Psi_{1,2}(F_2G_2)$ for any F_2 and G_2 in $\mathcal{N}_2$,

(ii) $(\Psi_{1,2}(F_2))^* = \Psi_{1,2}(F_2^*)$ for any F_2 in $\mathcal{N}_2$.

Then the following mathematical result is well known.

(a) $(\Psi_{1,2})_*(\mathfrak{S}^n((\mathcal{N}_1)_*)) \subseteq \mathfrak{S}^n((\mathcal{N}_2)_*)$.

Let $(\Psi_{1,2})_* : (\mathcal{N}_1)_* \to (\mathcal{N}_2)_*$ be the pre-dual operator[7] of a Markov operator $\Psi_{1,2} : \mathcal{N}_2 \to \mathcal{N}_1$, that is, it holds that

$$_{(\mathcal{N}_1)_*}\Big\langle \rho_1^n, \Psi_{1.2}F_2 \Big\rangle_{\mathcal{N}_1} = {}_{(\mathcal{N}_2)_*}\Big\langle (\Psi_{1.2})_*\rho_1^n, F_2 \Big\rangle_{\mathcal{N}_2} \quad (\forall \rho_1^n \in (\mathcal{N}_1)_*, \forall F_2 \in \mathcal{N}_2). \tag{9.15}$$

Suppose that $\mathcal{N}_1$ and $\mathcal{N}_2$ are commutative W^*-algebras, i.e., $\mathcal{N}_1 = L^\infty(\Omega_1, \mu_1)$ and $\mathcal{N}_2 = L^\infty(\Omega_2, \mu_2)$. Then, under the identification that $\mathfrak{S}^n(\mathcal{N}_1^*) = L^1_{+1}(\Omega_1, \mu_1)$ and $\mathfrak{S}^n((\mathcal{N}_2)_*) = L^1_{+1}(\Omega_2, \mu_2)$ (*cf.* Example 9.2), the above (a) implies that the pre-dual operator $(\Psi_{1,2})_*$ of a Markov operator Ψ_{12} can be identified with a *transition probability rule* $M(\omega_1, B_2)$, $(\omega_1 \in \Omega_1,\ B_2 \in \mathcal{B}_{\Omega_2})$, such that:

$$\int_{B_2}[(\Phi_{12})_*(\rho_1^n)](\omega_2)\mu_2(d\omega_2) = \int_{\Omega_1} M(\omega_1, B_2)\rho_1^n(\omega_1)\mu_1(d\omega_1) \quad (\forall \rho_1^n \in L^1_{+1}(\Omega_1, \mu_1), \forall B_2 \in \mathcal{B}_{\Omega_2}).$$

Also, it is well known that, a Markov operator $\Psi_{1,2} : L^\infty(\Omega_2, \mu_2) \to L^\infty(\Omega_1, \mu_1)$ is homomorphic, if and only if there exists a measurable map $\psi_{1,2}$ from Ω_1 into Ω_2 such that:

$$(\Psi_{1,2}f_2)(\omega_1) = f_2(\psi_{1,2}(\omega_1)) \qquad (\text{almost all } \mu_1) \tag{9.16}$$

for all $f_2 \in L^\infty(\Omega_2, \mu_2)$.

[7]The symbol $*$ is used in the three following ways (1) $\sim$ (v) in this book. (i) involution operator (e.g., F^*), (ii) dual operator (e.g., Ψ^*), (iii) dual space (e.g., $\mathcal{A}^*$), (iv) pre-dual operator (e.g., Ψ_*), (v) pre-dual space (e.g., $\mathcal{N}_*$).

Let $(T, \leq)$ be a tree-like partial ordered set, i.e., a partial ordered set such that "$t_1 \leq t_3$ and $t_2 \leq t_3$" implies "$t_1 \leq t_2$ or $t_2 \leq t_1$". Put $T^2_{\leq} = \{(t_1, t_2) \in T^2 : t_1 \leq t_2\}$. An element $t_0 \in T$ is called a *root* if $t_0 \leq t$ $(\forall t \in T)$ holds. Note that the sub-tree $T_{t_0} \equiv \{t \in T \mid t \geq t_0\}$ has the root t_0. Thus we always assume that the tree-like ordered set $(T, \leq)$ has a root. We assume that T is not always finite. $\Big($In the next Chapter 10, T is always assumed to be infinite.$\Big)$

Definition 9.11. [General systems]. *The pair* $\overline{\mathbf{S}}(\rho^n_{t_0}) \equiv [\overline{S}(\rho^n_{t_0}), \{\Phi_{t_1,t_2} : \mathcal{N}_{t_2} \to \mathcal{N}_{t_1}\}_{(t_1,t_2) \in T^2_{\leq}}]$ *is called a general system with an initial state* $\overline{S}(\rho^n_{t_0})$ *if it satisfies the following conditions* (i)$\sim$(iii).

(i) *With each* t $(\in T)$, *a* W^**-algebra* $\mathcal{N}_t$ *is associated.*

(ii) *Let* t_0 $(\in T)$ *be the root of* T. *And, assume that a system* S *has the normal state* $\rho^n_{t_0}$ $(\in \mathfrak{S}^n((\mathcal{N}_{t_0})_*))$ *at* t_0, *that is, the initial state is equal to* $\rho^n_{t_0}$.

(iii) *For every* $(t_1, t_2) \in T^2_{\leq}$, *Markov operator* $\Phi_{t_1,t_2} : \mathcal{N}_{t_2} \to \mathcal{N}_{t_1}$ *is defined such that* $\Phi_{t_1,t_2}\Phi_{t_2,t_3} = \Phi_{t_1,t_3}$ *holds for all* (t_1, t_2), $(t_2, t_3) \in T^2_{\leq}$.

The family $\{\Phi_{t_1,t_2} : \mathcal{N}_{t_2} \to \mathcal{N}_{t_1}\}_{(t_1,t_2) \in T^2_{\leq}}$ *is also called a "Markov relation among systems". Let an observable* $\overline{\mathbf{O}}_t \equiv (X_t, 2^{X_t}, F_t)$ *in a* W^**-algebra* $\mathcal{N}_t$ *be given for each* $t \in T$. *The pair* $[\{\overline{\mathbf{O}}_t\}_{t \in T}, \{\Phi_{t_1,t_2} : \mathcal{N}_{t_2} \to \mathcal{N}_{t_1}\}_{(t_1,t_2) \in T^2_{\leq}}]$ *is called a "sequential observable", which is denoted by* $[\overline{\mathbf{O}}_T]$, *i.e.,* $[\overline{\mathbf{O}}_T] \equiv [\{\overline{\mathbf{O}}_t\}_{t \in T}, \{\Phi_{t_1,t_2} : \mathcal{N}_{t_2} \to \mathcal{N}_{t_1}\}_{(t_1,t_2) \in T^2_{\leq}}]$.

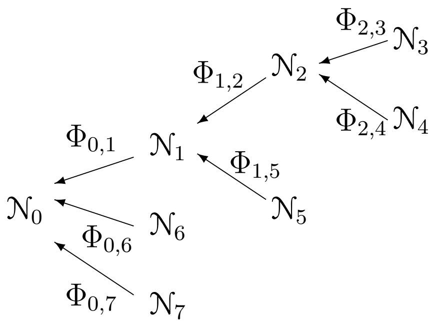
■

Before we explain Proclaim$^{W^*}$ 2, we prepare some notations. For simplicity, assume that T is finite, or a finite subtree of a whole tree. Let T $(= \{0, 1, ..., N\})$ be a tree with the root 0. Define the *parent map* $\pi : T \setminus \{0\} \to T$ such that $\pi(t) = \max\{s \in T : s < t\}$. It is clear that the tree $(T \equiv \{0, 1, ..., N\}, \leq)$ can be identified with the pair $(T \equiv \{0, 1, ..., N\}, \pi : T \setminus \{0\} \to T)$. Also, note that, for any $t \in T \setminus \{0\}$, there uniquely exists a natural number $h(t)$ (called the height of t) such that $\pi^{h(t)}(t) = 0$. Here, $\pi^2(t) = \pi(\pi(t))$, $\pi^3(t) = \pi(\pi^2(t))$, etc. Thus, the general system $\overline{\mathbf{S}}(\rho^n_0) \equiv [\overline{S}(\rho^n_0)$, $\{\Psi_{t_1,t_2} : \mathcal{N}_{t_2} \to \mathcal{N}_{t_1}\}_{(t_1,t_2) \in \{0,1,...,N\}^2_{\leq}}]$ is sometimes represented by $[\overline{S}(\rho^n_0)$, $\mathcal{N}_t \overset{\Psi_{\pi(t),t}}{\to} \mathcal{N}_{\pi(t)}$ ($t \in \{0, 1, ..., N\} \setminus \{0\})]$. Also, we define the $\Phi_{0,\tau} : \mathcal{N}_\tau \to \mathcal{N}_0$ such that $\Phi_{0,\tau} = \Psi_{0,\tau}$, that is,

$$\Phi_{0,\tau} = \Psi_{0,\pi^{h(\tau)-1}(\tau)} \Psi_{\pi^{h(\tau)-1}(\tau)),\pi^{h(\tau)-2}(\tau)} \cdots \Psi_{\pi^2(\tau),\pi(\tau)} \Psi_{\pi(\tau),\tau}. \tag{9.17}$$

Let $\overline{\mathbf{O}}_t \equiv (X_t, \mathcal{F}_t, F_t)$ be an observable in $\mathcal{N}_t$ $(\forall t \in T)$. The "measurement" of $\{\overline{\mathbf{O}}_t : t \in T\}$ for the $\overline{\mathbf{S}}(\rho^n_{t_0})$ is symbolically described by $\overline{\mathfrak{M}}(\{\overline{\mathbf{O}}_t\}_{t\in T}, \overline{\mathbf{S}}(\rho^n_{t_0}))$. The Markov relation $\{\Phi_{t_1,t_2} : \mathcal{N}_{t_2} \to \mathcal{N}_{t_1}\}_{(t_1,t_2)\in T^2_{\leq}}$ is also denoted by $\{\mathcal{N}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{N}_{\pi(t)}\}_{t\in T\setminus\{0\}}$

Example 9.12. [A simple general system. Compared to Examples 3.4 and 8.12]. Suppose that a tree $(T \equiv \{0, 1, ..., 6, 7\}, \pi)$ has an ordered structure such that $\pi(1) = \pi(6) = \pi(7) = 0$, $\pi(2) = \pi(5) = 1$, $\pi(3) = \pi(4) = 2$. (See the figure (9.18).) Consider a general system $\overline{\mathbf{S}}(\rho^n_0) \equiv [\overline{S}(\rho^n_0), \{\mathcal{N}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{N}_{\pi(t)}\}_{t\in T\setminus\{0\}}]$ with the initial system $\overline{S}(\rho^n_0)$.



(9.18)

Also, for each $t \in \{0, 1, ..., 6, 7\}$, consider an observable $\overline{\mathbf{O}}_t \equiv (X_t, 2^{X_t}, F_t)$ in a W^*-algebra $\mathcal{N}_t$. Thus, we have a sequential observable $[\{\overline{\mathbf{O}}_t\}_{t\in T}, \{\Phi_{t,\pi(t)} : \mathcal{N}_t \to \mathcal{N}_{\pi(t)}\}_{t\in T\setminus\{0\}}]$. Now we want to consider the following "measurement",

($\sharp$) for a system $\overline{S}((\rho^n_0))$ where $\rho^n_0 \in \mathfrak{S}^n((\mathcal{N}_0)_*)$, take a measurement of *"a sequential observable* $[\overline{\mathbf{O}}_T] \equiv [\{\overline{\mathbf{O}}_t\}_{t\in T}, \{\mathcal{N}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{N}_{\pi(t)}\}_{t\in T\setminus\{0\}}]$*,"* i.e., take a measurement of an observable $\overline{\mathbf{O}}_0$ at $0(\in T)$, and next, take a measurement of an observable $\overline{\mathbf{O}}_1$ at $1(\in T)$, $\cdots\cdots$, and finally take a measurement of an observable $\overline{\mathbf{O}}_7$ at $7(\in T)$,

which is symbolized by $\overline{\mathfrak{M}}(\{\overline{\mathbf{O}}_t\}_{t\in T}, \overline{\mathbf{S}}(\rho^n_0))$. Note that the $\overline{\mathfrak{M}}(\{\overline{\mathbf{O}}_t\}_{t\in T}, \overline{\mathbf{S}}(\rho^n_0))$ is merely a symbol since only one measurement is permitted (*cf.* §2.5 Remark (II)). In what follows let us describe the above ($\sharp$) $(= \overline{\mathfrak{M}}(\{\overline{\mathbf{O}}_t\}_{t\in T}, \overline{\mathbf{S}}(\rho^n_0)))$ precisely. Put

$$\widetilde{\mathbf{O}}_t = \overline{\mathbf{O}}_t \quad \text{and thus} \quad \widetilde{F}_t = F_t \quad (t = 3, 4, 5, 6, 7).$$

First we construct the quasi-product observable $\widetilde{\mathbf{O}}_2$ in $\mathcal{N}_2$ such as

$$\widetilde{\mathbf{O}}_2 = (X_2 \times X_3 \times X_4, 2^{X_2\times X_3\times X_4}, \widetilde{F}_2) \quad \text{where } \widetilde{F}_2 = F_2 \overset{\text{qp}}{\times} (\overset{\text{qp}}{\times}_{t=3,4} \Phi_{2,t}\widetilde{F}_t), \tag{9.19}$$

if it exists. Iteratively, we construct the following:

$$
\begin{array}{ccccc}
\mathcal{N}_0 & \xleftarrow{\Phi_{0,1}} & \mathcal{N}_1 & \xleftarrow{\Phi_{1,2}} & \mathcal{N}_2 \\
F_0 \overset{\text{qp}}{\times} \Phi_{0,6}\widetilde{F}_6 \overset{\text{qp}}{\times} \Phi_{0,7}\widetilde{F}_7 & & F_1 \overset{\text{qp}}{\times} \Phi_{1,5}\widetilde{F}_5 & & \\
\downarrow & & \downarrow & & \\
\underset{(F_0 \overset{\text{qp}}{\times} \Phi_{0,6}\widetilde{F}_6 \overset{\text{qp}}{\times} \Phi_{0,7}\widetilde{F}_7 \overset{\text{qp}}{\times} \Phi_{0,1}\widetilde{F}_1)}{\widetilde{F}_0} & \xleftarrow{\Phi_{0,1}} & \underset{(F_1 \overset{\text{qp}}{\times} \Phi_{1,5}\widetilde{F}_5 \overset{\text{qp}}{\times} \Phi_{1,2}\widetilde{F}_2)}{\widetilde{F}_1} & \xleftarrow{\Phi_{1,2}} & \underset{(F_2 \overset{\text{qp}}{\times} \Phi_{2,3}\widetilde{F}_3 \overset{\text{qp}}{\times} \Phi_{2,4}\widetilde{F}_4)}{\widetilde{F}_2}
\end{array}. \tag{9.20}
$$

That is, we get the quasi-product observable $\widetilde{\mathbf{O}}_1 \equiv (\prod_{t=1}^5 X_t, 2^{\prod_{t=1}^5 X_t}, \widetilde{F}_1)$ of $\overline{\mathbf{O}}_1$, $\Phi_{1,2}\widetilde{\mathbf{O}}_2$ and $\Phi_{1,5}\widetilde{\mathbf{O}}_5$, and finally, the quasi-product observable $\widetilde{\mathbf{O}}_0 \equiv (\prod_{t=0}^7 X_t, 2^{\prod_{t=0}^7 X_t}, \widetilde{F}_0)$ of $\overline{\mathbf{O}}_0$, $\Phi_{0,1}\widetilde{\mathbf{O}}_1$, $\Phi_{0,6}\widetilde{\mathbf{O}}_6$ and $\Phi_{0,7}\widetilde{\mathbf{O}}_7$, if it exists. Here, $\widetilde{\mathbf{O}}_0$ is called *the realization* (*or, the Heisenberg picture representation*) of *a sequential observable* $[\overline{\mathbf{O}}_T] \equiv [\{\overline{\mathbf{O}}_t\}_{t\in T}, \{\mathcal{N}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{N}_{\pi(t)}\}_{t\in T\setminus\{0\}}]$. Then, we have the measurement

$$
\overline{\mathbf{M}}_{\mathcal{N}_0}(\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, 2^{\prod_{t\in T} X_t}, \widetilde{F}_0), \overline{S}(\rho_0^n)), \tag{9.21}
$$

which is called *the realization* (*or, the Heisenberg picture representation*) *of the symbol* $\overline{\mathfrak{M}}(\{\overline{\mathbf{O}}_t\}_{t\in T}, \overline{\mathbf{S}}(\rho_0^n))$.

■

Examining Example 9.12, we have the following arguments. Let $(T \equiv \{0, 1, ..., N\}, \pi : T\setminus\{0\} \to T)$ be a tree with root 0 and let $\overline{\mathbf{S}}(\rho_0^n) \equiv [\overline{S}(\rho_0^n), \mathcal{N}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{N}_{\pi(t)}\ (t \in T\setminus\{0\})]$ be a general system with the initial system $\overline{S}(\rho_0^n)$. And, let an observable $\overline{\mathbf{O}}_t \equiv (X_t, \mathcal{F}_t, F_t)$ in a W^*-algebra $\mathcal{N}_t$ be given for each $t \in T$. For each s ($\in T$), define the observable $\widetilde{\mathbf{O}}_s \equiv (\prod_{t\in T_s} X_t, \prod_{t\in T_s} \mathcal{F}_t, \widetilde{F}_s)$ in $\mathcal{N}_s$ such that:

$$
\widetilde{\mathbf{O}}_s = \begin{cases} \overline{\mathbf{O}}_s & (\text{if } s \in T\setminus\pi(T)) \\ \overline{\mathbf{O}}_s \overset{\text{qp}}{\times} (\overset{\text{qp}}{\times}_{t\in\pi^{-1}(\{s\})} \Phi_{\pi(t),t}\widetilde{\mathbf{O}}_t) & (\text{if } s \in \pi(T)) \end{cases} \tag{9.22}
$$

if possible. Then, if an observable $\widetilde{\mathbf{O}}_0$ (i.e., the Heisenberg picture representation of the sequential observable $[\overline{\mathbf{O}}_T] \equiv [\{\overline{\mathbf{O}}_t\}_{t\in T}, \{\Phi_{t,\pi(t)} : \mathcal{N}_t \to \mathcal{N}_{\pi(t)}\}_{t\in T\setminus\{0\}}\ ])$ in $\mathcal{N}_0$ exists (such as in Example 9.12), we have the measurement

$$
\overline{\mathbf{M}}_{\mathcal{N}_0}(\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, \prod_{t\in T} \mathcal{F}_t, \widetilde{F}_0), \overline{S}(\rho_0^n)),
$$

which is called the Heisenberg picture representation of the symbol $\overline{\mathfrak{M}}(\{\overline{\mathbf{O}}_t\}_{t\in T}, \overline{\mathbf{S}}(\rho_0^n))$.

Summing up the essential part of the above argument, we can propose the following axiom, which corresponds to "the rule of the relation among systems" in SMT$^{W^*}$.

PROCLAIM$^{W^*}$2. [The Markov relation among systems, the Heisenberg picture] *The relation among systems is represented by a Markov relation* $\{\Phi_{t_1,t_2} : \mathcal{N}_{t_2} \to \mathcal{N}_{t_1}\}_{(t_1,t_2)\in T^2_{\leq}}$. *Let* $\overline{\mathbf{O}}_t$ $(\equiv (X_t, \mathcal{F}_t, F))$ *be an observable in* $\mathcal{N}_t$ *for each* t $(\in T)$. *If the procedure (9.22) is possible, a sequential observable* $[\overline{\mathbf{O}}_T] \equiv [\{\overline{\mathbf{O}}_t\}_{t\in T}, \{\Phi_{t_1,t_2} : \mathcal{N}_{t_2} \to \mathcal{N}_{t_1}\}_{(t_1,t_2)\in T^2_{\leq}}]$ *can be realized as the observable* $\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, \prod_{t\in T} \mathcal{F}_t, \widetilde{F}_0)$ *in* $\mathcal{N}_0$. (9.23)

Also, we must add the following statement:

- *Let* $\overline{\mathbf{S}}(\rho^n_{t_0}) \equiv [S(\rho^n_{t_0}), \{\Phi_{t_1,t_2} : \mathcal{N}_{t_2} \to \mathcal{N}_{t_1}\}_{(t_1,t_2)\in T^2_{\leq}}]$ *be a general system with an initial state* $\rho^n_{t_0}$ $(\in \mathfrak{S}^n((\mathcal{N}_{t_0})_*))$. *And then, a measurement represented by the symbol* $\overline{\mathfrak{M}}(\{\overline{\mathbf{O}}_t\}_{t\in T}, \overline{\mathbf{S}}(\rho^n_{t_0}))$ *can be realized by* $\overline{\mathfrak{M}}_{\mathcal{N}_0}(\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, \prod_{t\in T} \mathcal{F}_t, \widetilde{F}_0), \overline{S}(\rho^n_0))$, *if* $\widetilde{\mathbf{O}}_0$ *exists.*

which explains the relation between Proclaim$^{W^*}$1 and Proclaim$^{W^*}$2.

Remark 9.13. [How to read Proclaim$^{W^*}$2]. For completeness, we mention how to read Proclaim$^{W^*}$2 as follows: Recall Axiom 2 (3.26), that is,

- [Axiom 2. (The Markov relation among systems, the Heisenberg picture)] *The relation among systems is represented by a Markov relation* $\{\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2)\in T^2_{\leq}}$. *Let* $\mathbf{O}_t$ $(\equiv (X_t, \mathcal{F}_t, F))$ *be an observable in* $\mathcal{A}_t$ *for each* t $(\in T)$. *If the procedure (3.25) is possible, a sequential observable* $[\mathbf{O}_T] \equiv [\{\mathbf{O}_t\}_{t\in T}, \{\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2)\in T^2_{\leq}}]$ *can be realized as the observable* $\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, \prod_{t\in T} \mathcal{F}_t, \widetilde{F}_0)$ *in* $\mathcal{A}_0$.

Using this and the following correspondence, we can easily read the above Proclaim$^{W^*}$2.

Statistical measurement theory (9.24)

[SMT$^{C^*}$(C^*-algebraic formulation)]	$\longleftrightarrow$	[SMT$^{W^*}$(W^*-algebraic formulation)]
Proclaim 1 (8.10)	$\longleftrightarrow$	Proclaim$^{W^*}$1 (9.11)
$\mathfrak{S}^m(\mathcal{A}^*) \ni \rho^m$	$\longleftrightarrow$	$\rho^n \in \mathfrak{S}^n(\mathcal{N}_*)$
C^*-observable $\mathbf{O} \equiv (X, \mathcal{F}, F)$	$\longleftrightarrow$	W^*-observable $\overline{\mathbf{O}} \equiv (X, \mathcal{F}, F)$
$\mathbf{M}_{\mathcal{A}}(\mathbf{O} \equiv (X, \mathcal{F}, F), S(\rho^m))$	$\longleftrightarrow$	$\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{O}} \equiv (X, \mathcal{F}, F), \overline{S}(\rho^n))$
Axiom 2 (3.26)	$\longleftrightarrow$	Proclaim$^{W^*}$2 (9.23)
general system$\mathbf{S}(\rho^m)$ $\Big(=[S(\rho^m), \{\Psi_{t_1,t_2}: \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2)\in T^2_{\leq}}]\Big)$	$\longleftrightarrow$	general system$\overline{\mathbf{S}}(\rho^n)$ $\Big(=[\overline{S}(\rho^n), \{\Psi_{t_1,t_2}: \mathcal{N}_{t_2} \to \mathcal{N}_{t_1}\}_{(t_1,t_2)\in T^2_{\leq}}]\Big)$
sequential observable$[\mathbf{O}_T]$ $\Big(=[\{\mathbf{O}_t\}_{t\in T}, \{\Psi_{t_1,t_2}: \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2)\in T^2_{\leq}}]\Big)$	$\longleftrightarrow$	sequential observable$[\overline{\mathbf{O}}_T]$ $\Big(=[\{\overline{\mathbf{O}}_t\}_{t\in T}, \{\Psi_{t_1,t_2}: \mathcal{N}_{t_2} \to \mathcal{N}_{t_1}\}_{(t_1,t_2)\in T^2_{\leq}}]\Big)$
$\mathfrak{M}(\{\mathbf{O}_t\}_{t\in T}, \mathbf{S}(\rho^m))$	$\longleftrightarrow$	$\overline{\mathfrak{M}}(\{\overline{\mathbf{O}}_t\}_{t\in T}, \overline{\mathbf{S}}(\rho^n))$

Remark 9.14. [The C^*-algebraic and the W^*-algebraic formulations]. Now we have two formulations of SMT, i.e., the C^*-algebraic formulation and the W^*-algebraic formulation. Recall that any commutative C^*-algebra [resp. commutative W^*-algebra] is represented by some $C(\Omega)$ [resp. $L^\infty(\Omega, \mu)$]. Thus, we can say that the C^*-algebraic formulation and the W^*-algebraic formulation are respectively topological and measure theoretical. Therefore, from the mathematical point of view, the W^*-algebraic formulation is handy for us to deal with "limit" or "convergence". For example, this will be seen in Theorem 10.1 (the W^*-algebraic generalization of Kolmogorov's extension theorem).[8]

■

The following theorem is essentially the same as Theorem 3.7.

Theorem 9.15. [The measurability of a general system; Compared to Theorem 3.7]. *Let* $(T \equiv \{0, 1, ..., N\}, \pi : T \setminus \{0\} \to T)$ *be a tree with root* 0 *and let* $\overline{\mathbf{S}}(\rho_0^n) \equiv [\overline{S}(\rho_0^n), \mathcal{N}_t \overset{\Phi_{\pi(t),t}}{\to} \mathcal{N}_{\pi(t)}\ (t \in T \setminus \{0\})]$ *be a general system with the initial system* $\overline{S}(\rho_0^n)$. *And, let an observable* $\overline{\mathbf{O}}_t \equiv (X_t, \mathcal{F}_t, F_t)$ *in a* C^**-algebra* $\mathcal{N}_t$ *be given for each* $t \in T$. *For each* s

[8]If readers have some knowledge of Riemann integral (defined in terms of topology) and Lebesgue integral (defined in terms of measure, *cf.* [29]), they can easily understand the mathematical handiness of "measure theoretical approach".

($\in T$), define the observable $\widetilde{\mathbf{O}}_s \equiv (\prod_{t \in T_s} X_t, \prod_{t\in T_s} \mathcal{F}_t, \widetilde{F}_s)$ *in* $\mathcal{N}_s$ *such that:*

$$\widetilde{\mathbf{O}}_s = \begin{cases} \overline{\mathbf{O}}_s & (\textit{if } s \in T \setminus \pi(T)) \\ \overline{\mathbf{O}}_s \overset{\text{qp}}{\times} (\overset{\text{qp}}{\times}_{t \in \pi^{-1}(\{s\})} \Phi_{\pi(t),t} \widetilde{\mathbf{O}}_t) & (\textit{if } s \in \pi(T)) \end{cases}$$

if possible. Then, if an observable $\widetilde{\mathbf{O}}_0$ *(i.e., the Heisenberg picture representation of the sequential observable* $[\{\overline{\mathbf{O}}_t\}_{t\in T}, \{\Phi_{t,\pi(t)} : \mathcal{N}_t \to \mathcal{N}_{\pi(t)}\}_{t \in T\setminus\{0\}}\ ]$*) in* $\mathcal{N}_0$ *exists, we have the measurement*

$$\overline{\mathbf{M}}_{\mathcal{N}_0}(\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, \prod_{t \in T} \mathcal{F}_t, \widetilde{F}_0), \overline{S}(\rho_0^n)), \tag{9.25}$$

which is called the Heisenberg picture representation of the symbol $\overline{\mathfrak{M}}(\{\overline{\mathbf{O}}_t\}_{t\in T}, \overline{\mathbf{S}}\rho_{t_0}^n))$*. If the system is classical, i.e.,* $\mathcal{N}_t \equiv L^\infty(\Omega, \mu)$ $(\forall t \in T)$*, then the measurement always exists, while the uniqueness is not always guaranteed. Also, it should be noted that, for each* $s(\in T)$*, it holds that* $\Phi_{\pi(s),s}\widetilde{F}_s(\prod_{t\in T_s} \Xi_t) = \widetilde{F}_{\pi(s)}((\Pi_{t\in T_{\pi(s)}\setminus T_s} X_t) \times (\prod_{t \in T_s} \Xi_t))$ $(\forall \Xi_t \in \mathcal{F}_t \ (\forall t \in T))$*.*

Proof. The proof is the same as that of Theorem 3.7. □

Remark 9.16. [Summing up]. In Chapters 2 ∼ 8, we studied the C^*-algebraic formulation such that

$$\text{MT}^{C^*} \begin{cases} \text{PMT}^{C^*} = \underset{\text{[Axiom 1 (2.37)]}}{\text{measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} & \text{(In Chap. 2∼7)} \\ \text{SMT}^{C^*} = \underset{\text{[Proclaim 1 (8.10)]}}{\text{statistical measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} & \text{(In Chap. 8)} \end{cases}$$

In this chapter, we study the W^*-algebraic formulation as follows:

$$\text{MT}^{W^*} \begin{cases} \text{PMT}^{W^*} = \underset{\text{[Axiom}^{W^*}\text{1 (9.11)]}}{\text{measurement}} + \underset{\text{[Proclaim}^{W^*}\text{2 (9.23)]}}{\text{the relation among systems}} \ (\text{in } \mathcal{N}) \\ \text{SMT}^{W^*} = \underset{\text{[Proclaim}^{W^*}\text{1 (9.9)]}}{\text{statistical measurement}} + \underset{\text{[Proclaim}^{W^*}\text{2 (9.23)]}}{\text{the relation among systems}} \ (\text{in } \mathcal{N}) \end{cases} \tag{9.26}$$

Here we add the remarks as follows:

(i) MT$^{C^*}$ is fundamental,

(ii) MT$^{W^*}$ should be understood by an analogy of MT$^{C^*}$. Cf. Table (9.24).

(iii) From the mathematical point of view, SMT$^{W^*}$ is more handy than SMT$^{C^*}$. (Cf. Remark 9.14).

(iv) When $\mathcal{N} = B(V)$ or $\mathcal{N} = L^\infty(\Omega, \mu)$ (where Ω is finite or countable infinite), PMT^{W^*} is meaningful (*cf.* Example 9.1).

(v) Most results in MT^{C^*} hold in MT^{W^*}. However, we omit "Fisher's maximum likelihood method" and "Generalized Bayes theorem", etc. in MT^{W^*} since the proofs are the same.

■

9.3 Quantum mechanics in $B(L^2(\mathbf{R}))$

9.3.1 Schrödinger equation and Heisenberg kinetic equation

Recall the C^*-algebraic formulation (in $\mathcal{C}(L^2(\mathbf{R}))$) of quantum mechanics (*cf.* §3.1). However, as far as quantum mechanics, the W^*-algebraic formulation (in $B(L^2(\mathbf{R}))$) is more handy than the C^*-algebraic formulation (in $\mathcal{C}(L^2(\mathbf{R}))$). (Cf. [71].) Thus, in this section, we explain the W^*-algebraic formulation of quantum mechanics (*cf.* §3.1). though it is similar to the C^*-algebraic formulation of quantum mechanics,

We begin with the classical mechanics. For simplicity, consider the one dimensional case, i.e., $\mathbf{R}_q = \{q \mid q \in \mathbf{R}\}$. Thus $q(t)$, $-\infty < t < \infty$, means the particle's position at time t, and thus, $p(t)$ ($\equiv m\frac{dq(t)}{dt}$) means the particle's momentum at time t. Let $\mathbf{R}^2_{q,p}$ ($\equiv \{(q,p) \mid q,p \in \mathbf{R}\}$ be a phase space. Define a Hamiltonian $\mathcal{H} : \mathbf{R}^2_{q,p} \to \mathbf{R}$ such that:

$$\mathcal{H}(q,p) = \frac{p^2}{2m}(\text{=kinetic energy}) \quad + \quad V(q)(\text{=potential energy}). \tag{9.27}$$

Thus we see

$$\underset{\text{(total energy)}}{E} = \mathcal{H}(q,p) = \underset{\text{(kinetic energy)}}{\frac{p^2}{2m}} + \underset{\text{(potential energy)}}{V(q)} . \tag{9.28}$$

Put $H = L^2(\mathbf{R}_q, dq)$, i.e., the Hilbert space composed of all L^2-functions on $\mathbf{R}_q$. And put $\mathcal{N} = B(L^2(\mathbf{R}_q, dq))$. Applying the quantumization:

$$E \mapsto i\hbar\frac{\partial}{\partial t}, \quad p \mapsto -i\hbar\frac{\partial}{\partial q}, \quad q \mapsto q \tag{9.29}$$

to the (9.27), we obtain the Schrödinger equation:

$$i\hbar\frac{\partial}{\partial t} = \mathcal{H}(q, -i\hbar\frac{\partial}{\partial q}) = -\frac{\hbar^2\partial^2}{2m\partial q^2} + V(q) \tag{9.30}$$

or, precisely

$$i\hbar\frac{\partial}{\partial t}\psi(q,t) = -\frac{\hbar^2\partial^2}{2m\partial q^2}\psi(q,t) + V(q)\psi(q,t). \tag{9.31}$$

This solution is formally written by

$$\psi(q,t) = e^{-\frac{i}{\hbar}\mathcal{H}(q,-i\hbar\frac{\partial}{\partial q})t}\psi(q,0). \tag{9.32}$$

Put $U(t) = e^{-\frac{i}{\hbar}\mathcal{H}(q,-i\hbar\frac{\partial}{\partial q})t}$, and $\psi(\cdot,t) = \psi_t$. Then, we see,

$$\psi_t = U(t)\psi_0. \tag{9.33}$$

Thus, the time-evolution of the state $|\psi_t\rangle\langle\psi_t|$ is represented by

$$|\psi_t\rangle\langle\psi_t| = (\Phi_t^0)_*\Big(|\psi_0\rangle\langle\psi_0|\Big) = |U(t)\psi_0\rangle\langle U(t)\psi_0|$$

Let $\overline{\mathbf{O}}_0 = (X, \mathcal{F}, F_0)$ be a W^*-observable in $B(H)$. Then, the time-evolution of the observable $\overline{\mathbf{O}}_t = (X, \mathcal{F}, F_t)$ is represented by

$$(X, \mathcal{F}, F_t) = (X, \mathcal{F}, U(t)F_0U(t)^*) = (X, \mathcal{F}, \Phi_t^0 F_0). \tag{9.34}$$

Also, it should be note that it holds that

$$\frac{dF_t}{dt} = F_t\mathcal{H} - \mathcal{H}F_t, \tag{9.35}$$

which is the Heisenberg kinetic equation. Put $\Psi_{t_1,t_2} = \Phi^0_{t_2-t_1}$, And let ρ be any element in $Tr_{+1}^m(H)$, i.e, a normal state. Then, we get the general statistical system $[\overline{S}(\rho), \{\Psi_{t_1,t_2} : B(H) \to B(H)\}_{t_1\le t_2}]$. Also, let ρ_u be any element in $Tr_{+1}^p(H)$, i.e, $\rho_u = |u\rangle\langle u|$, a pure state. Then, we get the general system $[\overline{S}_{[\rho_u]}, \{\Psi_{t_1,t_2} : B(H) \to B(H)\}_{t_1\le t_2}]$.

Although the two formulations (i.e., the W^*-algebraic formulation (in $B(L^2(\mathbf{R}))$) and the C^*-algebraic formulation (in $\mathcal{C}(L^2(\mathbf{R}))$) are similar, it should be noted that the position observable and the momentum observable can not be represented in the C^*-algebraic formulation but the W^*-algebraic formulation (*cf.* Example 9.7).

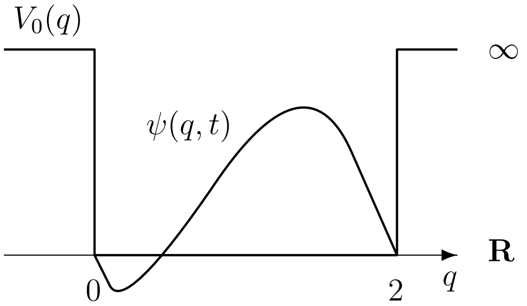
9.3.2 A simplest example of Schrödinger equation

Consider a particle with the mass m in the box (i.e., the closed interval $[0,2]$) in the one dimensional space $\mathbf{R}$. The motion of this particle (i.e., the wave function of the particle) is represented by the following Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}\psi(q,t) = -\frac{\hbar^2\partial^2}{2m\partial q^2}\psi(q,t) + V_0(q)\psi(q,t).$$

where

$$V_0(q) = \begin{cases} 0 & (0 \le q \le 2) \\ \infty & (\text{ otherwise }) \end{cases}$$



Put

$$\phi(q,t) = T(t)X(q) \qquad (0 \le q \le 2).$$

And consider the following equation:

$$i\hbar \frac{\partial}{\partial t}\phi(q,t) = -\frac{\hbar^2 \partial^2}{2m\partial q^2}\phi(q,t).$$

Then, we see

$$\frac{iT'(t)}{T(t)} = -\frac{X''(q)}{2mX(q)} = K(= \text{ constant }).$$

Then,

$$\phi(q,t) = T(t)X(q) = C_3 \exp(iKt)\Big(C_1 \exp(i\sqrt{2mK/\hbar}\, q) + C_2 \exp(-i\sqrt{2mK/\hbar}\, q).\Big)$$

Since $X(0) = X(2) = 0$ (perfectly elastic collision), putting $K = \frac{n^2\pi^2\hbar}{8m}$, we see

$$\phi(q,t) = T(t)X(q) = C_3 \exp(\frac{in^2\pi^2\hbar t}{8m}) \sin(n\pi q/2) \qquad (n = 1, 2, ...).$$

Assume the initial condition:

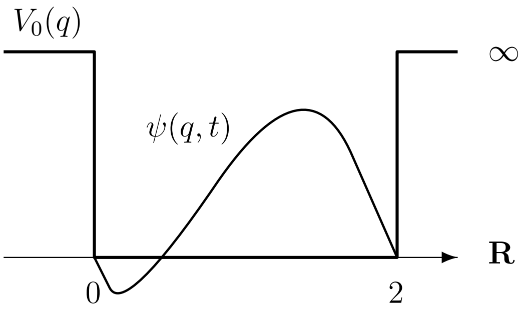
$$\psi(q,0) = c_1 \sin(\pi q/2) + c_2 \sin(2\pi q/2) + c_3 \sin(3\pi q/2) + \cdots .$$

where $\int_{\mathbf{R}} |\psi(q,0)|^2 dq = 1$. Then we see

$$\begin{aligned}&\psi(q,t)\\ =&c_1 \exp(\frac{i\pi^2\hbar t}{8m}) \sin(\pi q/2) + c_2 \exp(\frac{i4\pi^2\hbar t}{8m}) \sin(2\pi q/2) + c_3 \exp(\frac{i9\pi^2\hbar t}{8m}) \sin(3\pi q/2) + \cdots .\end{aligned}$$

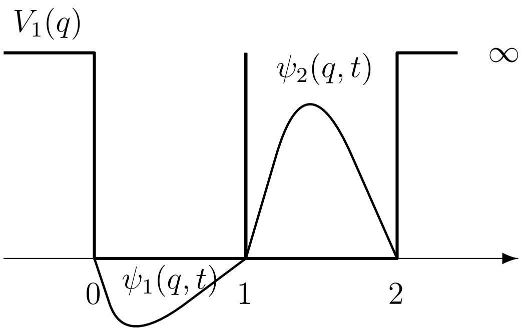
9.3.3 The de Broglie paradox

Consider the same situation in §9.3.2, i.e., a particle with the mass m in the box (i.e., the closed interval $[0,2]$) in the one dimensional space $\mathbf{R}$.

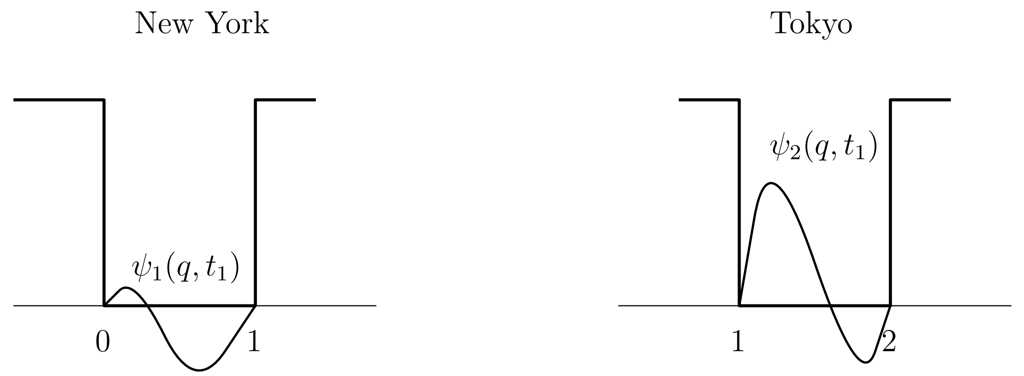


Now let us partition the box $[0,2]]$ into $[0,1]]$ and $[1,2]$. That is, we change $V_0(q)$ to $V_1(q)$, where

$$V_1(q) = \begin{cases} 0 & (0 \le q < 1) \\ \infty & (q = 1) \\ 0 & (1 < q \le 2) \\ \infty & (\text{ otherwise }) \end{cases}$$



Next, we carry the box $[0,1]$ $\big[$resp. the box $[1,2]\big]$ to New York (or, the earth) $\big[$resp. Tokyo (or, the polar star)$\big]$.



Note that the probability that we find the particle in the box $[0, 1]$ $\big[$resp. the box $[1, 2]\big]$ is given by $\int_{\mathbf{R}} |\psi_1(q, t_1)|^2 dq$ $\big[$resp. $\int_{\mathbf{R}} |\psi_2(q, t_1)|^2 dq\big]$. Here, we open the box $[0, 1]$ at New York. And assume that we find the particle in the box $[0, 1]$. Then, quantum mechanics says that at the moment the wave function ψ_2 vanishes.

Note that New York $\big[$resp. Tokyo$\big]$ may be the earth $\big[$resp. the polar star$\big]$. Thus, the above argument implies that there is something faster than light. This is called "the de Broglie paradox" (*cf.* §2.9.1, [78]).

9.4 The method of moments

9.4.1 The moment method

In this book we mainly devoted ourselves to Fisher's maximum likelihood method (*cf.* Corollary 5.6) in (pure) measurements, and Bayes' method (Cf. Theorem 6.6 and Theorem 8.13) in statistical measurements. In this section we study *"the method of moments"* (or, *the moment method*) in measurements theory (particularly, repeated measurements, cf.

Definition 2.27).

In what follows, we shall review "the method of moments" (*cf.* Definition 2.27). Let $\mathbf{M}_{\mathcal{A}}(\mathbf{O} \equiv (X, \mathcal{F}, F),\ S_{[\rho_0^p]})$ be a (pure) measurement, which may be constructed as in (8.13) of Remark 8.3. Assume the ρ_0^p (in $\mathbf{M}_{\mathcal{A}}(\mathbf{O},\ S_{[\rho_0^p]})$) is unknown. And further, we get the sample space $(X, \mathcal{F}, \nu_0)$ from the measured value $\widehat{x}$ $(= (x_1, x_2, ..., x_T) \in X^T)$ obtained by the repeated measurement $\otimes_{t=1}^T \mathbf{M}_{\mathcal{A}}(\mathbf{O} \equiv (X, \mathcal{F}, F),\ S_{[\rho_0^p]})$ $\Big(= \mathbf{M}_{\otimes \mathcal{A}}(\bigotimes_{t=1}^T \mathbf{O}, S_{[\otimes_{t=1}^T \rho_0^p]}) \Big)$. That is, $\nu_0 = \frac{1}{T} \sum_{t=1}^T \delta_{x_t}$ $\Big($i.e., $\nu_0(\Xi) = \frac{\sharp[\{k : x_t \in \Xi\}]}{T}\Big)$. Theorem 2.25 says that that $\rho^p(F(\Xi)) \approx \nu_0(\Xi)$ $(\forall \Xi \in \mathcal{F})$ if T is sufficiently large. Therefore,

- [Generalized moment method]; there is a very reason to infer the unknown ρ_0^p $(\in \mathfrak{S}^p(\mathcal{A}^*))$ such that:

$$\Delta(\nu_0, \rho_0^p(F(\,\cdot\,))\,) = \min_{\rho^p \in \mathfrak{S}^p(\mathcal{A}^*)} \Delta(\nu_0, \rho^p(F(\,\cdot\,))\,), \tag{9.36}$$

  where Δ is a certain semi-distance on $\mathcal{M}_{+1}^m(X)$.

This method is called "generalized moment method" or "moment method".

Note that the "semi-distance Δ on $\mathcal{M}_{+1}^m(X)$" is not always unique. In this sense, the moment method is somewhat artificial. If X is a finite set, it is usual to define the distance Δ on $\mathcal{M}_{+1}^m(X)$ such that:

$$\Delta(\nu_1, \nu_2) = \sum_{x \in X} |\nu_1(\{x\}) - \nu_2(\{x\})| \quad (\forall \nu_1, \nu_2 \in \mathcal{M}_{+1}^m(X)). \tag{9.37}$$

More generally, assume that X is an infinite set (and moreover, a metric space). Let $f_l : X \to \mathbf{R}$, $l = 1, 2, ..., L$, be a continuous function on X. Then, the semi-distance $\Delta_{\{f_l\}_{l=1}^L}$ on $\mathcal{M}_{+1}^m(X)$ is defined by

$$\Delta_{\{f_l\}_{l=1}^L}(\nu_1, \nu_2) = \sum_{l=1}^L \Big| \int_X f_l(x)(\nu_1(dx) - \nu_2(dx)) \Big| \quad (\forall \nu_1, \nu_2 \in \mathcal{M}_{+1}^m(X)). \tag{9.38}$$

The above argument is quite general. We usually use the following moment method.

Remark 9.17. [The simple case of (9.36)]. The minimization problem (9.36) may be somewhat troublesome. Thus, we often want to solve the equation $\Delta(\nu_0, \rho_0^p(F(\,\cdot\,))\,) = 0$ (i.e., the case of "$\min_{\rho^p \in \mathfrak{S}^p(\mathcal{A}^*)} \Delta(\nu_0, \rho^p(F(\,\cdot\,))\,) = 0$"). That is, our concern is to solve the following equation:

$$\sum_{l=1}^L \Big| \int_X f_l(x) \nu_0(dx) - \int_X f_l(x) \rho_0^p(F(dx)) \Big| = 0.$$

Or, equivalently,

$$\begin{cases} \int_X f_1(x)\nu_0(dx) = \int_X f_1(x)\rho_0^p(F(dx)) \\ \int_X f_2(x)\nu_0(dx) = \int_X f_2(x)\rho_0^p(F(dx)) \\ \qquad\qquad ..... \\ \int_X f_L(x)\nu_0(dx) = \int_X f_L(x)\rho_0^p(F(dx)). \end{cases} \tag{9.39}$$

This is usually called the method of moments.

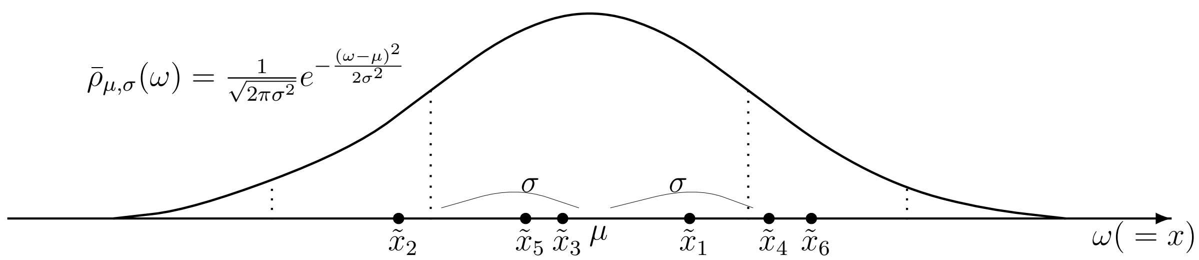
■

9.4.2 Example 1 [Normal distribution (= Gaussian distribution)]

Let $\overline{\rho}_{\mu,\sigma}$ be the Gaussian state in the commutative W^*-algebra $L^\infty(\mathbf{R}, d\omega)$ such that:

$$\rho_{\mu,\sigma}(\omega) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{(\omega-\mu)^2}{2\sigma^2}] \quad (\forall \omega \in \mathbf{R}),$$

where the average μ and the variance σ^2 are assumed to be unknown. Let $\overline{\mathbf{O}}_{\text{EXA}} \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, \chi_{(\cdot)})$ be the *exact observable* in $L^\infty(\mathbf{R}, d\omega)$ (*cf.* Example 9.4 (i)).



Consider the statistical measurement $\overline{\mathbf{M}}_{L^\infty(\mathbf{R},d\omega)}$ $(\overline{\mathbf{O}}_{\text{EXA}}, \overline{S}(\overline{\rho}_{\mu,\sigma}))$, which may be identified with the (pure) measurement $\mathbf{M}_{C_0(\mathbf{R}\times\mathbf{R}^+)}$ $(\mathbf{O}_{\text{G}} \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, G), S_{[\delta_{(\mu,\sigma)}]})$ in $C_0(\mathbf{R}\times\mathbf{R}^+)$ (*cf.* Remark 8.3 (hybrid measurements)), where $\mathbf{O}_{\text{G}}$ is defined by i.e.,

$$[G(\Xi)](\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_\Xi \exp[-\frac{(x-\mu)^2}{2\sigma^2}]dx \quad (\forall \Xi \in \mathcal{B}_{\mathbf{R}}, \forall (\mu,\sigma) \in \mathbf{R}\times\mathbf{R}^+). \tag{9.40}$$

Assume that we take the measurement $\overline{\mathbf{M}}_{L^\infty(\mathbf{R},d\omega)}$ $(\overline{\mathbf{O}}_{\text{EXA}}, \overline{S}(\overline{\rho}_{\mu,\sigma}))$ T times, that is, we take the measurement $\overline{\mathbf{M}}_{L^\infty(\mathbf{R}^T, \otimes_{t=1}^T d\omega)}$ $(\otimes_{t=1}^T \overline{\mathbf{O}}_{\text{EXA}}, \overline{S}(\otimes_{t=1}^T \overline{\rho}_{\mu,\sigma}))$, which may be identified with the (pure) measurement $\otimes_{t=1}^{T}\mathbf{M}_{C_0((\mathbf{R}\times\mathbf{R}^+))}$ $(\mathbf{O}_{\rm G} \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, G),\ S_{[\delta_{(\mu,\sigma)}]})$ $\Big($i.e., $\mathbf{M}_{C_0((\mathbf{R}\times\mathbf{R}^+)^T)}$ $(\otimes_{t=1}^{T}\mathbf{O}_{\rm G} \equiv (\mathbf{R}^T, \mathcal{B}_{\mathbf{R}^T}, \otimes_{t=1}^{T}G),\ S_{[\otimes_{t=1}^{T}\delta_{(\mu,\sigma)}]})$ in $C_0((\mathbf{R}\times\mathbf{R}^+)^T)$ (*cf.* Remark 8.3)$\Big)$. Again note that the average μ and variance σ^2 are assumed to be unknown. Here, we have the following problem:

(P) *Under the assumption that the measured value* $(\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_T)$ $(\in \mathbf{R}^T)$ *is obtained by the measurement* $\otimes_{t=1}^{T}\mathbf{M}_{C_0((\mathbf{R}\times\mathbf{R}^+))}$ $(\mathbf{O}_G \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, G),\ S_{[\delta_{(\mu,\sigma)}]})$, *infer the unknown average* μ *and variance* σ^2. (9.41)

[(i): Answer (Moment method)]. The problem (P) says that we have the sample space $(\mathbf{R}, \mathcal{B}_{\mathbf{R}}, \nu_0)$ such that:

$$\nu_0 = \frac{1}{T}\sum_{t\in T}\delta_{\tilde{x}_t}\Big(\in \mathcal{M}_{+1}^{m}(\mathbf{R})\Big). \tag{9.42}$$

Thus, it suffices to solve the following equation:

$$\Delta_{\{f_1,f_2\}}(\nu_0, [G(\,\cdot\,)](\mu_0,\sigma_0)\ ) = 0, \tag{9.43}$$

where $f_k : \mathbf{R} \to \mathbf{R}$ is usually defined by $f_1(x) = x$ and $f_2(x) = x^2$. That is, seeing (9.39), we have to solve

$$\begin{cases} (1). \quad \int_{\mathbf{R}} x\nu_0(dx) = \int_{\mathbf{R}} x[G(dx)](\mu_0,\sigma_0) \\ (2). \quad \int_{\mathbf{R}} x^2\nu_0(dx) = \int_{\mathbf{R}} x^2[G(dx)](\mu_0,\sigma_0). \end{cases} \tag{9.44}$$

The above (1) clearly implies that

$$\mu_0 = \frac{\tilde{x}_1 + \tilde{x}_2 + \cdots + \tilde{x}_T}{T}\Big(\equiv A_T \quad \text{say,}\Big) \tag{9.45}$$

Also, calculating (2)- (1)×(1), we get that

$$\sigma_0 = \sqrt{\frac{(\tilde{x}_1 - A_T)^2 + (\tilde{x}_2 - A_T)^2 + \cdots + (\tilde{x}_T - A_T)^2}{T}}. \tag{9.46}$$

This is the answer by the moment method.

[(ii): Answer (Fisher's maximum likelihood method)].

Next, we present the answer by Fisher's likelihood method. Note that the observable $\otimes_{t=1}^{T}\mathbf{O}_G = (\mathbf{R}^T, \mathcal{B}_{\mathbf{R}^T}^{\rm bd}, \otimes_{t=1}^{T}G \equiv \widehat{G})$ in $C_0((\mathbf{R}\times\mathbf{R}^+)^T)$ is represented by

$$[\widehat{G}(\Xi_1 \times \cdots \times \Xi_T)](\mu_1,\sigma_1,\mu_2,\sigma_2,\cdots,\mu_T,\sigma_T) = \Pi_{t=1}^{T}[G(\Xi_t)](\mu_t,\sigma_t).$$

Assume the condition in the above (P), and further add that

$$\Xi_t^\epsilon = [\tilde{x}_t - \epsilon, \tilde{x}_t + \epsilon], \quad \text{(for sufficiently small positive } \epsilon).$$

Since we take the (pure) measurement $\mathbf{M}_{C_0((\mathbf{R}\times\mathbf{R}^+)^T)}$ $(\otimes_{t=1}^T \mathbf{O}_{\rm G} \equiv (\mathbf{R}^T, \mathcal{B}_{\mathbf{R}^T}, \otimes_{t=1}^T G)$, $S_{[\otimes_{t=1}^T \delta_{(\mu,\sigma)}]})$ in $C_0((\mathbf{R}\times\mathbf{R}^+)^T)$, we see

$$\text{"maximum problem"} \;:\; \max_{(\mu,\sigma)\in\mathbf{R}\times\mathbf{R}^+} [\widehat{G}(\Xi_1^\epsilon \times \cdots \times \Xi_T^\epsilon)](\mu,\sigma,\mu,\sigma,\cdots,\mu,\sigma)$$

$$\Longleftrightarrow \text{"maximum problem"} \;:\; \max_{(\mu,\sigma)\in\mathbf{R}\times\mathbf{R}^+} \frac{1}{\sigma^T} \exp\Big[-\sum_{t=1}^T \frac{(\tilde{x}_t-\mu)^2}{2\sigma^2}\Big] \quad \text{(since } \epsilon \text{ is small)} \tag{9.47}$$

$$\Longleftrightarrow \begin{cases} \text{(i) } \mu = \frac{\tilde{x}_1+\tilde{x}_2+\cdots+\tilde{x}_T}{T} & (\leftarrow \quad \frac{\partial}{\partial\mu}(9.47)=0) \\ \text{(ii) } \sigma^2 = \frac{(\tilde{x}_1-\mu)^2+(\tilde{x}_2-\mu)^2+\cdots+(\tilde{x}_T-\mu)^2}{T} & (\leftarrow \quad \frac{\partial}{\partial\sigma}(9.47)=0) \\ \text{(where } \mu \text{ is defined by in the above (i))}. \end{cases} \tag{9.48}$$

Thus, Fisher's maximum likelihood method says that there is a reason to infer that

$$\mu = \frac{\tilde{x}_1+\tilde{x}_2+\cdots+\tilde{x}_T}{T} \equiv A_T, \qquad \sigma = \sqrt{\frac{(\tilde{x}_1-A_T)^2+(\tilde{x}_2-A_T)^2+\cdots+(\tilde{x}_T-A_T)^2}{T}}. \tag{9.49}$$

This is the answer by Fisher's likelihood method

9.4.3 Example 2 (measurement error model in SMT)

Put $\Omega_0 = \Omega_1 = \mathbf{R}, \Theta = \mathbf{R}^2$ and define the map $\psi^{(\theta_0,\theta_1)} : \Omega_0(\equiv \mathbf{R}) \to \Omega_1(\equiv \mathbf{R})$ such that:

$$\psi^{(\theta_0,\theta_1)}(\omega) = \theta_1\omega + \theta_0 \quad (\forall\omega \in \Omega_0(\equiv \mathbf{R}), \forall(\theta_0,\theta_1) \in \Theta \equiv \mathbf{R}^2). \tag{9.50}$$

Also, put $(X, \mathcal{F}, F) = (\mathbf{R}, \mathcal{B}_{\mathbf{R}}^{\rm bd}, G^{\sigma_1})$ in $C_0(\Omega_0)$ and $(Y, \mathcal{G}, G) = (\mathbf{R}, \mathcal{B}_{\mathbf{R}}^{\rm bd}, G^{\sigma_2})$ in $C_0(\Omega_1)$ (*cf.* Example 2.17 (Gaussian observable)), that is,

$$[G^{\sigma_i}(\Xi)](\omega) = \frac{1}{\sqrt{2\pi}\sigma_i} \int_\Xi \exp[-\frac{(x-\omega)^2}{2\sigma_i^2}]dx \qquad (\forall\Xi \in \mathcal{B}_{\mathbf{R}}^{\rm bd}, \quad \forall\omega\in\mathbf{R}, \quad i=1,2).$$

Define the product observable $\widetilde{\mathbf{O}}_{(\sigma_1,\sigma_2)}^{(\theta_0,\theta_1)} = (X\times Y, \mathcal{F}\times\mathcal{G}, H_{(\sigma_1,\sigma_2)}^{(\theta_0,\theta_1)} \equiv G^{\sigma_1} \times \Psi^{(\theta_0,\theta_1)} G^{\sigma_2})$ such that:

$$\begin{aligned} &[H_{(\sigma_1,\sigma_2)}^{(\theta_0,\theta_1)}(\Xi\times\Gamma)](\omega) \\ =&\frac{1}{(2\pi)^{2/2}\sigma_1\sigma_2} \int_{\Omega_0}\int_{\Xi\times\Gamma} \exp[-\frac{(x-\omega)^2}{2\sigma_1^2} - \frac{(y-(\theta_1\omega+\theta_0))^2}{2\sigma_2^2}]dxdyd\omega \end{aligned} \tag{9.51}$$

$$(\forall \Xi, \forall \Gamma \in \mathcal{B}_{\mathbf{R}}^{\mathrm{bd}}, \quad \forall \omega \in \Omega_0 \equiv \mathbf{R}),$$

where θ_0, θ_1 and σ_2 are assumed to be unknown, but σ_1 is known.

Let ν_{μ,σ_3} be the Gaussian state in $\mathcal{M}_{+1}^m(\Omega_0)$ such that:

$$\nu_{\mu,\sigma_3}(D) = \frac{1}{\sqrt{2\pi}\sigma_3} \int_D \exp[-\frac{(\omega-\mu)^2}{2\sigma_3^2}]d\omega \quad (\forall D \in \mathcal{B}_{\Omega_0}), \tag{9.52}$$

where the average μ and the variance $(\sigma_3)^2$ are assumed to be unknown.

Here we have the measurement $\mathbf{M}_{C_0(\Omega_0)}$ $(\widetilde{\mathbf{O}}_{(\sigma_1,\sigma_2)}^{(\theta_0,\theta_1)}, S(\overline{\rho}_{\mu,\sigma_3}))$. Define the observable $\widehat{\mathbf{O}}$ $= (X \times Y, \mathcal{F} \times \mathcal{G}, \widehat{H})$ in $C_0(\Theta \times ((\mathbf{R}^+)^3 \times \mathbf{R})$[9] such that:

$$[\widehat{H}(\Xi \times \Gamma)](\theta_0, \theta_1, \sigma_1, \sigma_2, \sigma_3, \mu) = {}_{\mathcal{M}(\Omega_0)}\Big\langle \nu_{\mu,\sigma_3}, H_{(\sigma_1,\sigma_2)}^{(\theta_0,\theta_1)}(\Xi \times \Gamma)\Big\rangle_{C_0(\Omega_0)}$$
$$= \frac{1}{(2\pi)^{3/2}\sigma_1\sigma_2\sigma_3} \int_{\Omega_0}\int_{\Xi\times\Gamma} \exp[-\frac{(x-\omega)^2}{2\sigma_1^2} - \frac{(y-(\theta_1\omega+\theta_0))^2}{2\sigma_2^2} - \frac{(\omega-\mu)^2}{2\sigma_3^2}]dxdyd\omega \tag{9.53}$$

$$(\forall \Xi, \forall \Gamma \in \mathcal{B}_{\mathbf{R}}^{\mathrm{bd}}, \quad \forall \omega \in \Omega_0 \equiv \mathbf{R}).$$

Thus we have the identification:

$$\mathbf{M}_{C_0(\Omega_0)}(\widetilde{\mathbf{O}}_{(\sigma_1,\sigma_2)}^{(\theta_0,\theta_1)}, S(\nu_{\mu,\sigma_3})) \longleftrightarrow \mathbf{M}_{C_0(\Theta\times((\mathbf{R}^+)^3\times\mathbf{R})}(\widehat{\mathbf{O}}, S_{[\delta_{(\theta_0,\theta_1,\sigma_1,\sigma_2,\sigma_3,\mu)}]}).$$

Thus, we have the sample space $(\mathbf{R}^2, \mathcal{B}_{\mathbf{R}^2}^{\mathrm{bd}}, \nu^{(\theta_0,\theta_1,\sigma_1,\sigma_2,\sigma_3,\mu)})$ such that:

$$\nu^{(\theta_0,\theta_1,\sigma_1,\sigma_2,\sigma_3,\mu)}(\Xi \times \Gamma) = [\widehat{H}(\Xi \times \Gamma)](\theta_0, \theta_1, \sigma_1, \sigma_2, \sigma_3, \mu) \quad (\forall \Xi, \forall \Gamma \in \mathcal{B}_{\mathbf{R}}). \tag{9.54}$$

Here, we have the following problem:

(P) *Assume that we take the measurement* $\mathbf{M}_{C_0(\Theta\times((\mathbf{R}^+)^3\times\mathbf{R})}(\widehat{\mathbf{O}}, S_{[\delta_{(\theta_0,\theta_1,\sigma_1,\sigma_2,\sigma_3,\mu)}]})$ *T-times, and get the measured value* $(\tilde{x}_1, \tilde{y}_1, \tilde{x}_2, \tilde{y}_2, ..., \tilde{x}_T, \tilde{y}_T)$ $(\in \mathbf{R}^{2T})$. *Here it is assumed that* $\theta_0, \theta_1, \sigma_2, \sigma_3$ *and* μ *are unknown (but* σ_1 *is known). Then, infer* θ_0 *and* θ_1 *(and moreover* σ_2, σ_3 *and* μ*) from the measured value* $(\tilde{x}_1, \tilde{y}_1, \tilde{x}_2, \tilde{y}_2, ..., \tilde{x}_T, \tilde{y}_T)$ $(\in \mathbf{R}^{2T})$ *and the known* σ_1. (9.55)

[(i): Answer (Moment method)].

[9] If $\Theta \times (\mathbf{R}^+)^3 \times \mathbf{R}$ is required to be compact, it suffices to consider $[-L, L]^2 \times [(1/L), L]^3 \times [-L, L]$ (for sufficiently large L) instead of $\Theta \times (\mathbf{R}^+)^3 \times \mathbf{R}$.

Under the notation in the problem (P), put

$$A_T^{\tilde{X}} = \frac{\widetilde{x}_1 + \widetilde{x}_2 + \cdots + \widetilde{x}_T}{T}, \qquad A_T^{\tilde{Y}} = \frac{\widetilde{y}_1 + \widetilde{y}_2 + \cdots + \widetilde{y}_T}{T}, \tag{9.56}$$

$$\overline{V}_T^{\tilde{X}\tilde{X}} = \frac{(\widetilde{x}_1 - A_T^{\tilde{X}})^2 + (\widetilde{x}_1 - A_T^{\tilde{X}})^2 + \cdots + (\widetilde{x}_T - A_T^{\tilde{X}})^2}{T}, \tag{9.57}$$

$$\overline{V}_T^{\tilde{Y}\tilde{Y}} = \frac{(\widetilde{y}_1 - A_T^{\tilde{X}})^2 + (\widetilde{y}_1 - A_T^{\tilde{X}})^2 + \cdots + (\widetilde{y}_T - A_T^{\tilde{X}})^2}{T}, \tag{9.58}$$

$$\overline{V}_T^{\tilde{X}\tilde{Y}} = \frac{(\widetilde{x}_1 - A_T^{\tilde{X}})(\widetilde{y}_1 - A_T^{\tilde{X}}) + (\widetilde{x}_2 - A_T^{\tilde{X}})(\widetilde{y}_2 - A_T^{\tilde{X}}) + \cdots + (\widetilde{x}_T - A_T^{\tilde{X}})(\widetilde{y}_T - A_T^{\tilde{X}})}{T}. \tag{9.59}$$

Recall (9.54), and put

$$\mu_X = \mu, \quad \sigma_{uu} = \sigma_1, \quad \sigma_{ee} = \sigma_2, \quad \sigma_{XX} = \sigma_3, \quad \mu_Y = \int_{\mathbf{R}} y[\widehat{H}(\mathbf{R} \times dy)]. \tag{9.60}$$

Then we see that

$$A_T^{\tilde{Y}} = \int_{\mathbf{R}} y[\widehat{H}(\mathbf{R} \times dy)](\equiv \mu_Y) = \theta_0 + \theta_1 \mu_X, \qquad A_T^{\tilde{X}} = \int_{\mathbf{R}} x[\widehat{H}(dx \times \mathbf{R})] = \mu_X, \tag{9.61}$$

and

$$\overline{V}_T^{\tilde{Y}\tilde{Y}} = \int_{\mathbf{R}} (y - \mu_Y)^2 [\widehat{H}(\mathbf{R} \times dy)] = \theta_1^2 \sigma_{XX}^2 + \sigma_{ee}^2, \tag{9.62}$$

$$\overline{V}_T^{\tilde{X}\tilde{Y}} = \int_{\mathbf{R}} (x - \mu_X)^2 [\widehat{H}(dx \times \mathbf{R})] = \theta_1 \sigma_{XX}^2, \tag{9.63}$$

$$\overline{V}_T^{\tilde{X}\tilde{X}} = \int_{\mathbf{R}^2} (x - \mu_X)(y - \mu_Y)[\widehat{H}(dx \times dy)] = \sigma_{XX}^2 + \sigma_{uu}^2, \tag{9.64}$$

which is easily solved. Thus, the moment method says that there is a reason to infer that

$$\theta_1 = (\overline{V}_T^{\tilde{X}\tilde{X}} - \sigma_1^2)^{-1} \overline{V}_T^{\tilde{X}\tilde{Y}}, \qquad \theta_0 = A_T^{\tilde{Y}} - (\overline{V}_T^{\tilde{X}\tilde{X}} - \sigma_1^2)^{-1} A_T^{\tilde{X}} \overline{V}_T^{\tilde{X}\tilde{Y}}. \tag{9.65}$$

[(ii): Answer (Fisher's maximum likelihood method)].

Next, we shall answer the problem (P) by Fisher's likelihood method. Put, for sufficiently small positive ϵ,

$$\Xi_t^\epsilon = [\tilde{x}_t - \epsilon, \tilde{x}_t + \epsilon], \quad \Gamma_t^\epsilon = [\tilde{y}_t - \epsilon, \tilde{y}_t + \epsilon] \quad (t = 1, 2, \cdots, T). \tag{9.66}$$

The probability that the measured value $(\tilde{x}_1, \tilde{y}_1, \tilde{x}_2, \tilde{y}_2, \ldots, \tilde{x}_T, \tilde{y}_T)$ $(\in \mathbf{R}^{2T})$ belongs to $\Pi_{t=1}^T (\Xi_t^\epsilon \times \Gamma_t^\epsilon)$ is given by

$$\Pi_{t=1}^T \Big[[\widehat{H}(\Xi_t^\epsilon \times \Gamma_t^\epsilon)](\theta_0, \theta_1, \sigma_1, \sigma_2, \sigma_3, \mu) \Big]. \tag{9.67}$$

Since ϵ is sufficiently small, we see, for some fixed σ_1, that

$$\max_{(\theta_0,\theta_1,\sigma_2,\sigma_3,\mu)\in\Theta\times(\mathbf{R}^+)^2\times\mathbf{R}} \Pi_{t=1}^T[\widehat{H}(\Xi_t^\epsilon \times \Gamma_t^\epsilon)](\theta_0,\theta_1,\sigma_1,\sigma_2,\sigma_3,\mu)$$

$$\Longleftrightarrow \max_{(\theta_0,\theta_1,\sigma_2,\sigma_3,\mu)\in\Theta\times(\mathbf{R}^+)^2\times\mathbf{R}} \Pi_{t=1}^T[\frac{1}{(2\pi)^{3/2}\sigma_1\sigma_2\sigma_3}\int_{\Omega_0(\equiv\mathbf{R})} e^{[-\frac{(\tilde{x}_t-\omega)^2}{2\sigma_1^2}-\frac{(\tilde{y}_t-(\theta_1\omega+\theta_0))^2}{2\sigma_2^2}-\frac{(\omega-\mu)^2}{2\sigma_3^2}]}d\omega]. \tag{9.68}$$

Thus, Fisher's maximum likelihood method says that it suffices to find the $(\theta_0,\theta_1,\sigma_2,\sigma_3,\mu)$ such that:

$$\Pi_{t=1}^T[\frac{1}{(2\pi)^{3/2}\sigma_1\sigma_2\sigma_3}\int_{\mathbf{R}} e^{[-\frac{(\tilde{x}_t-\omega)^2}{2\sigma_1^2}-\frac{(\tilde{y}_t-(\theta_1\omega+\theta_0))^2}{2\sigma_2^2}-\frac{(\omega-\mu)^2}{2\sigma_3^2}]}d\omega]$$

$$=\max_{(\theta_0,\theta_1,\sigma_2,\sigma_3,\mu)\in\Theta\times(\mathbf{R}^+)^2\times\mathbf{R}} \Pi_{t=1}^T[\frac{1}{(2\pi)^{3/2}\sigma_1\sigma_2\sigma_3}\int_{\mathbf{R}} e^{[-\frac{(\tilde{x}_t-\omega)^2}{2\sigma_1^2}-\frac{(\tilde{y}_t-(\theta_1\omega+\theta_0))^2}{2\sigma_2^2}-\frac{(\omega-\mu)^2}{2\sigma_3^2}]}d\omega]. \tag{9.69}$$

However, it may be difficult to solve it analytically. Thus, the numerical computation may be recommended.

Remark 9.18. Comparing (9.65) and (9.69), readers may consider that the moment method is simple and powerful. However, it should be noted that the moment method is somewhat artificial since the semi-distance is not unique. Summing up, we see,

	Inference	Example
(pure) measurement	Fisher's likelihood method (Theorem 5.3, Corollary 5.6)	Examples 5.8 and 5.9 Regression analysis (6.7), (6.48)
statistical measurement	Bayes' method (Theorem 8.13, Remark 8.14)	Example 8.6 Generalized Bayes theorem (Theorem 8.13)
repeated measurement (product sample space)	moment method (Definition 2.27) See §9.4.1	Normal distribution (§9.4.2) measurement error model (§9.4.3)

(9.70)

■

9.5 Principal components analysis in MT

Our present purpose is to study "principal components analysis" in the framework of MT.

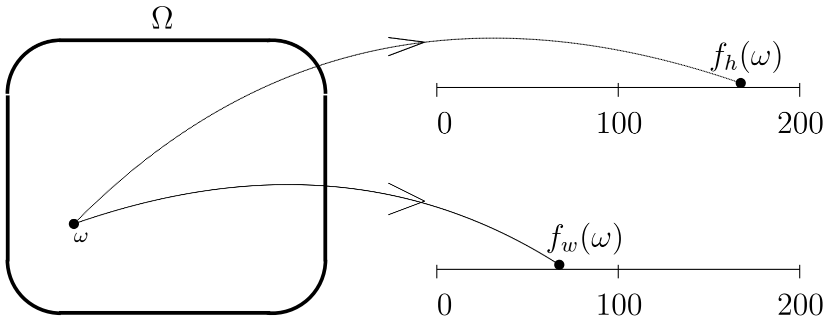
Consider the following two cases [I] and [II]:

[I: Homomorphic type]. Let Ω be a compact space. For each k $(=1,2,...,K)$, consider a continuous map $f_k:\Omega\to\mathbf{R}$. For example, we may consider that

(♯) the Ω $(=\{\omega_1,\omega_2,...,\omega_N\}$ represents the class of students in some high school. And further, assume that

(a) $f_h(\omega_n) \cdots$ the student ω_n's height

(b) $f_w(\omega_n) \cdots$ the student ω_n's weight

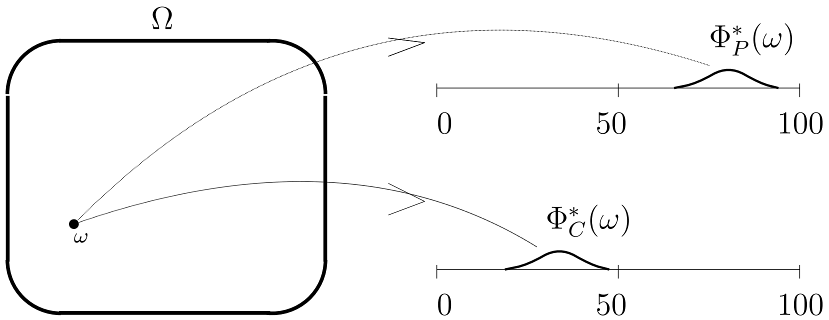


[II: Markov type]. Let Ω be a compact space. For each k $(= 1, 2, ..., K)$, consider a map $\Phi_k^* : \Omega \to \mathcal{M}_{+1}^m(\mathbf{R})$ in the C^*-algebraic formulation $\Big($or, $\Phi_k^* : \Omega \to L_{+1}^1(\mathbf{R}; dm)$ in the W^*-algebraic formulation$\Big)$. For example, we may consider that

($\sharp$) the Ω $(\equiv \{\omega_1, \omega_2, ..., \omega_N\})$ represents the set of students in some high school. And further, assume that

(a) $\Phi_P^*(\omega_n) \cdots$ the student ω_n's scholastic ability of physics (or, the distribution of the student ω_n's marks (e.g., deviation values) in physics)

(b) $\Phi_C^*(\omega_n) \cdots$ the student ω_n's scholastic ability of chemistry (or, the distribution of the student ω_n's marks (e.g., deviation values) in chemistry)



Here consider the following problem:

(P) What kind of relation among the height and weight in [I] (or, the scores of physics and chemistry in [II]) of the students of the high school can we find?

This problem (P) is usually studied by "principal components analysis". Thus, in what follows, we shall study it in the framework of PMT$^{W^*}$ (though it can be also studied in PMT$^{C^*}$ since a cyclic measurement is also formulated in PMT$^{C^*}$). Clearly the homomorphic type [I] is the special case of the Markov type [II]. Thus, from here, we devote ourselves to the Markov type [II].

Let Ω be a finite set, i.e., $\Omega = \{\omega_1, \omega_2, ..., \omega_N\}$, which is assumed to have the counting measure ν_c, that is, $\nu_c(A) = \sharp[A]$ $(\forall A \subseteq \Omega)$. For each k ($= 1, 2, ..., K$), consider a Markov operator $\Phi_k : L^\infty(\mathbf{R}, m) \to L^\infty(\Omega, \nu_c)$, where m is the Lebesgue measure on $\mathbf{R}$. Let $\mathbf{O} \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, E_{\rm EXA})$ be the exact observable in $L^\infty(\mathbf{R}, m)$. Define the observable $\widehat{\mathbf{O}} \equiv (\mathbf{R}^K, \mathcal{B}_{\mathbf{R}^K}, \times_{k=1}^K \Phi_k E_{\rm EXA})$ in $L^\infty(\Omega, \nu_c)$ such that

$$[\mathop{\times}_{k=1}^{K} \Phi_k E_{\rm EXA}](\Xi_1 \times \Xi_2 \times \cdots \times \Xi_K) = \mathop{\times}_{k=1}^{K} \Phi_k E_{\rm EXA}(\Xi_k) \quad (\Xi_1 \times \Xi_2 \times \cdots \times \Xi_K \in \mathcal{B}_{\mathbf{R}^K})$$

which is realization of the sequential observable $[\{\mathbf{O}\}_{k=1}^K, \{\Phi_k : L^\infty(\mathbf{R}, m) \to L^\infty(\Omega, \nu_c)\}_{k=1}^K]$. Thus we have the cyclic measurement $\otimes_{j=1}^{NL} \mathbf{M}_{L^\infty(\Omega,\nu_c)}(\widehat{\mathbf{O}}, \overline{S}(\bar{\rho}_{\omega_{{\rm mod}_N[j]}}))$, where $\bar{\rho}_{\omega_s} \in L^1_{+1}(\Omega, \nu_c)$, $(s = 1, 2, ..., N)$, is defined by $\bar{\rho}_{\omega_s}(\omega) = 1$ (if $\omega = \omega_s$), $= 0$ (if $\omega \neq \omega_s$).

Assume that, by the cyclic measurement $\otimes_{j=1}^{NL} \mathbf{M}_{L^\infty(\Omega,\nu_c)}(\widehat{\mathbf{O}}, \overline{S}(\bar{\rho}_{\omega_{{\rm mod}_N[j]}}))$ $\Big($or, the repeated measurement $\otimes_{j=1}^{LN} \mathbf{M}_{L^\infty(\Omega,\nu_c)}(\widehat{\mathbf{O}}, \overline{S}(1/N))$, *cf.* Example 8.7 (ii) $\Big)$, we get a measured value $(x_1, x_2, ..., x_{NL})$, where

$$\left.\begin{array}{l}
x_1 = (x_1^1, x_1^2, ..., x_1^K), \\
x_2 = (x_2^1, x_2^2, ..., x_2^K), \\
\quad \vdots \\
x_N = (x_N^1, x_N^2, ..., x_N^K) \\
x_{N+1} = (x_{N+1}^1, x_{N+1}^2, ..., x_{N+1}^K) \\
\quad \vdots \\
x_{2N} = (x_{2N}^1, x_{2N}^2, ..., x_{2N}^K) \\
\quad \vdots \\
x_{3N} = (x_{3N}^1, x_{3N}^2, ..., x_{3N}^K) \\
\quad \vdots \\
x_{LN} = (x_{NL}^1, x_{NL}^2, ..., x_{NL}^K)
\end{array}\right\} \tag{9.71}$$

Here, note that it holds:

$$\lim_{L\to\infty} \frac{\sharp[\{j \in \{1, 2, ..., NL\} \ : \ x_j \in \Xi_1 \times \Xi_2 \times \cdots \times \Xi_K\}]}{NL}$$
$$= \int_\Omega \frac{[\times_{k=1}^K \Phi_k E_{\rm EXA}(\Xi_k)](\omega)}{N} \nu_c(d\omega) \qquad (\forall \Xi_1 \times \Xi_2 \times \cdots \times \Xi_K \in \mathcal{B}_{\mathbf{R}^K})$$

Put

$$(\mu_1, \mu_2, ..., \mu_K) = \Big(\frac{\sum_{j=1}^{NL} x_j^1}{NL}, \frac{\sum_{j=1}^{NL} x_j^2}{NL}, \cdots, \frac{\sum_{j=1}^{NL} x_j^K}{NL} \Big) \tag{9.72}$$

and put

$$C_{pq} = \frac{\sum_{j=1}^{NL}(x_p^j - \mu_p)(x_q^j - \mu_q)}{NL - 1} \tag{9.73}$$

$\Big($For simplicity, here we are not concerned with the normalization, though it is reasonable.$\Big)$ Then, we have the correlation matrix C such that:

$$C = \Big[C_{pq}\Big]_{1 \le p,q \le K} = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1K} \\ C_{21} & C_{22} & \dots & C_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ C_{K1} & C_{K2} & \dots & C_{KK} \end{bmatrix}, \tag{9.74}$$

which is represented by

$$C = P \Lambda P^*$$

where Λ is a diagonal matrix such that:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_K \end{bmatrix} \quad (\lambda_1 \ge \lambda_2 \ge ... \ge 0)$$

and P is the orthonormal matrix such that:

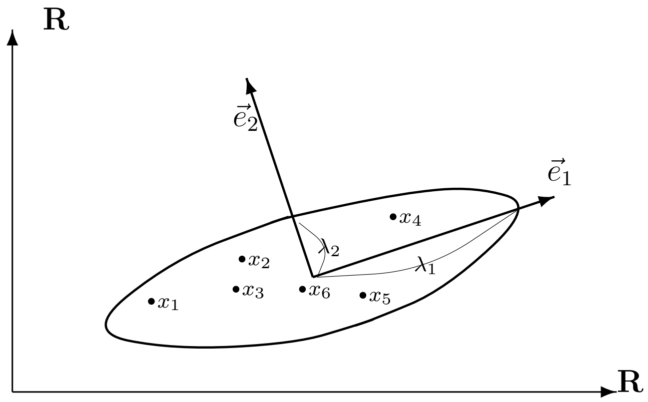
$$P = \Big[\vec{e}_1, \vec{e}_2, ..., \vec{e}_K\Big] = \begin{bmatrix} e_{11} & e_{12} & \dots & e_{1K} \\ e_{21} & e_{22} & \dots & e_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ e_{K1} & e_{K2} & \dots & e_{KK} \end{bmatrix}, \quad \vec{e}_k = \begin{bmatrix} e_{1k} \\ e_{1k} \\ \vdots \\ e_{1k} \end{bmatrix},$$

where

$$\langle \vec{e}_k, \vec{e}_{k'} \rangle_{\mathbf{R}^K} = \begin{cases} 1 & (\text{ if } k = k') \\ 0 & (\text{ if } k \ne k') \end{cases}.$$

Here, $\vec{e}_k$ is called the k-th principal component. Also, The k-contribution ratio is defined by $\frac{\lambda_k}{\sum_{i=1}^K \lambda_i}$.

Remark 9.19. [(i): Several interpretations of principal components analysis]. Principal components analysis (i.e., $\{(\vec{e}_k, \lambda_k)\}_{k=1}^K$) has several interpretations, which are important. For example, the following figure is frequently stated in usual books of statistics.



However, we are not concerned with it, because what we want to say here is the following (ii).

[(ii): Markov type and homomorphic type]. Note that the data (9.71) is obtained by the exact measurement. Thus the $\sqrt{C_{pp}}$ is not the error. In the case of Markov type, the following calculation is wrong. However, if $\Phi_k : L^\infty(\mathbf{R}, m) \to L^\infty(\Omega, \nu_c)$ is homomorphic, and if the observable $\widehat{\mathbf{O}}$ has the form such as $(\mathbf{R}^K.\mathcal{B}_{\mathbf{R}^K}, \times_{k=1}^K \Phi_k G^{\sigma_k})$ in $L^\infty(\Omega, \nu_c)$ where

$$G^{\sigma_k}_\Xi(\mu) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \int_\Xi e^{-\frac{(u-\mu)^2}{2\sigma_k^2}} du \qquad (\forall \mu \in \mathbf{R} \equiv \mathbf{R},\ \forall \Xi \in \mathcal{B}_{\mathbf{R}}). \quad (\sigma_k^2\text{: variance}),$$

(*cf.* Example 9.5 and Example 2.17), then the following calculation should be recommended: Put

$$\bar{x}_1 = \Big(\frac{\sum_{l=0}^{L-1} x^1_{1+lN}}{L}, \frac{\sum_{l=0}^{L-1} x^2_{1+lN}}{L}, \cdots, \frac{\sum_{l=0}^{L-1} x^K_{1+lN}}{L}\Big),$$
$$\bar{x}_2 = \Big(\frac{\sum_{l=0}^{L-1} x^1_{2+lN}}{L}, \frac{\sum_{l=0}^{L-1} x^2_{2+lN}}{L}, \cdots, \frac{\sum_{l=0}^{L-1} x^K_{2+lN}}{L}\Big),$$
$$\vdots$$
$$\bar{x}_N = \Big(\frac{\sum_{l=0}^{L-1} x^1_{N+lN}}{L}, \frac{\sum_{l=0}^{L-1} x^2_{N+lN}}{L}, \cdots, \frac{\sum_{l=0}^{L-1} x^K_{N+lN}}{L}\Big).$$

Put

$$(\bar{\mu}_1, \bar{\mu}_2, ..., \bar{\mu}_K) = \Big(\frac{\sum_{j=1}^N \bar{x}^1_j}{N}, \frac{\sum_{j=1}^N \bar{x}^2_j}{N}, \cdots, \frac{\sum_{j=1}^N \bar{x}^K_j}{N}\Big), \tag{9.75}$$

and put

$$\bar{C}_{pq} = \frac{\sum_{j=1}^N (\bar{x}^j_p - \bar{\mu}_p)(\bar{x}^j_q - \bar{\mu}_q)}{N}. \tag{9.76}$$

Then, we have the correlation matrix $\bar{C}$ such that:

$$\bar{C} = \Big[\bar{C}_{pq}\Big]_{1 \le p,q \le K} = \begin{bmatrix} \bar{C}_{11} & \bar{C}_{12} & \dots & \bar{C}_{1K} \\ \bar{C}_{21} & \bar{C}_{22} & \dots & \bar{C}_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{C}_{K1} & \bar{C}_{K2} & \dots & \bar{C}_{KK} \end{bmatrix}.$$

Thus, by a similar way, we can get the k-th principal component and the k-contribution ratio, etc.

Note that it holds:

$$(9.74) = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1K} \\ C_{21} & C_{22} & \dots & C_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ C_{K1} & C_{K2} & \dots & C_{KK} \end{bmatrix} = \begin{bmatrix} \bar{C}_{11} + (\sigma_1)^2 & \bar{C}_{12} & \dots & \bar{C}_{1K} \\ \bar{C}_{21} & \bar{C}_{22} + (\sigma_2)^2 & \dots & \bar{C}_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{C}_{K1} & \bar{C}_{K2} & \dots & \bar{C}_{KK} + (\sigma_K)^2 \end{bmatrix}$$

$$= (9.77) + \begin{bmatrix} (\sigma_1)^2 & 0 & \dots & 0 \\ 0 & (\sigma_2)^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (\sigma_K)^2 \end{bmatrix}$$

though the situations are different.

■

Chapter 10

Newtonian mechanics in measurement Theory

In the previous chapter, we propose the W^*-algebraic formulation of SMT:

$$\mathrm{SMT}^{W^*} = \underset{[\text{Proclaim}^{W^*}1\ (9.9)]}{\text{statistical measurement}} + \underset{[\text{Proclaim}^{W^*}2\ (9.23)]}{\text{the relation among systems}} \quad \text{in } W^*\text{-algebra} . \qquad (10.1)$$

As mentioned in Remark 1.1 (b), in this book, "Newtonian mechanics" in MT is called the "classical system theory (or dynamical system theory)". In this sense, we will study "Newtonian mechanics" in SMT^{W^*}. We first introduce "the W^*-algebraic generalization of Kolmogorov's extension theorem". This theorem is essential to MT just like Kolmogorov's extension theorem is so in his probability theory. Using this theorem, we can define "particle's trajectory" by "the sequence of measured values". And further we prove:

(i) the existence of "particle's trajectory" in Newtonian mechanics,

(ii) the existence of Brownian motion.

Thus, we can understand the difference between the concepts of "particle's trajectory" and "state's evolution" in both classical and quantum mechanics. Throughout this chapter, readers will see that, from the mathematical point of view, the W^*-algebraic formulation is more handy than the C^*-algebraic formulation.

10.1 Kolmogorov's extension theorem in W^*-algebra

In this section we study "Kolmogorov's extension theorem" in the (W^*-algebraic) Statistical MT. It is generally said that Kolmogorov's extension theorem is most fundamental in Kolmogorov's probability theory. That is because this theorem assures the existence of a probability space (i.e., sample space). On the other hand, our theorem (= Theorem 10.1, i.e., the W^*-algebraic generalization of Kolmogorov's extension theorem) assures the existence of a measurement (or, observable). Recall the our spirit $\Big($see Remark (in §2.3(I))$\Big)$:

($\sharp$) there is no probability without measurements.

Thus, in measurement theory, the concept of "measurement" is more fundamental than that of "sample space". Therefore, this theorem (i.e., the W^*-algebraic generalization of Kolmogorov's extension theorem) is very important in MT. That is, this theorem (= Theorem 10.1) is essential to MT just like Kolmogorov's extension theorem is so in his probability theory. Using this theorem, we can define "particle's trajectory" by "the sequence of measured values". And further we prove:

(i) the existence of "particle's trajectory" in Newtonian mechanics,

(ii) the existence of Brownian motion.

Thus, we can understand the difference between the concepts of "particle's trajectory" and "state's evolution" in both classical and quantum mechanics.

Let $\widehat{\Lambda}$ be an index set. For each $\lambda \in \widehat{\Lambda}$, consider a set X_λ. For any subsets $\Lambda_1 \subseteq \Lambda_2 (\subseteq \widehat{\Lambda})$, $\pi_{\Lambda_1,\Lambda_2}$ is the natural projection such that:

$$\pi_{\Lambda_1,\Lambda_2} : \mathop{\times}_{\lambda\in\Lambda_2} X_\lambda \longrightarrow \mathop{\times}_{\lambda\in\Lambda_1} X_\lambda.$$

Especially, put $\pi_\Lambda = \pi_{\Lambda,\widehat{\Lambda}}$. For each $\lambda \in \widehat{\Lambda}$, consider a W^*-observable $(X_\lambda, \mathcal{F}_\lambda, F_\lambda)$ in W^*-algebra $\mathcal{N}$. Note that the quasi-product observable $\overline{\mathbf{O}} \equiv (\times_{\lambda\in\widehat{\Lambda}} X_\lambda,\ \times_{\lambda\in\widehat{\Lambda}} \mathcal{F}_\lambda,\ F_{\widehat{\Lambda}})$ of $\{\ (X_\lambda, \mathcal{F}_\lambda,\ F_\lambda)\ \mid\ \lambda \in \widehat{\Lambda}\ \}$ is characterized as the observable such that:

$$F_{\widehat{\Lambda}}(\pi^{-1}_{\{\lambda\}}(\Xi_\lambda)) = F_\lambda(\Xi_\lambda) \qquad (\forall \Xi_\lambda \in \mathcal{F}_\lambda, \forall \lambda \in \widehat{\Lambda}), \tag{10.2}$$

though the existence and the uniqueness of a quasi-product observable are not guaranteed in general. The following theorem says something about the existence and uniqueness of the quasi-product observable.

Theorem 10.1. [W^*-algebraic generalization of Kolmogorov's extension theorem, *cf.* [43]]. *For each $\lambda \in \widehat{\Lambda}$, consider a Borel measurable space $(X_\lambda, \mathcal{F}_\lambda)$, where X_λ is a separable complete metric space. Define the set $\mathcal{P}_0(\widehat{\Lambda})$ such as* $\mathcal{P}_0(\widehat{\Lambda}) \equiv \{\Lambda \subseteq \widehat{\Lambda} \mid \Lambda \text{ is finite }\}$. *Assume that the family of the W^*-observables* $\{\ \overline{\mathbf{O}}_\Lambda \equiv (\ \times_{\lambda\in\Lambda} X_\lambda,\ \times_{\lambda\in\Lambda} \mathcal{F}_\lambda,\ F_\Lambda\ )\ \mid\ \Lambda \in \mathcal{P}_0(\widehat{\Lambda})\ \}$ *in a W^*-algebra $\mathcal{N}$ satisfies the following "W^*-algebraic consistency condition":*

- *for any* $\Lambda_1, \Lambda_2 \in \mathcal{P}_0(\widehat{\Lambda})$ *such that* $\Lambda_1 \subseteq \Lambda_2$,

$$F_{\Lambda_2}\big(\pi^{-1}_{\Lambda_1,\Lambda_2}(\Xi_{\Lambda_1})\big) = F_{\Lambda_1}\big(\Xi_{\Lambda_1}\big) \quad (\forall \Xi_{\Lambda_1} \in \mathop{\times}_{\lambda\in\Lambda_1} \mathcal{F}_\lambda). \tag{10.3}$$

Then, there uniquely exists the W^**-observable* $\widetilde{\mathbf{O}}_{\widehat{\Lambda}} \equiv \big(\times_{\lambda\in\widehat{\Lambda}} X_\lambda, \times_{\lambda\in\widehat{\Lambda}} \mathcal{F}_\lambda, \widetilde{F}_{\widehat{\Lambda}}\big)$ *in* $\mathcal{N}$ *such that:*

$$\widetilde{F}_{\widehat{\Lambda}}\big(\pi^{-1}_{\Lambda}(\Xi_{\Lambda})\big) = F_{\Lambda}\big(\Xi_{\Lambda}\big) \quad (\forall \Xi_{\Lambda} \in \mathop{\times}_{\lambda\in\Lambda} \mathcal{F}_\lambda, \ \forall \Lambda \in \mathcal{P}_0(\widehat{\Lambda})). \tag{10.4}$$

Proof. Let $\overline{\rho}$ be any normal state, i.e., $\overline{\rho} \in \mathfrak{S}^n(\mathcal{N}_*)$. Then, the $\overline{\rho}(F_\Lambda(\cdot))$ is a probability measure on the product measurable space $(\times_{\lambda\in\Lambda} X_\lambda, \times_{\lambda\in\Lambda}\mathcal{F}_\lambda)$ for all $\Lambda \in \mathcal{P}_0(\widehat{\Lambda})$. $\big($If $\mathcal{N} = L^\infty(\Omega,\mu)$, the existence is assured.$\big)$ It is clear that the family $\{\, (\times_{\lambda\in\Lambda} X_\lambda, \times_{\lambda\in\Lambda}\mathcal{F}_\lambda, \overline{\rho}(F_\Lambda(\,\cdot\,))) \;\mid\; \Lambda \in \mathcal{P}_0(\widehat{\Lambda}) \,\}$ satisfies the "usual consistency condition" in Kolmogorov's probability theory. Therefore, by Kolmogorov's extension theorem[56], there uniquely exists a probability measure $P^{\overline{\rho}}_{\widehat{\Lambda}}$ on the product measurable space $(\times_{\lambda\in\widehat{\Lambda}} X_\lambda, \times_{\lambda\in\widehat{\Lambda}}\mathcal{F}_\lambda)$ such that:

$$P^{\overline{\rho}}_{\widehat{\Lambda}}\big(\pi^{-1}_{\Lambda}(\Xi_{\Lambda})\big) = \overline{\rho}(F_{\Lambda}(\Xi_{\Lambda})) \quad (\forall \Xi_{\Lambda} \in \mathop{\times}_{\lambda\in\Lambda} \mathcal{F}_\lambda, \ \forall \Lambda \in \mathcal{P}_0(\widehat{\Lambda})). \tag{10.5}$$

Define the subfield $\times^{\sharp}_{\lambda\in\widehat{\Lambda}}\mathcal{F}_\lambda$ of $\times_{\lambda\in\widehat{\Lambda}}\mathcal{F}_\lambda$ such that:

$$\mathop{\times}^{\sharp}_{\lambda\in\widehat{\Lambda}} \mathcal{F}_\lambda = \{\pi^{-1}_{\Lambda}(\Xi_\Lambda) \mid \Xi_\Lambda \in \mathop{\times}_{\lambda\in\Lambda}\mathcal{F}_\lambda, \ \Lambda \in \mathcal{P}_0(\widehat{\Lambda})\}. \tag{10.6}$$

Then, we see, by (10.5), that there uniquely exists the countably additive function $F^{\sharp}_{\widehat{\Lambda}} : \times^{\sharp}_{\lambda\in\widehat{\Lambda}}\mathcal{F}_\lambda \to \mathcal{N}$ (in the sense of weak*-topology $\sigma(\mathcal{N},\mathcal{N}_*)$) such that:

$$P^{\overline{\rho}}_{\widehat{\Lambda}}(\Xi^{\sharp}_{\widehat{\Lambda}}) = \overline{\rho}(F^{\sharp}_{\widehat{\Lambda}}(\Xi^{\sharp}_{\widehat{\Lambda}})) \quad (\forall \Xi^{\sharp}_{\widehat{\Lambda}} \in \mathop{\times}^{\sharp}_{\lambda\in\widehat{\Lambda}}\mathcal{F}_\lambda). \tag{10.7}$$

Define the map $\widetilde{F}_{\widehat{\Lambda}} : \times_{\lambda\in\widehat{\Lambda}}\mathcal{F}_\lambda \to \mathcal{N}$ such that:

$$\widetilde{F}_{\widehat{\Lambda}}(\Xi_{\widehat{\Lambda}}) = \inf_{\{\Xi^{\sharp,k}_{\widehat{\Lambda}}\}_{k=1}^\infty \in Q(\Xi_{\widehat{\Lambda}})} \sum_{k=1}^{\infty} F^{\sharp}_{\widehat{\Lambda}}(\Xi^{\sharp,k}_{\widehat{\Lambda}}), \tag{10.8}$$

where $Q(\Xi_{\widehat{\Lambda}}) \equiv \Big\{\{\Xi^{\sharp,k}_{\widehat{\Lambda}}\}_{k=1}^\infty \mid \Xi_{\widehat{\Lambda}} \subseteq \cup_{k=1}^\infty \Xi^{\sharp,k}_{\widehat{\Lambda}}, \ \Xi^{\sharp,k}_{\widehat{\Lambda}} \in \times^{\sharp}_{\lambda\in\widehat{\Lambda}}\mathcal{F}_\lambda\Big\}$ $(\forall \Xi_{\widehat{\Lambda}} \in \times_{\lambda\in\widehat{\Lambda}}\mathcal{F}_\lambda)$. It clearly holds that

$$\widetilde{F}_{\widehat{\Lambda}}(\Gamma^1_{\widehat{\Lambda}} \cup \Gamma^2_{\widehat{\Lambda}}) \leq \widetilde{F}_{\widehat{\Lambda}}(\Gamma^1_{\widehat{\Lambda}}) + \widetilde{F}_{\widehat{\Lambda}}(\Gamma^2_{\widehat{\Lambda}}) \qquad (\forall \Gamma^1_{\widehat{\Lambda}}, \Gamma^2_{\widehat{\Lambda}} \in \mathop{\times}_{\lambda\in\widehat{\Lambda}}\mathcal{F}_\lambda).$$

Also, we see that, for any $\Xi_{\widehat{\Lambda}}$ in $\times_{\lambda \in \widehat{\Lambda}} \mathcal{F}_\lambda$,

$$
\begin{aligned}
P^{\overline{\rho}}_{\widehat{\Lambda}}(\Xi_{\widehat{\Lambda}}) &= \inf_{\{\Xi^{\sharp,k}_{\widehat{\Lambda}}\}_{k=1}^\infty \in Q(\Xi_{\widehat{\Lambda}})} \sum_{k=1}^\infty P^{\overline{\rho}}_{\widehat{\Lambda}}(\Xi^{\sharp,k}_{\widehat{\Lambda}}) && \text{(by Caratheodory theorem, } cf. \text{ [29])} \\
&= \inf_{\{\Xi^{\sharp,k}_{\widehat{\Lambda}}\}_{k=1}^\infty \in Q(\Xi_{\widehat{\Lambda}})} \sum_{k=1}^\infty \overline{\rho}(F^\sharp_{\widehat{\Lambda}}(\Xi^{\sharp,k}_{\widehat{\Lambda}})) && \text{(by (10.7))} \\
&\geq \overline{\rho}\Big(\inf_{\{\Xi^{\sharp,k}_{\widehat{\Lambda}}\}_{k=1}^\infty \in Q(\Xi_{\widehat{\Lambda}})} \sum_{k=1}^\infty F^\sharp_{\widehat{\Lambda}}(\Xi^{\sharp,k}_{\widehat{\Lambda}}) \Big) && \text{(by the property of } \mathcal{N}) \\
&= \overline{\rho}(\widetilde{F}_{\widehat{\Lambda}}(\Xi_{\widehat{\Lambda}})) && \text{(by (10.8)).}
\end{aligned}
$$

Similarly we see that $P^{\overline{\rho}}_{\widehat{\Lambda}}(\Xi^c_{\widehat{\Lambda}}) \geq \overline{\rho}(\widetilde{F}_{\widehat{\Lambda}}(\Xi^c_{\widehat{\Lambda}}))$ where $\Xi^c_{\widehat{\Lambda}} = (\times_{\lambda \in \widehat{\Lambda}} X_\lambda) \setminus \Xi_{\widehat{\Lambda}}$. Thus we see, by (10.9), that

$$
1 = P^{\overline{\rho}}_{\widehat{\Lambda}}(\Xi_{\widehat{\Lambda}}) + P^{\overline{\rho}}_{\widehat{\Lambda}}(\Xi^c_{\widehat{\Lambda}}) \geq \overline{\rho}(\widetilde{F}_{\widehat{\Lambda}}(\Xi_{\widehat{\Lambda}})) + \overline{\rho}(\widetilde{F}_{\widehat{\Lambda}}(\Xi^c_{\widehat{\Lambda}})) \geq \overline{\rho}(\widetilde{F}_{\widehat{\Lambda}}(\times_{\lambda \in \widehat{\Lambda}} X_\lambda)) = 1.
$$

This implies that $P^{\overline{\rho}}_{\widehat{\Lambda}}(\Xi_{\widehat{\Lambda}}) = \overline{\rho}(\widetilde{F}_{\widehat{\Lambda}}(\Xi_{\widehat{\Lambda}}))$. Thus we see that

$$
\overline{\rho}\Big(\widetilde{F}_{\widehat{\Lambda}}(\pi_\Lambda^{-1}(\Xi_\Lambda))\Big) = P^{\overline{\rho}}_{\widehat{\Lambda}}(\pi_\Lambda^{-1}(\Xi_\Lambda)) = \overline{\rho}\Big(F_\Lambda(\Xi_\Lambda)\Big) \quad (\forall \Xi_\Lambda \in \times_{\lambda \in \Lambda} \mathcal{F}_\lambda, \ \forall \Lambda \in \mathcal{P}_0(\widehat{\Lambda})),
$$

which implies (10.4). This completes the proof. □

10.2 The definition of "trajectories"

Now we shall propose the definition of the "trajectories" in SMT$^{W^*}$. Let $\overline{\mathbf{S}}(\overline{\rho}_0) \equiv [\overline{S}(\overline{\rho}_0), \{\Psi_{t_1,t_2} : \mathcal{N} \to \mathcal{N}\}_{(t_1,t_2) \in \mathbf{R}^2_{\leq}}]$ be a W^*-general system with an initial system $\overline{S}(\overline{\rho}_0)$. Let $\overline{\mathbf{O}} \equiv (X, \mathcal{F}, F)$ be a crisp observable in $\mathcal{N}$. For each time $t \in \overline{\mathbf{R}}^+ \equiv \{t \in \mathbf{R} \mid t \geq 0\}$, consider a W^*-observable $\overline{\mathbf{O}}_t \equiv (X_t, \mathcal{F}_t, F_t)$ in $\mathcal{N}$ such that:

- $(X_t, \mathcal{F}_t, F_t) = (X, \mathcal{F}, F)$ for all $t \in \overline{\mathbf{R}}^+$. (10.9)

Let us represent the "measurement $\overline{\mathfrak{M}}(\{\overline{\mathbf{O}}_t\}_{t \in \overline{\mathbf{R}}^+}, \overline{\mathbf{S}}(\overline{\rho}_0))$" in what follows. Let $\Lambda \in \mathcal{P}_0(\overline{\mathbf{R}}^+)$ $\Big(\equiv \{\Lambda_0 \in 2^{\overline{\mathbf{R}}^+} : \Lambda_0 \text{ is finite }\}\Big)$, that is, $\Lambda = \{t_1, t_2, ..., t_n\}$ where $0 \leq t_1 < t_2 < \cdots < t_n$. Then, we can uniquely define the observable $\overline{\mathbf{O}}_\Lambda \equiv (X^\Lambda, \mathcal{F}^\Lambda, F_\Lambda)$ at time 0 such that:

$$
F_\Lambda(\Xi_{t_1} \times \Xi_{t_2} \times \cdots \times \Xi_{t_n}) = \Psi_{0,t_1}\Big(F(\Xi_{t_1}) \cdots \Psi_{t_{n-2},t_{n-1}}\Big(F(\Xi_{t_{n-1}})(\Psi_{t_{n-1},t_n} F(\Xi_{t_n}))\Big) \cdots\Big), \tag{10.10}
$$

though the existence of $\overline{\mathbf{O}}_\Lambda$ is not always guaranteed except for the classical cases. $\Big($For the uniqueness, recall Theorem 9.8. $\Big)$ Assume that the observable $\overline{\mathbf{O}}_\Lambda$ exists for any $\Lambda \in \mathcal{P}_0(\overline{\mathbf{R}}^+)$. It is clear that the family $\{ \overline{\mathbf{O}}_\Lambda \mid \Lambda \in \mathcal{P}_0(\overline{\mathbf{R}}^+) \}$ satisfies the consistency condition (10.3). Thus, by Theorem 10.1 we have the observable $\widetilde{\mathbf{O}}_{\overline{\mathbf{R}}^+} \equiv (X^{\overline{\mathbf{R}}^+}, \mathcal{F}^{\overline{\mathbf{R}}^+}, \widetilde{F}_{\overline{\mathbf{R}}^+})$ in $\mathcal{N}$, which is called a *trajectory observable* (*concerning* $\overline{\mathbf{O}} \equiv (X, \mathcal{F}, F)$). Therefore, we get the Heisenberg picture representation $\overline{\mathbf{M}}_{\mathcal{N}}(\widetilde{\mathbf{O}}_{\overline{\mathbf{R}}^+}, \overline{S}(\overline{\rho}_0))$ of $\overline{\mathfrak{M}}(\{\overline{\mathbf{O}}_t\}_{t \in \overline{\mathbf{R}}^+}, \mathbf{S}(\overline{\rho}_0)\ )$.

Now we can propose the following definition, which is our main assertion in this chapter.

Definition 10.2. [Trajectory (= particle's trajectory)]. *Assume the above notations. The measured value obtained by the measurement* $\overline{\mathbf{M}}_{\mathcal{N}}(\widetilde{\mathbf{O}}_{\overline{\mathbf{R}}^+}, \overline{S}(\overline{\rho}_0))$ *is called a trajectory* (*concerning* $\overline{\mathbf{O}} \equiv (X, \mathcal{F}, F)$) *of the* W^**-general system* $\overline{\mathbf{S}}(\overline{\rho}_0) \equiv [\overline{S}(\overline{\rho}_0), \{\Psi_{t_1,t_2} : \mathcal{N} \to \mathcal{N}\}_{(t_1,t_2) \in \mathbf{R}^2_{\leq}}]$.

■

The difference of "particle's trajectory" and "state's evolution" is clear in Definition 10.2. That is,

$$\begin{cases} \text{"state's evolution"} & \cdots (\Psi_{0,t})_* \overline{\rho}_0, \quad (0 \leq t < \infty), \\ \text{"particle's trajectory"} & \cdots \text{ the measured value of } \overline{\mathbf{M}}_{\mathcal{N}}(\widetilde{\mathbf{O}}_{\overline{\mathbf{R}}^+}, \overline{S}(\overline{\rho}_0)). \end{cases} \tag{10.11}$$

Note that in quantum mechanics, the existence of $\widetilde{\mathbf{O}}_{\overline{\mathbf{R}}^+}$ is not usually guaranteed, and thus, the concept of "particle's trajectory" is meaningless in general (*cf.* [37, 40]).

Recall DST(1.2a), that is,

$$\boxed{\text{"dyn. syst. theor."}} = \begin{cases} \frac{dx(t)}{dt} = g(x(t), u_1(t), t), \ x(0) = \ x_0 & \cdots (\text{state equation}) , \\ y(t) = f(x(t), u_2(t), t) & (\text{ measurement equation}). \end{cases}$$

(10.12) (=(1.2a))

In order to compare (10.11) and (10.12), we add the following remark.

Remark 10.3. [(i): The case that $u_2 = 0$ in (10.12)] (The generalization of Definition 10.2). The condition (10.9) can be easily generalized as follows:

$$\bullet \quad (X_t, \mathcal{F}_t, F_t) \text{ is crisp for all } t \in \overline{\mathbf{R}}^+. \tag{10.13}$$

Under the condition, by a similar way of (10.10) we can easily define a *trajectory* (concerning $\{(X_t, \mathcal{F}_t, F_t) \mid t \in \overline{\mathbf{R}}^+\}$) of the W^*-general system $\overline{\mathbf{S}}(\overline{\rho}_0) \equiv [S(\overline{\rho}_0), \{\Psi_{t_1,t_2} : \mathcal{N} \to$

$\mathcal{N}\}_{(t_1,t_2)\in\mathbf{R}^2_{\leq}}]$. Here, consider classical cases, i.e., $\mathcal{N} = L^\infty(\Omega, \mu)$. And, for each $t \in \overline{\mathbf{R}}^+$, consider a measurable function $f_t : \Omega \to \mathbf{R}^m$, which can be identified with a crisp observable $(\mathbf{R}^m, \mathcal{B}_{\mathbf{R}^m}, F_t)$, (*cf.* (ii) in Example 9.4). Thus, by Theorem 10.1 we have the observable $\widetilde{\mathbf{O}}_{\overline{\mathbf{R}}^+} \equiv ((\mathbf{R}^m)^{\overline{\mathbf{R}}^+}, (\mathcal{B}_{\mathbf{R}^m})^{\overline{\mathbf{R}}^+}, \widetilde{F}_{\overline{\mathbf{R}}^+})$ in $\mathcal{N}$, which is called a *trajectory observable* (*concerning* $\{ \overline{\mathbf{O}}_t \equiv (\mathbf{R}^m, \mathcal{B}_{\mathbf{R}^m}, F_t) \mid t \in \overline{\mathbf{R}}^+\}$. Thus we can also define a *trajectory* (concerning $\{f_t \mid t \in \overline{\mathbf{R}}^+\}$) of the W^*-dynamical system $\overline{\mathbf{S}}(\overline{\rho}_0)$ as the trajectory concerning $\{ (\mathbf{R}^m, \mathcal{B}_{\mathbf{R}^m}, F_t) \mid t \in \overline{\mathbf{R}}^+\}$

[(ii): The case that $u_2 \neq 0$ in (10.12)] (The generalization of Definition 10.2). The condition (10.13) can be easily generalized as follows:

- $(X_t, \mathcal{F}_t, F_t)$ is not always crisp for all $t \in \overline{\mathbf{R}}^+$. (10.14)

By a similar way as in the above (i), we have the observable $\widetilde{\mathbf{O}}_{\overline{\mathbf{R}}^+} \equiv (\times_{t\in\overline{\mathbf{R}}^+} X_t, \times_{t\in\overline{\mathbf{R}}^+} \mathcal{F}_t, \widetilde{F}_{\overline{\mathbf{R}}^+})$ in $\mathcal{N}$ ($= L^\infty(\Omega; \mu)$), which is called a *trajectory observable* (*concerning* $\{ \overline{\mathbf{O}}_t \equiv (X_t, \mathcal{F}_t, F_t) \mid t \in \overline{\mathbf{R}}^+\}$).

■

10.3 Trajectories and Newtonian mechanics

In the previous section, we proposed Definition 10.2, in which the concept of "particle's trajectory" is characterized as a measured value of the measurement. Thus, our concern in this section is to show that the "particle's trajectory" is represented by the Newton equation. If it is true, we can completely understand "Newtonian mechanics" in measurement theory.

First we review Liouville's equation. Put $\mathcal{N} = L^\infty(\mathbf{R}^s_q \times \mathbf{R}^s_p, m^{2s})$ and $\mathcal{N}_* = L^1(\mathbf{R}^s_q \times \mathbf{R}^s_p, m^{2s})$, where $\mathbf{R}^s_q \times \mathbf{R}^s_p \equiv \{ (q,p) = (q_1, q_2, \cdots, q_s, p_1, p_2, \cdots, p_s) \mid q_j, p_j \in \mathbf{R}, j = 1, 2, \cdots, s \}$ and $(\mathbf{R}^s_q \times \mathbf{R}^s_p, \mathcal{B}(\mathbf{R}^s_q \times \mathbf{R}^s_p), m^{2s})$ is the $2s$-dimensional Lebesgue measure space. Liouville's equation with an initial density function $\overline{\rho}_0$ is as follows:

$$\frac{\partial \overline{\rho}_t(q,p)}{\partial t} = \sum_{j=1}^{s} \Big(\frac{\partial \mathcal{H}(q,p,t)}{\partial q_j} \frac{\partial \overline{\rho}_t(q,p)}{\partial p_j} - \frac{\partial \mathcal{H}(q,p,t)}{\partial p_j} \frac{\partial \overline{\rho}_t(q,p)}{\partial q_j} \Big), \tag{10.15}$$

$$\overline{\rho}_0 \in \mathfrak{S}^n(\mathcal{N}_*) \equiv \{\overline{\rho} \ : \ \|\overline{\rho}\|_{L^1} = 1, \ \overline{\rho} \geq 0\}, \tag{10.16}$$

where $\mathcal{H} : \mathbf{R}_q^s \times \mathbf{R}_p^s \times \mathbf{R} \to \mathbf{R}$ is a Hamiltonian. By using the solution of (10.15), we can define the operator $[\Psi_{t_1,t_2}]_* : L^1(\mathbf{R}_q^s \times \mathbf{R}_p^s, m^{2s}) \to L^1(\mathbf{R}_q^s \times \mathbf{R}_p^s, m^{2s})$ such that:

$$\Big([\Psi_{t_1,t_2}]_* \overline{\rho}_{t_1}\Big)(q,p) = \overline{\rho}_{t_2}(q,p) \quad \forall (q,p) \in \mathbf{R}_q^s \times \mathbf{R}_p^s, \ \forall (t_1,t_2) \in \mathbf{R}_{\leq}^2. \tag{10.17}$$

That is, the "state's evolution" is represented by the Schrödinger picture $\{[\Psi_{t_1,t_2}]_* \mid (t_1,t_2) \in \mathbf{R}_{\leq}^2\}$, which is induced by Liouville's equation (10.15) for states. And furthermore, putting $\Psi_{t_1,t_2} = ([\Psi_{t_1,t_2}]_*)^*$, we get the Heisenberg picture $\{\Psi_{t_1,t_2} \mid (t_1,t_2) \in \mathbf{R}_{\leq}^2\}$, which is also induced by Liouville's adjoint equation (i.e., Liouville's equation for observables). Thus, we get the W^*-dynamical system $\overline{\mathbf{S}}(\overline{\rho}_0) \equiv [S(\overline{\rho}_0), \{\Psi_{t_1,t_2} : \mathcal{N} \to \mathcal{N}\}_{(t_1,t_2)\in\mathbf{R}_{\leq}^2}]$. Also, it should be noted that the dynamical system $\overline{\mathbf{S}}(\overline{\rho}_0)$ is deterministic, i.e., each $\Psi_{t_1,t_2} : \mathcal{N} \to \mathcal{N}$ is (bijective) homomorphic.

It is well known that Liouville's equation is mathematically equivalent to the following Newton equation:

$$\frac{d}{dt}q_j(t) = \frac{\partial \mathcal{H}}{\partial p_j}(q(t),p(t),t), \quad \frac{d}{dt}p_j(t) = -\frac{\partial \mathcal{H}}{\partial q_j}(q(t),p(t),t), \quad j=1,2,\cdots,s \tag{10.18}$$

$$(q(0),p(0)) \in \mathbf{R}_q^s \times \mathbf{R}_p^s. \tag{10.19}$$

Using the solution of the Newton equation (10.18), we define the continuous map ψ_{t_1,t_2} : $\mathbf{R}_q^s \times \mathbf{R}_p^s \to \mathbf{R}_q^s \times \mathbf{R}_p^s$ such that:

$$\psi_{t_1,t_2}(q(t_1),p(t_1)) = (q(t_2),p(t_2)) \qquad (\forall (q(t_1),p(t_1)) \in \mathbf{R}_q^s \times \mathbf{R}_p^s). \tag{10.20}$$

Thus we can get the (bijective) homomorphism $\Psi_{t_1,t_2} : L^\infty(\mathbf{R}_q^s \times \mathbf{R}_p^s, m^{2s}) \to L^\infty(\mathbf{R}_q^s \times \mathbf{R}_p^s, m^{2s})$ such that:

$$(\Psi_{t_1,t_2}F)(q,p) = F(\psi_{t_1,t_2}(q,p)) \quad (\forall (q,p) \in \mathbf{R}_q^s \times \mathbf{R}_p^s, \forall F \in L^\infty(\mathbf{R}_q^s \times \mathbf{R}_p^s), \forall (t_1,t_2) \in \mathbf{R}_{\leq}^2). \tag{10.21}$$

Of course, this Ψ_{t_1,t_2} is the same as the Ψ_{t_1,t_2} derived from Liouville's equation. Since Liouville's equation and Newton equation are mathematically equivalent, there is a reason to say that the time evolution is also represented by Newton equation. However, note that the term "Newton equation" [resp. "Liouville's equation"] is, in this book, defined to be the equation that represents "particle's trajectory" [resp. "time evolution of states or observables"].

For simplicity, we put $(\Omega, \mathcal{B}, d\omega) = (\mathbf{R}_q^s \times \mathbf{R}_p^s, \mathcal{B}(\mathbf{R}_q^s \times \mathbf{R}_p^s), m^{2s})$. And, put $(\mathcal{N}, \mathcal{N}_*)$ $= (L^\infty(\Omega), L^1(\Omega))$. Consider the deterministic W^*-dynamical system $\overline{\mathbf{S}}(\overline{\rho}_0) \equiv [S(\overline{\rho}_0),$ $\{\Psi_{t_1,t_2} : \mathcal{N} \to \mathcal{N}\}_{(t_1,t_2)\in \mathbf{R}^2_{\leq}}]$, which is induced by Liouville's equation (10.15) and (10.16).

Define the *state space observable* (or, *exact observable*) $\overline{\mathbf{O}} \equiv (\Omega, \mathcal{B}, F)$ in $\mathcal{N}$ $\big(\equiv L^\infty(\Omega)\big)$ such that:

$$F(\Xi) = \chi_\Xi \quad \forall \Xi \in \mathcal{B}, \tag{10.22}$$

which is, of course, crisp. Thus, by the same arguments appearing above Definition 10.2, we can get the trajectory observable $\widetilde{\mathbf{O}}_{\overline{\mathbf{R}}^+} \equiv (\Omega^{\overline{\mathbf{R}}^+}, \mathcal{B}^{\overline{\mathbf{R}}^+}, \widetilde{F}_{\overline{\mathbf{R}}^+})$ concerning the state space observable $\overline{\mathbf{O}} \equiv (\Omega, \mathcal{B}, F)$. And therefore, we get the measurement $\overline{\mathbf{M}}_{L^\infty(\Omega)}(\widetilde{\mathbf{O}}_{\overline{\mathbf{R}}^+}, S(\overline{\rho}_0))$ (*cf.* Remark 10.3). Assume that

- a measured value $\widehat{\omega}$ $(= (\omega_t)_{t\in\overline{\mathbf{R}}^+} \in \Omega^{\overline{\mathbf{R}}^+})$ is obtained by $\overline{\mathbf{M}}_{L^\infty(\Omega)}(\widetilde{\mathbf{O}}_{\overline{\mathbf{R}}^+}, S(\overline{\rho}_0))$.

Note that the measured value $\widehat{\omega}$ is precisely the "particle's trajectory" in Definition 10.2.

Now we shall investigate the properties of the measured value $\widehat{\omega}$ $(= (\omega_t)_{t\in\overline{\mathbf{R}}^+} \in \Omega^{\overline{\mathbf{R}}^+})$, that is, we shall show that the trajectory $\widehat{\omega}$ is represented by the Newton equation (10.18) and (10.19). Let $D = \{t_0, t_1, t_2, \cdots, t_n\}$ be a finite subset of $\overline{\mathbf{R}}^+$, where $t_0 = 0 < t_1 < t_2 < \cdots < t_n$. Put $\widehat{\Xi} = \times^D_{t\in\overline{\mathbf{R}}^+} \Xi_t$ $\big(\in \mathcal{B}^{\overline{\mathbf{R}}^+}\big)$ where $\Xi_t = \Omega$ $(\forall t \notin D)$. Then, we see that

- the probability that $\widehat{\omega}(= (\omega_t)_{t\in\overline{\mathbf{R}}^+})$ belongs to the set $\widehat{\Xi} = \times^D_{t\in\overline{\mathbf{R}}^+} \Xi_t$ is given by

$$\begin{aligned}
\overline{\rho}_0(\widetilde{F}_{\overline{\mathbf{R}}^+}(\widehat{\Xi})) =& \overline{\rho}_0\Big(F(\Xi_0)\Psi_{0,t_1}\Big(F(\Xi_{t_1})\cdots\Psi_{t_{n-2},t_{n-1}}\Big(F(\Xi_{t_{n-1}})(\Psi_{t_{n-1},t_n}F(\Xi_{t_n}))\Big)\cdots\Big) \\
=& \overline{\rho}_0\Big(\Pi_{k=0}^n(\Psi_{0,t_k}F(\Xi_{t_k}))\Big) \qquad \text{(because each } \Psi_{t_{k-1},t_k} \text{ is homomorphic)} \\
=& \overline{\rho}_0\Big(\Pi_{k=0}^n F(\psi_{0,t_k}^{-1}(\Xi_{t_k}))\Big) \\
=& \int_\Omega \Big(\Pi_{k=0}^n \chi_{\psi_{0,t_k}^{-1}(\Xi_{t_k})}(\omega)\Big)\overline{\rho}_0(\omega)d\omega.
\end{aligned} \tag{10.23}$$

Let Ξ_0 be any element in $\mathcal{B}$ such that $\int_{\Xi_0} \overline{\rho}_0(\omega)d\omega \neq 0$. Thus, under the hypothesis that we know that $\omega_0 \in \Xi_0$, i.e., $\widehat{\omega}(= (\omega_t)_{t\in\overline{\mathbf{R}}^+}) \in \Xi_0 \times \Omega^{\mathbf{R}^+}$ (where $\mathbf{R}^+ = (0, \infty)$), we can calculate the following conditional probability:

$$\frac{\overline{\rho}_0\big(\widetilde{F}_{\overline{\mathbf{R}}^+}(\times^D_{t\in\overline{\mathbf{R}}^+}\Xi_t)\big)}{\overline{\rho}_0\big(\widetilde{F}_{\overline{\mathbf{R}}^+}(\Xi_0 \times \Omega^{\mathbf{R}^+})\big)} = \frac{\int_{\Xi_0}\Big(\Pi_{k=1}^n \chi_{\psi_{0,t_k}^{-1}(\Xi_{t_k})}(\omega)\Big)\overline{\rho}_0(\omega)d\omega}{\int_{\Xi_0}\overline{\rho}_0(\omega)d\omega}. \tag{10.24}$$

Thus, we see that

$$\lim_{\Xi_0 \to \{\omega_0\}} \frac{\overline{\rho}_0\big(\widetilde{F}_{\overline{\mathbf{R}}^+}(\times^D_{t \in \overline{\mathbf{R}}^+} \Xi_t)\big)}{\overline{\rho}_0\big(\widetilde{F}_{\overline{\mathbf{R}}^+}(\Xi_0 \times \Omega^{\mathbf{R}^+})\big)} = \begin{cases} 1 & \text{if } \omega_0 \in \cap_{k=1}^n \psi_{0,t_k}^{-1}(\Xi_{t_k}) \\ 0 & \text{otherwise.} \end{cases} \tag{10.25}$$

$\Big($Though the above argument is somewhat rough from the mathematical point of view, we can easily check it in mathematics.$\Big)$ This implies that

$$\omega_t = \psi_{0,t}(\omega_0) \qquad (\forall t \in \overline{\mathbf{R}}^+). \tag{10.26}$$

That is, the measured value $\widehat{\omega}$ $(= (\omega_t)_{t \in \overline{\mathbf{R}}^+} \in \Omega^{\overline{\mathbf{R}}^+})$ is the solution of the Newton equation. Also, note that the (10.25) is independent of the choice of the initial normal state $\overline{\rho}_0$.

Summing up, we see,

- In Newtonian mechanics, the state's evolution is represented by Liouville equation, and the existence of the trajectory (concerning the state space observable) is always guaranteed. That is, it can be represented by the Newton equation. Also, in quantum mechanics, the state's evolution is represented by Schrödinger equation. However, the existence of the trajectory is not always guaranteed.

That is,

	state's evolution	particle's trajectory
Newtonian mechanics	Liouville equation	Newton equation
quantum mechanics	Schrödinger equation	(meaningless)[1]

(10.27)

10.4 Brownian motions

As emphasized throughout this chapter, the concepts of "state's evolution" and "particle's trajectory" are completely different. This is, of course, a matter of common knowledge in quantum mechanics. And moreover, we can point out that the difference is clear in diffusion processes for classical systems. Therefore, in this section we examine diffusion processes in SMT$^{W^*}$. The examination will promote a better understanding of our theory.

[1]For the measurement theoretical model of Wilson chamber and its numerical analysis, see [37, 40].

Put $\mathcal{N} = L^\infty(\mathbf{R}_q, m)$ and $\mathcal{N}_* = L^1(\mathbf{R}_q, m)$, where $(\mathbf{R}_q, \mathcal{B}(\mathbf{R}_q), m)$ is the 1-dimensional Lebesgue measure space. The diffusion equation with an initial density function $\overline{\rho}_0$ at the time $t = 0$ is as follows:

$$\frac{\partial \overline{\rho}_t(q)}{\partial t} = \frac{\partial^2 \overline{\rho}_t(q)}{\partial q^2}, \tag{10.28}$$

$$\overline{\rho}_0 \in \{\overline{\rho} \in L^1(\mathbf{R}_q, m) \ : \ \|\overline{\rho}\|_{L^1} = 1, \ \overline{\rho} \geq 0\}. \tag{10.29}$$

By using the solution of (12.28), we can define the operator $[\Psi_{t_1,t_2}]_* : L^1(\mathbf{R}_q, m) \to L^1(\mathbf{R}_q, m)$ such that:

$$\Big([\Psi_{t_1,t_2}]_*(\overline{\rho}_{t_1})\Big)(q) = \overline{\rho}_{t_2}(q) = \int_{-\infty}^{\infty} \overline{\rho}_{t_1}(y) G_{t_2 - t_1}(q-y) m(dy), \quad (\forall (t_1,t_2) \in \mathbf{R}^2_{\leq}) \tag{10.30}$$

where $G_t(q)$ is the Gaussian function, that is, $G_t(q) = \frac{1}{\sqrt{2\pi t}} \exp\Big[- \frac{q^2}{2t}\Big]$. The "state's evolution" is, of course, represented by the Schrödinger picture $\{[\Psi_{t_1,t_2}]_* \mid (t_1,t_2) \in \mathbf{R}^2_{\leq}\}$.

For simplicity, we put $(\Omega, \mathcal{B}, d\omega) = (\mathbf{R}_q, \mathcal{B}(\mathbf{R}_q), m)$. And therefore, put $(\mathcal{N}, \mathcal{N}_*) = (L^\infty(\Omega), L^1(\Omega))$. Putting $\Psi_{t_1,t_2} = ([\Psi_{t_1,t_2}]_*)^*$, we get the Heisenberg picture $\{\Psi_{t_1,t_2} \mid (t_1,t_2) \in \mathbf{R}^2_{\leq}\}$, and consequently, the W^*-dynamical system $\overline{\mathbf{S}}(\overline{\rho}_0) \equiv [S(\overline{\rho}_0), \{\Psi_{t_1,t_2} : \mathcal{N} \to \mathcal{N}\}_{(t_1,t_2) \in \mathbf{R}^2_{\leq}}]$. Consider the state space observable $\overline{\mathbf{O}} \equiv (\Omega, \mathcal{B}, F)$ in $\mathcal{N}$ $\big(\equiv L^\infty(\Omega)\big)$ such as in Example 9.4.(i). Thus, by a similar way in the previous section, we get the measurement $\overline{\mathbf{M}}_{L^\infty(\Omega)}(\overline{\mathbf{O}}_{\overline{\mathbf{R}}^+}, S(\overline{\rho}_0))$. Assume that

- a measured value $\widehat{\omega}$ $(= (\omega_t)_{t \in \overline{\mathbf{R}}^+} \in \Omega^{\overline{\mathbf{R}}^+})$ is obtained by $\overline{\mathbf{M}}_{L^\infty(\Omega)}(\overline{\mathbf{O}}_{\overline{\mathbf{R}}^+}, S(\overline{\rho}_0))$.

Note that the measured value $\widehat{\omega}$ is precisely the "particle's trajectory" in Definition 10.2. Also, it may be usually called a "path".

By a similar way in the previous section, we shall investigate the properties of the measured value $\widehat{\omega}$ $(= (\omega_t)_{t \in \overline{\mathbf{R}}^+} \in \Omega^{\overline{\mathbf{R}}^+})$. Let $D = \{t_0, t_1, t_2, \cdots, t_n\}$ be a finite subset of $\overline{\mathbf{R}}^+$, where $t_0 = 0 < t_1 < t_2 < \cdots < t_n$. Put $\widehat{\Xi} = \times^D_{t \in \overline{\mathbf{R}}^+} \Xi_t$ $\big(\in \mathcal{B}^{\overline{\mathbf{R}}^+}\big)$ where $\Xi_t = \Omega$ $(\forall t \notin D)$. Then, by Proclaim$^{W^*}$2, we see

- the probability that $\widehat{\omega}(= (\omega_t)_{t \in \overline{\mathbf{R}}^+})$ belongs to the set $\widehat{\Xi} \equiv \times^D_{t \in \overline{\mathbf{R}}^+} \Xi_t$ is given by

$$\begin{aligned} &\overline{\rho}_0(\widetilde{F}_{\overline{\mathbf{R}}^+}(\widehat{\Xi})) = \overline{\rho}_0\Big(F(\Xi_0)\Psi_{0,t_1}\Big(F(\Xi_{t_1}) \cdots \Psi_{t_{n-2},t_{n-1}}\Big(F(\Xi_{t_{n-1}})\big(\Psi_{t_{n-1},t_n} F(\Xi_{t_n})\big)\Big) \cdots\Big) \\ = &\int_{\Xi_0} \overline{\rho}_0(\omega_0) \Big(\int_{\Xi_1} \Big(\cdots \Big(\int_{\Xi_{t_{n-1}}} \Big(\int_{\Xi_{t_n}} \prod_{k=1}^{n} G_{t_k - t_{k-1}}(\omega_k - \omega_{k-1}) d\omega_n \Big) d\omega_{n-1} \Big) \cdots \Big) d\omega_1 \Big) d\omega_0. \end{aligned} \tag{10.31}$$

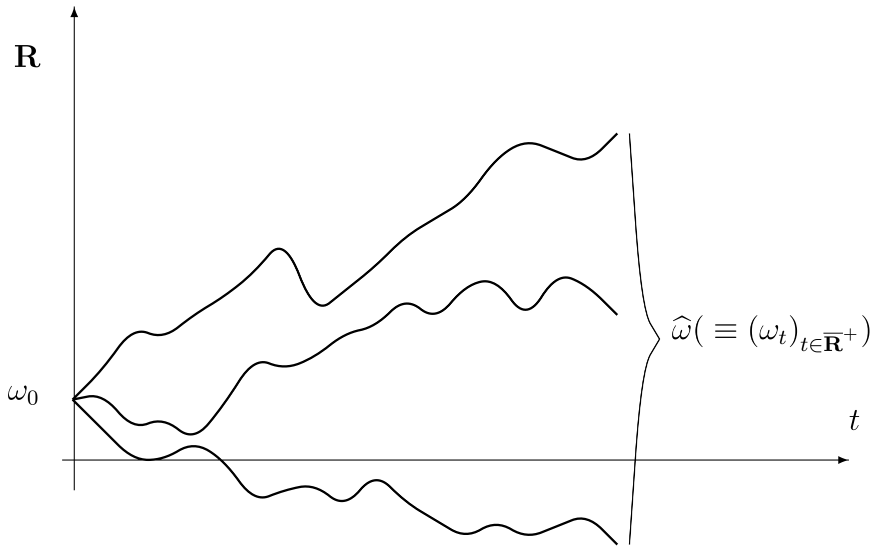
Let Ξ_0 be any element in $\mathcal{B}$ such that $\int_{\Xi_0} \overline{\rho}_0(\omega)d\omega \neq 0$. Suppose that we know that $\omega_0 \in \Xi_0$. i.e., $\widehat{\omega}(\equiv (\omega_t)_{t\in\overline{\mathbf{R}}^+}) \in \Xi_0 \times \Omega^{\mathbf{R}^+}$. Under the hypothesis, we can calculate the following conditional probability:

$$\frac{\overline{\rho}_0\big(\widetilde{F}_{\overline{\mathbf{R}}^+}(\times^D_{t\in\overline{\mathbf{R}}^+}\Xi_t)\big)}{\overline{\rho}_0\big(\widetilde{F}_{\overline{\mathbf{R}}^+}(\Xi_0 \times \Omega^{\mathbf{R}^+})\big)} = \frac{\int_{\Xi_0} \overline{\rho}_0(\omega_0)\left(\int_{\Xi_{t_1}} \cdots \int_{\Xi_{t_n}} \prod_{k=1}^n G_{t_k - t_{k-1}}(\omega_k - \omega_{k-1})d\omega_n \cdots d\omega_1\right) d\omega_0}{\int_{\Xi_0} \overline{\rho}_0(\omega_0)d\omega_0}. \tag{10.32}$$

And therefore, we see that

$$\lim_{\Xi_0 \to \{\omega_0\}} \frac{\overline{\rho}_0\big(\widetilde{F}_{\overline{\mathbf{R}}^+}(\times^D_{t\in\overline{\mathbf{R}}^+}\Xi_t)\big)}{\overline{\rho}_0\big(\widetilde{F}_{\overline{\mathbf{R}}^+}(\Xi_0 \times \Omega^{\mathbf{R}^+})\big)} = \int_{\Xi_{t_1}} \cdots \int_{\Xi_{t_n}} \prod_{k=1}^n G_{t_k - t_{k-1}}(\omega_k - \omega_{k-1})d\omega_n \cdots d\omega_1. \tag{10.33}$$

Thus, under the hypothesis that we know that $\widehat{\omega}(\equiv (\omega_t)_{t\in\overline{\mathbf{R}}^+}) \in \{\omega_0\} \times \Omega^{\mathbf{R}^+}$, the measured value $\widehat{\omega}(\equiv (\omega_t)_{t\in\overline{\mathbf{R}}^+})$ has the property like Brownian motion with the initial value ω_0. Also note that the (10.33) is independent of $\overline{\rho}_0$.



Remark 10.4. [Complex system theory]. Here I shall mention my opinion for the relation between Brownian motions and "complex system theory" (or, "chaotic system theory") as follows:

[(i): Chaotic system theory]. It is a matter of course that Brownian motion is used to analyze stochastic phenomena (*cf.* [32]). It should be noted that Brownian motion is, from the computational point of view, generated by "pseudo-random number". And moreover, it should be noted that random number generator is regarded as a kind of chaotic equation (*cf.* [19]). In this sense, we consider, from the computational point of view, that Brownian motion analysis is regarded as a kind of chaotic equation. However, chaotic theory (or complex system theory, cf [87]) should not be overestimated as "the third physics (i.e., relativity theory, quantum mechanics, complex system theory)" [2]. Chaotic theory is not such a theory. This is easily seen if chaotic theory is investigated in the framework of MT (in which "probability" (related to Axiom 1) is never born from "equations" (related to Axiom 2), *cf.* Chapter 4 ("staying time interpretation" and not "probabilistic interpretation") and Remark 8.4 (Bertrand's paradox)).

[(ii): Information compression]. Newtonian mechanics may be regarded as a kind of "information compression". In fact, if we want to know the motion of particles, it suffices to solve the Newtonian kinetic differential equation. Also, it should be noted that the differential equation is, numerically, solved by iteration method (= "loop (in computer programming)"). Thus, there is a reason to think that an iteration (= "loop"), which is mainly related to Axiom 2, is regarded as a kind of information compression method of "analytic function", "pseudo-random number", "self-similar figure (Julia and Mandelbrot set)", etc. In other words, *any figure (or graph) treated in mathematical science is always generated by iteration.* Thus, we assert that MT is also a kind of information compression method. That is, mathematical science always has the aspect such as *"mathematical method of information compression".*

[(iii): Butterfly effect]. "Butterfly effect" is mentioned as follows:

($\sharp$) *The flutter of a butterfly's wings in China could, in fact, actually effect weather patterns in New York City, thousands of miles away.*

It is impossible to test the above ($\sharp$). In this sense, we do not tell whether the ($\sharp$) is true or not. However, recall the spirit of the mechanical world view (1.12), i.e., *"at any rate, study every problem in the framework of MT".* Thus, if a certain differential equation

[2]This overestimation is like the proverb *"It's always darkest just beneath the lighthouse".* I have an opinion that Einstein's relativity theory, quantum mechanics and dynamical system theory (=DST(1.2)) are the most influential mathematical scientific theories in the 20th century, though DST is too familiar to us. The dropping of two atomic bombs (Einstein's relativity theory) is obviously one of the most tragic events in World War II. Also, Kalman filter (DST) and IC technology (quantum mechanics) lead the Apollo plan to success. This feat promoted the end of Cold War. And further, I think that this opinion is improved in this book (i.e., "quantum theory" + "DST" $\Longrightarrow$ "MT") and it is realized in Table (1.7), in which we may assert that "relativity theory (or, TOE)" $\leftrightarrow$ "the first physics", and "MT" $\leftrightarrow$"the second physics".

suggests the above fact (♯), we have to agree that there is a possibility that the above (♯) is true.

■

10.5 Conclusions

Summing up, we conclude (*cf.* [43]),

	state's evolution (≈Axiom 2)	particle's trajectory(sample space)
Newtonian mechanics	Liouville equation	Newton equation
quantum mechanics	Schrödinger equation	(meaningless)
diffusion process	diffusion equation	stochastic differential equation[3]

(10.34)

Thus there is a reason to say that the *state equation* in DST(1.2) should be called *"trajectory equation"*, though DST(1.2) is sometimes called *"state space method."* Therefore, in this book we say that DST(1.2) is the *"sample space method"*, in which the theory of differential equations and Kolmogorov's probability theory play essential roles.[4] Thus we can symbolically say:

$$\text{“MT”} \xleftarrow{\text{(our proposal)}} \underset{\text{(sample space method)}}{\text{“DST”} + \text{“statistics”}} \tag{10.35}$$

Here we have the following problem:

- Can we propose another mathematical scientific theory for data analysis? (*cf.* the third theory in Table (1.7))

I think that it is impossible to propose "the third theory" in mathematical science but computer science. Cf. Remark 1.5.

Also, recall we are not concerned with "Newtonian mechanics" in physics (which is represented in terms of differential geometry) but "Newtonian mechanics" in MT (which is represented in terms of operator algebra). Thus, it should be noted that our viewpoint (proposed in this book) is, of course, one-sided.

[3]Recall (1.2). It should be noted that the stochastic state equation (= stochastic differential equation) in (1.2) is not "state equation" but "trajectory equation (i.e., the equation that represents particle's trajectory)".

[4]I believe that "Kolmogorov's probability space" is essentially the same as "the sample space in MT".

Chapter 11

Measurement error

Let $\overline{\mathbf{Q}} \equiv (\mathbf{R}, \mathcal{B}, G)$ and $\overline{\mathbf{O}} \equiv (\mathbf{R}, \mathcal{B}, F)$ be respectively a crisp W^*-observable (i.e., quantity) and a W^*-observable in a W^*-algebra $\mathcal{N}$ such that $\overline{\mathbf{Q}}$ and $\overline{\mathbf{O}}$ commute. Under the assumption that $\overline{\mathbf{O}}$ is regarded as the approximation of $\overline{\mathbf{Q}}$, we define the measurement error $\Delta\Big(\overline{\mathbf{M}}_{\mathcal{N}}\ (\overline{\mathbf{Q}} \times \overline{\mathbf{O}}, \overline{S}(\overline{\rho}))\Big)$ by

$$\Delta\Big(\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{Q}} \times \overline{\mathbf{O}}, \overline{S}(\overline{\rho}))\Big) = \Big[\iint_{\mathbf{R}^2} |\lambda_1 - \lambda_2|^2 \overline{\rho}\big((G \times F)(d\lambda_1 d\lambda_2)\big)\Big]^{1/2}. \tag{11.1}$$

This is also called the *distance between* $\overline{\mathbf{Q}}$ *and* $\overline{\mathbf{O}}$ *concerning* $\overline{\rho}$. The purpose of this chapter is to investigate the measurement error. Readers will see that the $\Delta\Big(\overline{\mathbf{M}}_{\mathcal{N}}\ (\overline{\mathbf{Q}} \times \overline{\mathbf{O}}, \overline{S}(\overline{\rho}))\Big)$ is superior to the "conventional definition" such as |"true value" − "measured value" |.

11.1 Approximate measurements for quantities

Let $\mathcal{N}$ be a W^*-algebra. Let $\overline{\mathbf{Q}} \equiv (\mathbf{R}, \mathcal{B}, G)$ be a crisp W^*-observable (i.e., quantity) in $\mathcal{N}$. Let $\overline{\mathbf{O}} \equiv (\mathbf{R}, \mathcal{B}, F)$ be a W^*-observable in $\mathcal{N}$ such that $\overline{\mathbf{Q}}$ and $\overline{\mathbf{O}}$ commute. Let $\overline{\mathbf{Q}} \times \overline{\mathbf{O}} \equiv (\mathbf{R}^2, \mathcal{B}^2, G \times F)$ be the product observable of $\overline{\mathbf{Q}}$ and $\overline{\mathbf{O}}$. Consider the simultaneous measurement $\overline{\mathbf{M}}_{\mathcal{N}}\ (\overline{\mathbf{Q}} \times \overline{\mathbf{O}}, \overline{S}(\overline{\rho}))$. According to Proclaim$^{W^*}$1 (9.9), the probability that the measured value $(\lambda_1, \lambda_2)\ (\in \mathbf{R}^2)$ belong to $\Xi_1 \times \Xi_2\ (\in \mathcal{B}^2)$ is given by $\overline{\rho}((G \times F)(\Xi_1 \times \Xi_2))$. Thus, the variance of $|\lambda_1 - \lambda_2|$ is given by

$$\iint_{\mathbf{R}^2} |\lambda_1 - \lambda_2|^2 \overline{\rho}\big((G \times F)(d\lambda_1 d\lambda_2)\big) \tag{11.2}$$

Here we have the following definition.

Definition 11.1. [Error (or precisely, Measurement error), *cf.* [44]]. *Assume the above notations. And assume the situation that we hope to approximate* $\overline{\mathbf{Q}}$ $(\equiv (\mathbf{R}, \mathcal{B}, G))$ *by*

$\overline{\mathbf{O}}$ $(\equiv (\mathbf{R}, \mathcal{B}, F))$, *that is,* $\overline{\mathbf{O}}$ *is the approximation of* $\overline{\mathbf{Q}}$. *Then the measurement error,* $\Delta\Big(\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{Q}} \times \overline{\mathbf{O}}, \, \overline{S}(\overline{\rho}))\Big)$, *is defined by*

$$\Delta\Big(\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{Q}} \times \overline{\mathbf{O}}, \overline{S}(\overline{\rho}))\Big) = \Big[\iint_{\mathbf{R}^2} |\lambda_1 - \lambda_2|^2 \overline{\rho}\big((G \times F)(d\lambda_1 d\lambda_2)\big)\Big]^{1/2}. \tag{11.3}$$

This is also called the distance between $\overline{\mathbf{Q}}$ *and* $\overline{\mathbf{O}}$ *concerning* $\overline{\rho}$ (or, *the error of* $\overline{\mathbf{O}}$ *for* $\overline{\mathbf{Q}}$ *concerning* $\overline{\rho}$).

■

It should be noted that every measurement is *exact.* Thus the above definition is based on the following assumption:

($\sharp$) We want to take a measurement $\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{Q}}, \, \overline{S}(\overline{\rho}))$. But it is impossible for some reason. Thus, instead of the $\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{Q}}, \, \overline{S}(\overline{\rho}))$, we take a measurement $\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{O}}, \, \overline{S}(\overline{\rho}))$. In this sense, we regard $\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{O}}, \, \overline{S}(\overline{\rho}))$ as the approximation of $\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{Q}}, \, \overline{S}(\overline{\rho}))$.

The following examples will promote the understanding of Definition 11.1.

Example 11.2. [(i): Gaussian observables]. Consider the exact observable $\overline{\mathbf{O}}_{\text{EXA}} \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, \chi_{(\cdot)})$ and Gaussian observable $\overline{\mathbf{O}}_G \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, G^\sigma)$ in $\mathcal{N} \equiv L^\infty(\mathbf{R}, d\mu)$ such that:

$$[G^\sigma(\Xi)](\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_\Xi e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \qquad (\forall \mu \in \mathbf{R} \; \forall \Xi \in \mathcal{B}_{\mathbf{R}}), \tag{11.4}$$

(where σ^2 is a variance). Then we see, for each density function $\overline{\rho}$ $(\in L^1_{+1}(\mathbf{R}, d\mu))$,

$$\begin{aligned}
&\Delta\Big(\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{O}}_{\text{EXA}} \times \overline{\mathbf{O}}_G, \overline{S}(\overline{\rho}))\Big) = \Big[\iint_{\mathbf{R}^2} |\lambda_1 - \lambda_2|^2 \overline{\rho}\big((G \times G^\sigma)(d\lambda_1 d\lambda_2)\big)\Big]^{1/2} \\
=&\Big[\int_{\mathbf{R}^2} |\lambda_1 - \lambda_2|^2 \Big(\int_{\mathbf{R}} \chi_{d\lambda_1}(\mu) \frac{1}{\sqrt{2\pi\sigma^2}} \int_{d\lambda_2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \times \overline{\rho}(\mu) d\mu\Big)\Big]^{1/2} \\
=&\sigma,
\end{aligned} \tag{11.5}$$

which is independent of $\overline{\rho}$.

[(ii): Triangle observable, *cf.* Example 2.19]. Let ϵ be any positive number. Define the membership function (i.e., triangle function) $\mathcal{Z}_\epsilon : \mathbf{R} \to \mathbf{R}$ such that:

$$\mathcal{Z}_\epsilon(\omega) = \begin{cases} 1 - \frac{\omega}{\epsilon} & 0 \le \omega \le \epsilon \\ \frac{\omega}{\epsilon} + 1 & -\epsilon \le \omega \le 0 \\ 0 & \text{otherwise} . \end{cases} \tag{11.6}$$

Put $\mathbb{Z}_\epsilon \equiv \big\{\epsilon k \;:\; k \in \mathbb{Z} \equiv \{0, \pm 1, \pm 2, ...\}\big\}$. Define the W^*-observable $\overline{\mathbf{O}}_T \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, T^\epsilon_{(\cdot)})$ in the commutative W^*-algebra $L^\infty(\mathbf{R}, d\omega)$ such that $T^\epsilon_\Xi(\omega) = \sum_{x \in \Xi \cap \mathbb{Z}_\epsilon} \mathcal{Z}_\epsilon(\omega - x)$ $(\forall \Xi \in$

$\mathcal{B}_{\mathbf{R}}$, $\forall \omega \in \mathbf{R}$). This W^*-observable $\overline{\mathbf{O}}_T$ is called a *triangle observable* in $L^\infty(\mathbf{R}, d\omega)$. Consider the exact observable $\overline{\mathbf{O}}_{\text{EXA}} \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, \chi_{(\cdot)})$ and the triangle observable $\overline{\mathbf{O}}_T \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, T^\epsilon_{(\cdot)})$ in $\mathcal{N} \equiv L^\infty(\mathbf{R}, d\omega)$. Then we see, for each density function $\overline{\rho}$ ($\in L^1_{+1}(\mathbf{R}, d\omega)$),

$$\Delta\Big(\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{O}}_{\text{EXA}} \times \overline{\mathbf{O}}_T, \overline{S}(\overline{\rho}))\Big) = \epsilon\Big[\int_{\mathbf{R}} (\omega - [\omega]_G)(1 - [\omega]_G + \omega)\overline{\rho}(\omega) d\omega\Big]^{1/2} \leq \frac{\epsilon}{2}$$

where $[\omega]_G$ is the integer such that $[\omega]_G \leq \omega < [\omega]_G + 1$. ■

Example 11.3. [Self-adjoint operators]. Let A_1 and A_2 be commutative self-adjoint operators on a Hilbert space H. For each i ($= 1, 2$), consider the crisp observable $\overline{\mathbf{O}}_i \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, E_{A_i})$ in $B(H)$ which is the spectral measure of A_i, i.e., $A_i = \int_{\mathbf{R}} \lambda E_{A_i}(d\lambda)$. Then, we see that

$$\begin{aligned} \Delta\Big(\overline{\mathbf{M}}_{B(H)}(\overline{\mathbf{O}}_1 \times \overline{\mathbf{O}}_2, \overline{S}(|u\rangle\langle u|))\Big) &= \Big[\int_{\mathbf{R}^2} |\lambda_1 - \lambda_2|^2 \Big\langle u, E_{A_1}(d\lambda_1)E_{A_2}(d\lambda_2)u\Big\rangle\Big]^{1/2} \\ &= \|(A_1 - A_2)u\|^2. \end{aligned} \tag{11.7}$$

■

11.2 The estimation under loss function in statistics

Let $\overline{\mathbf{Q}} \equiv (\mathbf{R}, \mathcal{B}, G)$ and $\overline{\mathbf{O}} \equiv (X, \mathcal{F}, F)$ be a quantity (i.e., a crisp observable on $\mathbf{R}$) and a W^*-observable in a W^*-algebra $\mathcal{N}$ respectively. Consider the measurable map $h : X \to \mathbf{R}$, and the image observable $\overline{\mathbf{O}}_{[h]} \equiv (\mathbf{R}, \mathcal{B}, F(h^{-1}(\,\cdot\,))\,)$ in $\mathcal{N}$. This measurable map $h : X \to \mathbf{R}$ is called a *statistic*. Also assume that $\overline{\mathbf{Q}}$ and $\overline{\mathbf{O}}_{[h]}$ commute. Thus, the distance between $\overline{\mathbf{Q}}$ and $\overline{\mathbf{O}}_h$ (concerning $\overline{\rho} \in \mathfrak{S}^n(\mathcal{N}_*)$) is defined by $\Delta(\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{Q}} \times \overline{\mathbf{O}}_h, \overline{S}(\overline{\rho}))\,)$ as in the above definition.

Now we have the following problem:

Problem 11.4. [The estimation under loss function in statistics]. Assume the above notations. Then our present problem is as follows:

(♯) how to choose a proper image observable $\overline{\mathbf{O}}_{[h]}$ (i.e., $\overline{\mathbf{O}}$ ($\equiv (X, \mathcal{F}, F)$) and $h : X \to \mathbf{R}$) as the approximation of a quantity $\overline{\mathbf{Q}}$ ($\equiv (\mathbf{R}, \mathcal{B}, G)$).

Our interest is concentrated on the problem (♯), which is regarded as a kind of "inference". Note that this (♯) is entirely different from Fisher's spirit in Chapter 5, that is, *how to infer the unknown state from the measured data obtained by a measurement.*

Of course, it is desirable that $\overline{\mathbf{O}}$ and h in the above ($\sharp$) satisfy the following (A_1) and (A_2).

(A_1) (unbias condition). There exists a dense set D ($\in \mathfrak{S}^n(\mathcal{N}_*)$) such that:

$$\int_{\mathbf{R}} \lambda \ {}_{\mathcal{N}_*}\langle \rho, G(d\lambda)\rangle_{\mathcal{N}} = \int_{\mathbf{R}} \lambda \ {}_{\mathcal{N}_*}\langle \rho, F(h^{-1}(d\lambda))\rangle_{\mathcal{N}} \qquad (\forall \rho \in D)$$

(A_2) $\Delta(\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{Q}} \times \overline{\mathbf{O}}_{[h]}, \overline{S}(\overline{\rho}))$) is small (where $\overline{\mathbf{Q}}$ and $\overline{\mathbf{O}}_{[h]}$ commute).

In what follows, we shall study Problem 11.4 in Example 11.5 and Problem 11.6.

Example 11.5. [Heisenberg's uncertainty relation, *cf.* [31], [36], Chapter 12]. Let A_1 and A_2 be a position quantity and a momentum quantity respectively (i.e. A_1 and A_2 are self-adjoint operators on a Hilbert space H satisfying that $A_1A_2 - A_2A_1 = i\hbar$, $\hbar$ is "Plank constant" $/(2\pi)$). As mentioned before, we identify A_i with the spectral measure $\overline{\mathbf{A}}_i \equiv (\mathbf{R}, \mathcal{B}, G_i)$ in $B(H)$, i.e., $A_i = \int_{\mathbf{R}} \lambda G_i(d\lambda)$. Since A_1 and A_2 do not commute, the product observable does not exist. Therefore, consider an observable $\overline{\mathbf{O}} \equiv (X, \mathcal{F}, F)$ in $B(H)$ and measurable maps $h_i : X \to \mathbf{R}$, ($i = 1, 2$), and define the image observables $\overline{\mathbf{O}}_{[h_i]} \equiv (\mathbf{R}, \mathcal{B}, F(h_i^{-1}(\,\cdot\,)) \equiv F_i(\,\cdot\,))$ in $B(H)$. And furthermore, assume the conditions:

(i) There exists a set D ($\subset H$) such that $\overline{D}$ ($\equiv$"closure on D)$= \{u \in H \mid \|u\| = 1\}$ and it holds that $\langle u, A_i u\rangle_H = \int_{\mathbf{R}} \lambda \langle u, F_i(d\lambda))u\rangle_H$ ($\forall u \in D$, $i = 1, 2$

(ii) $\overline{\mathbf{Q}}_i$ and $\overline{\mathbf{O}}_{[h_i]}$ commute ($i = 1, 2$).

Then we get the following inequality:

$$\Delta\Big(\overline{\mathbf{M}}_{B(H)}(\overline{\mathbf{Q}}_1 \times \overline{\mathbf{O}}_{[h_1]}, \overline{S}(\overline{\rho}))\Big) \cdot \Delta\Big(\overline{\mathbf{M}}_{B(H)}(\overline{\mathbf{Q}}_2 \times \overline{\mathbf{O}}_{[h_2]}, \overline{S}(\overline{\rho}))\Big) \geq \hbar/2 \quad \text{for all } \overline{\rho} \in \mathrm{Tr}_{+1}(H). \tag{11.8}$$

This is just Heisenberg's uncertainty relation, of which non-mathematical representation was proposed by W. Heisenberg in the famous thought experiment of γ-rays microscope (*cf.* [31]). This will be discussed in Chapter 12.

■

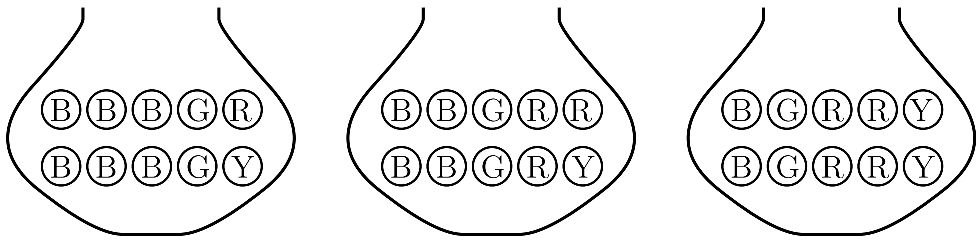
The following problem is a main part of this section. The reader should find "estimation under loss function in statistics" in the following problem.

Problem 11.6. [= Example 5.9 (Urn problem)]. Let U_j, $j = 1, 2, 3$, be urns that contain sufficiently many colored balls as follows:

	blue balls	green balls	red balls	yellow balls
urn U_1	60%	20%	10%	10%
urn U_2	40%	20%	30%	10%
urn U_3	20%	20%	40%	20%

Put $\mathbf{U} = \{U_1, U_2, U_3\}$. By the same argument in Example 5.9, we consider the state space Ω $\big(\equiv \{\omega_1, \omega_2, \omega_3\} \big)$ with the discrete topology, which is identified with $\mathbf{U}$, that is, $\mathbf{U} \ni U_j \leftrightarrow \omega_j \in \Omega \approx \mathcal{M}_{+1}^p(\Omega)$.

$U_1 \approx \omega_1$ $\qquad$ $U_2 \approx \omega_2$ $\qquad$ $U_3 \approx \omega_3$



Let Q be a quantity in $C(\Omega)$, i.e., $Q : \Omega\ (\approx \mathcal{M}_{+1}^p(\Omega)) \to \mathbf{R}$ is a real valued continuous function on Ω. For example we may consider in what follows. Assume that the weight of a blue ball is given by 10 (gram), and green 20, red 30 and yellow 10. $\Big($Thus, we can define the map $W : X \to \mathbf{R}$ such that $W(b) = 10$, $W(g) = 20$, $W(r) = 30$ and $W(y) = 10$.$\Big)$ Therefore, we can define the quantity $Q : \Omega \to [0, 50]$ such that the average weight $Q(\omega_1)$ of the balls in the urn U_1 is given by 14 $(= (10{\cdot}60+20{\cdot}20+30{\cdot}10+10{\cdot}10)/100)$, and similarly, $Q(\omega_2) = 18$ and $Q(\omega_3) = 20$. Define the observable $\mathbf{O} \equiv (X = \{b, g, r, y, \}, 2^X, F_{(\cdot)})$ in $C(\Omega)$ by the usual way. That is,

$$
\begin{array}{llll}
F_{\{b\}}(\omega_1) = 6/10 & F_{\{g\}}(\omega_1) = 2/10 & F_{\{r\}}(\omega_1) = 1/10 & F_{\{y\}}(\omega_1) = 1/10 \\
F_{\{b\}}(\omega_2) = 4/10 & F_{\{g\}}(\omega_2) = 2/10 & F_{\{r\}}(\omega_2) = 3/10 & F_{\{y\}}(\omega_2) = 1/10 \\
F_{\{b\}}(\omega_3) = 2/10 & F_{\{g\}}(\omega_3) = 2/10 & F_{\{r\}}(\omega_3) = 4/10 & F_{\{y\}}(\omega_3) = 2/10.
\end{array}
$$

Now consider the iterated measurement $\mathbf{M}_{C(\Omega)}(\times_{k=1}^2 \mathbf{O} \equiv (X^2, 2^{X^2}, \times_{k=1}^2 F), S_{[*]})$ where $(\times_{k=1}^2 F)_{\Xi_1 \times \Xi_2}(\omega) = F_{\Xi_1}(\omega) \cdot F_{\Xi_2}(\omega)$. Also, assume that

- the measured value (b, r) is obtained by the simultaneous measurement $\mathbf{M}_{C(\Omega)}(\times_{k=1}^2 \mathbf{O}, S_{[*]})$.

Now we have the following problem.

($\sharp$) *How do we infer* $Q(*)$ *from the measured value* (b, r) *obtained by the simultaneous measurement* $\mathbf{M}_{C(\Omega)}(\times_{k=1}^{2} \mathbf{O}, S_{[*]})$ *?*

■

In what follows, we provide four answers to the above problem.

Answer 1. [Fisher's method, *cf.* [44]]. Recall "[II]" in Example 5.8, in which we infer, by Fisher's method, that the unknown urn is U_2. That is, applying Fisher's method (*cf.* Corollary 5.6), we get the conclusion as follows: Put $E(\omega) = F_{\{b\}}(\omega) F_{\{r\}}(\omega)$. Clearly it holds that $E(\omega_1) = 6 \cdot 1/10^2 = 0.06$, $E(\omega_2) = 4 \cdot 3/10^2 = 0.12$ and $E(\omega_3) = 4 \cdot 2/10^2 = 0.08$. Therefore, there is a very reason to think that $[\,*\,] = \delta_{\omega_2}$, that is, the unknown urn is U_2. Since we inferred that $[\,*\,] = \delta_{\omega_2}$ ($\leftrightarrow \omega_2$) in Example 5.8(II), we can immediately conclude that (or more precisely, Regression analysis II (6.48))

$$Q(\,*\,) = Q(\omega_2) = 18.$$

Answer 2. [Moment method] Recall "Remark" in Example 5.8, in which we infer, by the moment method, that the unknown urn is U_2. Thus, we conclude that $Q(*) = Q(U_2) = 18$.

Answer 3. [Bayes' method in SMT$_{\rm PEP}$]. Next study the above problem ($\sharp$) in SMT$_{\rm PEP}$-method (*cf.* §8.6.2, and Theorem 11.12 later). Thus, we assume that the $[\,*\,]$ is chosen by a fair rule (e.g., a fair coin-tossing, a fair dice-throwing, etc.). Consider a statistical measurement $\mathbf{M}_{C(\Omega)}(\times_{k=1}^{2} \mathbf{O},\ S_{[*]}(\rho_0^m)\ )$, where we assume that $\rho_0^m = \rho_{\rm uni}^m$, i.e., $\rho_{\rm uni}^m = \frac{1}{3}\sum_{j=1}^{3} \delta_{\omega_j}$ on Ω. When we get the measured value (b, r) by the measurement $\mathbf{M}_{C(\Omega)}(\times_{k=1}^{2} \mathbf{O},\ S_{[*]}(\rho_0^m)\ )$, we infer, by Bayes' method (for example, (B_1) in Remark 8.14, or more precisely, Theorem 8.13), that the new state $\rho_{\rm new}^m$ is

$$\begin{aligned}\rho_{\rm new}^m &= \frac{1}{0.06+0.12+0.08}(0.06 \cdot \delta_{\omega_1} + 0.12 \cdot \delta_{\omega_2} + 0.08 \cdot \delta_{\omega_3}) \\ &= \frac{1}{6+12+8}(6 \cdot \delta_{\omega_1} + 12 \cdot \delta_{\omega_2} + 8 \cdot \delta_{\omega_3}).\end{aligned}$$

Thus there is a very reason to consider that

$$Q(\,*\,) \text{ is approximated by } \int_\Omega Q(\omega)\rho_{\rm new}^m(d\omega) = \tfrac{14 \cdot 6 + 18 \cdot 12 + 20 \cdot 8}{6+12+8} = 17.69\cdots.$$

Also, the variance σ^2 is given by

$$\sigma = \Big[\frac{(14-17.69)^2 \cdot 6 + (18-17.69)^2 \cdot 12 + (20-17.69)^2 \cdot 8}{6+12+8}\Big]^{1/2} = 2.19...$$

Answer 4. [The estimation under loss function in statistics, *cf.* [44]]. Let $\mathbf{M}_{C(\Omega)}$ ($\times_{k=1}^2 \mathbf{O}$, $S_{[*]}(\rho_0^m)$) and $Q : \Omega \to [0, 50]$ be as in Problem 11.6. Put $\mathbf{O} = (X = \{b, g, r, y\}$, 2^X, $F_{(\cdot)})$ in $C(\Omega)$ ($\equiv C(\{\omega_1, \omega_2, \omega_3\})$) and ρ_0^m is any mixed state $\in \mathcal{M}_{+1}^m(\Omega)$. Consider a measure ν on Ω, for example, $\nu(\{\omega_j\}) = 1$ $(j = 1, 2, 3)$. Define the W^*-observable $\overline{\mathbf{O}}$ in $L^\infty(\Omega, \nu)$ such that $\overline{\mathbf{O}} = \mathbf{O}$, and define the normal state $\overline{\rho}$ ($\in L_{+1}^1(\Omega, \nu)$) such that $\rho_0^m(B) = \int_B \overline{\rho}(\omega)\nu(d\omega)$ for all B ($\subseteq \Omega$). Then, we can identify $\mathbf{M}_{C(\Omega)}$($\times_{k=1}^2 \mathbf{O}$, $S_{[*]}(\rho_0^m)$) with $\overline{\mathbf{M}}_{L^\infty(\Omega,\nu)}$($\times_{k=1}^2 \overline{\mathbf{O}}$, $\overline{S}(\overline{\rho})$). Note that Q is equivalent to the crisp observable $\overline{\mathbf{Q}} \equiv (\mathbf{R}, \mathcal{B}, G^Q)$ in $L^\infty(\Omega, \nu)$ such that $G_\Xi^Q(\omega) = \chi_{\{\omega' \in \Omega : Q(\omega') \in \Xi\}}(\omega)$ for all $\Xi \in \mathcal{B}$ and all $\omega \in \Omega$. Define the map $h : X^2 \to \mathbf{R}$ such that:

$$h(x_1, x_2) = \frac{1}{2}\Big(W(x_1) + W(x_2)\Big) \qquad (\forall (x_1, x_2) \in X^2 \equiv \{b, g, r, y\}^2) \tag{11.9}$$

where $W(b) = 10$, $W(g) = 20$, $W(r) = 30$ and $W(y) = 10$. Consider the image observable $(\times_{k=1}^2 \overline{\mathbf{O}})_h \equiv (\mathbf{R}, \mathcal{B}, \widehat{F} = (\times_{k=1}^2 F)_{h^{-1}(\cdot)})$. Then, $\Delta\Big(\overline{\mathbf{M}}_{L^\infty(\Omega,\nu)}(\overline{\mathbf{Q}} \times (\times_{k=1}^2 \overline{\mathbf{O}})_h,\ \overline{S}(\overline{\rho}))\Big)$, the distance between $\overline{\mathbf{Q}}$ and $(\times_{k=1}^2 \overline{\mathbf{O}})_h$ concerning $\overline{\rho}$, is calculated as

$$\begin{aligned}
&\Delta\Big(\overline{\mathbf{M}}_{L^\infty(\Omega,\nu)}(\overline{\mathbf{Q}} \times (\times_{k=1}^2 \overline{\mathbf{O}})_h, \overline{S}(\overline{\rho}))\Big) = \Big[\iint_{\mathbf{R}^2} |\lambda_1 - \lambda_2|^2 \overline{\rho}((G^Q \times \widehat{F})(d\lambda_1 d\lambda_2))\Big]^{1/2} \\
&= \Big[\sum_{j=1}^{3} \sum_{(x_1,x_2) \in X^2} \overline{\rho}(\omega_j)|Q(\omega_j) - h(x_1, x_2)|^2 F_{\{x_1\}}(\omega_j) F_{\{x_2\}}(\omega_j)\Big]^{1/2} \\
&= \Big[22\overline{\rho}(\omega_1) + 38\overline{\rho}(\omega_2) + 38\overline{\rho}(\omega_3)\Big]^{1/2}.
\end{aligned} \tag{11.10}$$

Therefore, we see that (11.10) $\leq \sqrt{38} \approx 6.17$ for all $\overline{\rho} \in L_{+1}^1(\Omega, \nu)$. Now we can also answer the problem (♯) in Problem 11.6. That is, we see,

$$Q(\,*\,) = \tfrac{1}{2}(W(r) + W(b)) = (30 + 10)/2 = 20,$$

though it of course includes the error 6.17.

■

The map $h : X^n \to \mathbf{R}$, $(n = 2)$, in (11.9) may be chosen by the hint of "the law of large numbers". That is, if n is sufficiently large, the map $h : X^n \to \mathbf{R}$ (defined by $h(x_1, ..., x_n)$ $= \frac{1}{n}\sum_{k=1}^n W(x_k)$) has a proper property, i.e., $\lim_{n\to\infty} \Delta\Big(\overline{\mathbf{M}}_{L^\infty(\Omega,\nu)}(\overline{\mathbf{Q}} \times (\times_{k=1}^n \overline{\mathbf{O}})_h,\ \overline{S}(\overline{\rho}))\Big)$ $= 0$ for all $\overline{\rho} \in L_{+1}^1(\Omega, \nu)$. However, there are several ideas for the choice of h.

Definition 11.7. [Admissible]. *Let* $\overline{\mathbf{Q}} \equiv (\mathbf{R}, \mathcal{B}, G)$ *and* $\overline{\mathbf{O}} \equiv (X, \overline{\mathcal{F}}, F)$ *be a quantity and* W^**-observable in a* W^**-algebra* $\mathcal{N}$ *respectively. For each* $i = 1, 2$*, consider a measurable* *map* $h_i : X \to \mathbf{R}$, *and the image observable* $\overline{\mathbf{O}}_{h_i} \equiv (\mathbf{R}, \mathcal{B}, F(h_i^{-1}(\,\cdot\,))\,)$ *in* $\mathcal{N}$. *Also assume that* $\overline{\mathbf{Q}}$ *and* $\overline{\mathbf{O}}_{h_i}$ *commute.*

(i) *When it holds that*

$$\Delta(\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{Q}} \times \overline{\mathbf{O}}_{h_1}, \overline{S}(\overline{\rho}))\,) \le \Delta(\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{Q}} \times \overline{\mathbf{O}}_{h_2}, \overline{S}(\overline{\rho}))\,) \quad \forall \overline{\rho} \in \mathfrak{S}^n(\mathcal{N}_*), \tag{11.11}$$

we say that $\overline{\mathbf{O}}_{h_1}$ *is better* than $\overline{\mathbf{O}}_{h_2}$ as the approximation of $\overline{\mathbf{Q}}$.

(ii) *Also,* $\overline{\mathbf{O}}_{h_2}$ *is called admissible* as the approximation of $\overline{\mathbf{Q}}$, if there exists no h_1 that satisfies (11.11) and the following condition:

$$\Delta(\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{Q}} \times \overline{\mathbf{O}}_{h_1}, \overline{S}(\overline{\rho}_0))\,) < \Delta(\overline{\mathbf{M}}_{\mathcal{N}}(\overline{\mathbf{Q}} \times \overline{\mathbf{O}}_{h_2}, \overline{S}(\overline{\rho}_0))\,) \quad \text{for some } \overline{\rho}_0 \in \mathfrak{S}^n(\mathcal{N}_*). \tag{11.12}$$

■

As a well known result concerning "admissibility", we mention the following example.

Example 11.8. [Gaussian observable and admissibility]. Let $\overline{\mathbf{O}} \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, G^\sigma)$ be the Gaussian observable in $\mathcal{N} \equiv L^\infty(\mathbf{R}, d\mu)$, that is,

$$G^\sigma_\Xi(\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_\Xi e^{-\frac{(u-\mu)^2}{2\sigma^2}} du \qquad (\forall \mu, \in \mathbf{R},\ \forall \Xi \in \mathcal{B}_{\mathbf{R}}). \tag{11.13}$$

Consider the quantity $Q : \mathbf{R} \to \mathbf{R}$ such that $Q(\mu) = \mu$ $(\forall \mu \in \mathbf{R})$, which is identified with the observable $\overline{\mathbf{Q}} \equiv (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, F^Q_{(\cdot)})$ where $F^Q_\Xi(\mu) = \chi_\Xi(\mu)$. Consider the product observable $\times_{k=1}^n \overline{\mathbf{O}} \equiv (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, \times_{k=1}^n G^\sigma)$ in $L^\infty(\mathbf{R}, d\mu)$. Define the map $h : \mathbf{R}^n \to \mathbf{R}$ such that $\mathbf{R}^n \ni (\lambda_1, ..., \lambda_n) \overset{h}{\mapsto} \frac{\lambda_1 + \cdots + \lambda_n}{n} \in \mathbf{R}$. Then, it is well known (*cf.* [86]) that $(\times_{k=1}^n \overline{\mathbf{O}})_h$ is *admissible* as the approximation of $\overline{\mathbf{Q}}$.

■

11.3 Random observable

Recall the probabilistic measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}([\delta_{\omega_1}; p] \oplus [\delta_{\omega_2}; 1-p]))$ in Example 8.1 (8.8). Here, the symbol $[\delta_{\omega_1}; p] \oplus [\delta_{\omega_2}; 1-p]$ is called a "probabilistic state". The concept of "probabilistic state" urges us to propose the "random observable" as follows:

For simplicity, in this section we devote ourselves to the classical case (i.e., $C(\Omega)$ and $L^\infty(\Omega, \mu)$).

Let $\mathbf{O}_1 \equiv (X, \mathcal{F}, F_1)$, $\mathbf{O}_2 \equiv (X, \mathcal{F}, F_2)$, $\cdots$, $\mathbf{O}_N \equiv (X, \mathcal{F}, F_N)$ be observables in $C(\Omega)$. In a similar way in the procedures (P_1) and (P_2) of Example 8.1, define the "random observable" $\oplus_{n=1}^N[\mathbf{O}_n; p_n]$, where $\sum_{n=1}^N p_n = 1$ $(0 \le p_n \le 1$ $(n = 1, 2, ..., N))$. That is, we assume that:

- To take a measurement $\mathbf{M}_{C(\Omega)}(\oplus_{n=1}^N[\mathbf{O}_n; p_n],\ S_{[\delta_\omega]})$. $\Big($This measurement is called a *"random measurement"*.$\Big)$

$\Longleftrightarrow$

- To take one of $\{\mathbf{M}_{C(\Omega)}(\mathbf{O}_n, S_{[\delta_\omega]}) \mid n = 1, 2, ..., N\}$ according to the probabilistic rule $(p_1, p_2, ..., p_N)$. That is, to take the measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}_n, S_{[\delta_\omega]})$ with probability p_n.

Here, it should be noted that

- the statistical property of $\mathbf{M}_{C(\Omega)}(\oplus_{n=1}^N[\mathbf{O}_n; p_n],\ S_{[\delta_\omega]})$ is equal to that of $\mathbf{M}_{C(\Omega)}(\widehat{\mathbf{O}},\ S_{[\delta_\omega]})$, where $\widehat{\mathbf{O}} \equiv (X.\mathcal{F}, \widehat{F})$ is defined by $\widehat{F}(\Xi) = \sum_{n=1}^N p_n F_n(\Xi)$. That is, for each $\Xi(\in \mathcal{F})$ and ω $(\in \Omega)$,

$$
\begin{aligned}
&\text{"the probability that a measured value obtained by } \mathbf{M}_{C(\Omega)}(\oplus_{n=1}^N[\mathbf{O}_n; p_n],\ S_{[\delta_\omega]}) \\
&\quad \text{belongs to } \Xi\text{"} \\
=&\sum_{n=1}^N p_n [F(\Xi)](\omega) \qquad (11.14) \\
=&\text{"the probability that a measured value obtained by } \mathbf{M}_{C(\Omega)}(\widehat{\mathbf{O}},\ S_{[\delta_\omega]}) \text{ belongs} \\
&\quad \text{to } \Xi\text{ ",}
\end{aligned}
$$

which is easily seen by a similar argument such as stated in Example 8.1.

Again note that

(1) to take a random measurement $\mathbf{M}_{C(\Omega)}(\oplus_{n=1}^N[\mathbf{O}_n; p_n],\ S_{[\delta_\omega]})$ (11.15)

$\Longleftrightarrow$

to take a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}_n, S_{[\delta_\omega]})$ with probability p_n $(n = 1, 2, ..., N)$.

(2) to take a probabilistic measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O},\ S_{[*]}(\oplus_{n=1}^N [\delta_{\omega_n}; p_n]))$ (11.16)

$\Longleftrightarrow$

to take a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_n}]})$ with probability p_n $(n = 1, 2, ..., N)$.

In the case that $N = \infty$. it suffices to prepare a probability space $(\Lambda, \mathcal{F}(\Lambda), \nu)$. And, for each $\lambda(\in \Lambda)$, consider an observable $\mathbf{O}_\lambda$ $(\equiv (X, \mathcal{F}, F_\lambda))$ in $C(\Omega)$. Then, the random observable $\oplus_{n=1}^{N}[\mathbf{O}_n; p_n]$ is generalized as $\oint_\Lambda \mathbf{O}_\lambda \nu(d\lambda)$ $\Big(\equiv (X, \mathcal{F}, \oint_\Lambda F_\lambda \nu(d\lambda))\Big)$.

The following example is typical (though the description is due to the W^*-algebraic formulation).

Example 11.9. [Gaussian observable as a random observable]. For each $\lambda(\in \mathbf{R}(\equiv \Lambda))$, consider an observable $\mathbf{O}_\lambda$ $(\equiv (\mathbf{R}(\equiv X), \mathcal{B}_\mathbf{R}, E_\lambda))$ in $L^\infty(\mathbf{R}(\equiv \Omega), d\omega)$ such that

$$[F_\lambda(\Xi)](\omega) = \chi_\Xi(\omega - \lambda) \quad (\forall \Xi \in \mathcal{B}_\mathbf{R}(\subseteq 2^X), \forall \omega \in \mathbf{R}(\equiv \Omega), \forall \lambda \in \mathbf{R}(\equiv \Lambda)).$$

Define the probability space $(\mathbf{R}(\equiv \Lambda), \mathcal{B}_\mathbf{R}, \nu)$ such that:

$$\nu(S) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_S e^{-\frac{\lambda^2}{2\sigma^2}} d\lambda \qquad (\forall S \in \mathcal{B}_\mathbf{R}). \tag{11.17}$$

Thus, we have the random observable:

$$\oint_\Lambda \mathbf{O}_\lambda \nu(d\lambda) \Big(\equiv (\mathbf{R}(\equiv X), \mathcal{B}_\mathbf{R}(\equiv \mathcal{F}), \oint_\Lambda F_\lambda \nu(d\lambda))\Big) \tag{11.18}$$

which the probabilistic form of the Gaussian observable $(\mathbf{R}(\equiv X), \mathcal{B}_\mathbf{R}(\equiv \mathcal{F}), G^\sigma)$ in $L^\infty(\mathbf{R}(\equiv \Omega), d\omega)$ such that:

$$\begin{aligned}[G^\sigma(\Xi)](\omega) &= \int_\Lambda [F_\lambda(\Xi)](\omega)\nu(d\lambda) = \int_\Lambda \chi_\Xi(\omega - \lambda) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\lambda^2}{2\sigma^2}} d\lambda \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_\Xi e^{-\frac{(x-\omega)^2}{2\sigma^2}} dx \qquad (\forall \omega \in \mathbf{R}(\equiv \Omega)\ \forall \Xi \in \mathcal{B}_\mathbf{R}(\subseteq 2^X)),\end{aligned} \tag{11.19}$$

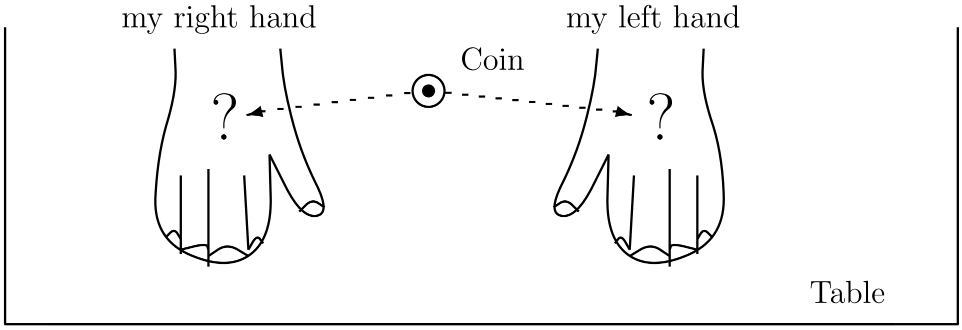
(Cf. Example 11.8.)

■

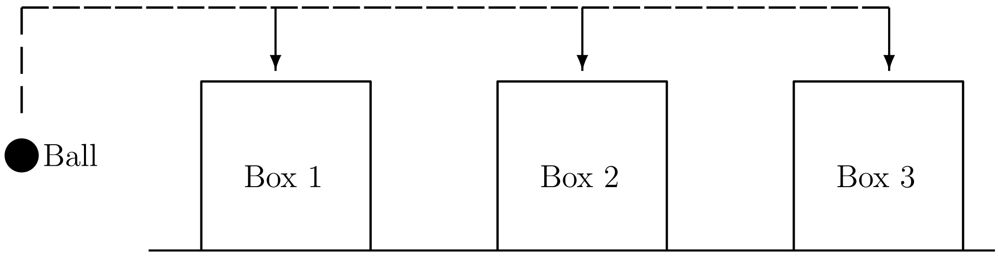
Although the following problem is easy, its measurement theoretical answer is quite important.

Problem 11.10. [Which hand is the coin under?]. The following problems (P_1) and (P_2) are essentially the same.

(P_1) A coin is, intentionally or unintentionally, put under my right hand or my left hand. Suppose that you do not know which hand the coin is under, and you choose one of my hands which you guess that the coin is under. Is it reasonable to believe that the probability that the ball is under the hand you choose is equal to 1/2. How do you think about it?



(P_2) There are three boxes (i.e., Box 1, Box 2 and Box 3) and a ball. A ball is, intentionally or unintentionally, put in one box (i.e., Box 1 or Box 2 or Box 3). Suppose that you do not know which box contains the ball, and you choose one of three boxes which you guess the ball is in. In this case, it is often believed that the probability that the ball is in Box 1 [resp. in Box 2; in Box 3] is 1/3 [resp. 1/3; 1/3]. How do you think about it?



•[The experimental answer to Problem (P_1)]. We can easily say "Yes", that is,

(A_1) the probability that the ball is under the hand you choose is equal to 1/2.

In fact, it can be easily tested experimentally. For example, it suffices to ask to 1000 persons "Which hand is the coin under?". About 500 persons will say "Right hand", and the other persons will say "Left hand". In either case, about 500 persons' guess is hit. Thus the above (A_1) is true. Although this (P_1) is the easiest problem throughout this book, what I want to say is the measurement theoretical answer mentioned in what follows.

•[The measurement theoretical answer to Problem (P_2)]. Since the two (P_1) and (P_2) are essentially the same, it suffices to answer Problem (P_2) from the measurement theoretical point of view. When the conclusion is said first, we can say that:

(A_2) the probability that the ball is in your chosen box is equal to 1/3.

In what follows we shall explain it. Put $\Omega = \{\omega_1, \omega_2, \omega_3\}$, where ω_1 [resp. ω_2, ω_3] means the state that the ball is in Box 1 [resp. Box 2; Box 3]. First we consider the case ω_1, that is, the ball is in Box 1.

[(i): The case ω_1, that is, the ball is in Box 1]. Define three observables $\mathbf{O}_1^e$ ($= (\{0,1\}, 2^{\{0,1\}}, F_1^e)$), $\mathbf{O}_2^e$ ($= (\{0,1\}, 2^{\{0,1\}}, F_2^e)$), $\mathbf{O}_2^e$ ($= (\{0,1\}, 2^{\{0,1\}}, F_3^e)$) such that:

$$
\begin{aligned}
&[F_1^e(\{0\})](\omega_1) = 0, \quad [F_1^e(\{0\})](\omega_2) = 1, \quad [F_1^e(\{0\})](\omega_3) = 1,\\
&[F_1^e(\{1\})](\omega_1) = 1, \quad [F_1^e(\{1\})](\omega_2) = 0, \quad [F_1^e(\{1\})](\omega_3) = 0,
\end{aligned} \tag{11.20}
$$

$$
\begin{aligned}
&[F_2^e(\{0\})](\omega_1) = 1, \quad [F_2^e(\{0\})](\omega_2) = 0, \quad [F_2^e(\{0\})](\omega_3) = 1,\\
&[F_2^e(\{1\})](\omega_1) = 0, \quad [F_2^e(\{1\})](\omega_2) = 1, \quad [F_2^e(\{1\})](\omega_3) = 0,
\end{aligned} \tag{11.21}
$$

$$
\begin{aligned}
&[F_3^e(\{0\})](\omega_1) = 1, \quad [F_3^e(\{0\})](\omega_2) = 1, \quad [F_3^e(\{0\})](\omega_3) = 0,\\
&[F_3^e(\{1\})](\omega_1) = 0, \quad [F_3^e(\{1\})](\omega_2) = 0, \quad [F_3^e(\{1\})](\omega_3) = 1.
\end{aligned} \tag{11.22}
$$

Note that we identify the following (S_1^1) and (S_2^1):

(S_1^1) We take a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}_1^e, S_{[\delta_{\omega_1}]})$. And we obtain a measured value 1. (Or, we obtain a measured value 0.) (11.23)

(S_2^1) We open Box 1. And we find the ball. (Or, we do not find the ball.) (11.24)

Similarly, we see the following identification:

(S_1^{23}) We take a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}_2^e, S_{[\delta_{\omega_1}]})$ [resp. $\mathbf{M}_{C(\Omega)}(\mathbf{O}_3^e, S_{[\delta_{\omega_1}]})$]. And we obtain a measured value 1. (Or, we obtain a measured value 0.)

(S_2^{23}) We open Box 2. [resp. Box 3.]. And we find the ball. (Or, we do not find the ball.)

Since "the state ω_1" = "the case that the ball is in Box 1", we can assume that

- the measured value obtained by $\mathbf{M}_{C(\Omega)}(\mathbf{O}_1^e, S_{[\delta_{\omega_1}]})$ [resp. $\mathbf{M}_{C(\Omega)}(\mathbf{O}_2^e, S_{[\delta_{\omega_1}]})$; $\mathbf{M}_{C(\Omega)}(\mathbf{O}_3^e, S_{[\delta_{\omega_1}]})$] is 1 [resp. 0; 0].

Since you have no information about the [∗], your choice is the same as the choice by a fair coin-tossing. That is, we assume that

$$
\text{"decision without having information"} \iff \text{"decision by a fair coin-tossing"}, \tag{11.25}
$$

which is the fundamental spirit of "the principle of equal probability" in the following section. Thus, it is reasonable to consider that

$$\text{the probability that Box 1 is opened} = \text{the probability that Box 2 is opened}$$
$$= \text{the probability that Box 3 is opened} = 1/3. \tag{11.26}$$

Therefore, we see that

(a) the probability that the measured value obtained by $\mathbf{M}_{C(\Omega)}(\oplus_{k=1}^{3}[\mathbf{O}_k^e; 1/3], S_{[\delta_{\omega_1}]})$ is 1 [resp. 0] is given by 1/3 [resp. 2/3].

[(ii): The case ω_2, that is, the ball is in Box 2]. Similarly we see that

(b) the probability that the measured value obtained by the "measurement" $\mathbf{M}_{C(\Omega)}(\oplus_{k=1}^{3}[\mathbf{O}_k^e; 1/3], S_{[\delta_{\omega_2}]})$ is 1 [resp. 0] is given by 1/3. [resp. 2/3].

[(iii): The case ω_3, that is, the ball is in Box 3]. Similarly we see that

(c) the probability that the measured value obtained by the "measurement" $\mathbf{M}_{C(\Omega)}(\oplus_{k=1}^{3}[\mathbf{O}_k^e; 1/3], S_{[\delta_{\omega_3}]})$ is 1 [resp. 0] is given by 1/3. [resp. 2/3].

[(iv): The case that we do not know which box contains the ball]. By the above (a), (b) and (c), we see that

- the probability that the measured value obtained by the "measurement" $\mathbf{M}_{C(\Omega)}(\oplus_{k=1}^{3}[\mathbf{O}_k^e; 1/3], S_{[*]})$ is 1 [resp. 0] is given by 1/3 [resp. 2/3].

Note that "measured value 1 is obtained" $\Leftrightarrow$ "open the box that contains the ball". Thus, we can believe that the probability that the ball is in Box 1 [resp. in Box 2; in Box 3] is 1/3 [resp. 1/3 ; 1/3].

[Remark]. Recall BMT (in §8.6). Then, the system in Problem (P_2) is clearly represented by $S_{[*]}((\nu_u))_{bw}$, *cf.* §8.6.1. Here, $\nu_u(\{\omega_k\}) = 1/3$ $(k = 1, 2, 3)$. However, in the above argument, we conclude that the "probability" that the ball is in Box 1 [resp. in Box 2; in Box 3] is 1/3 [resp. 1/3; 1/3]. Therefore, we have the following question:

- Is the system represented by $S_{[*]}(\nu_u)$ (as well as $S_{[*]}((\nu_u))_{bw}$)?

This will be discussed in the following section. ■

11.4 The principle of equal probability

Consider a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[*]})$, where Ω is finite, i.e., $\Omega \equiv \{\omega_1, \omega_2, ..., \omega_N\}$. There may be several definitions of "Having no information about the [∗]". As mentioned in §8.6, in this book we introduce three definitions of "Having no information about the [∗]" such as:

$$\begin{cases} \text{(a). iterative likelihood function method in §5.6,} \\ \text{(b). SMT}_{\text{PEP}} \text{ in SMT in this section and §11.4,} \\ \text{(c). BMT in §8.6.} \end{cases}$$

We want to change $S_{[*]}((\nu_u))_{bw}$ (belief weight) to $S_{[*]}(\nu_u)$ (statistical state). This will be done according to the spirit (11.25), that is,

$$\text{"decision without having information"} \iff \text{"decision by a fair coin-tossing"},$$

which assures that the principle of equal probability holds. This is the purpose of this section.

Let Ω be a finite set, i.e., $\Omega = \{\omega_1, \omega_2, ..., \omega_N\}$. A map $\phi : \Omega \to \Omega$ is said to be ergodic, if it is a bijection and if it holds that $\Omega = \{\phi^n(\omega) \mid n = 0, 1, ..., N-1\}$ for any ω $(\in \Omega)$. Also, a homomorphism $\Phi : C(\Omega) \to C(\Omega)$ is said to be ergodic, if there exists an ergodic bijection $\phi : \Omega \to \Omega$ such that

$$(\Phi f)(\omega) = f(\phi(\omega)) \quad (\forall f \in C(\Omega), \forall \omega \in \Omega). \tag{11.27}$$

Theorem 11.12. [The principle of equal probability (="PEP"), SMT$_{\text{PEP}}$ method]. *Consider a measurement* $\mathbf{M}_{C(\Omega)}(\mathbf{O} \equiv (X, \mathcal{F}, F), S_{[*]})$, *where* Ω *is finite, i.e.,* $\Omega \equiv \{\omega_1, \omega_2, ..., \omega_N\}$. *And consider the measurement* $\mathbf{M}_{C(\Omega)}(\oplus_{n=0}^{N-1} [\Phi^n \mathbf{O}; 1/N], S_{[*]})$ *(where* $\Phi : C(\Omega) \to C(\Omega)$ *is ergodic), which is called an unintentional random measurement.*[1] *Then we see*

$$\mathbf{M}_{C(\Omega)}(\oplus_{n=0}^{N-1} [\Phi^n \mathbf{O}; 1/N], S_{[*]}) \underset{\text{identification}}{\iff} \mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\oplus_{n=1}^{N} [\delta_{\omega_n}; 1/N])) \tag{11.28}$$

and

$$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\oplus_{n=1}^{N} [\delta_{\omega_n}; 1/N])) \underset{\text{statistical form}}{\overset{\text{probabilistic form}}{\rightleftarrows}} \mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_u)) \tag{11.29}$$

(=(8.9))

[1] Also, it is called a "completely random measurement", "coin-tossing measurement", "no information measurement".

where $\nu_u = \frac{1}{N}\sum_{n=1}^{N} \delta_{\omega_n}$. *That is, we can assert that:*

$$\mathbf{M}_{C(\Omega)}(\oplus_{n=0}^{N-1} [\Phi^n \mathbf{O}; 1/N], S_{[*]}) \underset{identification}{\Longleftrightarrow} \mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_u)). \tag{11.30}$$

Proof. Let $\omega \in \Omega$. Then we see that:

to take an unintentional random measurement $\mathbf{M}_{C(\Omega)}(\oplus_{n=0}^{N-1} [\Phi^n \mathbf{O}; 1/N], S_{[\delta_\omega]})$

$\Longleftrightarrow$

to take a measurement $\mathbf{M}_{C(\Omega)}(\Phi^n \mathbf{O}, S_{[\delta_\omega]})$
with probability $1/N$, $(n = 1, 2, ..., N)$

$\Longleftrightarrow$

to take a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\phi^n(\omega)}]})$ with probability $1/N$
$(n = 0, 1, 2, ..., N-1)$

$\Longleftrightarrow$ (Note that $\Omega = \{\phi^n(\omega) \mid n = 0, 1, ..., N-1\}$.)

to take a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_n}]})$ with probability $1/N$, $(n = 1, 2, ..., N)$

$\Longleftrightarrow$

to take a probabilistic measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\oplus_{n=0}^{N-1} [\delta_{\omega_n}; 1/N]))$

$\Longleftrightarrow$

to take a measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_u))$.

Thus we see that:

$$\mathbf{M}_{C(\Omega)}(\oplus_{n=0}^{N-1} [\Phi^n \mathbf{O}; 1/N], S_{[*]}) \underset{\text{identification}}{\Longleftrightarrow} \mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_u)). \tag{11.31}$$

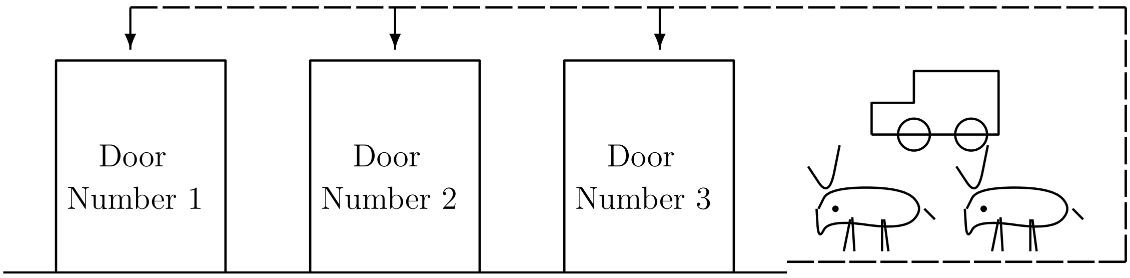
□

Problem 11.13. [Monty Hall problem, *cf.*[33]].
The Monty Hall problem is as follows (*cf.* Problem 5.12, Remark 5.13, Problem 8.8) :

(P) Suppose you are on a game show, and you are given the choice of three doors (i.e., "number 1", "number 2", "number 3"). Behind one door is a car, behind the others, goats.

(C) The host knows the fact that the probability that the car was set behind the k-th door (i.e., "number k") is given by p_k $(k = 1, 2, 3)$, for example, $p_1 = 3/7$, $p_2 = 1/7$, $p_3 = 3/7$. But you do not know this fact.

You pick a door (strictly speaking, you pick a door at random), say number 1, and the host, who knows what's behind the doors, opens another door, say "number 3", which has a goat. He says to you, "Do you want to pick door number 2?" Is it to your advantage to switch your choice of doors?



[*Answer*]. Put $\Omega = \{\omega_1, \omega_2, \omega_3\}$, where ω_1 [resp. ω_2, ω_3] means the state that the car is behind the door number 1 [resp. the door number 2, the door number 3]. Define the observable $\mathbf{O} \equiv (\{1,2,3\}, 2^{\{1,2,3\}}, F)$ in $C(\Omega)$ such that

$$
\begin{aligned}
&[F(\{1\})](\omega_1) = 0.0, \quad && [F(\{2\})](\omega_1) = 0.5, \quad && [F(\{3\})](\omega_1) = 0.5,^{2} \\
&[F(\{1\})](\omega_2) = 0.0, && [F(\{2\})](\omega_2) = 0.0, && [F(\{3\})](\omega_2) = 1.0, \\
&[F(\{1\})](\omega_3) = 0.0, && [F(\{2\})](\omega_3) = 1.0, && [F(\{3\})](\omega_3) = 0.0,
\end{aligned}
\tag{11.32}
$$

Thus, we have the unintentional random measurement $\mathbf{M}_{C(\Omega)}(\oplus_{n=0}^{2}[\Phi^n \mathbf{O}; 1/3], S_{[*]})$ (where $\Phi : C(\Omega) \to C(\Omega)$ is ergodic). Theorem 11.12 says that

$$
\mathbf{M}_{C(\Omega)}(\oplus_{n=0}^{2} [\Phi^n \mathbf{O}; 1/3], S_{[*]}) \Longleftrightarrow \mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_u)) \tag{11.33}
$$

where $\nu_u(\{\omega_1\}) = \nu_u(\{\omega_2\}) = \nu_u(\{\omega_3\}) = 1/3$. Thus, it suffices to consider the statistical measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\nu_u))$. Here, note that

- By the statistical measurement $\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_u))$, you obtain a measured value 3,

which corresponds to the fact that the host said "Door (number 3) has a goat". Then, the posttest state ν_{post} ($\in \mathcal{M}_{+1}^m(\Omega)$) is given by

$$
\nu_{\text{post}} = \frac{F(\{3\}) \times \nu_u}{\langle \nu_u, F(\{3\}) \rangle}. \tag{11.34}
$$

[2]Strictly speaking, $F(\{1\})(\omega_1) = 0.5$ and $F(\{2\})(\omega_1) = 0.5$ should be assumed in the problem (P)

That is,

$$\nu_{\text{post}}(\{\omega_1\}) = 1/3, \quad \nu_{\text{post}}(\{\omega_2\}) = 2/3, \quad \nu_{\text{post}}(\{\omega_3\}) = 0, \tag{11.35}$$

and thus, *you should pick door number 2.*

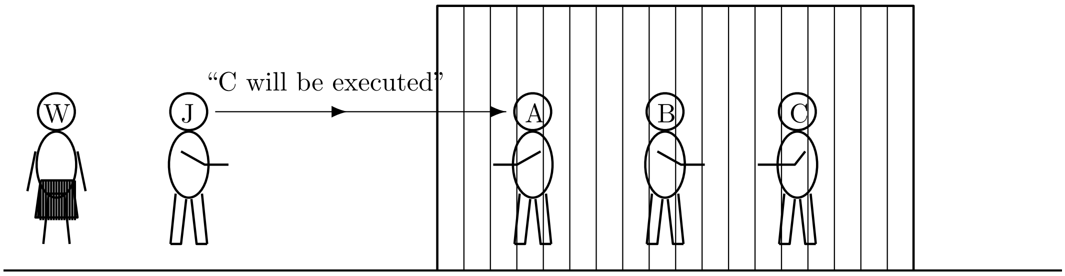
■

Remark 11.14. [Four answers to Monty Hall problem]. In this book four answers to the Monty Hall problem are presented in Problem 5.12, Remark 5.13, Problem 8.8, Problem 11.13. However, I believe that the Monty Hall problem in Problem 11.13 is the most natural.

■

Problem 11.15. [The problem of three prisoners, *cf.* Problem 8.10 and Remark 8.11]. Consider the following problem:

(P) Three men, A, B, and C were in jail. A knew that one of them was to be set free and the other two were to be executed. But he did not know who was the one to be spared. To the jailer who did know, A said, "Since two out of the three will bee executed, it is certain that either B or C will be, at least. You will give me no information about my own chances if you give me the name of one man, B or C, who is going to be executed." Accepting this argument after some thinking, the jailer said, "C will be executed." Thereupon A felt happier because now either he or C would go free, so his chance had increased from 1/3 to 1/2. This prisoner's happiness may or may not be reasonable. What do you think?



(Q) (Continued from the above (P)). There is a woman, who was proposed to by the three prisoners A, B and C. She listened to the conversation between A and the jailer. Thus, assume that she has the same information as A has. Then, we have the following problem:

(♯) Whose proposal should she accept?

[*Answer to* (P)]. Let Ω ($\equiv \{\omega_a, \omega_b, \omega_c\}$) and $\mathbf{O} \equiv (X \equiv \{x_A, x_B, x_C\}, 2^{\{x_A, x_B, x_C\}}, F)$ be as in Problem 8.10. Since A has no information, the unintentional random measurement $\mathbf{M}_{C(\Omega)}(\oplus_{k=0}^{2}[\Phi^k \mathbf{O}; 1/3], S_{[*]}(\nu_0))$ (where $\Phi : C(\Omega) \to C(\Omega)$ is ergodic) is considered. Theorem 11.12 asserts the following identification:

$$\mathbf{M}_{C(\Omega)}(\oplus_{k=0}^{2}[\Phi^k \mathbf{O}; 1/3], S_{[*]}) \underset{\text{identification}}{\Longleftrightarrow} \mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0)) \tag{11.36}$$

where ν_0 ($\in \mathcal{M}_{+1}^m(\Omega)$) is defined by

$$\nu_0(\{\omega_a\}) = 1/3, \quad \nu_0(\{\omega_b\}) = 1/3, \quad \nu_0(\{\omega_c\}) = 1/3. \tag{11.37}$$

Thus, we can assume that the (P) in the above is the same as the (P) in Problem 8.10. Therefore, we get that

$$\nu_{\text{post}}(\{\omega_a\}) = \frac{\frac{\nu_0(\{\omega_a\})}{2}}{\frac{\nu_0(\{\omega_a\})}{2} + \nu_0(\{\omega_b\})} = 1/3, \quad \nu_{\text{post}}(\{\omega_b\}) = \frac{\nu_0(\{\omega_b\})}{\frac{\nu_0(\{\omega_a\})}{2} + \nu_0(\{\omega_b\})} = 2/3,$$
$$\nu_{\text{post}}(\{\omega_c\}) = 0. \tag{11.38}$$

Therefore, we conclude that

- *the prisoner's happiness is not reasonable.* That is because $\nu_0(\{\omega_a\}) = 1/3 = \nu_{\text{post}}(\{\omega_a\})$.

[*Answer to* (Q)]. In the above (11.38), we see that

$$\nu_{\text{post}}(\{\omega_a\}) = 1/3, \quad \nu_{\text{post}}(\{\omega_b\}) = 2/3, \quad \nu_{\text{post}}(\{\omega_c\}) = 0. \tag{11.39}$$

Thus, we conclude that

- *she should choose the prisoner B.* That is because

$$\nu_{\text{post}}(\{\omega_c\}) = 0 < \nu_{\text{post}}(\{\omega_a\}) = 1/3 < \nu_{\text{post}}(\{\omega_b\}) = 2/3. \tag{11.40}$$

■

Chapter 12

Heisenberg's uncertainty relation

Quantum mechanics is surely one of the most successful theories in all science. In fact, most of the Nobel prizes of physics and chemistry are due to quantum mechanics. Also, as recent topics (particularly, related to measurements), we see quantum computer [80], quantum cryptography [91], quantum teleportation [10], etc. Although these are quite interesting and promising, in this chapter, we devote ourselves to Heisenberg's uncertainty relation, which is the most fundamental in quantum mechanics.

Heisenberg's uncertainty relation (*cf.* [31]).

(i) *The particle position q and momentum p can be measured "simultaneously", if the "errors" $\Delta(q)$ and $\Delta(p)$ in determining the particle position and momentum are permitted to be non-zero.*

(ii) *Moreover, for any $\epsilon > 0$, we can take the above "approximate simultaneous" measurement of the position q and momentum p such that $\Delta(q) < \epsilon$ (or $\Delta(p) < \epsilon$).*

(iii) *However, the following Heisenberg's uncertainty relation holds:*

$$\Delta(q) \cdot \Delta(p) \geq \frac{\hbar}{2}, \tag{12.1}$$

for all "approximate simultaneous" measurements of the particle position and momentum.

However, it should be noted that some ambiguous terms (i.e., "approximate simultaneous", "error") are included in the above statement, Thus, we believe that it is not a scientific statement but a "catch phrase" that was used to promote the paradigm shift from classical mechanics to quantum mechanics. Thus, in this last chapter[1] we try to describe this uncertainty relation precisely in terms of mathematics and further to derive it in the framework of the W^*-algebraic formulation of MT. For this, we first give the mathematical definitions of "$\Delta(q)$" (or "$\Delta(p)$") and "approximate simultaneous measurement", etc. in terms of MT.

---

[1]Every result mentioned in this chapter was published in [36], which was the oldest result in our study of "measurement theory". That is, our research of "measurement theory" starts from the paper [36]. On the other hand, the philosophical assertion mentioned in Chapter 1 is the latest result in our study. In this sense, the progress of our research is symbolically summarized as

$$\underset{\text{(in Chapter 12)}}{\text{"quantum" (physics)}} \longrightarrow \underset{\text{(in Chapters 2}\sim\text{11)}}{\text{"classical" (engineering)}} \longrightarrow \underset{\text{(in Chapter 1)}}{\text{"philosophical" (epistemology)}}$$

.

12.1 Introduction

Although the uncertainty relation (discovered by Heisenberg in 1927) has a long history, the various discussions about its interpretations are even now continued. Mainly there are two interpretations of uncertainty relations. One is the statistical interpretation. By repeating the exact (i.e. the "error" $\Delta(q) = 0$) measurements of the position q of particles with same states, we can obtain its average value $\bar{q}$ and its variance $var(q)$. Also, by repeating the exact (i.e. the "error" $\Delta(p) = 0$) measurements of the momentum p of the same particles, we can similarly get its average value $\bar{p}$ and its variance $var(p)$. From the simple mathematical deduction, we can obtain the following uncertainty relation:

$$[var(q)]^{\frac{1}{2}} \cdot [var(p)]^{\frac{1}{2}} \geq \frac{\hbar}{2}, \tag{12.2}$$

where $\hbar =$"Plank's constant"$/2\pi$. This is the statistical aspect of the uncertainty relation. The mathematical derivation of the uncertainty relation (12.2) was proposed by Kennard in 1927 (or more generally, Robertson 1n 1929). Cf. [54, 73]. Thus, this inequality (12.2) is called Robertson's uncertainty relation.

On the other hand, Heisenberg's uncertainty relation is rather individualistic. Most physicists will agree that the content of Heisenberg's uncertainty relation is roughly as stated in the following proposition (though it includes some ambiguous sentences as well as some ambiguous words, i.e. "approximate simultaneous" and "error").

Proposition 12.1. [Heisenberg's uncertainty relation, *cf.* [31]].[2]

(i) *The particle position q and momentum p can be measured "approximately" and "simultaneously", if the "errors" $\Delta(q)$ and $\Delta(p)$ in determining the particle position and momentum are permitted to be non-zero.*

(ii) *Moreover, for any $\epsilon > 0$, we can take the "approximate simultaneous" measurement of the position q and momentum p such that $\Delta(q) < \epsilon$ (or $\Delta(p) < \epsilon$).*

---

[2]It may be usually considered that the (12.2) is the mathematical representation of the (12.3). However, it is not true. In fact, in [84], J. von Neumann pointed out the difference between Robertson's uncertainty relation (= (12.2)) and Heisenberg's uncertainty relation (= (12.3)).

(iii) *However, the following Heisenberg's uncertainty relation holds:*

$$\Delta(q) \cdot \Delta(p) \geq \frac{\hbar}{2}, \tag{12.3}$$

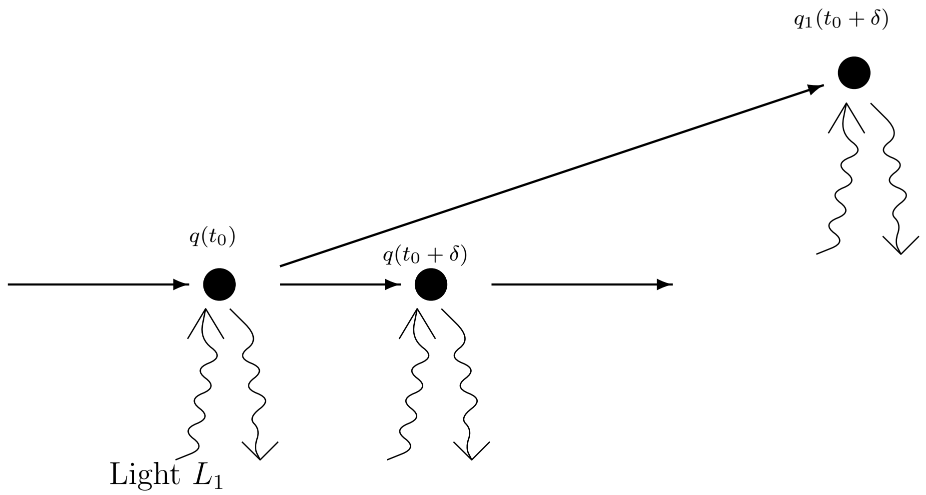
for all "approximate simultaneous" measurements of the particle position and momentum.

■

It should be noted that the above "proposition (= Heisenberg's assertion)" is ambiguous, that is, *it is not a scientific statement but a "catch phrase" that was used to promote the paradigm shift from classical mechanics to quantum mechanics.* In fact, the above "proposition" is powerless to solve the paradox (i.e., the paradox between EPR-experiment and Heisenberg's uncertainty relation), *cf.* §12.7.

Several authors have contributed to the problem to deduce Heisenberg's uncertainty relation. In [2] (Ali and Emach, 1974), [3] (Ali and Prugovečki, 1977), these were done by means of the concept of (generalized) observable which has been developed by E.B. Davies [17] (*cf.* Definition 9.3 for $B(V)$). Hence, a certain part of this problem has been already solved. In particular, the statements (i) and (ii) in the above Proposition 12.1 were deduced satisfactorily. However, concerning the statement (iii), there still seems to be some questions. The mathematical formulation and derivation of the Heisenberg's uncertainty relation (iii) (in the above Proposition 12.1) was proposed by M. Ozawa [67], S. Ishikawa [36] independently. We believe that this is the final version of Heisenberg's uncertainty relation concerning measurement errors. Thus, in this chapter we shall introduce this formulation and derivation of the above Proposition 12.1.

Remark 12.2. [(i): A classical understanding of Heisenberg's uncertainty relation]. Let us explain the classical understanding of Heisenberg's uncertainty relation (which is essentially equal to the thought experiment of γ-rays microscope (*cf.* [31])). In order to know the position $q(t_0)$ and momentum $p(t_0)$ of a particle A at time t_0, it suffices to measure the position $q(t_0)$ of a particle A at time t_0 (i.e., light L_1 is irradiated at the particle at time t_0), and continuously (i.e., after δ seconds), measure the position $q(t_0+\delta)$ at time $t_0+\delta$. That is because $(q(t_0), p(t_0)(\equiv \frac{mdq}{dt}(t_0)))$ is approximately calculated by $(q(t_0), \frac{m(q(t_0+\delta)-q(t_0))}{\delta})$.



[a]. However, if we want to know the exact position $q(t_0)$ (i.e., if we want $\Delta q \approx 0$), the wavelength λ of the light L_1 must be short (i.e., the energy $(= \frac{\text{“Plank constant”} \times \text{“lightspeed”}}{\lambda})$ of the light L_1 must be large), and therefore, the particle A is strongly perturbed. Thus, the position of the particle A at time $t_0+\delta$ will be changed to $q_1(t_0+\delta)$. Thus we observe that the momentum of the particle A at time t_0 is equal to $\frac{m(q_1(t_0+\delta)-q(t_0))}{\delta}$, which is away from $p(t_0)(\equiv \frac{mdq}{dt}(t_0) \approx \frac{m(q(t_0+\delta)-q(t_0))}{\delta})$ (i.e., Δp is large).

[b]. Also, if we want to know the exact momentum $p(t_0)$ (i.e., if we want $\Delta p \approx 0$), the wavelength λ of the light L_1 must be long, and therefore, the particle A is weakly perturbed. Although the position of the particle A at time $t_0+\delta$ will be changed to $q_1(t_0+\delta)$, it is almost the same as $q(t_0+\delta)$. Thus we observe that the momentum of the particle A at time t_0 is equal to $\frac{m(q_1(t_0+\delta)-q(t_0))}{\delta}$, which is near $p(t_0)(\equiv \frac{mdq}{dt}(t_0) \approx \frac{m(q(t_0+\delta)-q(t_0))}{\delta})$ (i.e., Δp is small) if δ is large. However it should be noted that Δq is large since the wavelength λ of the light L_1 is long.

[c]. Therefore, $\Delta p \approx 0$ and $\Delta q \approx 0$ are not compatible, that is, the inequality "$\Delta p \cdot \Delta q >$ constant" always holds. Although this explanation is, of course, rough, there is something thought-provoking in the above argument.

[(ii): EPR-experiment [22]]. Let A and B be particles with the same masses m. Consider the situation described in the following figure:

where "the velocity of A" $= -$"the velocity of B". The position q_A of the particle A can be measured, and moreover, the velocity of v_B of the particle B can be measured. Thus, we can conclude that the position and momentum of the particle A are respectively equal to q_A and $-mv_B$. Is this contradictory to Heisenberg's uncertainty relation? This question is significant though their (i.e. Einstein, Podolosky and Rosen) interest is concentrated on "the reality of physics".

■

12.2 Example due to Arthurs-Kelley

Here, we mainly consider the following identification:

$$\underset{(\|u\|_{L^2(\mathbf{R},dx)}=1,\ u\approx e^{i\theta}u)}{L^2(\mathbf{R},dx)\ni u} \quad \underset{\text{identification}}{\longleftrightarrow} \quad |u\rangle\langle u| \in Tr^p_{+1}(L^2(\mathbf{R},dx)).$$

We first introduce Robertson's uncertainty relation, which generally seems to be understood (or, misunderstood) as the mathematical representation of Heisenberg's uncertainty relation. By repeating the exact (i.e. the "error" $\Delta(q)=0$) measurements of the position q of particles with same states, we can obtain its average value $\bar{q}$ and its variance $var(q)$. Also, by repeating the exact (i.e. the "error" $\Delta(p)=0$) measurements of the momentum p of the same particles, we can similarly get its average value $\bar{p}$ and its variance $var(p)$. A simple calculation shows:

$$\bar{q} = \int_{\mathbf{R}} x\Big|u(x)\Big|^2 dx \quad \text{and} \quad \bar{p} = \int_{\mathbf{R}} \overline{u(x)}\Big[\frac{\hbar d}{idx}u(x)\Big]dx \quad \Big(= \int_{\mathbf{R}} p\Big|\tilde{u}(p)\Big|^2 dp\Big) \tag{12.4}$$

where $\tilde{u}$ is the Fourier transform of u, $\Big($that is, $\tilde{u}(p) = \sqrt{\frac{\hbar}{2\pi}}\int_{\mathbf{R}} u(x)e^{-i\hbar xp}dx\Big)$. And further, we see,

$$\begin{aligned} var(q) &= \int_{\mathbf{R}} |x-\bar{q}|^2\Big|u(x)\Big|^2 dx = \int_{\mathbf{R}} |x|^2\Big|u(x)\Big|^2 dx - \bar{q}^2, \\ var(p) &= \int_{\mathbf{R}} |p-\bar{p}|^2\Big|\tilde{u}(p)\Big|^2 dp = \int_{\mathbf{R}} |\frac{\hbar d}{idx}u(x)|^2 dx - \bar{p}^2. \end{aligned} \tag{12.5}$$

Immediately after Heisenberg's discovery (="Proposition 12.1", 1927), Kennard, by a simple calculation, showed the following uncertainty relation:

$$[var(q)]^{\frac{1}{2}} \cdot [var(p)]^{\frac{1}{2}} \geq \frac{\hbar}{2}. \tag{12.6}$$
(=(12.2))

(*cf.* Lemma 12.13 later). Of course, it is clear that there is a great gap between Heisenberg's uncertainty relation (12.3) and Kennard's uncertainty relation (12.6).

Next we shall introduce the nice idea by Arthurs-Kelly [7], that is, a certain approximate simultaneous measurement of the position q and the momentum p of a particle A in one dimensional real line $\mathbf{R}$, which has a state function $u(x)$ $(\in L^2(\mathbf{R}), \|u\|_{L^2(\mathbf{R})} = 1)$.

Note that the position observable $Q(\equiv x)$ and the momentum observable $P(\equiv \frac{\hbar d}{idx})$ do not commute, that is,

$$QP - PQ = i\hbar\Big(\neq 0\Big). \tag{12.7}$$

Therefore, any simultaneous measurement of the position observable x and the momentum observable $\frac{\hbar d}{idx}$ for a particle "A" can not be realized. However, Arthurs-Kelly's idea is excellent as follows: We first prepare another particle "B" with the state $u_0(y)$ such that:

$$\int_{\mathbf{R}} y\Big|u_0(y)\Big|^2 dy = \int_{\mathbf{R}} \overline{u_0(y)}\Big[\frac{\hbar d}{idy}u_0(y)\Big]dy = 0 \tag{12.8}$$

for example, $u_0(y) = \frac{1}{(\pi\hbar)^{1/4}}\exp(-\frac{y^2}{2\hbar})$. Further we regard these two particles "A" and "B" as a "particle C" in two dimensional Euclidean space $\mathbf{R}^2$ with the state $u(x)u_0(y)$ $(\in L^2(\mathbf{R}^2), \|u \cdot u_0\|_{L^2(\mathbf{R}^2)} = 1)$. Now consider the self-adjoint operators $(x - y)$ and $\frac{\hbar\partial}{i\partial x} + \frac{\hbar\partial}{i\partial y}$ in $L^2(\mathbf{R}^2)$, which commute, that is, it holds that:

$$(\frac{\hbar\partial}{i\partial x} + \frac{\hbar\partial}{i\partial y})(x-y) = (x-y)(\frac{\hbar\partial}{i\partial x} + \frac{\hbar\partial}{i\partial y}) \tag{12.9}$$

That is because we can easily calculate:

$$\begin{aligned}
&[(\frac{\hbar\partial}{i\partial x} + \frac{\hbar\partial}{i\partial y})(x-y)]f(x,y)\\
=&\frac{\hbar}{i}f(x,y) + x\frac{\hbar\partial}{i\partial x}f(x,y) - y\frac{\hbar\partial}{i\partial x}f(x,y) + x\frac{\hbar\partial}{i\partial y}f(x,y) - \frac{\hbar}{i}f(x,y) - y\frac{\hbar\partial}{i\partial y}f(x,y)\\
=&[(x-y)(\frac{\hbar\partial}{i\partial x} + \frac{\hbar\partial}{i\partial y})]f(x,y).
\end{aligned}$$

Thus the simultaneous measurement of observables $(x - y)$ and $\frac{\hbar\partial}{i\partial x} + \frac{\hbar\partial}{i\partial y}$ for a "particle C" (= "A" + "B") can be realized. Moreover, we can easily calculate these expectations as follows:

$$\iint_{\mathbf{R}^2} \overline{u(x)u_0(y)}\Big[(x-y)u(x)u_0(y)\Big]dxdy = \int_{\mathbf{R}} x\Big|u(x)\Big|^2 dx \tag{12.10}$$

and

$$\iint_{\mathbf{R}^2} \overline{u(x)u_0(y)}\Big[(\frac{\hbar\partial}{i\partial x} + \frac{\hbar\partial}{i\partial y})u(x)u_0(y)\Big]dxdy = \int_{\mathbf{R}} \overline{u(x)}\Big[\frac{\hbar d}{idx}u(x)\Big]dx. \tag{12.11}$$

By the reason that the equalities (12.10)$= \bar{q}$ and (12.11)$= \bar{p}$ hold, we may say that

($\sharp$) An "approximate simultaneous measurement" of the position observable $Q(\equiv x)$ and the momentum observable $P(\equiv \frac{\hbar d}{idx})$ can be realized.

Here, the variances $var_{asm}(q)$ and $var_{asm}(p)$ in the *approximate simultaneous measurement* of the position q and the momentum p of a particle "C" are given respectively by:

$$
\begin{aligned}
var_{asm}(q) &= \iint_{\mathbf{R}^2} \Big[(x-y)u(x)u_0(y)\Big]^2 dxdy - \Big|\iint_{\mathbf{R}^2} \overline{u(x)u_0(y)}\Big[(x-y)u(x)u_0(y)\Big]dxdy\Big|^2 \\
&= \int_{\mathbf{R}} \Big|xu(x)\Big|^2 dx - \Big|\int_{\mathbf{R}} x\Big|u(x)\Big|^2 dx\Big|^2 + \Big|\int_{\mathbf{R}} \Big|yu_0(y)\Big|^2 dy\Big|^2
\end{aligned} \tag{12.12}
$$

and

$$
var_{asm}(p) = \int_{\mathbf{R}} \Big|\frac{\hbar d}{idx}u(x)\Big|^2 dx - \Big|\int_{\mathbf{R}} \overline{u(x)}\Big[\frac{\hbar d}{idx}u(x)\Big]dx\Big|^2 + \Big|\int_{\mathbf{R}} \overline{u(y)}\Big[\frac{\hbar d}{idy}u_0(y)\Big]dy. \tag{12.13}
$$

Hence, we can get, by the arithmetic-geometric inequality and the well-known uncertainty relation (Robertson uncertainty relation, *cf.* Lemma 12.13 later), the following simultaneous uncertainty relation;

$$
\begin{aligned}
&[var_{asm}(q)]^{1/2} \cdot [var_{asm}(p)]^{1/2} \\
=&2\Big[\int_{\mathbf{R}} \Big|xu(x)\Big|^2 dx - \Big|\int_{\mathbf{R}} x\Big|u(x)\Big|^2 dx\Big|^2\Big]^{1/4} \times \Big[\Big|\int_{\mathbf{R}} \Big|yu_0(y)\Big|^2 dy\Big|\Big]^{1/4} \\
&\times \Big[\int_{\mathbf{R}} \Big|\frac{\hbar d}{idx}u(x)\Big|^2 dx - \Big|\int_{\mathbf{R}} \overline{u(x)}\Big[\frac{\hbar d}{idx}u(x)\Big]dx\Big|^2\Big]^{1/4} \times \Big[\Big|\int_{\mathbf{R}} \overline{u_0(y)}\Big[\frac{\hbar d}{idy}u_0(y)\Big]dy\Big|^2\Big]^{1/4} \\
\geq&\hbar.
\end{aligned} \tag{12.14}
$$

This is Arthurs-Kelly's idea. We believe that Arthurs-Kelly's discovery (12.14) is the first great step to the understanding of Heisenberg's uncertainty relation.

12.3 Approximate simultaneous measurement

Since our main purpose in this chapter is to describe Proposition 12.1 in terms of mathematics and further to prove it, we must clarify the ambiguous words (i.e., "approximate simultaneous", "error") in Proposition 12.1. For this, we prepare several definitions in this section.

According to the well-known spectral representation theorem (*cf.* [92]), there is a bijective correspondence of a crisp observable $(\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, E)$ in $B(H)$ to an n-tuple (A_1, ..., A_n) of commutative (unbounded) self-adjoint operators on H such that $A_i = \int_{\mathbf{R}^n} \lambda_i$ $E(d\lambda_1 ... d\lambda_n)$. That is,

$$\underset{\text{(commutative self-adjoint operators on } H)}{(A_1, A_2, ..., A_n)} \quad \underset{A_i = \int_{\mathbf{R}^n} \lambda_i E(d\lambda_1 ... d\lambda_n)}{\longleftrightarrow} \quad \underset{\text{(crisp observable in } B(V))}{(\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, E)} \tag{12.15}$$

In particular, we frequently identify a crisp observable $(\mathbf{R}, \mathcal{B}_{\mathbf{R}}, E)$ in $B(H)$ with a (unbounded) self-adjoint operator $A \Big(= \int_{\mathbf{R}} \lambda \ E(d\lambda) \Big)$ on H.

Note that Proclaim$^{W^*}$1 (9.9) $\Big($or, Axiom$^{W^*}$1 (9.11)$\Big)$ says as follows:

[♯] *Let* $\overline{\mathbf{O}} \equiv (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, F)$ *be an observable in* $B(H)$. *And consider a measurement* $\overline{\mathbf{M}}_{B(H)}(\overline{\mathbf{O}} \equiv (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, F), \overline{S}_{[\rho_u]})$, *where* $\rho_u = |u\rangle\langle u|$. *When we take a measurement* $\overline{\mathbf{M}}_{B(H)}(\mathbf{O} \equiv (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, F), \overline{S}_{[\rho_u]})$, *the probability that the measured value* $\lambda(\in \mathbf{R}^n)$ *belongs to a set* Ξ $(\in \mathcal{B}_{\mathbf{R}^n})$ *is given by*

$$\langle u, F(\Xi) u \rangle_H \ \Big(= tr[\rho_u F(\Xi)] \Big). \tag{12.16}$$

Therefore, the expectation $\mathbf{E}\Big[\overline{\mathbf{M}}_{B(H)}(\overline{\mathbf{O}}, \overline{S}_{[\rho_u]})\Big] \left(\equiv \Big(\mathbf{E}^{(i)} \Big[\overline{\mathbf{M}}_{B(H)}(\overline{\mathbf{O}}, \overline{S}_{[\rho_u]}) \Big] \Big)_{i=1}^n \right)$ of the measured value obtained by the measurement $\overline{\mathbf{M}}_{B(H)}(\overline{\mathbf{O}} \equiv (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, F), \overline{S}_{[\rho_u]})$ is given by

$$\begin{aligned} & \mathbf{E}^{(i)} \Big[\overline{\mathbf{M}}_{B(H)}(\overline{\mathbf{O}}, \overline{S}_{[\rho_u]}) \Big] \\ = & \int_{\mathbf{R}^n} \lambda_i \langle u, F(d\lambda_1 \cdots d\lambda_n) u \rangle_H \qquad i = 1, 2, ..., n, \end{aligned} \tag{12.17}$$

where $\rho_u = |u\rangle\langle u|$. Further, its variance $var\Big[\overline{\mathbf{M}}_{B(H)}(\overline{\mathbf{O}}, \ \overline{S}_{[\rho_u]})\Big] \left(\equiv \Big(var^{(i)} \Big[\overline{\mathbf{M}}_{B(H)}(\overline{\mathbf{O}}, \overline{S}_{[\rho_u]}) \Big] \Big)_{i=1}^n \right)$ is given by

$$\begin{aligned} & var^{(i)} \Big[\overline{\mathbf{M}}_{B(H)}(\overline{\mathbf{O}}, \overline{S}_{[\rho_u]}) \Big] \\ = & \int_{\mathbf{R}^n} \Big| \lambda_i - \mathbf{E}^{(i)} \Big[\overline{\mathbf{M}}_{B(H)}(\overline{\mathbf{O}}, \overline{S}_{[\rho_u]}) \Big] \Big|^2 \langle u, F(d\lambda_1 \cdots d\lambda_n) u \rangle_H & (12.18) \\ = & \int_{\mathbf{R}^n} |\lambda_i|^2 \langle u, F(d\lambda_1 \cdots d\lambda_n) u \rangle_H - \Big| \int_{\mathbf{R}^n} \lambda_i \langle u, F(d\lambda_1 \cdots d\lambda_n) u \rangle_H \Big|^2 & (12.19) \end{aligned}$$

$(i = 1, 2, ..., n)$.

We begin with the following definition.

Definition 12.3. *Let H be a Hilbert space with the inner product $\langle \cdot, \cdot \rangle_H$.*
(1). *A triplet $\widehat{\mathbf{O}}^{tnsr}_{H\otimes K} = (K, s, (X, \mathcal{F}, \widehat{F}))$ is called a "tensor observable" (or precisely, "tensor represented observable") in $B(H \otimes K)$, if it satisfies the following conditions* (i) and (ii):

(i) *K is a Hilbert space and s is an element in K such that $\|s\| = 1$,*

(ii) *$(X, \mathcal{F}, \widehat{F})$ is a crisp observable in $B(H \otimes K)$, where $H \otimes K$ is a tensor Hilbert space with the inner product $\langle \cdot, \cdot \rangle_{H\otimes K}$.*

(2). *Let $(X, \mathcal{F}, F)$ be any observable in $B(H)$. A tensor observable $\widehat{\mathbf{O}}^{tnsr}_{H\otimes K} = (K, s, (X, \mathcal{F}, \widehat{F}))$ is called a realization of the observable $(X, \mathcal{F}, F)$ in tensor Hilbert space $H \otimes K$, if it holds that*

$$\langle u \otimes s, \widehat{F}(\Xi)(u \otimes s) \rangle_{H\otimes K} = \langle u, F(\Xi)u \rangle_H \quad (\forall u \in H, \forall \Xi \in \mathcal{F}). \tag{12.20}$$

■

The following proposition is essential to our argument.

Proposition 12.4. [Holevo [34]]. *Let $(X, \mathcal{F}, F)$ be an observable in $B(H)$. Then, there exists a tensor observable $\widehat{\mathbf{O}}^{tnsr}_{H\otimes K} = (K, s, (X, \mathcal{F}, \widehat{F}))$ that is the realization of $(X, \mathcal{F}, F)$, that is, it holds that*

$$\langle u \otimes s, \widehat{F}(\Xi)(u \otimes s) \rangle_{H\otimes K} = \langle u, F(\Xi)u \rangle_H \quad (u \in H, \Xi \in \mathcal{F}). \tag{12.21}$$

Conversely any crisp observable $(X, \mathcal{F}, \widehat{F})$ in $B(H \otimes K)$ and any $s(\in K, \|s\|_K = 1)$ give rise to the unique observable $(X, \mathcal{F}, F)$ in $B(H)$ satisfying (12.21).

■

We shall use the following notations.

Notation 12.5. [Domain]. *Let $A \left(= \int_{\mathbf{R}} \lambda \, E_A(d\lambda)\right.$, the spectral representation of $\left. A \right)$ be a (unbounded) self-adjoint operator on H. Then, we define the* Dom(A), *the domain of A, by*

$$\mathrm{Dom}(A) := \{u \in H : \int_{\mathbf{R}} |\lambda|^2 \langle u, E_A(d\lambda)u \rangle < \infty\}.$$

Let $\overline{\mathbf{O}} \equiv (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, F)$ *and* $\widehat{\mathbf{O}}^{tnsr}_{H\otimes K} = (K, s, (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, \widehat{F}))$ *be an observable and a tensor observable in* $B(H)$ *and in* $B(H \otimes K)$ *respectively. Then, we define that*

$$[\overline{\mathbf{O}}]^{mar}_{(k)} := (\mathbf{R}, \mathcal{B}_{\mathbf{R}}, [F]^{mar}_{(k)}) \qquad \textit{(it will be called the } k\textit{th marginal observable of } \overline{\mathbf{O}}) ,$$

where

$$[F]^{mar}_{(k)}(\Xi) := F(\underbrace{\mathbf{R} \times \cdots \times \mathbf{R}}_{k-1 \text{ times}} \times \Xi \times \underbrace{\mathbf{R} \times \cdots \times \mathbf{R}}_{n-k \text{ times}}) \qquad (\forall \Xi \in \mathcal{B}_{\mathbf{R}})$$

Further, define that

$$\mathrm{Dom}([\overline{\mathbf{O}}]^{mar}_{(k)})\Big(\equiv \mathrm{Dom}([F]^{mar}_{(k)})\Big) := \{u \in H : \int_{\mathbf{R}^n} |\lambda_k|^2 \langle u, F(d\lambda_1 ... d\lambda_n) u\rangle < \infty\},$$

$$\mathrm{Dom}([\widehat{\mathbf{O}}]^{mar}_{(k)})\Big(\equiv \mathrm{Dom}([\widehat{F}]^{mar}_{(k)})\Big) := \{\hat{v} \in H \otimes K : \int_{\mathbf{R}^n} |\lambda_k|^2 \langle \hat{v}, \widehat{F}(d\lambda_1 ... d\lambda_n) \hat{v}\rangle_{H\otimes K} < \infty\},$$

$$\mathrm{Dom}_{\otimes s}([\widehat{\mathbf{O}}^{tnsr}_{H\otimes K}]^{mar}_{(k)})\Big(\equiv \mathrm{Dom}_{\otimes s}([\widehat{F}]^{mar}_{(k)})\Big)$$
$$:= \{u \in H : \int_{\mathbf{R}^n} |\lambda_k|^2 \langle u \otimes s, \widehat{F}(d\lambda_1 ... d\lambda_n)(u \otimes s)\rangle_{H\otimes K} < \infty\}, \qquad (12.22)$$

where $\mathrm{Dom}([\overline{\mathbf{O}}]^{mar}_{(k)})$ *(or* $\mathrm{Dom}([\widehat{\mathbf{O}}]^{mar}_{(k)})$*) is called the* k*-th domain of* $\overline{\mathbf{O}}$ *(or* $\widehat{\mathbf{O}}$*).*

■

Now we have the following main definition.

Definition 12.6. [Approximate simultaneous observable]. *Let* $A_1, ..., A_n$ *be (unbounded) self-adjoint operators in* H*. An observable* $\overline{\mathbf{O}}^{ASO}_{[A_l]^n_{l=1}} \equiv (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, F)$ *in* $B(H)$ *is called the approximate simultaneous observable* of $A_1, ..., A_n$, if it satisfies the following conditions

(i) *(domain condition) for each* i $(= 1, 2, ..., n)$, $\mathrm{Dom}([\overline{\mathbf{O}}^{ASO}_{[A_l]^n_{l=1}}]^{mar}_{(i)}) \cap \mathrm{Dom}(A_i)$ *is dense in* H

(ii) *(unbias condition) for each* i $(= 1, 2, ..., n)$,

$$\langle u, A_i u\rangle = \int_{\mathbf{R}} \lambda \langle u, [F]^{mar}_{(i)}(d\lambda) u\rangle, \qquad (u \in \mathrm{Dom}([\overline{\mathbf{O}}^{ASO}_{[A_l]^n_{l=1}}]^{mar}_{(i)}) \cap \mathrm{Dom}(A_i)). \quad (12.23)$$

■

Remark 12.7. [1]. As seen later (*cf.* Lemma 12.14(iii)), it holds that $\mathrm{Dom}([\overline{\mathbf{O}}^{ASO}_{[A_l]^n_{l=1}}]^{mar}_{(i)}) \subseteq \mathrm{Dom}(A_i)$ holds. Thus, $\mathrm{Dom}([\overline{\mathbf{O}}^{ASO}_{[A_l]^n_{l=1}}]^{mar}_{(i)}) \cap \mathrm{Dom}(A_i) = \mathrm{Dom}([\overline{\mathbf{O}}^{ASO}_{[A_l]^n_{l=1}}]^{mar}_{(i)})$

[2]. There is a very reason to assume the following condition (iii) or (iv) instead of the above (i). $\Big($(iii) and (iv) are stronger than (i), more precisely, (iv) $\Longrightarrow$ (iii) $\Longrightarrow$ (i).$\Big)$

(iii) (*self-adjointness*) *for each* i $(= 1, 2, ..., n)$, A_i *is essentially self-adjoint on* $\mathrm{Dom}([\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}]^{mar}_{(i)}) \cap \mathrm{Dom}(A_i)$,

or

(iv) (*commutative condition*) *for each* i $(= 1, 2, ..., n)$, $A_i \Big(= \int_{\mathbf{R}} \lambda \, E_i(d\lambda)\Big)$ *and* $[\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}]^{mar}_{(i)}$ *commute.*

Although each of (i), (iii) and (iv) has merit and demerit respectively, the physical meaning of the (iv) is the clearest. (Continued on Remark 12.12.)
[3]. Also, see the condition (i) in Example 11.5. This condition is equivalent to

- the formula (12.23) holds on a dense set $\cap_{i=1}^n \Big(\mathrm{Dom}([\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}]^{mar}_{(i)} \cap \mathrm{Dom}(A_i)\Big)$.

■

Definition 12.8. [Approximate simultaneous tensor observable]. *Let* $A_1, ..., A_n$ *be (unbounded) self-adjoint operators in* H*. A tensor observable* $\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n} = (K, s, (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, \widehat{F}))$ *is called an approximate simultaneous tensor observable of* $A_1, ..., A_n$*, if* $\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n} = (K, s, (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, \widehat{F}))$ *satisfies the following conditions:*

(i) (*domain condition*) *for each* i $(= 1, 2, ..., n)$, $\mathrm{Dom}_{\otimes s}([\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}]^{mar}_{(i)}) \cap \mathrm{Dom}(A_i)$ *is dense in* H

(ii) (*unbias condition*) *for each* i $(= 1, 2, ..., n)$,

$$\langle u, A_i u\rangle = \int_{\mathbf{R}^n} \lambda_i \langle u \otimes s, \widehat{F}(d\lambda_1 \cdots d\lambda_n)(u \otimes s)\rangle \qquad (12.24)$$
$$(u \in \mathrm{Dom}_{\otimes s}([\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}]^{mar}_{(i)}) \cap \mathrm{Dom}(A_i),).$$

■

The relation between $\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}$ and $\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}$ is characterized by the following proposition.
Proposition 12.9. *Let* $A_1, ..., A_n$ *be (unbounded) self-adjoint operators in* H.

(i) *Let* $\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n} \equiv (K, s, (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, \widehat{F})$ *be an approximate simultaneous tensor observable of* $A_1, ..., A_n$ *in* H*. Then, there exists an approximate simultaneous observable* $\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n} \equiv (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, F)$ *such as* $\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}$ *is a realization of* $\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}$.

(ii) *Let* $\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n} \equiv (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, F)$ *be an approximate simultaneous observable of* $A_1, ..., A_n$ *in* H*. Then, there exists a approximate simultaneous tensor observable* $\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n} \equiv (K, s, (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, \widehat{F})$ *such as it is a realization of* $\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}$.

(iii) *Let* $\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n} \equiv (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, F)$ *be an approximate simultaneous observable of* $A_1, ..., A_n$ *in* H*. Let* $\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n} \equiv (K, s, (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, \widehat{F})$ *be an approximate simultaneous tensor observable of* $A_1, ..., A_n$ *in* H*. And assume that* $\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}$ *is a realization of* $\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}$*. Then, for each* i $(= 1, 2, ..., n)$,

$$\mathrm{Dom}([\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}]^{mar}_{(i)}) = \mathrm{Dom}_{\otimes s}([\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}]^{mar}_{(i)}) \subseteq \mathrm{Dom}(A_i). \tag{12.25}$$

Proof. The statement (i) is trivial. Also the statement (ii) and the equality "=" in (12.25) immediately follow from Proposition 12.4. Also, the inclusion "⊆" in (12.25) is proved in Lemma 12.14(iii) later. □

Definition 12.10. [Uncertainty] *Let* $A_1, ..., A_n$ *be (unbounded) self-adjoint operators on a Hilbert space* H*.*

[I]. *Let* $\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n} = (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, F)$) *be an approximate simultaneous observable of* $A_1, \ldots, A_n$.

(i). Then, the uncertainty $\Big(\Delta_{\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}}(A_i, u)\Big)_{i=1}^n$ *of* $\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}$ *for a state* u $(\|u\|_H = 1)$ *is defined by*

$$\Delta_{\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}}(A_i, u) = \int_{\mathbf{R}^n} \lambda_i^2 \langle u, F(d\lambda_1 \cdots d\lambda_n)u\rangle - \int_{\mathbf{R}} \lambda^2 \langle u, A_i(d\lambda)u\rangle \qquad (u \in H \text{ such that } \|u\| = 1\ ), \tag{12.26}$$

where (12.26) *should be interpreted that* $\Delta_{\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}}(A_i, u) = \infty$ *for* $u \notin \mathrm{Dom}([F]^{mar}_{(i)})$ *(cf.* $\mathrm{Dom}([F]^{mar}_{(i)}) \subseteq \mathrm{Dom}(A_i)$ *in (12.25)).* $\Big($*"*$\Delta_{\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}}(A_i, u) \geq 0$*" will be shown in Theorem 12.15 later.*$\Big)$

(ii). Also the i*-th variance* $var_{(i)}[\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}, u]$ *is defined by*

$$var_{(i)}[\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}, u] = \int_{\mathbf{R}^n} |\lambda_i - \langle u, A_i u\rangle|^2 \langle u, F(d\lambda_1 \cdots d\lambda_n)u\rangle_H \tag{12.27}$$

$$(i = 1, 2, ..., n),$$

[II] Let $\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n} = (K, s, (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, \widehat{F}))$ be an approximate simultaneous tensor observable of $A_1, \ldots, A_n$.

(i). Then, the uncertainty $\Big(\Delta_{\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}}(A_i, u)\Big)_{i=1}^n$ of $\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}$ for a state u ($\|u\|_H = 1$) is defined by

$$\Delta_{\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}}(A_i, u) = \int_{\mathbf{R}^n} \lambda_i^2 \langle u \otimes s, \widehat{F}(d\lambda_1 \cdots d\lambda_n)(u \otimes s)\rangle - \int_{\mathbf{R}} \lambda^2 \langle u, A_i(d\lambda)u\rangle \tag{12.28}$$
$$(u \in H \text{ such that } \|u\| = 1 \ ),$$

where (12.28) *should be interpreted that* $\Delta_{\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}}(A_i, u) = \infty$ *for* $u \notin \mathrm{Dom}_{\otimes s}([\widehat{F}]^{mar}_{(i)})$ *(cf.* $\mathrm{Dom}_{\otimes s}([\widehat{F}]^{mar}_{(i)}) \subseteq \mathrm{Dom}(A_i)$ *in (12.25)).* $\Big($*"*$\Delta_{\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}}(A_i, u) \geq 0$*" will be shown in Theorem 12.15 later.*$\Big)$

(ii). *Also the* i*-th variance* $var_{(i)}[\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}, u]$ *is defined by*

$$var_{(i)}[\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}, u] = \int_{\mathbf{R}^n} |\lambda_i - \langle u, A_i u\rangle|^2 \langle u \otimes s, \widehat{F}(d\lambda_1 \cdots d\lambda_n)(u \otimes s)\rangle_{H \otimes K} \tag{12.29}$$
$$(i = 1, 2, ..., n).$$

■

Proposition 12.11. *Let* $A_1, ..., A_n$ *be (unbounded) self-adjoint operators on a Hilbert space* H*. Assume that* $\widehat{\mathbf{O}}^{ASTO}_{[A_i]_{i=1}^n} = (K, s, (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, \widehat{F}))$ *is a realization of* $\overline{\mathbf{O}}^{ASO}_{[A_i]_{i=1}^n} = (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, F)$*. Let* $u \in H$ ($\|u\|_H = 1$)*. Then it holds that*

$$\Delta_{\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}}(A_i, u) = \Delta_{\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}}(A_i, u) \tag{12.30}$$

and

$$var_{(i)}[\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}, u] = var_{(i)}[\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}, u]. \tag{12.31}$$

Proof. This immediately follows from Definition 12.10. □

Remark 12.12. [Continued from Remark 12.7]. Again note that, if the commutative condition (iv) in Remark 12.7 is assumed in the Definition 12.10, we can define $\Delta\Big(\overline{\mathbf{M}}_{B(H)}(A_i \times [\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}]^{mar}_{(i)}, \overline{S}(\rho_u))\Big)$, the distance between A_i and $[\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}]^{mar}_{(i)}$, *cf.* Definition 11.1. And further we see that

$$\underset{(\text{"error" defined in Definition 11.1})}{\Delta\Big(\overline{\mathbf{M}}_{B(H)}(A_i \times [\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}]^{mar}_{(i)}, \overline{S}(\rho_u))\Big)} = \underset{(\text{"uncertainty" defined in Definition 12.10})}{\Delta_{\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}}(A_i, u)} \tag{12.32}$$

Thus, in this case, the physical meaning of "uncertainty" is clear.

■

12.4 Lemmas

In this section, we shall prepare some Lemmas.

Lemma 12.13. [Robertson's uncertainty relation]. *Let* A_1 *and* A_2 *be any symmetric operators on a Hilbert space* H*. Then, it holds that*

$$\Big[\|A_1u\|^2 - |\langle u, A_1u\rangle|^2\Big]^{1/2} \cdot \Big[\|A_2u\|^2 - |\langle u, A_2u\rangle|^2\Big]^{1/2} \geq \frac{1}{2}\Big|\Big\langle A_1u, A_2u\Big\rangle - \Big\langle A_2u, A_1u\Big\rangle\Big| \tag{12.33}$$

for all $u \in \mathrm{Dom}(A_1) \cap \mathrm{Dom}(A_2)$.

Proof. Using Schwartz inequality, we see

$$\begin{aligned} &|\langle A_1u, A_2u\rangle - \langle A_2u, A_1u\rangle| \\ =&\Big|\Big\langle A_1u - \langle u, A_1u\rangle u, A_2u - \langle u, A_2u\rangle u\Big\rangle - \Big\langle A_2u - \langle u, A_2u\rangle u, A_1u - \langle u, A_1u\rangle u\Big\rangle\Big| \\ \leq& 2\Big[\|A_1u\|^2 - |\langle u, A_1u\rangle|^2\Big]^{1/2} \cdot \Big[\|A_2u\|^2 - |\langle u, A_2u\rangle|^2\Big]^{1/2}. \end{aligned} \tag{12.34}$$

□

Lemma 12.14. *Let* $A_1, \cdots, A_n$ *be any (unbounded) self-adjoint operators in a Hilbert space* H*. Let* $(K, s, (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, \widehat{F})$ *be an approximate simultaneous tensor observable for* $A_1, \cdots, A_n$*. Put* $\widehat{A}_k = \int_{\mathbf{R}^n} \lambda_k \widehat{F}(d\lambda_1 d\lambda_2 \cdots d\lambda_n)$ $\Big(\equiv \int_{\mathbf{R}} \lambda[\widehat{F}] = (i)^{mar}(d\lambda)\Big)$ $(k = 1, 2, ..., n)$*. Then, the following equalities* (i) $\sim$ (iii) *hold*

(i)

$$\langle v, A_k u\rangle = \langle v \otimes s, \widehat{A}_i(u \otimes s)\rangle = \int_{\mathbf{R}^2} \lambda_k \langle v \otimes s, \widehat{F}(d\lambda_1 d\lambda_2 \cdots d\lambda_n)(u \otimes s)\rangle \tag{12.35}$$

for all $u \in \mathrm{Dom}_{\otimes s}(\widehat{A}_k)$ *and all* $v \in H$ $(k = 1, 2, ..., n)$,

(ii)

$$\begin{aligned} &\int_{\mathbf{R}^n} \lambda_i \lambda_j \langle u \otimes s, \widehat{F}(d\lambda_1 d\lambda_2 \cdots d\lambda_n)(u \otimes s)\rangle \\ &= \langle \widehat{A}_i(u \otimes s), \widehat{A}_j(u \otimes s)\rangle \end{aligned}$$

$$= \langle A_i u, A_j u \rangle + \langle (\widehat{A}_i - A_i \otimes I)(u \otimes s), (\widehat{A}_j - A_j \otimes I)(u \otimes s) \rangle \tag{12.36}$$

for all $i \neq j$ *and all* $u \in \mathrm{Dom}_{\otimes s}(\widehat{A}_i) \cap \mathrm{Dom}_{\otimes s}(\widehat{A}_j)$,

(iii)

$$\begin{aligned} &\int_{\mathbf{R}^2} |\lambda_k|^2 \langle u \otimes s, \widehat{F}(d\lambda_1 d\lambda_2 \cdots d\lambda_n)(u \otimes s) \rangle \\ &= \|\widehat{A}_k(u \otimes s)\|^2 = \|A_k u\|^2 + \|(\widehat{A}_k - A_k \otimes I)(u \otimes s)\|^2 \geq \|A_k u\|^2 \end{aligned} \tag{12.37}$$

for all $u \in \mathrm{Dom}_s(\widehat{A}_k)$ $(k = 1, 2, ..., n)$. *Thus, it holds that, for each* i $(= 1, 2, ..., n)$,

$$\mathrm{Dom}([\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}]^{mar}_{(i)}) = \mathrm{Dom}_{\otimes s}([\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}]^{mar}_{(i)}) \subseteq \mathrm{Dom}(A_i). \tag{12.38}$$

Proof. First we prove (i). Fix $k \in \{1, 2\}$. We can see that, for any $v, u \in \mathrm{Dom}_{\otimes s}(\widehat{A}_k)$.

$$\begin{aligned} &\langle v, A_k u \rangle \\ &= \frac{1}{4}\{\langle (v+u), A_k(v+u) \rangle - \langle (v-u), A_k(v-u) \rangle \\ &\quad - i\langle (v+iu), A_k(v+iu) \rangle + i\langle (v-iu), A_k(v-iu) \rangle\} \\ &= \frac{1}{4}\{\langle (v+u) \otimes s, \widehat{A}_k((v+u) \otimes s) \rangle - \langle (v-u) \otimes s, \widehat{A}_k((v-u) \otimes s) \rangle \\ &\quad - i\langle (v+iu) \otimes s, \widehat{A}_k((v+iu) \otimes s) \rangle + i\langle (v-iu) \otimes s, \widehat{A}_k((v-iu) \otimes s) \rangle\} \\ &= \langle v \otimes s, \widehat{A}_k(u \otimes s) \rangle \\ &= \langle v \otimes s, \int_{\mathbf{R}^n} \lambda_k \widehat{F}_k(d\lambda_1 d\lambda_2 \cdots d\lambda_n)(u \otimes s) \rangle = \int_{\mathbf{R}^n} \lambda_k \langle v \otimes s, \widehat{F}_k(d\lambda_1 d\lambda_2 \cdots d\lambda_n)(u \otimes s) \rangle. \end{aligned} \tag{12.39}$$

Since $\mathrm{Dom}_{\otimes s}(\widehat{A}_k)$ is dense in H, we see that

$$\langle v, A_k u \rangle = \langle v \otimes s, \widehat{A}_k(u \otimes s) \rangle = \int_{\mathbf{R}^n} \lambda_k \langle v \otimes s, \widehat{A}_k(d\lambda_1 d\lambda_2 \cdots d\lambda_n)(u \otimes s) \rangle \tag{12.40}$$

for all $u \in \mathrm{Dom}_{\otimes s}(\widehat{A}_k)$ and all $v \in H$. This completes the proof of (i).

Next, we prove (ii). Without loss of generality, we put $i = 1$ and $j = 2$. Let u be any element in $\mathrm{Dom}_{\otimes s}(\widehat{A}_1) \cap \mathrm{Dom}_{\otimes s}(\widehat{A}_2)$. Then, we see, by the above (i), that

$$\begin{aligned} &\int_{\mathbf{R}^n} \lambda_1 \lambda_2 \langle u \otimes s, \widehat{F}(d\lambda_1 d\lambda_2 \cdots d\lambda_n)(u \otimes s) \rangle \\ &= \langle \int_{\mathbf{R}^n} \lambda_1 \widehat{F}(d\lambda_1 d\lambda_2 \cdots d\lambda_n)(u \otimes s), \int_{\mathbf{R}^n} \lambda_2 \widehat{F}(d\lambda_1 d\lambda_2 \cdots d\lambda_n)(u \otimes s) \rangle \end{aligned}$$

$$
\begin{aligned}
&= \langle \widehat{A}_1(u\otimes s), \widehat{A}_2(u\otimes s)\rangle \\
&= \langle (\widehat{A}_1 - A_1\otimes I)(u\otimes s) + (A_1 u\otimes s), (\widehat{A}_2 - A_2\otimes I)(u\otimes s) + (A_2 u\otimes s)\rangle \\
&= \langle (\widehat{A}_1 - A_1\otimes I)(u\otimes s), (\widehat{A}_2 - A_2\otimes I)(u\otimes s)\rangle \\
&\quad + \langle (\widehat{A}_1 - A_1\otimes I)(u\otimes s), A_2 u\otimes s\rangle \\
&\quad + \langle A_1 u\otimes s, (\widehat{A}_2 - A_2\otimes I)(u\otimes s)\rangle + \langle A_1 u\otimes s, A_2 u\otimes s\rangle \\
&= \langle (\widehat{A}_1 - A_1\otimes I)(u\otimes s), (\widehat{A}_2 - A_2\otimes I)(u\otimes s)\rangle \\
&\quad + \langle \widehat{A}_1(u\otimes s), A_2 u\otimes s\rangle - \langle A_1 u, A_2 u\rangle \\
&\quad + \langle A_1 u\otimes s, \widehat{A}_2(u\otimes s)\rangle - \langle A_1 u, A_2 u\rangle + \langle A_1 u, A_u\rangle \\
&= \langle (\widehat{A}_1 - A_1\otimes I)(u\otimes s), (\widehat{A}_2 - A_2\otimes I)(u\otimes s)\rangle - \langle A_1 u, A_2 u\rangle \\
&\quad + \int_{\mathbf{R}^n} \lambda_2 \langle A_1 u\otimes s, \widehat{F}(d\lambda_1 d\lambda_2)(u\otimes s)\rangle + \int_{\mathbf{R}^n} \lambda_1 \langle \widehat{F}(d\lambda_1 d\lambda_2)(u\otimes s), A_2 u\otimes s\rangle \\
&= \langle A_1 u, A_2 u\rangle + \langle (\widehat{A}_1 - A_1\otimes I)(u\otimes s), (\widehat{A}_2 - A_2\otimes I)(u\otimes s)\rangle. \qquad (12.41)
\end{aligned}
$$

Hence, the proof of (ii) is completed. Also, the proof of (12.37) is carried out just in a similar way. Lastly, we can easily see that (12.37) implies (12.38) since we see that $\mathrm{Dom}([\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}]^{mar}_{(i)}) = \mathrm{Dom}_{\otimes s}([\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}]^{mar}_{(i)})$ in (12.25). □

Now we have the following theorem, which is one of our main results.

Theorem 12.15. *Let* $\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n} = (K, s, (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, \widehat{F})\ )$ *be a realization of an approximate simultaneous tensor observable* $\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n} = (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, \widehat{F}\ )$ *of* $A_1, \ldots, A_n$*. Put* $\widehat{A}_i = \int_{\mathbf{R}} \lambda [\widehat{F}]^{mar}_{(i)}(d\lambda)$*. Then, we see that*

$$
\begin{aligned}
\Delta_{\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}}(A_i, u) =& \Delta_{\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}}(A_i, u) = \int_{\mathbf{R}} \lambda^2 \langle u, [F]^{mar}_{(i)}(d\lambda) u\rangle - \int_{\mathbf{R}} \lambda^2 \langle u, A_i(d\lambda) u\rangle \\
=& \|\widehat{A}_i(u\otimes s)\|^2 - \|A_i u\|^2 \qquad (12.42) \\
=& \|(\widehat{A}_i - A_i\otimes I)(u\otimes s)\|^2 \qquad (\forall u \in H \text{ such that } \|u\| = 1 \qquad (12.43)
\end{aligned}
$$

Proof. It immediately follows from Lemma 12.14. □

12.5 Existence theorem

Now we shall mention the following theorem, which assures the existence of an approximate simultaneous tensor observable of arbitrary observables $A_1, \ldots, A_n$. For two observables A_1 and A_2, the similar theorem was proved by P. Busch, *et al.* [15, 14].

Theorem 12.16. [Cf.[36]] *Let $A_1, ..., A_n$ be (unbounded) self-adjoint operators on a Hilbert space H. Let $a_1, ..., a_n$ be any positive numbers such that $\sum_{i=1}^n (1+{a_i}^2)^{-1} = 1$. Then, we see,*

(i) there exists an approximate simultaneous tensor observable $\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n} \equiv (K, s, (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, \widehat{F}))$ of $A_1, ..., A_n$ such that:

$$\Delta_{\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}}(A_i, u) = a_i \|A_i u\| \quad (u \in \mathrm{Dom}_{\otimes s}([\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}]^{mar}_{(i)}) \quad i = 1, 2, ..., n). \tag{12.44}$$

and equivalently,

(ii) there exists an approximate simultaneous observable $\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n} \equiv (\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, F)$ of $A_1, ..., A_n$ such that:

$$\Delta_{\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}}(A_i, u) = a_i \|A_i u\| \quad (u \in \mathrm{Dom}([\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^n}]^{mar}_{(i)}) \quad i = 1, 2, ..., n). \tag{12.45}$$

Proof. By Proposition 12.11, it suffices to prove (i). Put $K = \mathbf{C}^n = \{z = (z_1, ..., z_n) : z_i \in \mathbf{C} \ (i = 1, 2, ..., n)\}$, i.e., the n-dimensional Hilbert space with the norm $\|z\|_n = [\Sigma_{i=1}^n |z_i|^2]^{1/2}$. Put $e_1 = (1, 0, ..., 0)$, $e_2 = (0, 1, 0, ..., 0)$,...., $e_n = (0, 0, ..., 1) \in \mathbf{C}^n$. Put $s = e_1$. And put $P_i : \mathbf{C}^n \to \mathbf{C}^n$, $(i = 1, 2, ..., n)$, a projection such that $P_i e_i = e_i, P_i e_k = 0 (k \neq i)$, that is, $P_i = |e_i\rangle\langle e_i|$. Put $b_i = (1 + {a_i}^2)^{1/2}$ and $B_i = {b_i}^2 A_i$ $(i = 1, 2, ..., n)$. Consider the spectral representations

$$A_i = \int_{\mathbf{R}} \lambda E_{A_i}(d\lambda), \quad B_i = \int_{\mathbf{R}} \lambda E_{B_i}(d\lambda), \quad 0 = \int_{\mathbf{R}} \lambda E_0(d\lambda) \quad \text{in } H$$

and

$$P_i = \int_{\mathbf{R}} \lambda E^{\mathbf{C}^n}_{P_i}(d\lambda), \quad I = \int_{\mathbf{R}} \lambda E^{\mathbf{C}^n}_{I}(d\lambda) \quad \text{in } \mathbf{C}^n.$$

Note that $E_{A_i}(d(\lambda/{b_i}^2)) = E_{B_i}(d\lambda)$. Define the unitary operator $\widehat{U} : H \otimes \mathbf{C}^n \to H \otimes \mathbf{C}^n$ by $\widehat{U} = I \otimes U$ where a unitary operator U on $\mathbf{C}^n$ satisfies that $U e_1 = \Sigma_{i=1}^n e_i / b_i$. And define the crisp observable $(\mathbf{R}, \mathcal{B}_{\mathbf{R}}, \widehat{E}_{\widehat{A}_i})$ in $B(H \otimes \mathbf{C}^n)$ by

$$\widehat{E}_{\widehat{A}_i}(d\xi) = \widehat{U}^* [E_{B_i}(d\xi) \otimes P_i + E_0(d\xi) \otimes (I - P_i)] \widehat{U} \quad (i = 1, 2, ..., n). \tag{12.46}$$

Since $\widehat{E}_{\widehat{A}_1}, ..., \widehat{E}_{\widehat{A}_n}$ commute, we can define a crisp observable $(\mathbf{R}^n, \mathcal{B}_{\mathbf{R}^n}, \widehat{E}_{\widehat{A}})$ in $B(H \otimes \mathbf{C}^n)$ such that:

$$\widehat{E}_{\widehat{A}}(d\xi_1 d\xi_2 ... d\xi_n) = \prod_{i=1}^n \widehat{E}_{\widehat{A}_i}(d\xi_i). \tag{12.47}$$

Now, we shall show that the tensor observable $\widehat{\mathbf{O}}^{tnsr}_{H\otimes K} = (\mathbf{C}^n, e_1, (\mathbf{R}^n, B_n, \widehat{E}_{\widehat{A}}))$ is an approximate simultaneous tensor observable of $A_1, ..., A_n$. Put $\widehat{A}_i = \int_{\mathbf{R}^n} \xi_i \, \widehat{E}_{\widehat{A}} \, (d\xi_1 d\xi_2 ... d\xi_n)$ $(i = 1, ..., n)$. Then we see that,

$$
\begin{aligned}
&\int_{\mathbf{R}^n} |\xi_i|^2 \langle u \otimes e_1, \widehat{E}_{\widehat{A}}(d\xi_1 d\xi_2 ... d\xi_n)(u \otimes e_1)\rangle \\
&= \int_{\mathbf{R}} |\xi_i|^2 \langle u \otimes e_1, \widehat{E}_{\widehat{A}_i}(d\xi_i)(u \otimes e_1)\rangle \\
&= \int_{\mathbf{R}} |\xi_i|^2 \langle u \otimes e_1, [(I \otimes U^*)\Big(E_{B_i}(d\xi_i) \otimes P_i + E_0(d\xi_i) \otimes (I - P_i)\Big)(I \otimes U)](u \otimes e_1)\rangle \\
&= \int_{\mathbf{R}} |\xi|^2 \langle u, E_{B_i}(d\xi)u\rangle \cdot \langle e_1, U^* P_i U e_1\rangle \\
&= \int_{\mathbf{R}} |\xi|^2 \langle u, E_{B_i}(d\xi)u\rangle \cdot \langle \Sigma_{j=1}^n \frac{e_j}{b_j}, P_i \Sigma_{k=1}^n \frac{e_k}{b_k}, \rangle \\
&= |b_i|^{-2} \int_{\mathbf{R}} |\lambda|^2 \langle u, E_{B_i}((d\lambda)u\rangle = |b_i|^2 \int_{\mathbf{R}} |\lambda|^2 \langle u, E_{A_i}(d\lambda)u\rangle.
\end{aligned}
\tag{12.48}
$$

Hence, $\mathrm{Dom}_{\otimes s}(\widehat{A}_i) = \mathrm{Dom}(A_i)$ (where $s = e_1$). Similarly we see

$$
\begin{aligned}
&\int_{\mathbf{R}^n} \xi_i \langle u \otimes e_1, \widehat{E}_{\widehat{A}}(d\xi_1 d\xi_2 ... d\xi_n)(u \otimes e_1)\rangle \\
&= |b_i|^{-2} \int_{\mathbf{R}} \lambda \langle u, E_{B_i}((d\lambda)u\rangle = \int_{\mathbf{R}} \lambda \langle u, E_{A_i}(d\lambda)u\rangle.
\end{aligned}
\tag{12.49}
$$

Thus, $\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}$ satisfies the condition (ii) in Definition 12.6. Also, noting that $I(d\lambda) = I(1 \in d\lambda), = 0(1 \notin d\lambda)$, we also see that, for each i $(i = 1, 2, ..., n)$ and $\Xi_k \in \mathcal{B}$,

$$
\begin{aligned}
&\widehat{E}_{\widehat{A}_i}(\Xi_1) \cdot (E_{A_i}(\Xi_2) \otimes I) \\
&= (I \otimes U^*)\Big(E_{B_i}(\Xi_1) \otimes P_i + E_0(\Xi_1) \otimes (I - P_i)\Big)(I \otimes U)(E_{A_i}(\Xi_2) \otimes I) \\
&= (E_{A_i}(\Xi_2) \otimes I)(I \otimes U^*)\Big(E_{B_i}(\Xi_1) \otimes P_i + E_0(\Xi_1) \otimes (I - P_i)\Big)(I \otimes U) \\
&= (E_{A_i}(\Xi_2) \otimes I) \cdot \widehat{E}_{\widehat{A}_i}(\Xi_1).
\end{aligned}
\tag{12.50}
$$

So, $\widehat{A}_i$ and $A_i \otimes I$ commute since $\widehat{A}_i = \int_{\mathbf{R}} \xi \; E_{\widehat{A}_i}(d\xi)$ and $A_i \otimes I = \int_{\mathbf{R}} \xi \; (E_{A_i}(d\xi) \otimes I)$. Hence, $\widehat{A}_i - A_i \otimes I$ on $\mathrm{Dom}(\widehat{A}_i) \cap \mathrm{Dom}(A_i \otimes I)$ has the unique self-adjoint extension $[\widehat{A}_i - A_i \otimes I]$, which has the spectral representation

$$
[\widehat{A}_i - A_i \otimes I] = \int_{\mathbf{R}^2} (\xi_1 - \xi_2) \widehat{E}_{\widehat{A}_i}(d\xi_1)(E_{A_i}(d\xi_2) \otimes I). \tag{12.51}
$$

Then, we see that

$$
\|[\widehat{A}_i - A_i \otimes I](u \otimes e_1)\|^2 \tag{12.52}
$$

$$
\begin{aligned}
&= \int_{\mathbf{R}^2} |\xi_1 - \xi_2|^2 \langle u \otimes e_1, E_{\widehat{A}_i}(d\xi_1)(E_{A_i}(d\xi_2) \otimes I)(u \otimes e_1)\rangle \\
&= \int_{\mathbf{R}} |\xi|^2 \langle u \otimes e_1, E_{\widehat{A}_i}(d\xi_1)(u \otimes e_1)\rangle \\
&\quad - 2 \int_{\mathbf{R}^2} \xi_1 \xi_2 \langle u \otimes e_1, E_{\widehat{A}_i}(d\xi_1)(E_{A_i}(d\xi_2) \otimes I)(u \otimes e_1)\rangle \\
&\quad + \int_{\mathbf{R}} |\xi_2|^2 \langle u \otimes e_1, (E_{A_i}(d\xi_2) \otimes I)(u \otimes e_1)\rangle \\
&= (|b_i|^2 - 2 + 1) \int_{\mathbf{R}} |\xi|^2 \langle u, E_{A_i}(d\xi) u\rangle \\
&= |a_i|^2 \|A_i u\|^2,
\end{aligned}
\tag{12.53}
$$

which implies that $\mathrm{Dom}_{\otimes s}([\widehat{A}_i - A_i \otimes I]) = \mathrm{Dom}(A_i)$ (where $s = e_1$) and $\Delta_{\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^n}}(A_i, u)$ $= a_i \|A_i u\|$. Therefore, the proof of theorem is completed. □

Remark 12.17. In the above proof, the following statements were also proved:

(i) $\widehat{A}_i$ and $A_i \otimes I$ commute, so $\widehat{A}_i - A_i \otimes I$ on $\mathrm{Dom}(\widehat{A}_i) \cap \mathrm{Dom}(A_i \otimes I)$ has a unique self-adjoint extension $[\widehat{A}_i - A_i \otimes I]$ $(i = 1, 2)$,

(ii) $\mathrm{Dom}_{\otimes s}(\widehat{A}_i) = \mathrm{Dom}_{\otimes s}([\widehat{A}_i - A_i \otimes I]) = \mathrm{Dom}(A_i)$ $(i = 1, 2)$.

Thus the commutative condition (iv) in Remark 12.7 is satisfied.

■

12.6 Uncertainty relations

Now we propose the following theorem, which is our main result in this chapter. We believe that this theorem is the final version of Heisenberg's uncertainty relation concerning measurement errors.

Theorem 12.18. [Heisenberg's uncertainty relation, *cf.* [36, 67]]. *Let A_1 and A_2 be any (unbounded) self-adjoint operators on a Hilbert space H. Then, we see,*

(i) for any approximate simultaneous tensor observable $\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^2} \equiv (K, s, (\mathbf{R}^2, \mathcal{B}_{\mathbf{R}^2}, \widehat{F}))$ of A_1 and A_2, the following inequality holds:

$$
\Delta_{\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^2}}(A_1, u) \cdot \Delta_{\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^2}}(A_2, u) \geq \frac{1}{2} |\langle A_1 u, A_2 u\rangle - \langle A_2 u, A_1 u\rangle| \tag{12.54}
$$

for all $u \in H$ *such that* $\|u\| = 1$, *where the left hand side of* (12.54) *is defined by* ∞ *if* $\Delta_{\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^2}}(A_i, u) = \infty$ *for some* i,

and equivalently,

(ii) *for any approximate simultaneous observable* $\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2} \equiv (\mathbf{R}^2, \mathcal{B}_{\mathbf{R}^2}, F)$ *of* A_1 *and* A_2, *the following inequality holds:*

$$\Delta_{\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2}}(A_1, u) \cdot \Delta_{\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2}}(A_2, u) \geq \frac{1}{2}|\langle A_1 u, A_2 u\rangle - \langle A_2 u, A_1 u\rangle| \tag{12.55}$$

for all $u \in H$ *such that* $\|u\| = 1$, *where the left hand side of* (12.55) *is defined by* ∞ *if* $\Delta_{\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2}}(A_i, u) = \infty$ *for some* i.

Proof. By Proposition 12.11, it suffices to prove (i). Put $\widehat{A}_i = \int_{\mathbf{R}^2} \lambda_i \widetilde{F}(d\lambda_1 d\lambda_2)$ $(i = 1, 2)$. Let $u \in D(A_1) \cap D(A_2)$. If $u \notin \mathrm{Dom}_{\otimes s}(\widehat{A}_i)$ for some i, we see, by the definition of the uncertainty, that $\Delta_{\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^2}}(A_i, u) = \infty$, so (12.55) clearly holds. Hence, it is sufficient to prove (12.55) for $u \in \mathrm{Dom}_{\otimes s}(\widehat{A}_1) \cap \mathrm{Dom}_{\otimes s}(\widehat{A}_2)$. Let u be any element in $u \in \mathrm{Dom}_{\otimes s}(\widehat{A}_1) \cap \mathrm{Dom}_{\otimes s}(\widehat{A}_2)$. We see, by the part (ii) of Lemma 12.14, that

$$\begin{aligned}
&\langle A_1 u, A_2 u\rangle + \langle(\widehat{A}_1 - A_1 \otimes I)(u \otimes s), (\widehat{A}_2 - A_2 \otimes I)(u \otimes s)\rangle \\
&= \int_{\mathbf{R}^2} \lambda_1 \lambda_2 \langle u \otimes s, \widetilde{F}(d\lambda_1 d\lambda_2)(u \otimes s)\rangle \\
&= \langle A_2 u, A_1 u\rangle + \langle(\widehat{A}_2 - A_2 \otimes I)(u \otimes s), (\widehat{A}_1 - A_1 \otimes I)(u \otimes s)\rangle
\end{aligned} \tag{12.56}$$

from which, we get, by Schwarz inequality, that

$$\begin{aligned}
&\frac{1}{2}|\langle A_1 u, A_2 u\rangle - \langle A_2 u, A_1 u\rangle| \\
&= \frac{1}{2}|\langle(\widehat{A}_1 - A_1 \otimes I)(u \otimes s), (\widehat{A}_2 - A_2 \otimes I)(u \otimes s)\rangle \\
&\qquad - \langle(\widehat{A}_2 - A_2 \otimes I)(u \otimes s), (\widehat{A}_1 - A_1 \otimes I)(u \otimes s)\rangle| \\
&\leq \|(\widehat{A}_1 - A_1 \otimes I)(u \otimes s)\| \cdot \|(\widehat{A}_2 - A_2 \otimes I)(u \otimes s)\|.
\end{aligned} \tag{12.57}$$

Hence (by Theorem 12.15), the proof is completed. □

The following theorem was first discovered by Arthurs and Goodman [6]. However we did not know their discovery in the preparation of [36].

Theorem 12.19. [Approximate simultaneous uncertainty relation, cf [6]]. *Let* A_1 *and* A_2 *be any (unbounded) self-adjoint operators on a Hilbert space* H. *Then, we see,*

(i) *for any approximate simultaneous tensor observable* $\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^2} = (K, s, (\mathbf{R}^2, \mathcal{B}_{\mathbf{R}^2}, \widehat{F}),$ *) of* (A_1, A_2)*, the following inequality holds:*

$$(var[\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^2}, u]_1)^{1/2} \cdot (var[\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^2}, u]_2)^{1/2} \geq |\langle A_1 u, A_2 u\rangle - \langle A_2 u, A_1 u\rangle| \tag{12.58}$$

for all $u \in H$ *such that* $\|u\| = 1$*, where the left hand side of* (12.58) *is defined by* ∞ *if* $var[\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^2}, u]^{mar}_{(i)} = \infty$ *for some* i*, also the right hand side of* (12.58) *is defined by* ∞ *if* $u \notin \mathrm{Dom}(A_1) \cap \mathrm{Dom}(A_2)$*,*

and equivalently

(ii) *for any approximate simultaneous observable* $\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2} = (\mathbf{R}^2, \mathcal{B}_{\mathbf{R}^2}, \widehat{F}\ )$ *of* (A_1, A_2)*, the following inequality holds:*

$$(var[\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2}, u]_1)^{1/2} \cdot (var[\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2}, u]_2)^{1/2} \geq |\langle A_1 u, A_2 u\rangle - \langle A_2 u, A_1 u\rangle| \tag{12.59}$$

for all $u \in H$ *such that* $\|u\| = 1$*, where the left hand side of* (12.59) *is defined by* ∞ *if* $var[\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^2}, u]^{mar}_{(i)} = \infty$ *for some* i*, also the right hand side of* (12.59) *is defined by* ∞ *if* $u \notin \mathrm{Dom}(A_1) \cap \mathrm{Dom}(A_2)$*.*

Proof. By Proposition 12.11, it suffices to prove (i). Put $\widehat{A}_i = \int_{\mathbf{R}^2} \lambda_i \widehat{F}(d\lambda_1 d\lambda_2)$ $(i = 1, 2)$. If $u \notin \mathrm{Dom}_{\otimes s}(\widehat{A}_i)$ for some i, we see, by the definition of the variance, that $var[\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^2}, u]^{mar}_{(i)} = \infty$, so, (12.58) clearly holds. Hence, it is sufficient to prove (12.58) in the case that $u \in \mathrm{Dom}_{\otimes s}(\widehat{A}_1) \cap \mathrm{Dom}_{\otimes s}(\widehat{A}_2)$. Let u be any element in $\mathrm{Dom}_{\otimes s}(\widehat{A}_1) \cap \mathrm{Dom}_{\otimes s}(\widehat{A}_2)$. Then, we see, by (iii) in Lemma 12.14, that

$$var[\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^2}, u]^{mar}_{(i)} = \|\widehat{A}_i(u \otimes s)\|^2 - |\langle u \otimes s, \widehat{A}_i(u \otimes s)\rangle|^2 \tag{12.60}$$

$$= \|A_i u\|^2 + \|(\widehat{A}_i - A_i \otimes I)(u \otimes s)\|^2 - |\langle u, A_i u\rangle|^2$$

$$\leq 2(\|A_i u\|^2 - |\langle u, A_i u\rangle|^2)^{1/2} \cdot \|(\widehat{A}_i - A_i \otimes I)(u \otimes s)\| \quad (i = 1, 2), \tag{12.61}$$

therefore, by Lemma 12.13 and Theorem 12.18 we get,

$$\begin{aligned} & var[\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^2}, u]_1 \cdot var[\widehat{\mathbf{O}}^{ASTO}_{[A_l]_{l=1}^2}, u]_2 \\ \geq\ & 4(\|A_1 u\|^2 - |\langle u, A_1 u\rangle|^2)^{1/2} \cdot (\|A_2 u\|^2 - |\langle u, A_2 u\rangle|^2)^{1/2} \\ & \quad \cdot \|(\widehat{A}_1 - A_1 \otimes I)(u \otimes s)\| \cdot \|(\widehat{A}_2 - A_2 \otimes I)(u \otimes s)\| \\ \geq\ & |\langle A_1 u, A_2 u\rangle - \langle A_2 u, A_1 u\rangle|^2. \end{aligned} \tag{12.62}$$

Hence, the proof is completed. □

Now we have the following corollary[3].

Corollary 12.20. [Uncertainty relations concerning a pair of conjugate observables]. *Let A_1 and A_2 be a pair of conjugate observables in a Hilbert space H.*

(i: *cf.* [7]) *There exists an approximate simultaneous observable $\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2} = (\mathbf{R}^2, \mathcal{B}_{\mathbf{R}^2}.F\ )$ of A_1 and A_2. Thus, we can take an approximate simultaneous measurement $\overline{\mathbf{M}}_{B(H)}(\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2}, \overline{S}_{[|u\rangle\langle u|]})$.*

(ii: *cf.* [36]) *For any positive number ϵ and any $k(=1,2)$, there exists an approximate simultaneous observable $\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2} = (\mathbf{R}^2, \mathcal{B}_{\mathbf{R}^2}.F\ )$ of A_1 and A_2 such that:*

$$\Delta_{\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2}}(A_k, u) \le \epsilon \|A_k u\|_H \qquad (\forall u \in H \text{ such that } \|u\| = 1),$$

(iii: *cf.* [36, 67]) (*Heisenberg's uncertainty relation*) *However the following inequality holds*

$$\Delta_{\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2}}(A_1, u) \cdot \Delta_{\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2}}(A_2, u) \ge \hbar/2 \tag{12.63}$$

for all $u \in H$ $\quad (\|u\|_H = 1)$,

(iv: *cf.* [6]) *The following inequalities hold:* (*approximate simultaneous uncertainty relation*)

$$(var[\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2}, u]_1)^{1/2} \cdot (var[\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2}, u]_2)^{1/2} \ge \hbar \tag{12.64}$$

for all $u \in H$ $\quad (\|u\|_H = 1)$.

Proof. Note that $\langle A_1 u, A_2 u\rangle - \langle A_2 u, A_1 u\rangle = i\hbar$ $(u \in \mathrm{Dom}(A_1) \cap \mathrm{Dom}(A_2), \|u\|_H = 1)$. Then, the above assertions (i) and (ii) are consequences of Theorem 12.16. Also, the above assertions (iii) and (iv) are respectively consequences of Theorem 12.18 and Theorem 12.19. □

[3] There are other uncertainty relations, For the recent variants, see [68].

12.7 EPR-experiment and Heisenberg's uncertainty relation

Now we have the complete form of Heisenberg's uncertainty relation as Corollary 12.20. To be compared with Corollary 12.20, we should note that the conventional Heisenberg's uncertainty relation (= Proposition 12.1) is ambiguous. Wrong conclusions are sometimes derived from the ambiguous statement (= Proposition 12.1). For example, in some books of physics, it is concluded that EPR-experiment (Einstein, Podolosky and Rosen [22]) contradicts with Heisenberg's uncertainty relation. That is,

(I) Heisenberg's uncertainty relation says that the position and the momentum of a particle can not be measured simultaneously and exactly.

On the other hand,

(II) EPR-experiment says that the position and the momentum of a certain "particle" can be measured simultaneously and exactly.

Thus someone may conclude that the above (i) and (ii) includes a paradox, and therefore, EPR-experiment contradicts with Heisenberg's uncertainty relation. Of course, this is a misunderstanding. This "paradox" was solved in [36]. Now we shall explain the solution of the paradox.

[Concerning the above (I)] Put $H = L^2(\mathbf{R}_q)$. Consider two-particles system in $H \otimes H = L^2(\mathbf{R}^2_{(q_1,q_2)})$. In the EPR problem, we, for example, consider the state u_s $(\in H \otimes H = L^2(\mathbf{R}^2_{(q_1,q_2)}))$ $\Big($or precisely, $|u_s\rangle\langle u_s|\Big)$ such that:

$$u_s(q_1,q_2) = \sqrt{\frac{1}{2\pi\epsilon\sigma}} e^{-\frac{1}{8\sigma^2}(q_1-q_2-a)^2-\frac{1}{8\epsilon^2}(q_1+q_2-b)^2} \cdot e^{i\phi(q_1,q_2)} \tag{12.65}$$

where ϵ is assumed to be a sufficiently small positive number and $\phi(q_1,q_2)$ is a real-valued function. This is the quantum form of EPR-experiment in Remark 12.2(ii). Let $A_1 : L^2(\mathbf{R}^2_{(q_1,q_2)}) \to L^2(\mathbf{R}^2_{(q_1,q_2)})$ and $A_2 : L^2(\mathbf{R}^2_{(q_1,q_2)}) \to L^2(\mathbf{R}^2_{(q_1,q_2)})$ be self-adjoint operators such that

$$A_1 = q_1, \qquad A_2 = \frac{\hbar\partial}{i\partial q_1}. \tag{12.66}$$

Then, Corollary 12.20 (i) says that there exists an approximate simultaneous observable $\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2} = (\mathbf{R}^2, \mathcal{B}_{\mathbf{R}^2}.F\ )$ of A_1 and A_2. Thus we can take an approximate simultaneous measurement $\overline{\mathbf{M}}_{B(H)}(\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2}, \overline{S}_{[|u_s\rangle\langle u_s|]})$. And thus, the following Heisenberg's uncertainty relation (= Corollary 12.20 (iii)) holds,

$$\Delta_{\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2}}(A_1, u_s) \cdot \Delta_{\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2}}(A_2, u_s) \geq \hbar/2 \tag{12.67}$$

[Concerning the above (II)] However, it should be noted that, in the above situation we assume that the state u_s is known before the measurement. In such a case, we may take another measurement as follows: Define the self-adjoint operators $\widehat{A}_1 : L^2(\mathbf{R}^2_{(q_1,q_2)}) \to L^2(\mathbf{R}^2_{(q_1,q_2)})$ and $\widehat{A}_2 : L^2(\mathbf{R}^2_{(q_1,q_2)}) \to L^2(\mathbf{R}^2_{(q_1,q_2)})$ such that

$$\widehat{A}_1 = b - q_2, \qquad \widehat{A}_2 = A_2 = \frac{\hbar\partial}{i\partial q_1} \tag{12.68}$$

Note that these operators commute. Therefore,

($\sharp$) we can take an exact simultaneous measurement of $\widehat{A}_1$ and $\widehat{A}_2$ (for the state u_s).

And moreover, we can easily calculate as follows (*cf.* Definition 11.1 and Remark 12.12).

$$\begin{aligned}
&\Delta\Big(\overline{\mathbf{M}}_{B(H)}(A_1 \times \widehat{A}_1, \overline{S}(\rho_{u_s}))\Big) = \|\widehat{A}_1 u_s - A_1 u_s\| \\
=&\Big[\iint_{\mathbf{R}^2} \Big|((b-q_2)-q_1)\sqrt{\frac{1}{2\pi\epsilon\sigma}} e^{-\frac{1}{8\sigma^2}(q_1-q_2-a)^2-\frac{1}{8\epsilon^2}(q_1+q_2-b)^2} \cdot e^{i\phi(q_1,q_2)}\Big|^2 dq_1 dq_2\Big]^{1/2} \\
=&\Big[\iint_{\mathbf{R}^2} \Big|((b-q_2)-q_1)\sqrt{\frac{1}{2\pi\epsilon\sigma}} e^{-\frac{1}{8\sigma^2}(q_1-q_2-a)^2-\frac{1}{8\epsilon^2}(q_1+q_2-b)^2}\Big|^2 dq_1 dq_2\Big]^{1/2} \\
=&\sqrt{2}\epsilon,
\end{aligned} \tag{12.69}$$

and

$$\Delta\Big(\overline{\mathbf{M}}_{B(H)}(A_2 \times \widehat{A}_2, \overline{S}(\rho_{u_s}))\Big) = \|\widehat{A}_2 u_s - A_2 u_s\| = 0. \tag{12.70}$$

Thus we see

$$\Delta\Big(\overline{\mathbf{M}}_{B(H)}(A_1 \times \widehat{A}_1, \overline{S}(\rho_{u_s}))\Big) \cdot \Delta\Big(\overline{\mathbf{M}}_{B(H)}(A_2 \times \widehat{A}_2, \overline{S}(\rho_{u_s}))\Big) = 0. \tag{12.71}$$

Since ϵ (> 0) can be taken sufficiently small, the above measurement ($\sharp$) is superior to the approximate simultaneous measurement $\overline{\mathbf{M}}_{B(H)}(\overline{\mathbf{O}}^{ASO}_{[A_l]_{l=1}^2}, \overline{S}_{[|u_s\rangle\langle u_s|]})$. $\Big($Here, $\overline{S}_{[|u_s\rangle\langle u_s|]}$ is identified with $\overline{S}(|u_s\rangle\langle u_s|)$ since $|u_s\rangle\langle u_s|$ is a pure state.$\Big)$ However it should be again noted that, the measurement ($\sharp$) is made from the knowledge of the state u_s.

[**(I) and (II) are consistent,** *cf.* [36]] The above conclusion (12.71) does not contradicts with Heisenberg's uncertainty relation (12.67), since the measurement (♯) is not an approximate simultaneous measurement of A_1 and A_2.

■

In the above arguments, note that Theorem 12.19 (approximate simultaneous uncertainty relation) is powerless to solve the paradox (i.e., the paradox between EPR-experiment and Heisenberg's uncertainty relation). That is because the concept "error" (or "uncertainty") is not explicit in Theorem 12.19.

POSTSCRIPT

In this book I propose "measurement theory", that is,

> an epistemology that is considered to be the mathematical representation of "the mechanical world view".

In this sense, I may not deny that this book is regarded as the book of philosophy.§

I surmise that a "postscript" is the part that is firstly (and most frequently) read throughout a book. Thus, in what follows I would like to enumerate important new results (≈ my favorite results) in this book.

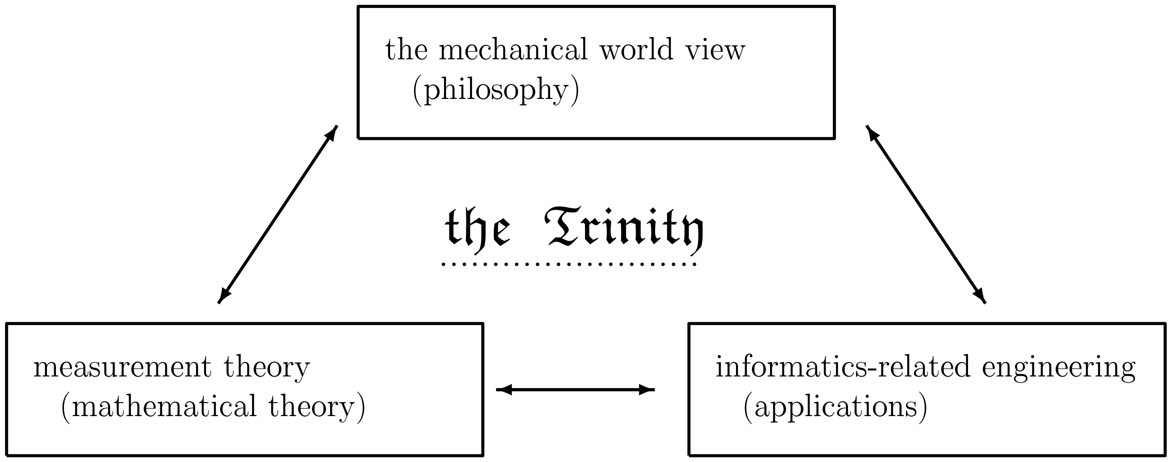
(1) MT (= measurement theory) is the mathematical representation of the epistemology called "the mechanical world view", and thus, it is also called GDST (= general dynamical system theory). I hope that the following assertion (= Table (1.7)) will be generally accepted.

Table (1.7′)

- **Sci. Theor.**
  - **Mathematics** "Foundations of math." — logic, number theory, topology, differential geometry, complex analysis, real analysis, operator algebra, differential equation, probability theory, etc. (C_1)
  - **Math. Sci. Theor.**
    - **Theor. Physics** 'TOE' — Newtonian mechanics; quantum mechanics; Maxwell's electromagnetic theory; Einstein's relativity theory; Weinberg-Salam theory; quantum chromodynamics; etc. (C_2)
    - [My proposal in this book] **Theor. Informatics** MT (= GDST) — dynamical system theory; quantum system theory; practical logic; statistics, circuit theory; control theory; multivariate analysis; information theory; chaotic system theory; automata theory; OR, game theory, etc. (C_3)
  - **Usual Sci. Theor.** — economics, chemistry, biology, medicine, psychology, statistical mechanics, fluid mechanics, engineering (also, see (I_7) and (I_8) in §1.2), etc. (C_5)

§In fact, this book can not be read and understood without Chapter 1 (the philosophy of measurement theory).

We assume that "measurement", "its philosophy" and "its applications ($\approx$ informatics-related engineering)" should be regarded as "the Trinity" as follows:



(2) I propose the characterization of Bell's inequality in the framework of PMT (i.e., Axioms 1 and 2), *cf.* §3.7. I conclude that:

- if we admit PMT (= "Axiom 1 + Axiom 2 (Markov relation)"), we must admit the fact that there is something faster than light. (3.49)

This assertion is, of course, one of the most profound scientific assertions in all science. As mentioned in the footnote below §3.7.1, my understanding of Bell's inequality may be shallow. Thus, I think that the most of originality may not be due to me but great pioneers (i.e., de Broglie, A. Einstein, J.S. Bell, etc.).

(3) I assert that equilibrium statistical mechanics should be due to "STI" (= "staying time interpretation of statistical mechanics in (4.28)") and not "PI" (= "probabilistic interpretation of statistical mechanics in (4.30)"]) in Chapter 4. That is, under the "STI" (which is nearly regarded as common sense), equilibrium statistical mechanics can be understood in classical PMT as follows:

$$\text{“equilibrium statistical mechanics”} = \underbrace{\underset{((A_1)(=\text{ Axiom 1}))}{\text{“probabilistic rule”}} + \underset{((T^1)\text{ and }(T^2))\text{ under (EH))}}{\text{“Newton equation”}}}_{\text{(STI } (\approx \text{ “common sense”}))} \qquad \begin{matrix}(4.28)\\ (=(4.4))\end{matrix}$$

Also, see the other proposals (4.29) and (4.31).

(4) I stress the following correspondence:

$$\text{Axiom 1 (measurement) in PMT} \leftrightarrow \text{ Fisher's likelihood method in statistics}$$

That is, Fisher's likelihood method is one of aspects of Axiom 1 (measurement). Cf. Theorem 5.3.

(5) Regression analysis II (6.48) (and not Regression analysis I (6.7)) in Chapter 6. This and the above (4) imply that Fisher's statistics is "theoretically true", (*cf.* Declaration 1.11).

(6) In §7.1, I assert that "measurement", "inference" and "control" are the different aspects of the same thing. Also, since "(practical) logic" is a qualitative aspect of "inference", there is a reason to consider that "(practical) logic" [resp. "inference"] is used in rough [resp. precise] arguments.

(7) Theorem 7.19 (practical logic in classical measurements). This theorem justifies the following famous saying;

- *Since Socrates is a man and all men are mortal, it follows that Socrates is mortal.*

Also, the following strange logic is proposed:

$$\text{"SWEET"} \Rightarrow \text{"RIPE" and "RIPE"} \Rightarrow \text{"RED" implies "RED"} \Rightarrow \text{"SWEET" (in some sense)} \tag{7.38}$$

(8) If Zadeh's assertion is that *system theory has a logical aspect,* I agree with him. In fact, practical logic is discussed in the framework of GDST (= MT) in Chapter 7. However, I think that Zadeh's fuzzy sets theory overstates many things. Thus, in §7.5, I assert "Zadeh's fuzzy sets theory can not be completely formulated in MT". That is, his theory is not completely "theoretically true" (*cf.* Declaration (1.11) in Chapter 1). And thus we do not add Zadeh's fuzzy sets theory to (C_3) in (1.7). His "theory" sholud be regarded as one of empirical methods in MT. However, the fashion of his theory gave me the original motivation of our theory (*cf.* the footnote below Problem 1.2 in Chapter 1).

(9) The measurement theoretical formulation of Kalman filter in §8.4 (though it is merely a simple corollary of the generalized Bayes theorem (= Theorem 6.6 or Theorem 8.13)).

(10) The entropy of a measurement (particularly, Examples 8.17 and 8.18).

(11) Theorem 8.20 (Bayes theorem for belief measurements). It should be noted that belief measurements have no samples spaces. Thus, the proof is different from the proof of Bayes theorem for statistical measurements.

(12) Bertrand's paradox is clear in MT (*cf.* §8.7). It is obvious that we encounter Bertrand's paradox if "invariant state" is unreasonably regarded as "statistical state". It should be noted that "invariant state" and "statistical state" are not directly related in MT.

(13) The generalized moment method in §9.4. I want to compare Fisher's likelihood method (Theorem 5.3), Bayes' method (Theorem 8.13, Remark 8.14) and the moment method in the framework of measurement theory. In order to do so, we have to propose the generalized moment method (in §9.4).

(14) The definition of "particle's trajectories" due to Theorems 10.1 [W^*-algebraic generalization of Kolmogorov's extension theorem]. Particularly, the definition of Brownian motion $B(t)$ in §10.4. Since Brownian motion is not a "motion" but "measured values", we can understand the fact: the velocity "$\frac{dB(t)}{dt}$ does not exist".

(15) The definition of "measurement error" in §11.1. This is superior to the "conventional definition" such as | "true value" − "measured value" |. Also, this is essential to the characterization of Heisenberg's uncertainty relation (*cf.* Chapter 12).

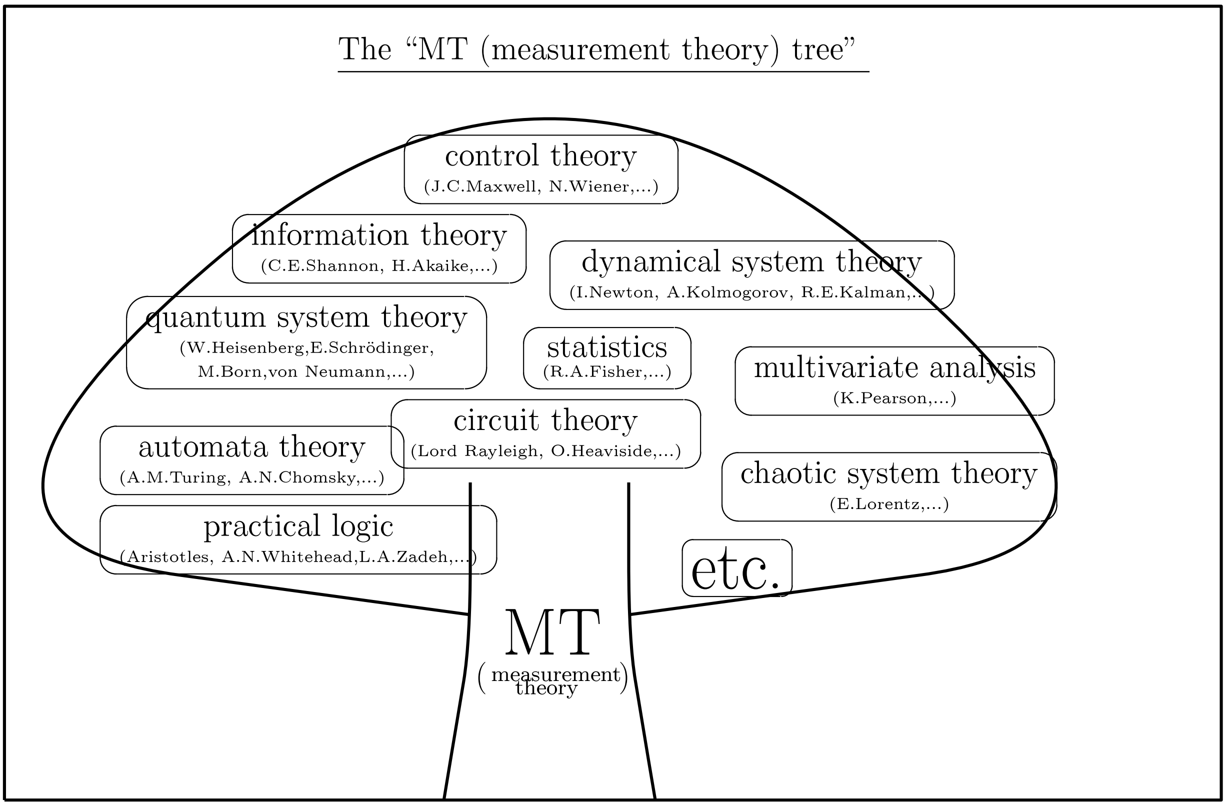
(16) Theorem 11.12 (The principle of equal probability, SMT_{PEP}-method), which makes Bayes theorem quite applicable. That is, I consider that this theorem (=Theorem 11.12) and the generalized Bayes theorem (= Theorem 8.13) are the most important in SMT.

(17) Four answers to the Monty Hall problem (i.e., Problem 5.12, Remark 5.13, Problem 8.8, Problem 11.13) are presented in this book. Although these are all reasonable, the answer in Problem 11.13 may be the most natural.

(18) I assert the mathematical representation of Heisenberg's uncertainty relation in §12.7. This solves the paradox between Heisenberg's uncertainty relation and EPR-experiment in §12.7.

Note that "*the mechanical world view*" (due to I. Newton, "Principia";1687, [66]) is one of the most successful epistemologies in the history of science as well as mechanics. This is the histrical fact. And therefore, I am convinced that our proposal (i.e., "measurement theory" (=the mathematical representation of "*the mechanical world view*")) has a gerat power to understand and analyze every phenomenon.

I hope that "MT tree" will grow more and more.



S. Ishikawa¶

¶For the further information (development, errata, etc.) of our theory, see "http://www.keio-up.co.jp/kup/mfomt/"

Bibliography

[1] Abu-Zeid, O. *On the existence of joint probability distribution for arbitrary incompatible observables in quantum theory,* Phys. Lett. **A125**, 162-164 (1987)

[2] Ali, S.T. and Emch, G.G. *Fuzzy observables in quantum mechanics,* J. Math. Phys. **15(2)**, 176-182 (1974)

[3] Ali, S.T. and Prugovečki, E. *Systems of im primitivity and representations of quantum mechanics on fuzzy phase spaces,* J. Math. Phys. **18(2)**, 219-228 (1976)

[4] Achiezer, N.I. and Glazman, I.M. *The theory of linear operators in Hilbert space,* Akademie-Verlag, Berlin (1954)

[5] Amari, S. *Differential geometrical methods in statistics,* Springer Lecture Notes in statistics, (28) (1985)

[6] Arthurs, E and Goodman, M.S. *Quantum correlations- A generalized Heisenberg uncertainty relation,* Physical Review Letters **60**, (24) 2447-2449 (1988)

[7] Arthurs, E. and Kelly, J.L.,Jr. *On the simultaneous measurement of a pair of conjugate observables,* Bell System Tech. J. **44**, 725-729 (1965)

[8] Aspect, A, Dallibard, J. and Roger, G. *Experimental test of Bell inequalities time-varying analysis,* Physical Review Letters **49**, 1804-1807 (1982)

[9] Bell, J.S. *On the Einstein-Podolosky-Rosen Paradox,* Physics **1**, 195-200 (1966)

[10] Bennett, C.H. et all. *Teleporting an unknown quantum state via dual classic and Einstein-Podolsky-Rosen channels,* Physical Review Letters **70**, 1895-1899 (1993)

[11] Billingsley, P. *Ergodic theory and information,* John Wiley (1965)

[12] Bohr, N. *Can quantum-mechanical description of physical reality be considered complete?,* Phys. Rev. (48) 696-702 1935

[13] Born, M. *Zur Quantenmechanik der Stoßprozesse (Vorläufige Mitteilung),* Z. Phys. (37) 863-867 1926

[14] Busch, P. *Indeterminacy relations and simultaneous measurements in quantum theory,* International J. Theor. Phys. **24**, 63-92 (1985)

[15] Busch, P. and Lahti, P. *On various joint measurements of position and momentum observables in quantum theory,* Phys. Rev. **D29**, 1634-1646 (1984)

[16] Cheng, Chi-Lun and Van Ness, J.W. *Statistical regression with measurement error,* (Kendall's Library of statistics) Arnold 1999

[17] Davies, E.B. *Quantum theory of open systems,* Academic Press 1976

[18] Davies, P.C.W. and Brown, J.R. *The ghost in the atom,* Cambridge University Press 1986

[19] Devaney, L.R. *An introduction to chaotic dynamical systems,* Benjaminn 1986

[20] de Broglie, L. *L'interpretation de la mecanique ondulatoire,* Journ. Phys. Rad. **20**, 963 (1959)

[21] de Finetti, B. *Theory of probability,* Vols. 1 and 2, Wiley, New York 1974, 1975

[22] Einstein, A., Podolosky, B. and Rosen, N. *Can quantum-mechanical description of reality be considered completely?* Physical Review Ser 2(47) 777-780 (1935)

[23] Feynman, R.P. and Hibbs, A.R. *Quantum mechanics and path integrals,* McGraw-Hill Inc. (1965)

[24] Frieden, F. *Fisher information in physics,* Cambridge University Press (1999)

[25] Fuller, W.A. *Measurement error models,* John Wiley & Sons (1987)

[26] Gibbs, J.W. *Elementary principles in statistical mechanics,* Yale Univ. Press 1902 (Dover 1960)

[27] Gillies, D. *Philosophical Theories of Probability,* Publisher:Routledge Published (2000)

[28] Green, M.B., Schwarz, J. M., Witten, E. *Superstring theory,* Cambridge university press (1987)

[29] Halmos, P.R. *Measure theory,* van Northland (1953)

[30] Hand, D.J. *Measurement theory and Practice,* Arnord (2004)

[31] Heisenberg, W. *Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik,* Z. Phys. **43**, 172-198 (1927)

[32] Hida, T. *Brownian motions,* Springer-Verlag 1975

[33] Hoffman, P. *The Man Who Loved Only Numbers, The story of Paul Erdös and the search for mathematical truth,* Hyperion, New York 1998

[34] Holevo, A.S. *Probabilistic and statistical aspects of quantum theory,* North-Holland publishing company (1982)

[35] Isaac, R. *The pleasures of probability,* Springer-Verlag (Undergraduate texts in mathematics) 1995

[36] Ishikawa, S. *Uncertainty relation in simultaneous measurements for arbitrary observables,* Rep. Math. Phys. **29**, 257-273 (1991)

[37] Ishikawa, S. *Uncertainties and an interpretation of nonrelativistic quantum theory,* Internat. J. Theoret. Phys. **30** 401-417 (1991)

[38] Ishikawa, S. *Heisenberg's uncertainty relation and a new measurement axiom in quantum theory* (In Japanese) Soryusironkenkyu (Kyoto University) **83**,6. F58-F79 (1991) (This is translated to English in "Keio Science and Technology Reports",(1992), Vol 45, No.1, 1-22)

[39] Ishikawa, S. *A Consistent History approach to compatibilities and commutativities in non-relativistic quantum mechanics,* Interference between physics and mathematics, (Zhejian University, 1993), World Sci. Publishing, River Edge NJ. (1994) 300-314

[40] Ishikawa, S., Arai, T. and Kawai, T. *Numerical Analysis of trajectories of a quantum particle in two-slit experiment,* Internat. J. Theoret. Phys. **33** 1265-1274 (1994)

[41] Ishikawa,S. *Fuzzy inferences by algebraic method,* Fuzzy Sets and Systems **87**, 181-200 (1997)

[42] Ishikawa, S. *Fuzzy logic in measurements,* Fuzzy Sets and Systems **100** 291-300 (1998)

[43] Ishikawa, S., Arai, T. and Takamura, T. *A dynamical system theoretical approach to Newtonian mechanics,* Far east journal dynamical system **1**, 1-34 (1999)

[44] Ishikawa, S. *Statistics in measurements,* Fuzzy Sets and Systems **116**, 141-154 (2000)

[45] Ishikawa, S. *A dynamical system theoretical approach to equilibrium statistical mechanics,* Far east journal of dynamical system **3**, 9-28 (2001)

[46] Ishikawa, S. *The existence theorem of Kalman filter in measurement theory,* Far east journal of dynamical system **3**, 175-204 (2001)

[47] Ishikawa, S. *A system theoretical characterization between Fisher's and Bayes' methods,* Far east journal of theoretical statistics **7**, 19-33 (2002)

[48] Ishikawa, S. *Fundamentals of measurements (Plenary lecture),* The 2002 annual conference of the Japan Society for Industrial and Applied mathematics (20th Sep. 2002) http://www.jstage.jst.go.jp/article/pjsiam/2002/0/2002_16/

[49] Jaynes, E.T. *The well-posed problem,* Foundations of Physics **3**, 477-493 (1973)

[50] Kadison, R. V. and Ringrose, J. R. *Fundamentals of Theory of Operator Algebras I, III,* Academic press (1986, 1991)

[51] Kalman, R. E. *A new approach to linear filtering and prediction problems,* Trans. ASME, J. Basic Eng. **82**, 35 (1960)

[52] Kalman, R. E. and Bucy, R.S. *New results in linear filtering and prediction,* Trans. ASME, J. **83**, 95-108 (1961)

[53] Kalman, R.E. and Falb, P.L. and Arbib, M.A. *Topics in mathematical system theory,* TATA McGrow-Hill Publishing Company LTD, New Delhi (1969)

[54] Kennard, E.H. *Zur Quantenmechanik ein Facher Bewegungstypen,* Zeitschrift fur Physik, (1927)

[55] Kikuchi, K. and Ishikawa, S. *Regression analysis, Kalman filter and measurement error model in measurement theory,* Research report, KSTS/RR-06/005 (Keio University) , (2006)

[56] Kolmogorov, A. *Foundations of probability (translation),* Chelsea Publishing Co. 1950

[57] Krengel. U. *Ergodic Theorems,* Walter de Gruyter. Berlin, New York 1985

[58] Lee, R. C. K. *Optimal Estimation, Identification, and Control,* M.I.T. Press 1964

[59] Lehmann, E.L. *Testing statistical hypothesis,* John Wiley&Sons Inc., New York 1959

[60] Lehmann, E.L. *Theory of point estimation,* John Wiley&Sons Inc., New York 1983

[61] Lindley, D.V. *The philosophy of statistics,* Journal of the royal statistical society series D the statistician 49, 293-319, part 3, 2000

[62] Mandelbrot, B,B. *The fractal geometry of nature,* WH Freeman 1982

[63] Maxwell, J.C. *On governors,* Proc.R.Soc.London (16) 270-283, (1868)

[64] McNeil, D. *Fuzzy logic,* Simon & Schuster 1994

[65] Misra, B. and Sudarshan, E.C.G. *The Zeno's paradox in quantum theory,* J. Math. Phys. (18) 756-763, (1977)

[66] Newton. I. *Philosophiae Naturalis Principia Mathematica (1687),* Andrew Motte's English translation of 1729, revised by Florian Cajori, University of California Press 1960

[67] Ozawa, M. *Quantum limits of measurements and uncertainty principle,* in Quantum Aspects of Operational Communication edited by Bendjaballah et all. Springer, Berlin, 3-17, (1991)

[68] Ozawa, M. *Uncertainty relations for noise and disturbance in generalized quantum measurement* Annals of Physics (311) 350-416, (2004)

[69] Ozawa, M. *New development of Heisenberg's uncertainty relation — new interpretation and old interpretation,* (Japanese), Mathematical Sciences (Saiensu-sha press), No.508, 5-17, September (2005)

[70] Popper, K.R. *The Logic of Scientific Discovery,* New York: Harper and Row. (1959).

[71] Prugovečki, E. *Quantum mechanics in Hilbert space,* Academic Press, New York. (1981).

[72] Rao, C.R. *Linear statistical inference and its applications,* John Wiley. (1973)

[73] Robertson, H.P. *The uncertainty principle,* Phys. Rev. **34**, 163 (1929)

[74] Rudin, W. *Function analysis,* McGraw-Hill Book Company 1973

[75] Ruelle, D. *Statistical mechanics, rigorous results,* W.A. Benjamin 1969

[76] Sakai, S. *C^*-algebras and W^*-algebras,* Ergebnisse der Mathematik und ihrer Grenzgebiete (Band 60), Springer-Verlag, Berlin, Heidelberg, New York 1971

[77] Schrödinger, E. *Die gegenwartige Situation in der Quantenmechanik,* Naturwissenschaftern (23) 807-812; 823-823, 844-849, (1935)

[78] Selleri, F. *Die Debatte um die Quantentheorie,* Friedr. Vieweg&Sohn Verlagsgesellscvhaft MBH, Braunschweig (1983)

[79] Shannon, C.E. *A mathematical theory of communication,* Bell Syst. Tech.J. **27** 379-423, 623-656, (1948)

[80] Shor, P.W. *Algorithms for Quantum Computation: Discrete Logarithms and Factoring,* In Proceeding of 35th IEEE FOCS 124-134 (1994)

[81] Shimojo, S., Ichikawa, S. *Intuitive reasoning about probability: theoretical and experimental analyses of the "problem of three prisoners"* Cognition **32** 1-24 (1989)

[82] Takesaki, M. *Theory of operator algebras,* New York; Springer-Verlag , 1979

[83] Toda, M., Kubo, R. and Saito, N. *Statistical physics,* Springer series in solid-state sciences ; 30 Berlin ; New York; Springer-Verlag , 1983

[84] von Neumann, J. *Mathematical foundations of quantum mechanics,* Springer Verlag, Berlin (1932)

[85] von Neumann, J. and Morgenstern, O. *Theory of games and economic behavior,* Princeton Univ. Press (1944)

[86] Wald, A. *Statistical decision functions,* John Wiley 1950

[87] Waldrop, M.M. , *Complexity, The emerging science at Edge of order and chaos,* Sterling Lord Literistic Inc., New York, (1992)

[88] Weyl, H , *Gruppentheorie Und Quantenmechanik,* Hirzel, Leipzig,, (1928)

[89] Whitehead, A.N., Russell, B. *Principia Mathematica, 3vols,* Cambridge University Press, 1910, 1912, and 1913. Second edition, 1925 (Vol. 1), 1927 (Vols 2, 3). Abridged as Principia Mathematica to *56, Cambridge University Press, 1962

[90] Wiener, N. *Cybernetics or Control and Communication in the Animal and the Machine,* Paris, Hermann et Cie - MIT Press, Cambridge, MA (1948)

[91] Wiesner, S. *Conjugate coding,* Sigact News **15**, 78-88 (1983)

[92] Yosida, K. *Functional analysis,* Springer-Verlag (Sixth Edition) 1980

[93] Zadeh,L.A. *Fuzzy Sets*, Information and control, **8**, 338-353 1965

Index

a priori probability, 182
admissible (Definition) , 281
almost always, 80
almost experimentally true, 9, 35, 39, 86
almost independent sequence with the identical distribution, 82
approximate exact observable, 41
approximate simultaneous measurement, 299
Aristotles, 9, 149
automata theory, 65
Avogadro constant, 76
Axiom 1 [Measurement axiom], 29
Axiom 2 [The Markov relation among systems, the Heisenberg picture], 65
Axiom$^{W^*}$1 [The W^*-algebraic formulation of Axiom 1], 234

$\mathcal{B}^{\mathrm{bd}}_{\mathbf{R}^d}$, 40
Banach algebra, 16
$B(V)$, 18
Bayes operator (Bayes-Kalman operator), 98, 134
Bayes theorem, 98
Bayes' method, 196, 197
Bayes' postulate (= the principle of equal weight), 214
BMT (=belief measurement theory), 208
belief weight (approval rate, conviction degree), 210
Bell's inequality, 69
Bell's thought experiment, 49
Bertrand's paradox, 182, 218
Boltzmann constant, 81
bookmaker, 217
$\mathcal{B}_{\mathbf{R}}$, 28
Borel field, 28
Born, M., 16
bounded linear operator (= continuous linear operator), 2, 18
Brownian motion, 269
butterfly effect, 272

$C(\Omega)$, $C_0(\Omega)$, 17
$\mathcal{C}(V)$, 18
chaotic system theory, 272
chaotic theory, 272
characteristic function, 46
circuit theory, 65
collapse (reduction) of wave packet, 198
commutativity condition, 24
completely random measurement, 288
complex system theory, 272
conditional probability, 31
consistent condition , 157
contraposition , 156
control, 150

covering family of observables, 157
crisp observable, 22
C^*-algebra (Definition), 16
cyclic measurement, 187, 257

δ_ω, 18
Dirac notation, 19
discrete Gaussian observable, 40
discrete topology, 20
dynamical system theory (= DST), 2

entropy(U-entropy), 86
entropy(of measurement), 204
entropy(Shannon's entropy), 207
EPR-experiment (Einstein, Podolosky and Rosen), 315
equilibrium state, 81
equilibrium statistical mechanics, 75
Erdös, 190
ergodic homomorphism, 288
ergodic hypothesis, 78, 84
ergodic theorem, 78
estimator, 107
exact observable, 41, 229
experimentally true or false, 7
extensive interpretation, 13, 212

fair coin-tossing, fair dice-throwing, 208
falsifiability, 9
field, 21
Fisher's estimator, 95
Fisher's maximum likelihood method in classical and quantum measurements, 94
Fisher's maximum likelihood method in classical measurements, 96, 101
flow, 78
form, 29
foundationalism, 3
Fourier transform, 231
frequency probability, 45
fuzzy decomposition, 27
fuzzy numbers observable, 40
fuzzy observable, 22, 228, 230
fuzzy sets theory, 169

game theory, 29
γ-rays microscope, 295
Gaussian observable, 39, 40, 103, 229, 276, 282
GDST(= general dynamical system theory), 12
general system (Definition), 59
Glauber-Sudarshan representation, 231
GNS-construction, 19

Hamilton equation, 55
Hamiltonian, 55, 56, 78, 244
Heisenberg kinetic equation, 57, 245
Heisenberg picture, 66
Hilbert space, 18
HMT (= hybrid measurement theory), 180
homomorphism, C^*-homomorphism, 58
homomorphism, W^*-homomorphism, 237
hybrid measurement theory, 180
hybrid state, 180
hysteresis, 65

idea, 22, 29
idealism, 9

identity, 16, 20, 227
image observable (Definition), 23
implication, 156
inference, 150
inference interval, 108
information compression, 272
information quantity, 204
invariant state, 79
involution, 16
iterated measurement (=simultaneous measurement), 31
iteration (= loop), 272
iterative likelihood function method, 121

χ^2-observable, 44
kinetic energy, 56, 244
Kolmogorov's extension theorem, 262

label set (=measured value set), 22
law of large numbers, 45, 90, 281
likelihood ratio test, 111
$L^p(\Omega, \mu)$,$L^1(\Omega, \mu)$, $L^\infty(\Omega, \mu)$,$L^1_{+1}(\Omega, \mu)$, 227
logic (practical logic), 149
logic (pure logic = mathematical logic), 149
loss function , 277

$\mathcal{M}(\Omega)$,$\mathcal{M}^p_{+1}(\Omega)$, $\mathcal{M}^m_{+1}(\Omega)$, 17
Mandelbrot set, 272
marginal observable, 23
Markov operator, 58
Markov relation among systems (Definition 3.3) , 59
matter, 29
measurability theorem, 67
$\mathbf{M}_{\mathcal{A}}\big(\mathbf{O}, S_{[\rho^p]}\big)$, 29
$\mathfrak{M}(\{\mathbf{O}_t\}_{t\in T}, \mathbf{S}_{[\rho^p_{t_0}]})$, 60
$\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[*]})$, 94
$\mathbf{M}_{C(\Omega)}\big(\mathbf{O}, S_{[*]}((kG))_{lq}\big)$, 123
$\mathbf{M}_{\mathcal{A}}(\mathbf{O}, S_{[*]}(\rho^m)\ )$, 178
$\mathfrak{M}(\{\mathbf{O}_t\}_{t\in T},\ S(\rho^m_0))$, 194
$\mathbf{M}_{C(\Omega)}\big(\mathbf{O}, S_{[*]}((\nu))_{bw}\big)$, 213
measured value setq 柩 abel set), 22
measurement, 29
measurement error (Definition), 275
measurement error model in PMT, 117
measurement error model in SMT, 252
mechanical world view, 12
membership function, 169
metaphysics, 9
metric (or distance), 47
mixed state, 17
moment method (generalized moment method), 48, 248, 249
Monty Hall problem, 105, 107, 188, 289
MT(= measurement theory), 2
MT tree, 323
multiple Markov properties, 65

Neyman-Pearson theorem, 112
no information quantity, 124
normal state (= density state), 226
normalized invariant measure, 79
null hypothesis, 111
null observable, 42

observable, 29
observable (Definition), 22
observable in W^*-algebra, 228

operator algebra, 218
OR (= operation research), 29

$\times_{k=1}^{n}$, $\prod_{k=1}^{n}$, 24
paradox (de Broglie paradox), 51
paradox (Schrödinger's cat paradox), 32
paradox(EPR-paradox), 50, 52
paradox(Monty Hall problem), 105, 107, 188, 289
paradox(the problem of three prisoners), 190, 291
parallel measurement, 45
parameter, 92
parent map, 60
parimutuel betting, 210, 218
phase space (= state space), 56, 244
Plank constant, 56
Plato, 9
PMT(= pure measurement theory), 3
$\mathrm{PMT}_{\mathrm{PEW}}$, 196
point measure, 18
population, 92
posttest state, 98
potential energy, 56, 244
$\mathcal{P}(X)$ $(=2^X)$, 21
pretest state, 98
principal components analysis, 255
principle of equal a priori probability, 79
principle of equal probability, 208, 288
principle of equal weight (= Bayes' postulate), 214
probabilistic interpretation of statistical mechanics, 79
probabilistic measurement, 177
probabilistic state, 177
probability space, 26
Proclaim 1 (the probabilistic interpretation of mixed states), 179
Proclaim$^{W^*}$1 [Statistical measurements in the W^*-algebraic formulation, 233
Proclaim$^{W^*}$2 [The Markov relation among systems, the Heisenberg picture], 241
pure state, 17

quantity in C^*-algebra (= system theoretical quantity), 29
quantity in W^*-algebra, 229
quantity(momentum quantity), 230
quantity(position quantity), 230
quantum computer, 293
quantum crytography, 293
quantum mechanics, 2
quantum teleportation, 293
quantum Zeno effect, 198
quasi-product observable (Definition), 23

Radon-Nikodým derivative, 94
random measurement, 282
random number, 271
random observable, 282
random walk, 68
realism, 9
realization of $\mathfrak{M}(\{\mathbf{O}_t\}_{t\in T}, \mathbf{S}_{[\rho^p_{t_0}]})$, 64
realization of observable in tensor space, 301
realization of sequential observable, 61, 63, 195
regression analysis I, 131

regression analysis II, 145
repeated measurement, 45
resolution of the identity, 26
ring, 21
root, 59
round error, 40

S-state (= Schrödinger state), 138
$\mathfrak{S}^p(\mathcal{A}^*)$, $\mathfrak{S}^m(\mathcal{A}^*)$, 17
sample space, 31
Schrödinger equation, 57, 244
Schrödinger's cat paradox, 32
Schrödinger picture, 66
self-similar figure, 272
semi-distance, 47
sequential observable (Definition), 59
$\sharp[\,\cdot\,]$, 45
simultaneous measurement (=iterated measurement), 31
singlet state, 50, 52
SMT(=statistical measurement theory), 3
SMT_{PEP}-method, 208, 218, 280, 288
spin observable, 33
state, 17, 29
state space (= phase space), 17
state space observable, 268
statistic, 277
statistical mechanics, 75
statistical stat, 175
staying time interpretation of statistical mechanics, 79
Stern and Gerlach's experiment, 33
stochastic differential equation (= stochastic state equation) , 2
string theory, 4
structure(series structure), 60
structure(parallel structure), 61
subjective Bayesian statistics, 208
syllogism (practical syllogism), 162
syllogism (standard syllogism), 168

$Tr(V)$,$Tr^p_{+1}(V)$,$Tr^m_{+1}(V)$, 228
tensor C^*-algebras, 20
tensor Hilbert space, 21, 50, 52, 301
tensor observable, 301
testing statistical hypothesis, 111
theoretically true or false, 7
thermodynamical condition, 81
three prisoners problem, 190, 291
TOE (Theory of Everything), 4
trajectory (Definition), 265
tree, 59

uncertainty relation (approximate simultaneous uncertainty relation), 312
uncertainty relation (Heisenberg's uncertainty relation), 294, 311
uncertainty relation (Robertson's uncertainty relation), 306
unintentional random measurement, 288
unital C^*-algebra, 16
urn problem, 35, 92, 99, 100, 109, 124, 176, 205, 214, 278

W^*-algebra, 226

Mathematical Foundations of Measurement Theory

2006年12月1日　初版第1刷発行

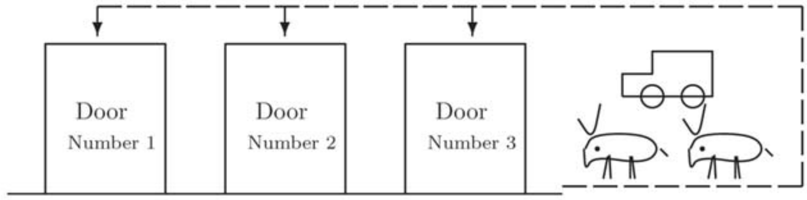
著者／発行者 — 石川史郎（Shiro Ishikawa）
制作・発売 —— 慶應義塾大学出版会株式会社
〒108-8346　東京都港区三田2-19-30
TEL〔編集部〕03-3451-0931
〔営業部〕03-3451-3584〈ご注文〉
〃　03-3451-6926
FAX〔営業部〕03-3451-3122
振替　00190-8-155497
URL　http://www.keio-up.co.jp/
印刷・製本 —— 株式会社太平印刷社

Printed in Japan　ISBN978-4-7664-1398-4

MATHEMATICAL FOUNDATIONS OF MEASUREMENT THEORY

「測定理論の数学的基礎」

[Monty Hall problem]. Suppose you are on a game show, and you are given the choice of three doors (i.e., "number 1", "number 2", "number 3"). Behind one door is a car, behind the others, goats. You pick a door, say number 1, and the host, who knows what's behind the doors, opens another door, say "number 3", which has a goat. He says to you, "Do you want to pick door number 2?" Is it to your advantage to switch your choice of doors?



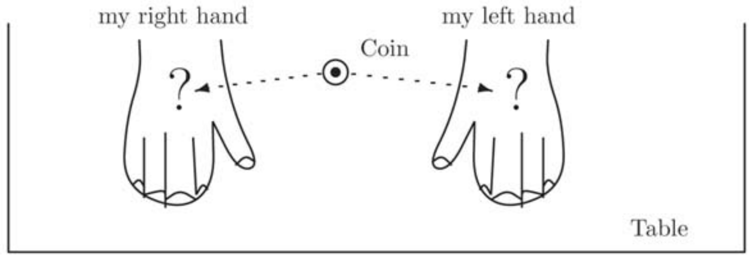
Four answers (Problem 5.12, Remark 5.13, Problem 8.8, Problem 11.13) are presented in this book.

The following old statement

[♯] *Since Socrates is a man and all men are mortal, it follows that Socrates is mortal,*

is, of course, famous. However, we have the question: "Is the syllogism [♯] true or not?"
Or, can you prove it? (See Theorem 7.19)

A coin is, at random, put under my right hand or my left hand. Suppose that you do not know which hand the coin is under, and you choose one of my hands which you guess that the coin is under. Then, the probability that the ball is under the hand you choose is, of course, equal to 1/2. Next, consider the case that the condition: "at random" is not assumed in this problem. How do you think about this case?



(See Problem 11.10)

[The problem concerning EPR-experiment]. Let A and B be particles with the same masses m. Consider the situation described in the following figure:

where "the velocity of A" = −"the velocity of B".

The position q_A (at time t_0) of the particle A can be exactly measured, and moreover, the velocity of v_B (at time t_0) of the particle B can be exactly measured. Thus, we can conclude that the position and momentum (at time t_0) of the particle A are respectively equal to q_A and $-mv_B$. Is this fact contradictory to Heisenberg's uncertainty relation?

(See §12.7)

Keio University Press Inc.