*Applied Mathematics*, 2012, 3, 283-292
http://dx.doi.org/10.4236/am.2012.33044 Published Online March 2012 (http://www.SciRP.org/journal/am)

# A Measurement Theoretical Foundation of Statistics

**Shiro Ishikawa**
Department of Mathematics, Faculty of Science and Technology,
Keio University, Yokohama, Japan
Email: ishikawa@math.keio.ac.jp

Received January 6, 2012; revised February 14, 2012; accepted February 21, 2012

## ABSTRACT

It is a matter of course that Kolmogorov's probability theory is a very useful mathematical tool for the analysis of statistics. However, this fact never means that statistics is based on Kolmogorov's probability theory, since it is not guaranteed that mathematics and our world are connected. In order that mathematics asserts some statements concerning our world, a certain theory (so called "world view") mediates between mathematics and our world. Recently we propose measurement theory (*i.e.*, the theory of the quantum mechanical world view), which is characterized as the linguistic turn of quantum mechanics. In this paper, we assert that statistics is based on measurement theory. And, for example, we show, from the pure theoretical point of view (*i.e.*, from the measurement theoretical point of view), that regression analysis can not be justified without Bayes' theorem. This may imply that even the conventional classification of (Fisher's) statistics and Bayesian statistics should be reconsidered.

**Keywords:** The Copenhagen Interpretation; Operator Algebra; Quantum and Classical Measurement Theory; Fisher Maximum Likelihood Method; Regression Analysis; Philosophy of Statistics

## 1. Introduction

For example, consider Newtonian mechanics. It is natural to understand that Newton mechanics is based on Newton's three laws of motion, though the mathematical theory of differential equations is a useful tool for the analysis of Newtonian mechanics. That is because any mathematical theory is a closed logical system derived from set theory, and thus, it is not qualified to assert statements concerning our world without laws. If it is so, and, if Kolmogorov's probability theory [1] is a mathematical theory, we think that the foundation of statistics does not yet established. Thus, the following problem is natural:

(A) What kind of law is statistics based on? Or, propose a foundation of statistics!

The purpose of this paper is to answer this problem.

Although in a series of our research [2-8] we have been concerned with this problem (A), in this paper we give a decisive answer to the problem (A) in the light of our final version [7,8] of measurement theory. Here, as mentioned in Section 2 later, measurement theory (*i.e.*, the theory of the quantum mechanical world view) is characterized as the linguistic turn of quantum mechanics. Hence, note that measurement theory is not physics but a kind of language, and thus, the "law" in (A) is called "axiom" in this paper.

## 2. Measurement Theory (Axioms and Interpretation)

### 2.1. Mathematical Preparations

In this section, we prepare mathematics, which is used in measurement theory (or in short, MT).

Measurement theory ([2-8]) is, by an analogy of quantum mechanics (or, as a linguistic turn of quantum mechanics), constructed as the scientific theory formulated in a certain $C^*$-algebra $\mathcal{A}$ (*i.e.*, a norm closed subalgebra in the operator algebra $B(H)$ composed of all bounded operators on a Hilbert space *H*, cf. [9,10]). MT is composed of two theories (*i.e.*, pure measurement theory (or, in short, PMT] and statistical measurement theory (or, in short, SMT). That is, we see:

(B) MT (measurement theory)

$$
=\begin{cases}
(\mathrm{B}_1):[\mathrm{PMT}] \\
\qquad =\Big[\underset{(\mathrm{Axiom}^{\mathrm{P}}\ 1)}{(\text{pure})\,\text{measurement}}\Big]+\Big[\underset{(\text{Axiom } 2)}{\text{causality}}\Big] \\
(\mathrm{B}_2):[\mathrm{SMT}] \\
\qquad =\Big[\underset{(\mathrm{Axiom}^{\mathrm{S}}\ 1)}{(\text{statistical})\,\text{measurement}}\Big]+\Big[\underset{(\text{Axiom } 2)}{\text{causality}}\Big]
\end{cases}
$$

where Axiom 2 is common in PMT and SMT. For completeness, note that measurement theory (B) (*i.e.*, ($B_1$) and ($B_2$)) is a kind of language based on the quantum mechanical world view, (*cf.* [8]). It may be understandable to consider that

(C) PMT and SMT is related to Fisher's statistics and Bayesian statistics respectively.

Also, as mentioned in Section 2.6 latter, our concern in this paper is to give an answer to the question "Which is fundamental, PMT or SMT?".

When $\mathcal{A} = B_c(H)$, the $C^*$-algebra composed of all compact operators on a Hilbert space *H*, the (B) is called quantum measurement theory (or, quantum system theory), which can be regarded as the linguistic aspect of quantum mechanics. Also, when $\mathcal{A}$ is commutative (that is, when $\mathcal{A}$ is characterized by $C_0(\Omega)$, the $C^*$-algebra composed of all continuous complex-valued functions vanishing at infinity on a locally compact Hausdorff space $\Omega$ (*cf.* [9])), the (B) is called classical measurement theory. Thus, we have the following classification:

$$\text{(D)}\quad MT\begin{cases}\text{quantum MT}\left(\text{when }\mathcal{A} = B_c(H)\right)\\ \text{classical MT}\left(\text{when }\mathcal{A} = C_0(\Omega)\right)\end{cases}$$

In this paper, we mainly devote ourselves to classical MT (*i.e.*, classical PMT and classical SMT).

Now we shall explain the measurement theory (B). Let $\mathcal{A}(\subseteq B(H))$ be a $C^*$-algebra, and let $\mathcal{A}^*$ be the dual Banach space of $\mathcal{A}$. That is, $\mathcal{A}^* = \{\rho \mid \rho$ is a continuous linear functional on $\mathcal{A}\}$, and the norm $\|\rho\|_{\mathcal{A}^*}$ is defined by

$$\sup\left\{|\rho(F)| : F \in \mathcal{A} \text{ such that } \|F\|_{\mathcal{A}}\left(=\|F\|_{B(H)}\right) \le 1\right\}.$$

The bi-linear functional $\rho(F)$ is also denoted by ${}_{\mathcal{A}^*}\langle\rho, F\rangle_{\mathcal{A}}$, or in short $\langle\rho, F\rangle$. Define the mixed state $\rho(\in \mathcal{A}^*)$ such that $\|\rho\|_{\mathcal{A}^*} = 1$ and $\rho(F) \ge 0$ for all $F \in \mathcal{A}$ satisfying $F \ge 0$. And put

$$\mathfrak{S}^m(\mathcal{A}^*) = \left\{\rho \in \mathcal{A}^* \mid \rho \text{ is a mixed state}\right\}.$$

A mixed state $\rho\left(\in \mathfrak{S}^m(\mathcal{A}^*)\right)$ 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 $\rho = \rho_1 = \rho_2$. Put

$$\mathfrak{S}^p(\mathcal{A}^*) = \left\{\rho \in \mathfrak{S}^m(\mathcal{A}^*) \mid \rho \text{ is a pure state}\right\},$$

which is called a *state space*. The Riese theorem (*cf.* [11]) says that

$$C_0(\Omega)^* = M(\Omega) = \left\{\rho \mid \rho \text{ is a signed measure on } \Omega\right\},$$

$$\mathfrak{S}^m\left(C_0(\Omega)^*\right) = M_{+1}^m(\Omega)$$
$$= \left\{\rho \mid \rho \text{ is a measure on } \Omega \text{ such that } \rho(\Omega) = 1\right\}.$$

Also, it is well known (*cf.* [9]) that

$$\mathfrak{S}^p\left(B_c(H)^*\right) = \left\{|u\rangle\langle u| \left(\textit{i.e.}, \text{ the Dirac notation}\right) \,\middle|\, \|u\|_H = 1\right\},$$

and

$$\mathfrak{S}^p\left(C_0(\Omega)^*\right) = M_{+1}^p(\Omega)$$
$$= \left\{\delta_{\omega_0} \,\middle|\, \delta_{\omega_0} \text{ is a point measure at } \omega_0 \in \Omega\right\},$$

where $\int_\Omega f(\omega)\delta_{\omega_0}(\mathrm{d}\omega) = f(\omega_0)\left(\forall f \in C_0(\Omega)\right)$. The latter implies that $\mathfrak{S}^p\left(C_0(\Omega)^*\right)$ can be also identified with $\Omega$ (called a spectrum space or maximal ideal space) such as

$$\underset{\text{(state space)}}{\mathfrak{S}^p\left(C_0(\Omega)^*\right)} \ni \delta_\omega \leftrightarrow \omega \in \underset{\text{(spectrum space)}}{\Omega}$$

Here, assume that the $C^*$-algebra $\mathcal{A}(\subseteq B(H))$ is unital, *i.e.*, it has the identity *I*. This assumption is not unnatural, since, if $I \notin \mathcal{A}$, it suffices to reconstruct the $\mathcal{A}$ such that it includes $\mathcal{A} \cup \{I\}$.

According to the noted idea (*cf.* [12]) in quantum mechanics, an observable $O \equiv (X, \mathcal{F}, F)$ in $\mathcal{A}$ is defined as follows:

($E_1$) [Field] *X* is a set, $\mathcal{F}$ ($\subseteq 2^X$, the power set of *X*) is a field of *X*, that is, "$\Xi_1, \Xi_2 \in \mathcal{F} \Rightarrow \Xi_1 \cup \Xi_2 \in \mathcal{F}$", "$\Xi \in \mathcal{F} \Rightarrow X \setminus \Xi \in \mathcal{F}$".

($E_2$) [Countably additivity] *F* is a mapping from $\mathcal{F}$ to $\mathcal{A}$ satisfying: 1) for every $\Xi \in \mathcal{F}$, $F(\Xi)$ is a non-negative element in $\mathcal{A}$ such that $0 \le F(\Xi) \le I$, 2) $F(\emptyset) = 0$ and $F(X) = I$, where 0 and *I* is the 0-element and the identity in $\mathcal{A}$ respectively. 3): for any countable decomposition $\{\Xi_1, \Xi_2, \cdots\}$ of $\Xi \in \mathcal{F}$ (*i.e.*, $\Xi_k, \Xi \in \mathcal{F}$ such that $\bigcup_{k=1}^\infty \Xi_k = \Xi$, $\Xi_i \cap \Xi_j = \emptyset (i \ne j)$), it holds that

$$\lim_{K\to\infty}\rho\left(F\left(\bigcup_{k=1}^K \Xi_k\right)\right) = \rho\left(F(\Xi)\right) \quad \left(\forall \rho \in \mathfrak{S}^m(\mathcal{A}^*)\right) \tag{1}$$

(*i.e.*, in the sense of weak convergence).

*Remark* 1. By the Hopf extension theorem (*cf.* [11]), we have the mathematical probability space ($X$, $\overline{\mathcal{F}}$, $\rho^m(F(\cdot))$) where $\overline{\mathcal{F}}$ is the smallest $\sigma$-field such that $\mathcal{F} \subseteq \overline{\mathcal{F}}$. For the other formulation (*i.e.*, $W^*$-algebraic formulation), see the appendix in [7].

## 2.2. Pure Measurement Theory in ($B_1$)

In what follows, we shall explain PMT in ($B_1$).

With any *system S*, a $C^*$-algebra $\mathcal{A}(\subseteq B(H))$ can be associated in which the pure measurement theory ($B_1$) of that system can be formulated. A *state* of the system *S* is represented by an element $\rho\left(\in \mathfrak{S}^p(\mathcal{A}^*)\right)$ and an *observable* is represented by an observable $O = (X, \mathcal{F}, F)$ in $\mathcal{A}$. Also, the *measurement of the observable O for* *the system S with the state* $\rho$ is denoted by $M_{\mathcal{A}}\left(O, S_{[\rho]}\right)$ (or more precisely, $M_A\left(O=\left(X, \mathcal{F}, F\right), S_{[\rho]}\right)$). An observer can obtain a measured value $x(\in X)$ by the measurement $M_A\left(O, S_{[\rho]}\right)$.

The Axiom[P] 1 presented below is a kind of mathematical generalization of Born's probabilistic interpretation of quantum mechanics. And thus, it is a statement without reality.

**Axiom[P] 1.** [Pure Measurement]. *The probability that a measured value* $x(\in X)$ *obtained by the measurement* $M_A\left(O \equiv\left(X, \mathcal{F}, F\right), S_{[\rho_0]}\right)$ *belongs to a set* $\Xi(\in \mathcal{F})$ *is given by* $\rho_0\left(F(\Xi)\right)$.

Next, we explain Axiom 2 in (B). Let $(T, \leq)$ be a tree, *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 $t_2 \leq t_1$. In this paper, we assume that *T* is finite (*cf.* Remark 9 in Section 7 later). Assume that there exists an element $t_0 \in T$, called the *root* of *T*, such that $t_0 \leq t$ ($\forall t \in T$) holds. Put $T_{\leq}^2=\left\{\left(t_1, t_2\right) \in T^2 \mid t_1 \leq t_2\right\}$. The family $\left\{\Phi_{t_1, t_2}: \mathcal{A}_{t_2} \rightarrow \mathcal{A}_{t_1}\right\}_{\left(t_1, t_2\right) \in T_{\leq}^2}$ is called a *causal relation* (*due to the Heisenberg picture*), if it satisfies the following conditions ($F_1$) and ($F_2$).

($F_1$) With each $t \in T$, a $C^*$-algebra $\mathcal{A}_t$ is associated.

($F_2$) For every $\left(t_1, t_2\right) \in T_{\leq}^2$, a Markov operator $\Phi_{t_1, t_2}: \mathcal{A}_{t_2} \rightarrow \mathcal{A}_{t_1}$ is defined (*i.e.*, $\Phi_{t_1, t_2} \geq 0$, $\Phi_{t_1, t_2}\left(I_{\mathcal{A}_{t_2}}\right)=I_{\mathcal{A}_{t_1}}$). And it satisfies that $\Phi_{t_1, t_2} \Phi_{t_2, t_3}=\Phi_{t_1, t_3}$ holds for any $\left(t_1, t_2\right)$, $\left(t_2, t_3\right) \in T_{\leq}^2$.

The family of dual operators $\left\{\Phi_{t_1, t_2}^*: \mathfrak{S}^m\left(\mathcal{A}_{t_1}^*\right) \rightarrow \mathfrak{S}^m\left(\mathcal{A}_{t_2}^*\right)\right\}_{\left(t_1, t_2\right) \in T_{\leq}^2}$ is called a *dual causal relation* (*due to the Schrödinger picture*). When $\Phi_{t_1, t_2}^*\left(\mathfrak{S}^p\left(\mathcal{A}_{t_1}^*\right)\right) \subseteq\left(\mathfrak{S}^p\left(\mathcal{A}_{t_2}^*\right)\right)$ holds for any $\left(t_1, t_2\right) \in T_{\leq}^2$, the causal relation is said to be deterministic.

Now Axiom 2 in the measurement theory (B) is presented as follows:

**Axiom 2.** [Causality]. *The causality is represented by a causal relation* $\left\{\Phi_{t_1, t_2}: \mathcal{A}_{t_2} \rightarrow \mathcal{A}_{t_1}\right\}_{\left(t_1, t_2\right) \in T_{\leq}^2}$.

## 2.3. Interpretation

Next, we have to study how to use the above axioms as follows. That is, we present the following interpretation (*G*) [= ($G_1$) – ($G_3$)], which is characterized as a kind of linguistic turn of so-called Copenhagen interpretation (cf. [7,8]). That is, we propose:

($G_1$) Consider the dualism composed of observer and system (= measuring object). And therefore, observer and system must be absolutely separated.

($G_2$) Only one measurement is permitted. And thus, the state after a measurement is meaningless since it can not be measured any longer. Also, the causality should be assumed only in the side of system, however, a state never moves. Thus, the Heisenberg picture should be adopted, and thus, the Schrödinger picture should be prohibited.

($G_3$) Also, the observer does not have the space-time. Thus, the question: "When and where is a measured value obtained?" is out of measurement theory. And thus, Schrödinger's cat is out of measurement theory, and so on.

## 2.4. Sequential Causal Observable and Its Realization

For each $k=1,2, \cdots, K$, consider a measurement $M_{\mathcal{A}}\left(O_k \equiv\left(X_k, \mathcal{F}_k, F_k\right), S_{[\rho]}\right)$. However, since the ($G_2$) says that only one measurement is permitted, the measurements $\left\{M_{\mathcal{A}}\left(O_k, S_{[\rho]}\right)\right\}_{k=1}^K$ should be reconsidered in what follows. Under the commutativity condition such that

$$F_i\left(\Xi_i\right) F_j\left(\Xi_j\right)=F_j\left(\Xi_j\right) F_i\left(\Xi_i\right) \quad \left(\forall \Xi_i \in \mathcal{F}_i, \forall \Xi_j \in \mathcal{F}_j, i \neq j\right), \tag{2}$$

we can define the product observable $\times_{k=1}^K O_k=\left(\times_{k=1}^K X_k, \boxtimes_{k=1}^K F_k, \times_{k=1}^K F_k\right)$ in $\mathcal{A}$ such that

$$\left(\times_{k=1}^K F_k\right)\left(\times_{k=1}^K \Xi_k\right)=F_1\left(\Xi_1\right) F_2\left(\Xi_2\right) \cdots F_K\left(\Xi_K\right) \quad \left(\forall \Xi_k \in \mathcal{F}_k, \forall k=1, \cdots, K\right).$$

Here, $\boxtimes_{k=1}^K F_k$ is the smallest field including the family $\left\{\times_{k=1}^K \Xi_k: \Xi_k \in \mathcal{F}_k\ k=1,2, \cdots, K\right\}$. Then, the above $\left\{M_{\mathcal{A}}\left(O_k, S_{[\rho]}\right)\right\}_{k=1}^K$ is, under the commutativity condition (2), represented by the simultaneous measurement $M_A\left(\times_{k=1}^K O_k, S_{[\rho]}\right)$.

Consider a tree $\left(T \equiv\left\{t_0, t_1, \cdots, t_n\right\}, \leq\right)$ with the root $t_0$. This is also characterized by the map $\pi: T \backslash\left\{t_0\right\} \rightarrow T$ such that $\pi(t)=\max \{s \in T \mid s<t\}$. Let $\left\{\Phi_{t, t'}: \mathcal{A}_{t'} \rightarrow \mathcal{A}_t\right\}_{(t, t') \in T_{\leq}^2}$ be a causal relation, which is also represented by $\left\{\Phi_{\pi(t), t}: \mathcal{A}_t \rightarrow \mathcal{A}_{\pi(t)}\right\}_{t \in T \backslash\left\{t_0\right\}}$. Let an observable $O_t \equiv\left(X_t, \mathcal{F}_t, F_t\right)$ in the $\mathcal{A}_t$ be given for each $t \in T$. Note that $\Phi_{\pi(t), t} O_t\left(\equiv\left(X_t, \mathcal{F}_t, \Phi_{\pi(t), t} F_t\right)\right)$ is an observable in the $\mathcal{A}_{\pi(t)}$.

The pair $\left[\mathbb{O}_T\right]=\left[\left[O_t\right]_{t \in T}\left\{\Phi_{t, t'}: \mathcal{A}_{t'} \rightarrow \mathcal{A}_t\right\}_{(t, t') \in T_{\leq}^2}\right]$, is called a *sequential causal observable*. For each $s \in T$, put $T_s=\{t \in T \mid t \geq s\}$. And define the observable $\hat{O}_s \equiv\left(\times_{t \in T_s} X_t, \boxtimes_{t \in T_s} \mathcal{F}_t, \hat{F}_s\right)$ in $\mathcal{A}_s$ as follows:

$$\hat{O}_s = \begin{cases} O_s & \text{if } s \in T \setminus \pi(T) \\ O_s \times \left( \times_{t \in \pi^{-1}(\{s\})} \Phi_{\pi(t),t} \hat{O}_t \right) & \text{if } s \in \pi(T) \end{cases} \quad (3)$$

if the commutativity condition holds (*i.e.*, if the product observable $O_s \times \left( \times_{t \in \pi^{-1}(\{s\})} \Phi_{\pi(t),t} \hat{O}_t \right)$ exists) for each $s \in \pi(T)$. Using (3) iteratively, we can finally obtain the observable $\hat{O}_{t_0}$ in $\mathcal{A}_{t_0}$. The $\hat{O}_{t_0}$ is called the realization (or, realized causal observable) of $[\mathbb{O}_T]$.

### 2.5. Statistical Measurement Theory in (B$_2$)

We shall introduce the following notation: it is usual to consider that we do not know the pure state $\rho_0^p \left( \in \mathfrak{S}^p \left( \mathcal{A}^* \right) \right)$ when we take a measurement $M_{\mathcal{A}}\left( O, S_{\left[\rho_0^p\right]} \right)$. That is because we usually take a measurement $M_{\mathcal{A}}\left( O, S_{\left[\rho_0^p\right]} \right)$ in order to know the state $\rho_0^p$. Thus, when we want to emphasize that we do not know the state $\rho_0^p$, $M_{\mathcal{A}}\left( O, S_{\left[\rho_0^p\right]} \right)$ is denoted by $M_{\mathcal{A}}\left( O, S_{[*]} \right)$. Also, when we know the distribution $\rho_0^m \left( \in \mathfrak{S}^m \left( \mathcal{A}^* \right) \right)$ of the unknown state $\rho_0^p$, the $M_{\mathcal{A}}\left( O, S_{\left[\rho_0^p\right]} \right)$ is denoted by $M_{\mathcal{A}}\left( O, S_{[*]}\left( \left\{ \rho_0^m \right\} \right) \right)$. The $\rho_0^m$ is called a mixed state. And further, if we know that a mixed state $\rho_0^m$ belongs to a compact set $K \left( \subseteq \mathfrak{S}^m \left( \mathcal{A}^* \right) \right)$, the $M_{\mathcal{A}}\left( O, S_{\left[\rho_0^p\right]} \right)$ is denoted by $M_{\mathcal{A}}\left( O, S_{[*]}(K) \right)$.

The Axiom$^{\rm S}$ 1 presented below is a kind of mathematical generalization of Axiom$^{\rm P}$ 1.

**Axiom$^{\rm S}$ 1.** [Statistical measurement]. *The probability that a measured value* $x (\in X)$ *obtained by the measurement* $M_{\mathcal{A}}\left( O \equiv (X, \mathcal{F}, F), S_{[*]}\left( \left\{ \rho_0^m \right\} \right) \right)$ *belongs to a set* $\Xi (\in \mathcal{F})$ *is given by* $\rho_0^m \left( F(\Xi) \right) \left( = {}_{\mathcal{A}^*}\left\langle \rho_0^m, F(\Xi) \right\rangle_{\mathcal{A}} \right)$.

Thus, we can propose the statistical measurement theory (B$_2$), in which Axiom 2 and Interpretation (G) are common.

Let $\hat{O}\left( X \times Y, \mathcal{F} \boxtimes \mathcal{G}, H \right)$ be an observable in a $C^*$-algebra $\mathcal{A}$. Assume that we know that the measured value $(x, y) (\in X \times Y)$ obtained by a statistical measurement $M_{\mathcal{A}}\left( \hat{O}, S_{[*]}\left( \left\{ \rho_0^m \right\} \right) \right)$ belongs to $\Xi \times Y \left( \in \mathcal{F} \boxtimes \mathcal{G} \right)$. Then, there is a reason to infer that the unknown measured value $y (\in Y)$ is distributed under the conditional probability $P_{\Xi}\left( G(\Gamma) \right)$, where

$$P_{\Xi}\left( G(\Gamma) \right) = \frac{{}_{\mathcal{A}^*}\left\langle \rho_0^m, H\left( \Xi \times \Gamma \right) \right\rangle_{\mathcal{A}}}{{}_{\mathcal{A}^*}\left\langle \rho_0^m, H\left( \Xi \times Y \right) \right\rangle_{\mathcal{A}}} \quad \left( \forall \Gamma \in \mathcal{G} \right) \qquad (4)$$

Thus, by a hint of Fisher's maximum likelihood method, we have the following theorem, which is the most fundamental in this paper.

**Theorem 1.** [Fisher's maximum likelihood method in general $\mathcal{A}$]. *Let* $\hat{O}\left( X \times Y, \mathcal{F} \boxtimes \mathcal{G}, H \right)$ *be an observable in a* $C^*$*-algebra* $\mathcal{A}$. *Let* $K \left( \subseteq \mathfrak{S}^m \left( \mathcal{A}^* \right) \right)$ *be a compact set. Assume that we know that the measured value* $(x, y) (\in X \times Y)$ *obtained by a measurement* $M_{\mathcal{A}}\left( \hat{O}, S_{[*]}(K) \right)$ *belongs to* $\Xi \times Y \left( \in \mathcal{F} \boxtimes \mathcal{G} \right)$. *Then, there is a reason to infer that the unknown measured value* $y (\in Y)$ *is distributed under the conditional probability* $P_{\Xi}\left( G(\Gamma) \right)$, *where*

$$P_{\Xi}\left( G(\Gamma) \right) = \frac{{}_{\mathcal{A}^*}\left\langle \rho_0^m, H\left( \Xi \times \Gamma \right) \right\rangle_{\mathcal{A}}}{{}_{\mathcal{A}^*}\left\langle \rho_0^m, H(\Xi \times Y) \right\rangle_{\mathcal{A}}} \quad \left( \forall \Gamma \in \mathcal{G} \right). \qquad (5)$$

Here, $\rho_0^m \left( \in K \subseteq \mathfrak{S}^m \left( \mathcal{A}^* \right) \right)$ is defined by

$${}_{\mathcal{A}^*}\left\langle \rho_0^m, H\left( \Xi \times Y \right) \right\rangle_{\mathcal{A}} = \max_{\rho^m \in K_{\mathcal{A}^*}} \left\langle \rho^m, H\left( \Xi \times Y \right) \right\rangle_{\mathcal{A}}.$$

*Remark* 2. Theorem 1 is new throughout our research [2-8], though, in a particular case that $K \subseteq \mathfrak{S}^p \left( \mathcal{A}^* \right)$, Theorem 1 was proposed in [7] where we devoted ourselves to PMT.

### 2.6. Our Concern in This Paper

Note that

(H$_1$) $M_{A}\left( O, S_{\left[\rho_0^p\right]} \right) = M_{A}\left( O, S_{[*]}\left( \left\{ \rho_0^p \right\} \right) \right)$ for $\rho_0^p \in \mathfrak{S}^p \left( \mathcal{A}^* \right)$, therefore, we see that [PMT] $\subset$ [SMT].

However, we have the following problem:

(H$_2$) Which is fundamental, PMT or SMT?

Recalling the (C), most readers may consider that PMT is more fundamental than SMT. In fact, throughout our research [2-8], we have believed in the fundamentality of PMT. However, in this paper, we assert that Theorem 1 in SMT is the most fundamental as far as inference. In fact, every result in this paper is regarded as one of the corollaries of Theorem 1. And hence, we shall conclude that SMT is proper as the answer to the problem (A). Also, our proposal has a merit such that the philosophy of statistics is naturally induced by the philosophy of measurement theory (*cf*. [8]).

## 3. Fisher-Bayes Method in Classical $C(\Omega)$

### 3.1. Notations

We shall devote ourselves to classical case (*i.e.*, $\mathcal{A} = C_0(\Omega)$ ). From here, $C_0(\Omega)$ (or, commutative unital $C^*$-algebra that includes $C_0(\Omega)$) is, for simplicity, denoted by $C_0(\Omega)$. Thus, we put

$$\mathcal{A}^* = C(\Omega)^* = \mathcal{M}(\Omega),$$

$$\mathfrak{S}^m(\mathcal{A}^*) = \mathfrak{S}^m(C(\Omega)^*) = \mathcal{M}_{+1}^m(\Omega),$$

and

$$\mathfrak{S}^p(\mathcal{A}^*) = \mathfrak{S}^p(C(\Omega)^*) = \mathcal{M}_{+1}^p(\Omega) \approx \Omega .$$

And, for any mixed state $\nu \in \mathcal{M}_{+1}^m(\Omega)$ and any observable $O \equiv (X, \mathcal{F}, F)$ in $C(\Omega)$, we put:

$$\begin{aligned}\nu(F(\Xi)) &= {}_{C(\Omega)^*}\langle \nu, F(\Xi)\rangle_{C(\Omega)} = {}_{\mathcal{M}(\Omega)}\langle \nu, F(\Xi)\rangle_{C(\Omega)} \\ &= \int_\Omega [F(\Xi)](\omega)\nu(\mathrm{d}\omega).\end{aligned} \quad (6)$$

Also, put $\nu(D) = \int_D \nu(\mathrm{d}\omega)$ ($\forall D \in \mathcal{B}_\Omega$: Borel $\sigma$-field). In order to avoid the confusion between $\nu(F(\Xi))$ in (6) and $\nu(D)$, we do not use $\nu(F(\Xi))$. Also, for any $\delta_{\omega_0} \in \mathcal{M}_{+1}^p(\Omega) \approx \Omega$, we put:

$$\begin{aligned}&{}_{C(\Omega)^*}\langle \delta_{\omega_0}, F(\Xi)\rangle_{C(\Omega)} = {}_{\mathcal{M}(\Omega)}\langle \delta_{\omega_0}, F(\Xi)\rangle_{C(\Omega)} \\ &= \int_\Omega [F(\Xi)](\omega)\delta_{\omega_0}(\mathrm{d}\omega) = [F(\Xi)](\omega_0).\end{aligned}$$

### 3.2. Bayes Method in Classical $C(\Omega)$

Let $O_1 \equiv (X, \mathcal{F}, F)$ be an observable in a commutative $C^*$-algebra $C(\Omega)$. And let $O_2 \equiv (Y, \mathcal{G}, G)$ be any observable in $C(\Omega)$. Consider the product observable $O_1 \times O_2 \equiv (X \times Y, \mathcal{F} \boxtimes \mathcal{G}, F \times G)$ in $C(\Omega)$. The existence will be shown in Section 7 (Appendix).

Assume that we know that the measured value $(x, y)$ obtained by a simultaneous measurement $M_{C(\Omega)}(O_1 \times O_2, S_{[*]}(\{\nu_0\}))$ belongs to $\Xi \times Y (\in \mathcal{F} \boxtimes \mathcal{G})$. Then, by (4), we can infer that

(I) the probability $P_\Xi(G(\Gamma))$ that *y* belongs to $\Gamma (\in \mathcal{G})$ is given by

$$P_\Xi(G(\Gamma)) = \frac{\int_\Omega [F(\Xi)G(\Gamma)](\omega)\nu_0(\mathrm{d}\omega)}{\int_\Omega [F(\Xi)](\omega)\nu_0(\mathrm{d}\omega)} \quad (\forall \Gamma \in \mathcal{G}).$$

Thus, we can assert that:

**Theorem 2.** [Bayes method, *cf.* [4,5]]. *When we know that a measured value obtained by a measurement* $M_{C(\Omega)}(O_1 \equiv (X, \mathcal{F}, F), S_{[*]}(\{\nu_0\}))$ *belongs to* $\Xi$, *there is a reason to infer that the mixed state after the measurement is equal to* $\nu_0^a (\in \mathcal{M}_{+1}^m(\Omega))$, *where*

$$\nu_0^a(D) = \frac{\int_D [F(\Xi)](\omega)\nu_0(\mathrm{d}\omega)}{\int_\Omega [F(\Xi)](\omega)\nu_0(\mathrm{d}\omega)} \quad (\forall D \in \mathcal{B}_\Omega).$$

*Proof.* Note that we can regard that $P_\Xi \in \mathcal{M}_{+1}^m(\Omega) (\subseteq C(\Omega)^*)$. That is, there exists $\nu_0^a (\in C(\Omega)^*)$ such that

$$P_\Xi(G(\Gamma)) = \int_\Omega [G(\Gamma)](\omega)\nu_0^a(\mathrm{d}\omega) \quad (\forall \Gamma \in \mathcal{G}) \quad (7)$$

Then, Axiom[S] 1 says that the probability that a measured value $y (\in Y)$ obtained by the measurement $M_{C(\Omega)}(O_2 \equiv (Y, \mathcal{G}, G), S_{[*]}(\{\nu_0^a\}))$ belongs to a set $\Gamma (\in \mathcal{G})$ is given by $\int_\Omega [G(\Gamma)](\omega)\nu_0^a(\mathrm{d}\omega)$, which is equal to $P_\Xi(G(\Gamma))$ in (7). Since $O_2 \equiv (Y, \mathcal{G}, G)$ is arbitrary, we obtain Theorem 2.

*Remark* 3. The above (I) is, of course, fundamental. However, in the sense mentioned in the above proof, we admit Theorem 2 as the equivalent statement of the (I). That is, in spite of Interpretation ($G_2$), we admit the wavefunction collapse such as

$$\text{(J)} \quad \underset{(\in \mathcal{M}_{+1}^m(\Omega))}{\overset{\text{(pretest state)}}{\nu_0}} \xrightarrow[\text{Theorem 2}]{\text{Bayes}} \underset{(\in \mathcal{M}_{+1}^m(\Omega))}{\overset{\text{(posttest state)}}{\nu_0^a}}$$

Theorem 2 was, for the first time, proposed in [4,5] without the conscious understanding of Interpretation ($G_2$). Also, note that,

(K) in Theorem 2, if $\nu_0 = \delta_{\omega_0} (\in \mathcal{M}_{+1}^p(\Omega))$, then it clearly holds that $\nu_0^a = \delta_{\omega_0}$.

Also, for our opinion concerning the wavefunction collapse in quantum mechanics, see [7].

### 3.3. Fisher-Bayes Method in Classical $C(\Omega)$

Combining Theorem 1 (Fisher's method) and Theorem 2 (Bayes' method), we get the following corollary.

**Corollary 1.** [Fisher-Bayes method (*i.e.*, Regression analysis in a narrow sense)]. *When we know that a measured value obtained by a measurement* $M_{C(\Omega)}(O_1 \equiv (X, \mathcal{F}, F), S_{[*]}(K))$ *belongs to* $\Xi$, *there is a reason to infer that the state after the measurement is equal to* $\nu_0^a (\in \mathcal{M}_{+1}^m(\Omega))$ *such that*

$$\nu_0^a(D) = \frac{\int_D [F(\Xi)](\omega)\nu_0(\mathrm{d}\omega)}{\int_\Omega [F(\Xi)](\omega)\nu_0(\mathrm{d}\omega)} \quad (\forall D \in \mathcal{B}_\Omega)$$

where the $\nu_0 (\in K)$ is defined by

$$\int_\Omega [F(\Xi)](\omega)\nu_0(\mathrm{d}\omega) = \max_{\nu \in K} \int_\Omega [F(\Xi)](\omega)\nu(\mathrm{d}\omega).$$

*Remark* 4. As mentioned in the above, note that Corollary 1 is composed of the following two procedure:

$$\text{(L)} \quad \underset{(\subseteq \mathcal{M}_{+1}^m(\Omega))}{K} \xrightarrow[\text{Theorem 1}]{\text{Fisher}} \underset{(\in K)}{\nu_0} \xrightarrow[\text{Theorem 2}]{\text{Bayes}} \underset{\in K}{\nu_0^a}$$

### 3.4. A Simple Example of Fisher-Bayes Method (Regression Analysis in a Narrow Sense)

In this section, we examine Corollary 1 in a simple example. Readers will find that Corollary 1 can be regarded as regression analysis in a narrow sense.

We have a rectangular water tank filled with water. Assume that the height of water at time *t* is given by the following function $h(t)$:

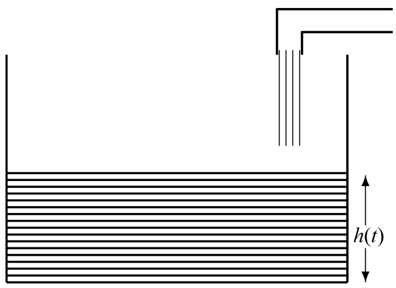
$$h(t) = \alpha_0 + \beta_0 t, \tag{8}$$

where $\alpha_0$ and $\beta_0$ are unknown fixed parameters such that $\alpha_0$ is the height of water filling the tank at the beginning and $\beta_0$ 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

$$h_m(t) = \alpha_0 + \beta_0 t + e(t), \tag{9}$$

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 $t = 0,1,2$ as follows:

$$h_m(0) = 0.5, \quad h_m(1) = 1.6, \quad h_m(2) = 3.3. \tag{10}$$




Under this setting, we shall study the following problem:

(M) [Inference]: when measured data (10) is obtained, infer the unknown parameter $(\alpha_0, \beta_0)$ in (9).

In what follows, from the measurement theoretical point of view, we shall answer the problem (M). Let $T = \{0,1,2\}$ be a series ordered set such that the parent map $\pi : T \setminus \{0\} \to T$ is defined by $\pi(t) = t-1$ $(t = 0,1,2)$. Put $\Omega_0 = [0,2]\times[0,2]$, $\Omega_1 = [0,4]\times[0,2]$, $\Omega_2 = [0,6]\times[0,2]$. For each $t = 1,2$, consider a continuous map $\phi_{\pi(t),t} : \Omega_{\pi(t)} \to \Omega_t$ such that

$$\begin{aligned} &\phi_{0,1}(\alpha,\beta) = (\alpha+\beta,\beta)\left(\forall \omega_0 = (\alpha,\beta)\in\Omega_0\right) \\ &\phi_{1,2}(\alpha,\beta) = (\alpha+\beta,\beta)\left(\forall \omega_1 = (\alpha,\beta)\in\Omega_1\right). \end{aligned} \tag{11}$$

Then, we get the deterministic causal operators hus, $\left\{\Phi_{\pi(t),t} : C(\Omega_t) \to C(\Omega_{\pi(t)})\right\}_{t\in\{1,2\}}$ such that

$$\begin{aligned} &(\Phi_{0,1} f_1)(\omega_0) = f_1(\phi_{0,1}(\omega_0))\left(\forall f_1 \in C(\Omega_1), \forall \omega_0 \in \Omega_0\right) \\ &(\Phi_{1,2} f_2)(\omega_1) = f_2(\phi_{1,2}(\omega_1))\left(\forall f_2 \in C(\Omega_2), \forall \omega_1 \in \Omega_1\right). \end{aligned} \tag{12}$$

Thus, we have the causal relation as follows.

$$C(\Omega_0) \xleftarrow{\Phi_{0,1}} C(\Omega_1) \xleftarrow{\Phi_{1,2}} C(\Omega_2).$$

Put $\phi_{0,2}(\omega_0) = \phi_{1,2}(\phi_{0,1}(\omega_0))$, $\Phi_{0,2} = \Phi_{0,1}\cdot\Phi_{1,2}$.

Let $\mathbb{R}$ be the set of real numbers. Fix $\sigma > 0$. For each $t = 0,1,2$, define the *normal observable* $O_t \equiv (\mathbb{R}, \mathcal{B}_{\mathbb{R}}, G_\sigma^n)$ in $C(\Omega_t)$ such that

$$\begin{aligned} &\left[G_\sigma^n(\Xi)\right](\omega_t) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_\Xi \exp\left(-\frac{(x-\alpha)^2}{2\sigma^2}\right)\mathrm{d}x \\ &\left(\forall \Xi \in \mathcal{B}_{\mathbb{R}}, \forall \omega_t = (\alpha,\beta) \in \Omega_t = [0,2t+2]\times[0,2]\right). \end{aligned} \tag{13}$$

Thus, we get the sequential deterministic causal observable

$$[\mathbb{O}_T] = \left[\{O_t\}_{t=0,1,2}, \left\{\Phi_{\pi(t),t} : C(\Omega_t) \to C(\Omega_{\pi(t)})\right\}_{t=1,2}\right].$$

Then, the realized causal observable $\hat{O}_0 \equiv (\mathbb{R}^3, \mathcal{B}_{\mathbb{R}^3}, \hat{F}_0)$ in $C(\Omega_0)$ is, by (3) and (12), obtained as follows:

$$\begin{aligned} &\left[\hat{F}_0(\Xi_0\times\Xi_1\times\Xi_2)\right](\omega_0) \\ &= \left[\left(G_\sigma^n(\Xi_0)\Phi_{0,1}\left(G_\sigma^n(\Xi_1)\Phi_{1,2}\left(G_\sigma^n(\Xi_2)\right)\right)\right)\right](\omega_0) \\ &= \left[G_\sigma^n(\Xi_0)\right](\omega_0)\cdot\left[G_\sigma^n(\Xi_1)\right](\phi_{0,1}(\omega_0)) \\ &\quad \cdot\left[G_\sigma^n(\Xi_2)\right](\phi_{0,2}(\omega_0)) \\ &\left(\forall \Xi_0,\Xi_1,\Xi_2 \in \mathcal{B}_{\mathbb{R}}, \forall \omega_0 = (\alpha,\beta)\in\Omega_0\right). \end{aligned} \tag{14}$$

Putting $K = \mathcal{M}_{+1}^p(\Omega_0)$, we have the measurement $M_{C(\Omega_0)}\left(\hat{O}_0, S_{[*]}\left(\mathcal{M}_{+1}^p(\Omega_0)\right)\right)$. Recall the (10), that is, the measured value $(x_0, x_1, x_2)$ obtained by the measurement $M_{C(\Omega_0)}\left(\hat{O}_0, S_{[*]}\left(\mathcal{M}_{+1}^p(\Omega_0)\right)\right)$ is equal to

$$(0.5, 1.6, 3.3)\left(\in \mathbb{R}^3\right). \tag{15}$$

Define the closed interval $\Xi_t\ (t = 0,2,3)$ such that

$$\Xi_0 = \left[0.5 - \frac{1}{2N}, 0.5 + \frac{1}{2N}\right],$$

$$\Xi_1 = \left[1.6 - \frac{1}{2N}, 1.6 + \frac{1}{2N}\right],$$

$$\Xi_2 = \left[3.3 - \frac{1}{2N}, 3.3 + \frac{1}{2N}\right],$$

for sufficiently large *N*. Here, Fisher's method (Theorem 1) says that it suffices to solve the problem.

(N) Find $(\alpha_0, \beta_0)$ such as

$$\max_{(\alpha,\beta)\in\Omega_0}\left[\hat{F}_0(\Xi_0\times\Xi_1\times\Xi_2)\right](\alpha,\beta) \tag{16}$$

Putting

$$U(x_0,x_1,x_2,\alpha,\beta) = \sum_{k=0}^{2}\left(x_k - (\alpha+k\beta)\right)^2$$

we have the following problem that is equivalent to (N):

(O) Find $(\alpha_0, \beta_0)$ such as

$$\min_{(\alpha,\beta)\in\Omega_0} \exp\left(-\frac{U(x_0,x_1,x_2,\alpha,\beta)}{2\sigma^2}\right)$$

$$\Leftrightarrow \max_{(\alpha,\beta)\in\Omega_0} U(x_0,x_1,x_2,\alpha,\beta).$$

Calculating

$$\frac{\partial}{\partial\alpha}U(0.5,1.6,3.3,\alpha,\beta)=0,$$

$$\frac{\partial}{\partial\beta}U(0.5,1.6,3.3,\alpha,\beta)=0,$$

we get

$$(\alpha,\beta)=(0.4,1.4) \tag{17}$$

Thus, we see, by the statement ($K$), that

(P) $\underset{(\subseteq\mathcal{M}_{+1}^m(\Omega))}{\mathcal{M}_{+1}^p(\Omega_0)} \xrightarrow[\text{Theorem 1}]{\text{Fisher}} \underset{(\in K)}{\delta_{(0.4,1.4)}} \xrightarrow[\text{Theorem 2}]{\text{Bayes}} \underset{(\in K)}{\delta_{(0.4,1.4)}}$

This (*i.e.*, $(\alpha_0,\beta_0)=(0.4,1.4)$) is the answer to the problem (M).

**Problem 1.** Since the above example is quite easy, the validity of Bayes' theorem in (P) may not be clear. If it is so, instead of the problem (M), we should present the following simple problem.

(Q) Infer the water level at time 1.

Some may calculate and conclude as follows:

$$h(1)=\alpha_0+\beta_0\times 1=0.4+1.4=1.8 \tag{18}$$

However, this calculation is based on the Schrödinger picture, and thus, the justification of this calculation (18) is not assured. That is because measurement theory (particularly, Interpretation ($G_2$)) says that the Heisenberg picture should be adopted. Therefore, in order to answer the problem (Q), we must prepare Corollary 2 (*i.e.*, regression analysis in a wide sense) in the following section.

*Remark* 5. It should be noted that the following two are equivalent:

($R_1$) [=(M); Inference]: when measured data (10) is obtained, infer the unknown parameter $(\alpha_0,\beta_0)$.

($R_2$) [Control]: Settle the parameter $(\alpha_0,\beta_0)$ such that measured data (10) will be obtained.

That is, we see that

"inference" = "control".

Hence, from the measurement theoretical point of view, we consider that

"Statistics" = "Dynamical system theory",

though these are superficially different in applications.

# 4. Causal Fisher-Bayes Method in Classical $C(\Omega)$

## 4.1. Causal Bayes Method in Classical $C(\Omega)$

Let $t_0$ be the root of a tree $T$. Let

$$\left[\mathbb{O}_T^\times\right]=\left[\left\{O_t^\times\left(\equiv\left(X_t\times Y_t,\mathcal{F}_t\boxtimes\mathcal{G}_t,F_t\times G_t\right)\right)\right\}_{t\in T},\right.$$
$$\left.\left\{\Phi_{t_1,t_2}:C(\Omega_{t_2})\to C(\Omega_{t_1})\right\}_{(t_1,t_2)\in T_{\leq}^2}\right]$$

be a sequential causal observable with the realization $\hat{O}_{t_0}^\times\left(\times_{t\in T}(X_t\times Y_t),\boxtimes_{t\in T}(\mathcal{F}_t\boxtimes\mathcal{G}_t),\hat{H}_{t_0}\right)$ in $C(\Omega_{t_0})$. Thus we have the statistical measurement $M_{C(\Omega_{t_0})}\left(\hat{O},S_{[*]}(\{\nu_0\})\right)$, where $\nu_0\in\mathcal{M}_{+1}^m(\Omega_{t_0})$. Assume that we know that the measured value $(x,y)\left(=\left((x_t)_{t\in T},(x_t)_{t\in T}\right)\in(\times_{t\in T}X_t)\times(\times_{t\in T}Y_t)\right)$ obtained by the measurement $M_{C(\Omega_{t_0})}\left(\hat{O},S_{[*]}(\{\nu_0\})\right)$ belongs to $(\times_{t\in T}\Xi_t)\times(\times_{t\in T}Y_t)\left(\in(\boxtimes_{t\in T}\mathcal{F}_t)\boxtimes(\boxtimes_{t\in T}Y_t)\right)$. Then, by (4), we can infer that

(S) the probability $P_{\times_{t\in T}\Xi_t}\left(\left(G_t(\Gamma_t)\right)_{t\in T}\right)$ that $y$ belongs to $\times_{t\in T}\Gamma_t\left(\in\boxtimes_{t\in T}\mathcal{G}_t\right)$ is given by

$$P_{\times_{t\in T}\Xi_t}\left(\left(G_t(\Gamma_t)\right)_{t\in T}\right)$$
$$=\frac{\int_\Omega\left[\hat{H}_{t_0}\left((\times_{t\in T}\Xi_t)\times(\times_{t\in T}\Gamma_t)\right)\right](\omega)\nu_0(\mathrm{d}\omega)}{\int_\Omega\left[\hat{H}_{t_0}\left(\times_{t\in T}\Xi_t\right)\times(\times_{t\in T}Y_t)\right](\omega)\nu_0(\mathrm{d}\omega)} \tag{19}$$
$$\left(\forall\Gamma_t\in\mathcal{G}_t,t\in T\right).$$

Note that we can regard that $P_{\times_{t\in T}\Xi_t}\in\mathcal{M}_{+1}^m(\times_{t\in T}\Omega_t)\left(\subseteq C(\times_{t\in T}\Omega_t)^*\right)$. That is, there uniquely exists $\nu_T^a\in\mathcal{M}_{+1}^m(\times_{t\in T}\Omega_t)$ such that

$$P_{\times_{t\in T}\Xi_t}\left(\left(G_t(\Gamma_t)\right)_{t\in T}\right)=\int_{\times_{t\in T}\Omega_t}\left[\otimes_{t\in T}G_t(\Gamma_t)\right](\omega)\nu_T^a(\mathrm{d}\omega) \tag{20}$$

for any observable $(Y_t,\mathcal{G}_t,G_t)$ in $C(\Omega_t)$ $(t\in T)$. Here, we used the following notation:

$$\left[\otimes_{t\in T}G_t(\Gamma_t)\right](\omega)=\underset{t\in T}{\times}\left[G_t(\Gamma_t)\right](\omega_t)$$
$$\left(\forall\omega=(\omega_t)_{t\in T}\in\underset{t\in T}{\times}\Omega_t\right).$$

Define the observable $\hat{O}_{t_0}\equiv\left(\times_{t\in T}X_t,\boxtimes_{t\in T}F_t,\hat{F}_{t_0}\right)$ such that

$$\hat{F}_{t_0}\left(\underset{t\in T}{\times}\Xi_t\right)=\hat{H}_{t_0}\left(\left(\underset{t\in T}{\times}\Xi_t\right)\times\left(\underset{t\in T}{\times}Y_t\right)\right).$$

Then, we can define the Bayes operator $\left[B_{\hat{O}_{t_0}}(\times_{t\in T}\Xi_t)\right]:\mathcal{M}_{+1}^m(\Omega_{t_0})\to\mathcal{M}_{+1}^m(\times_{t\in T}\Omega_t)$ by (20).

Thus, as the generalization of Theorem 2, we have:

**Theorem 3.** [Causal Bayes' theorem in classical measurements]. *Let* $t_0$ *be the root of a tree T. Let*

$$[\mathbb{O}_T] = \left[ \left\{ O_t \left( \equiv \left( X_t, \mathcal{F}_t, F_t \right) \right) \right\}_{t\in T}, \left\{ \Phi_{t_1,t_2} : C\left(\Omega_{t_2}\right) \to C\left(\Omega_{t_1}\right) \right\}_{(t_1,t_2)\in T^2_{\leq}} \right]$$

*be a sequential causal observable with the realization* $\hat{O}_{t_0} \equiv \left( \times_{t\in T} X_t, \boxtimes_{t\in T} F_t, \hat{F}_{t_0} \right)$. *Thus we have the statistical measurement* $M_{C\left(\Omega_{t_0}\right)}\left( \hat{O}_{t_0}, S_{[*]}\left(\{\nu_0\}\right)\right)$, *where* $\nu_0 \in \mathcal{M}^m_{+1}\left(\Omega_{t_0}\right)$. *Assume that we know that a measured value obtained by the statistical measurement* $M_{C\left(\Omega_{t_0}\right)}\left( \hat{O}_{t_0}, S_{[*]}\left(\{\nu_0\}\right)\right)$ *belongs to* $\times_{t\in T} \Xi_t$. *Then, there is a reason to infer that the mixed state* $\nu^a_T \left( \in \mathcal{M}^m_{+1}\left(\times_{t\in T}\Omega_t\right)\right)$ *after the statistical measurement* $M_{C\left(\Omega_{t_0}\right)}\left( \hat{O}_{t_0}, S_{[*]}\left(\{\nu_0\}\right)\right)$ *is given by*

$$\left[ B_{\hat{O}_{t_0}}\left(\times_{t\in T}\Xi_t\right)\right](\nu_0) \left( \in \mathcal{M}^m_{+1}\left(\times_{t\in T}\Omega_t\right)\right).$$

*Proof.* The proof is similar to the proof of Theorem 2. Thus, we omit it.

*Remark* 6. In Theorem 3, we see that

$$\text{(T)} \quad \underset{\left(\in\mathcal{M}^m_{+1}\left(\Omega_{t_0}\right)\right)}{\overset{\text{(pretest state)}}{\nu_0}} \xrightarrow[\text{Theorem 2}]{\text{Bayes}} \underset{\left(\in\mathcal{M}^m_{+1}\left(\times_{t\in T}\Omega_t\right)\right)}{\overset{\text{(posttest state)}}{\nu^a_T}}$$

which is the generalization of the (J).

The following example promotes the understanding of Theorem 3.

**Example 1.** [The simple case such that $T = \{0,1,2\}$ ]. Consider a particular case such that $T = \{0,1,2\}$ is series ordered set, *i.e.*, $\pi(t) = t-1$ $\left(\forall t \in T \setminus \{0\}\right)$. And consider a causal relation $\left\{ C\left(\Omega_t\right) \xrightarrow{\Phi_{\pi(t),t}} C\left(\Omega_{\pi(t)}\right)\right\}_{t\in T\setminus\{0\}}$, that is,

$$C\left(\Omega_0\right) \xleftarrow{\Phi_{0,1}} C\left(\Omega_1\right) \xleftarrow{\Phi_{1,2}} C\left(\Omega_2\right).$$

Further consider sequential causal observable

$$[\mathbb{O}_T] = \left[ \{O_t\}_{t\in T}, \left\{ \Phi_{t,\pi(t)} : C\left(\Omega_t\right) \to C\left(\Omega_{\pi(t)}\right)\right\}_{t\in T\setminus\{0\}}\right].$$

Let $\hat{O}_0 \equiv \left( \times_{t\in T} X_t, \times_{t\in T} F_t, \hat{F}_0\right)$ be its realization. Note, by the Formula (3), that,

$$\hat{F}_0\left(\Xi_0 \times \Xi_1 \times \Xi_2\right) = \Phi_{0,1}\left(F_0\left(\Xi_0\right)\left(\Phi_{1,2}F_1\left(\Xi_1\right)\left(\Phi_{1,2}\left(F_2\left(\Xi_2\right)\right)\right)\right)\right) \quad \left(\Xi_t \in \mathcal{F}_t\left(\forall t \in T\right)\right).$$

Putting $K = \{\nu_0\}$, we have the measurement

$$M_{C(\Omega_0)}\left(\hat{O}_0 \equiv \left(\times_{t\in T} X_t, \boxtimes_{t\in T} F_t, \hat{F}_0\right), S_{[*]}\left(\{\nu_0\}\right)\right). \tag{21}$$

Let $\nu^a_T \left( \in \mathcal{M}^m_{+1}\left(\Omega_0 \times \Omega_1 \times \Omega_2\right)\right)$ be the posttest state in (T), that is, $\nu^a_T \left[ B_{\hat{O}_0}\left(\times_{t\in T}\Xi_t\right)\right](\nu_0)$. Define $\nu^a_{\{1\}} \left(\in \mathcal{M}^m_{+1}\left(\Omega_1\right)\right)$ such that

$$\nu^a_{\{1\}}\left(D_1\right) = \nu^a_T\left(\Omega_0 \times D_1 \times \Omega_2\right) \quad \left(\forall D_1 \in \mathcal{B}_{\Omega_1}\right).$$

Then, we see that

$$\nu^a_{\{1\}} = \frac{\left(F_1\left(\Xi_1\right)\right)\left(\Phi_{1,2}F_2\left(\Xi_2\right)\right)\left(\Phi^*_{0,1}\left(F_0\left(\Xi_0\right)\nu_0\right)\right)}{\left\langle \nu_0, F_0\left(\Xi_0\right)\Phi_{0,1}\left(F_1\left(\Xi_1\right)\Phi_{1,2}\left(F_2\left(\Xi_2\right)\right)\right)\right\rangle}.$$

That is because we see that, for any observable $\left(Y_1, \mathcal{G}_1, G_1\right)$ in $C\left(\Omega_1\right)$,

$$\begin{aligned}
&\left\langle \nu^a_{\{1\}}, G_1\left(\Gamma_1\right)\right\rangle \\
&= \frac{\left\langle \nu_0, F_0\left(\Xi_0\right)\Phi_{0,1}\left(F_1\left(\Xi_1\right)G_1\left(\Gamma_1\right)\Phi_{1,2}\left(F_2\left(\Xi_2\right)\right)\right)\right\rangle}{\left\langle \nu_0, F_0\left(\Xi_0\right)\Phi_{0,1}\left(F_1\left(\Xi_1\right)G_1\left(Y_1\right)\Phi_{1,2}\left(F_2\left(\Xi_2\right)\right)\right)\right\rangle} \\
&= \frac{\left\langle \left(F_1\left(\Xi_1\right)\right)\left(\Phi_{1,2}F_2\left(\Xi_2\right)\right)\left(\Phi^*_{0,1}\left(F_0\left(\Xi_0\right)\nu_0\right)\right), G_1\left(\Gamma_1\right)\right\rangle}{\left\langle \nu_0, F_0\left(\Xi_0\right)\Phi_{0,1}\left(F_1\left(\Xi_1\right)\Phi_{1,2}\left(F_2\left(\Xi_2\right)\right)\right)\right\rangle} \\
&\left(\forall \Gamma_1 \in \mathcal{G}_1\right).
\end{aligned} \tag{22}$$

**Example 2.** [Continued from the above example]. For each $t = 1,2$, assume that $\Phi_{\pi(t),t} : C\left(\Omega_t\right) \to C\left(\Omega_{\pi(t)}\right)$ is deterministic, that is, there exists a continuous map $\phi_{\pi(t),t} : \Omega_{\pi(t)} \to \Omega_t$ satisfying (12). And, putting $K = \left\{\delta_{\omega_0}\right\}$, consider the measurement

$$M_{C(\Omega_0)}\left(\hat{O}_0 \equiv \left(\times_{t\in T} X_t, \boxtimes_{t\in T} F_t, \hat{F}_0\right), S_{[*]}\left(\left\{\delta_{\omega_0}\right\}\right)\right).$$

Then, we see, by (22), that, for any $g_1$ in $C\left(\Omega_1\right)$,

$$\begin{aligned}
\left\langle \nu^a_{\{1\}}, g_1 \right\rangle &= \frac{\left\langle \delta_{\omega_0}, F_0\left(\Xi_0\right)\Phi_{0,1}\left(F_1\left(\Xi_1\right)g_1\Phi_{1,2}\left(F_2\left(\Xi_2\right)\right)\right)\right\rangle}{\left\langle \delta_{\omega_0}, F_0\left(\Xi_0\right)\Phi_{0,1}\left(F_1\left(\Xi_1\right)\Phi_{1,2}\left(F_2\left(\Xi_2\right)\right)\right)\right\rangle} \\
&= \frac{\left[F_0\left(\Xi_0\right)\right]\left(\omega_0\right)\left[F_1\left(\Xi_1\right)g_1\Phi_{1,2}\left(F_2\left(\Xi_2\right)\right)\right]\left(\phi_{0,1}\left(\omega_0\right)\right)}{\left[F_0\left(\Xi_0\right)\right]\left(\omega_0\right)\left[F_1\left(\Xi_1\right)\Phi_{1,2}\left(F_2\left(\Xi_2\right)\right)\right]\left(\phi_{0,1}\left(\omega_0\right)\right)} \\
&= g_1\left(\phi_{0,1}\left(\omega_0\right)\right) = \left\langle \delta_{\phi_{0,1}(\omega_0)}, g_1\right\rangle.
\end{aligned}$$

Thus, we see that

$$\nu^a_{\{1\}} = \delta_{\phi_{0,1}(\omega_0)}. \tag{23}$$

Further we easily see that

$$\nu^a_T = \left[B_{\hat{O}_0}\left(\times_{t\in T}\Xi_t\right)\right]\left(\delta_{\omega_0}\right) = \delta_{\left(\omega_0, \phi_{0,1}(\omega_0), \phi_{0,2}(\omega_0)\right)}\left(\in \mathcal{M}^p_{+1}\left(\Omega_0 \times \Omega_1 \times \Omega_2\right)\right).$$

## 4.2. Causal Fisher-Bayes Method in Classical $C(\Omega)$

Now we can present Corollary 2 (*i.e.*, regression analysis in a wide sense) as follows.

(U) $\left[\text{Corollary 2}\right]=\left[\underset{\text{(Fisher's method)}}{\text{Theorem 1}}\right]+\left[\underset{\text{(Bayes' method)}}{\text{Theorem 3}}\right]$

**Corollary 2.** [Causal Fisher-Bayes method (*i.e.*, Regression analysis in a wide sense)]. *Let* $t_0$ *be the root of a tree T. Let*

$$[\mathbb{O}_T]=\left[\left\{O_t\left(\equiv\left(X_t,\mathcal{F}_t,F_t\right)\right)\right\}_{t\in T},\right.$$
$$\left.\left\{\Phi_{t_1,t_2}:C\left(\Omega_{t_2}\right)\to C\left(\Omega_{t_1}\right)\right\}_{(t_1,t_2)\in T^2_{\leq}}\right]$$

*be a sequential causal observable with the realization* $\hat{O}_{t_0}\equiv\left(\times_{t\in T}X_t,\boxtimes_{t\in T}F_t,\hat{F}_{t_0}\right)$ *Assume the statistical measurement* $M_{C\left(\Omega_{t_0}\right)}\left(\hat{O}_{t_0},S_{[*]}(K)\right)$*. And assume that we know that a measured value obtained by the measurement* $M_{C\left(\Omega_{t_0}\right)}\left(\hat{O}_{t_0},S_{[*]}(K)\right)$ *belongs to* $\times_{t\in T}\Xi_t$*. Then, there is a reason to infer that the mixed state* $\nu_T^a\left(\in\mathcal{M}_{+1}^m\left(\times_{t\in T}\Omega_t\right)\right)$ *after the measurement* $M_{C\left(\Omega_{t_0}\right)}\left(\hat{O}_{t_0},S_{[*]}(K)\right)$ *is given by* $\left[B_{\hat{O}_{t_0}}\left(\times_{t\in T}\Xi_t\right)\right](\nu_0)$.

*Here, the* $\nu_0(\in K)$ *is defined by*

$$\begin{aligned}&\int_\Omega\left[\hat{F}_{t_0}\left(\underset{t\in T}{\times}\Xi_t\right)\right](\omega)\nu(\mathrm{d}\omega)\\&=\max_{\nu\in K}\int_\Omega\left[\hat{F}_{t_0}\left(\underset{t\in T}{\times}\Xi_t\right)\right](\omega)\nu(d\omega).\end{aligned}\qquad(24)$$

*Remark* 7. Note that Fisher maximum likelihood method and Bayes' theorem are hidden in Corollary 2. That is, Corollary 2 includes the following procedure:

(V) $\underset{\left(\in\mathcal{M}_{+1}^m\left(\Omega_{t_0}\right)\right)}{K}\xrightarrow[\text{Theorem 1}]{\text{Fisher}}\underset{(\in K)}{\nu_0}\xrightarrow[\text{Theorem 3}]{\text{Bayes}}\underset{\left(\in\mathcal{M}_{+1}^m\left(\times_{t\in T}\Omega_t\right)\right)}{\nu_T^a}$

which is the generalization of the (L).

**Answer 1.** [Answer to Problem 1 (Q)]. Now we can answer Problem 1 (Q) as follows. The (17) says that $\nu_0=\delta_{(\alpha_0,\beta_0)}=\delta_{(0.4,1.4)}$. Thus, using (23), we see that $\nu_{\{1\}}^a=\delta_{\alpha_0+\beta_0}=\delta_{1.8}$. Also, note that (17) and (23) are consequences of Corollary 2. Hence, the calculation (18) is justified by Corollary 2.

*Remark* 8. As mentioned in Section 1, in our research [2-8], we have been concerned with the problem (A). Particularly, in [6], we discussed Corollary 2 in the commutative $W^*$-algebra $L^\infty(\Omega)$. However, this was somewhat shallow, since "max" is not proper in $L^\infty(\Omega)$ but $C(\Omega)$. Now we believe that fundamental statements concerning statistics should be always asserted in the framework of $C(\Omega)$. Also, note that Corollary 2 is the natural generalization of Theorem 6.3 in [5].

## 5. Conclusions

In this paper, we devote ourselves to the problem (A) in the light of the quantum mechanical word view (cf. [7,8]). And, we show that regression analysis, which is the most fundamental in statistics, is formulated as Corollary 2 in SMT (*i.e.*, statistical measurement theory). We believe that Corollary 2 is the finest formulation of regression analysis, since no clear formulation can be presented without the answer to the problem (A). Also, note that Corollary 2 (or, the (U)) implies that even the conventional classification of (Fisher's) statistics and Bayesian statistics should be reconsidered.

We expect that there is a great possibility that our proposal (*i.e.*, statistics is based on statistical measurement theory) will be generally accepted. We of course know that the conventional statistics methodology can be good applied in many fields. Hence, we hope that our methodology in the light of the quantum mechenical word view should be examined from various points of view.

## REFERENCES


[1] A. Kolmogorov, "Foundations of the Theory of Probability (Translation)," 2nd Edition, Chelsea Pub Co., New York, 1960,

[2] S. Ishikawa, "Fuzzy Inferences by Algebraic Method," *Fuzzy Sets and Systems*, Vol. 87, No. 2, 1997, pp. 181-200. doi:10.1016/S0165-0114(96)00035-8

[3] S. Ishikawa, "A Quantum Mechanical Approach to Fuzzy Theory," *Fuzzy Sets and Systems*, Vol. 90, No. 3, 1997, pp. 277-306. doi:10.1016/S0165-0114(96)00114-5

[4] S. Ishikawa, "Statistics in Measurements," *Fuzzy Sets and Systems*, Vol. 116, No. 2, 2000, pp. 141-154. doi:10.1016/S0165-0114(98)00280-2

[5] S. Ishikawa, "Mathematical Foundations of Measurement Theory," Keio University Press Inc., Tokyo, 2006, pp. 1-335. http://www.keio-up.co.jp/kup/mfomt/

[6] S. Ishikawa, "Fisher's Method, Bayes' Method and Kalman Filter in Measurement Theory," *Far East Journal of Theoretical Statistics*, Vol. 29, No. 1, 2009, pp. 9-23.

[7] S. Ishikawa, "A New Interpretation of Quantum Mechanics," *Journal of Quantum Information Science*, Vol. 1, No. 2, 2011, pp. 35-42. doi:10.4236/jqis.2011.12005

[8] S. Ishikawa, "Quantum Mechanics and the Philosophy of Language: Reconsideration of Traditional Philosophies," *Journal of Quantum Information Science*, Vol. 2, No. 1, 2012, pp. 2-9.

[9] G. J. Murphy, "$C^*$-Algebras and Operator Theory," Academic Press, London, 1990.

[10] J. von Neumann, "Mathematical Foundations of Quantum Mechanics," Springer Verlag, Berlin, 1932.

[11] K. Yosida, "Functional Analysis," 6th Edition, Springer-

Verlag, Berlin, 1980.

[12] E. B. Davies, "Quantum Theory of Open Systems," Academic Press, London, 1976.

## Appendix

As mentioned in Section 3.1, we have to prove the following theorem.

**Theorem 4.** [Existence theorem of product observable]. Let $O_1 \equiv (X, \mathcal{F}, F)$ and $O_2 \equiv (Y, \mathcal{G}, G)$ be observables in a $C^*$-algebra $A = C(\Omega)$. Then, there exists the product observable $O_1 \times O_2 \equiv (X \times Y, \mathcal{F} \boxtimes \mathcal{G}, F \times G)$ in $C(\Omega)$.

*Proof*. Let $\bar{\mathcal{F}}$ [resp. $\bar{\mathcal{G}}$; $\overline{\mathcal{F} \boxtimes \mathcal{G}}$] be the smallest $\sigma$-field including $\mathcal{F}$ [resp. $\mathcal{G}$; $\mathcal{F} \boxtimes \mathcal{G}$]. That is, for each $k = 1, 2, \cdots$, consider $\Xi_k \times \Gamma_k \ (\in \mathcal{F} \boxtimes \mathcal{G})$ such that

$$(\Xi_i \times \Gamma_i) \cap (\Xi_j \times \Gamma_j) = \varnothing (i \neq j)$$

and

$$\bigcup_{k=1}^{\infty} (\Xi_k \times \Gamma_k) \in \mathcal{F} \boxtimes \mathcal{G}.$$

Note, by the Hopf extension theorem (*cf.* Remark 1), that it suffices to show that, for any $\nu \in \mathcal{M}_{+1}^m(\Omega)$, it holds:

$$\int_\Omega \left[ (F \times G) \left( \bigcup_{k=1}^{\infty} (\Xi_k \times \Gamma_k) \right) \right] (\omega) \nu(\mathrm{d}\omega)$$
$$= \lim_{K \to \infty} \int_\Omega \sum_{k=1}^{K} \left[ F(\Xi_k) \right](\omega) \cdot \left[ G(\Gamma_k) \right](\omega) \nu(\mathrm{d}\omega)$$

which is equivalent to the following equality. That is, for any $\omega \in \Omega$, it holds:

$$\left[ (F \times G) \left( \bigcup_{k=1}^{\infty} (\Xi_k \times \Gamma_k) \right) \right] (\omega) = \lim_{K \to \infty} \sum_{k=1}^{K} \left[ F(\Xi_k) \right](\omega) \cdot \left[ G(\Gamma_k) \right](\omega). \quad (25)$$

However, it is easily seen since $(X, \bar{\mathcal{F}}, [F(\cdot)](\omega))$ and $(Y, \bar{\mathcal{G}}, [G(\cdot)](\omega))$ can be regarded as probability spaces. And therefore, we have the product probability space $(X \times Y, \overline{\mathcal{F} \boxtimes \mathcal{G}}, [(F \times G)(\cdot)](\omega))$. This imlies that the equality (25) holds. This completes the proof.

*Remark* 9. The above proof is applicable to the realization of a sequential causal observable $[\mathbb{O}_T]$ in the case of an infinite *T* under a similar condition such that the Kolmogorov extension theorem holds (cf. [1]). Also, in quantum case (*i.e.*, $A = B_c(H)$), it is well known that the weak convergence (1) in $B_c(H)$ can be identified with the weak convergence in $B(H)$, therefore, we see, by a usual way (cf. [10,11]), that Theorem 4 holds under the commutativity condition (2).