Title	Chapter 8 : Statistical measurements in C*-algebraic formulation
Sub Title	
Author	石川, 史郎(Ishikawa, Shiro)
Publisher	Keio University Press Inc.
Publication year	2006
Jtitle	Mathematical Foundations of Measurement Theory (測定理論の数学的基礎). (2006.) ,p.175- 223
JaLC DOI	
Abstract	
Notes	
Genre	Book
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO52003001-00000000-0175

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:

$$\text{PMT} = \underset{\text{[Axiom 1 (2.37)]}}{\text{measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} \quad \text{in } C^*\text{-algebra} \, . \tag{8.2} \quad (=(1.4))$$

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:

$$\boxed{\begin{matrix}\text{"observable"}\\ (X,\mathcal{F},F) \text{ in } \mathcal{A}\end{matrix}} \quad \text{and} \quad \boxed{\begin{matrix}\text{"state"}\\ \rho^p\in\mathfrak{S}^p(\mathcal{A}^*)\end{matrix}} \quad \underset{\text{measurement}}{\Longrightarrow} \quad \boxed{\begin{matrix}\text{"sample space"}\\ (X,\mathcal{F},P(\cdot)\equiv\rho^p(F(\cdot)))\end{matrix}}.$$

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".

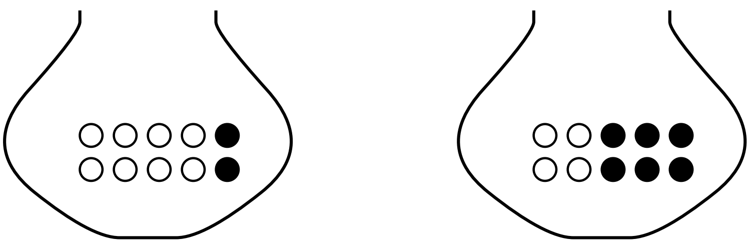
Example 8.1. [Coin-tossing and urn problem]. There are two urns U_1 and U_2. The urn U_1 [resp. U_2] contains 8 white and 2 black balls [resp. 4 white and 6 black balls]. Under the following identification (*cf.* (5.16) in Example 5.8):

$$U_1 \approx \omega_1, \qquad U_2 \approx \omega_2,$$

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

$$[F(\{w\})](\omega_1) = 0.8, \qquad [F(\{b\})](\omega_1) = 0.2,$$
$$[F(\{w\})](\omega_2) = 0.4, \qquad [F(\{b\})](\omega_2) = 0.6. \tag{8.6}$$

$$U_1 \approx \omega_1 \qquad\qquad U_2 \approx \omega_2$$



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 $\nu_0(\in \mathcal{M}^m_{+1}(\Omega))$ 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

$$_{\mathcal{M}(\Omega)}\langle \delta_{\omega_1}, F(\{x\}) \rangle_{C(\Omega)} = 0.8 \ \text{(if } x = w), \quad = 0.2 \ \text{(if } x = b),$$
$$\Big[\text{resp. } _{\mathcal{M}(\Omega)}\langle \delta_{\omega_2}, F(\{x\}) \rangle_{C(\Omega)} = 0.4 \ \text{(if } x = w), \quad = 0.6 \ \text{(if } x = b) \quad \Big].$$

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

$$P(\{x\}) = \int_\Omega {}_{\mathcal{M}(\Omega)}\langle \delta_\omega, F(\{x\}) \rangle_{C(\Omega)} \nu_0(d\omega) = {}_{\mathcal{M}(\Omega)}\langle \nu_0, F(\{x\}) \rangle_{C(\Omega)}$$
$$= \begin{cases} 0.8p + 0.4(1-p) & \text{(if } x = w), \\ 0.2p + 0.6(1-p)) & \text{(if } x = b). \end{cases}$$

This is the answer to the above question (Q). Summing up, we see:

(♯) 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 $\nu_0 (\in \mathcal{M}^m_{+1}(\Omega))$ 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

$$ {}_{C(\Omega)^*}\Big\langle \nu_0, F(\{x\}) \Big\rangle_{C(\Omega)} \ \Big(\equiv \int_\Omega {}_{C(\Omega)^*}\Big\langle \delta_\omega, F(\{x\}) \Big\rangle_{C(\Omega)} \nu_0(d\omega) \Big). $$

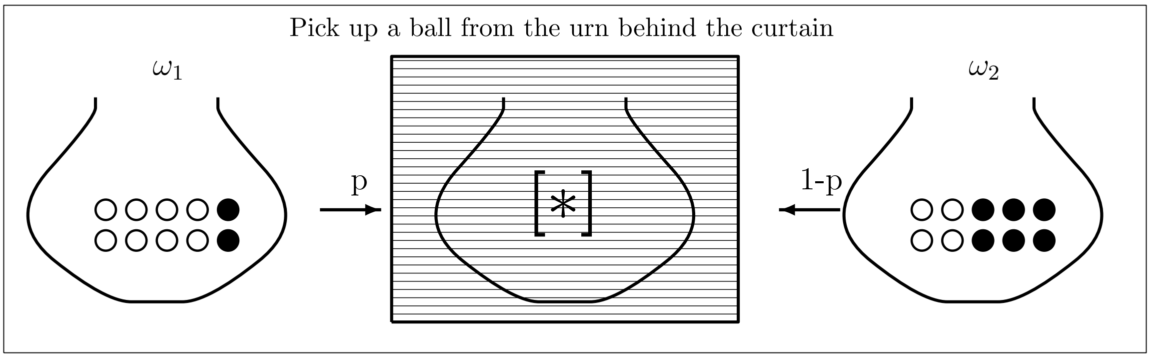
Thus we consider that

$$ S_{[*]}([\delta_{\omega_1}; p] \oplus [\delta_{\omega_2}; 1-p])) \ \underset{\text{statistical form}}{\overset{\text{probabilistic form}}{\longleftrightarrow}} \ S_{[*]}(\nu_0) \tag{8.9} $$

That is, the statistical state ν_0 is *the mixed state with probabilistic interpretation,* or, *the mixed state generated by coin-tossing.*

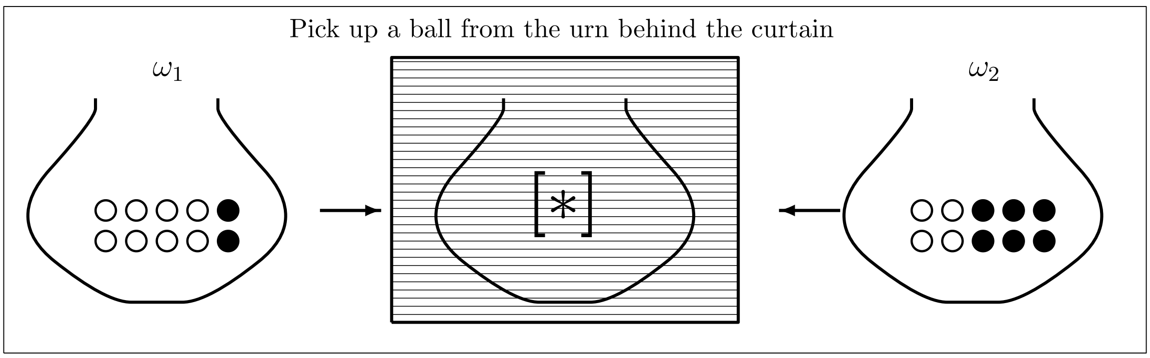
Thus, we see

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

$$[\sharp]\ \text{"statistical state"} = \underset{\text{(mathematics)}}{\text{"mixed state"}} + \underset{\text{(such as coin-tossing)}}{\text{"probabilistic interpretation"}}\ ^{2} \tag{8.11}$$

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., $\{\Phi_{t_1,t_2} : \mathcal{A}_{t_2} \to \mathcal{A}_{t_1}\}_{(t_1,t_2)\in T^2_{\leq}}$)" 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., $m \in \mathcal{M}^m_{+1}([0,1] \times [0,1])$), 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:

$$\underset{\text{(pure) state}\rho_0^p}{[\mathcal{A}_0]} \overset{\Psi_{0,1}}{\longleftarrow} [\mathcal{A}_1]. \tag{8.14}$$

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:

$$\underset{\text{(meaningful in the sense of Axiom 1)}}{\overset{\text{[the representation by Heisenberg picture]}}{\mathbf{M}_{\mathcal{A}_0}(\Psi_{0,1}\mathbf{O}_1, S_{[\rho_0^p]})}} \overset{\text{identification}}{\longleftrightarrow} \underset{\text{(meaningful in the sense of Proclaim 1)}}{\overset{\text{[the representation by Schödinger picture]}}{\mathbf{M}_{\mathcal{A}_1}(\mathbf{O}_1, S(\Psi_{0,1}^*\rho_0^p))}} \tag{8.17}$$

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.

$$\underset{\text{statistical state}\rho_1^m}{[\mathcal{A}_1]} \overset{\Psi_{0,1}}{\longleftarrow} [\mathcal{A}_2]. \tag{8.18}$$

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]

$$\underset{\text{(meaningful in the sense of Proclaim 1)}}{\overset{\text{[the representation by Heisenberg picture]}}{\mathbf{M}_{\mathcal{A}_1}(\Psi_{1,2}\mathbf{O}_2, S(\rho_1^m))}} \overset{\text{identification}}{\longleftrightarrow} \underset{\text{(meaningful in the sense of Proclaim 1)}}{\overset{\text{[the representation by Schödinger picture]}}{\mathbf{M}_{\mathcal{A}_2}(\mathbf{O}_2, S(\Psi_{1,2}^*\rho_1^m))}} \tag{8.22}$$

in which the both sides are understood in Proclaim 1.

■

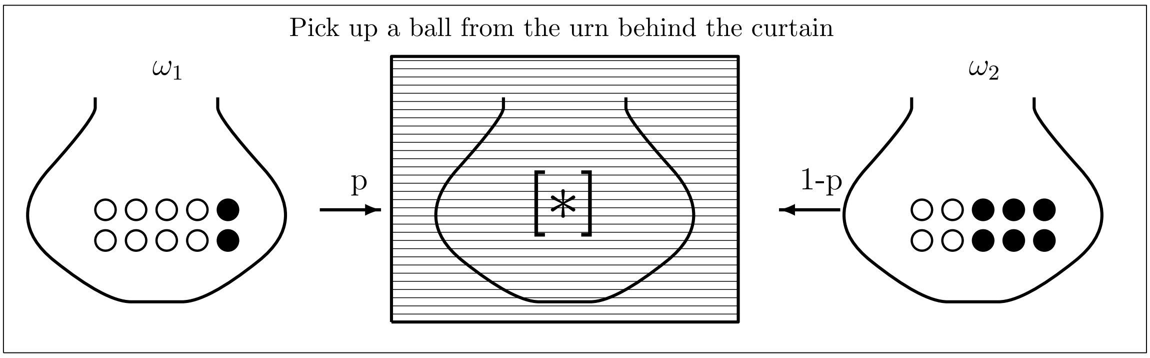
The statistical state also appears in Bayes theorem, which was already studied in Chapter 6.

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

$$\big\langle \nu_0, F(\{w\}) \times G(\Gamma) \big\rangle.$$

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

$$\frac{{}_{\mathcal{M}(\Omega)}\big\langle \nu_0, F(\{w\}) \times G(\Gamma) \big\rangle_{C(\Omega)}}{{}_{\mathcal{M}(\Omega)}\big\langle \nu_0, F(\{w\}) \big\rangle_{C(\Omega)}} \Big(= {}_{\mathcal{M}(\Omega)}\Big\langle \frac{F(\{w\}) \times \nu_0}{{}_{\mathcal{M}(\Omega)}\big\langle \nu_0, F(\{w\}) \big\rangle_{C(\Omega)}}, G(\Gamma) \Big\rangle_{C(\Omega)} \Big). \tag{8.23}$$

Since $\mathbf{O}_1$($\equiv (Y, \mathcal{G}, G)$) is arbitrary observable in $C(\Omega)$, this implies the following state-reduction:

$$\underset{\text{before "white" is obtained in } (P_2)}{\text{pretest state "}\nu_0\text{"}} \longrightarrow \underset{\text{after "white" is obtained in } (P_2)}{\text{posttest state "}\nu_1\text{"}} \Big(= \frac{F(\{w\}) \times \nu_0}{\big\langle \nu_0, F(\{w\}) \big\rangle} \Big). \tag{8.24}$$

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

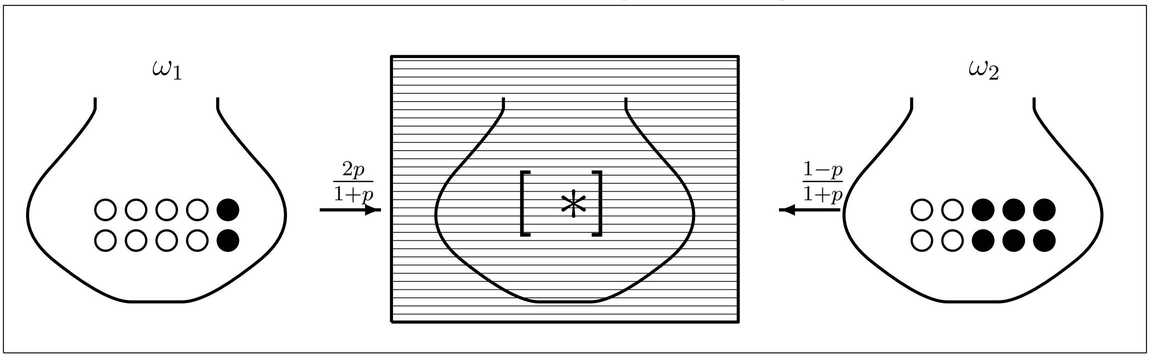
$$\begin{aligned} \nu_1(\{\omega_1\}) &= \frac{0.8p}{0.8p + 0.4(1-p)} = \frac{2p}{1+p}, \\ \nu_1(\{\omega_2\}) &= \frac{0.4(1-p)}{0.8p + 0.4(1-p)} = \frac{1-p}{1+p}. \end{aligned} \tag{8.27}$$

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 ······	measured value	measurement ······	measured value
$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_1}]})$ ······	x_1	$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\frac{\delta_{\omega_1}+\delta_{\omega_2}}{2}))$ ······	y_1
$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_2}]})$ ······	x_2	$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\frac{\delta_{\omega_1}+\delta_{\omega_2}}{2}))$ ······	y_2
$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_1}]})$ ······	x_3	$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\frac{\delta_{\omega_1}+\delta_{\omega_2}}{2}))$ ······	y_3
$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_2}]})$ ······	x_4	$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\frac{\delta_{\omega_1}+\delta_{\omega_2}}{2}))$ ······	y_4
······	···	······	···
$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_1}]})$ ······	x_{2N-1}	$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\frac{\delta_{\omega_1}+\delta_{\omega_2}}{2}))$ ······	y_{2N-1}
$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[\delta_{\omega_1}]})$ ······	x_{2N}	$\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\frac{\delta_{\omega_1}+\delta_{\omega_2}}{2}))$ ······	y_{2N}

■

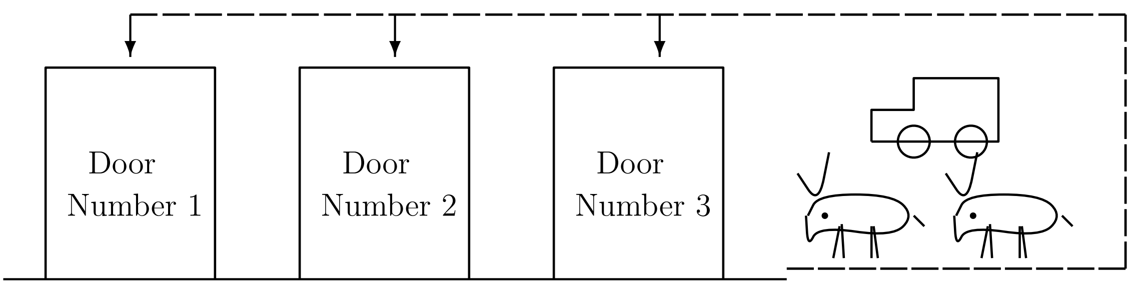
8.1.3 Problems (statistical measurements)

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

$$
\begin{aligned}
&[F(\{1\})](\omega_1) = 0.0, \quad && [F(\{2\})](\omega_1) = 0.5, \quad && [F(\{3\})](\omega_1) = 0.5,^{7} \\
&[F(\{1\})](\omega_2) = 0.0, \quad && [F(\{2\})](\omega_2) = 0.0, \quad && [F(\{3\})](\omega_2) = 1.0, \\
&[F(\{1\})](\omega_3) = 0.0, \quad && [F(\{2\})](\omega_3) = 1.0, \quad && [F(\{3\})](\omega_3) = 0.0.
\end{aligned}
\tag{8.31}
$$

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 $\nu_{\text{post}}(\{\omega_1\}) = 1/3$, $\nu_{\text{post}}(\{\omega_2\}) = 2/3$, $\nu_{\text{post}}(\{\omega_3\}) = 0$, *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* $\nu_{\text{post}}(\{\omega_1\}) = 3/5$, $\nu_{\text{post}}(\{\omega_2\}) = 2/5$, $\nu_{\text{post}}(\{\omega_3\}) = 0$, *and thus, you should not pick Door (number 2).*

Also, more generally, we can say that

$$\begin{cases} \text{if } \nu_{\text{post}}(\{\omega_1\}) \le \nu_{\text{post}}(\{\omega_2\})(\text{i.e.,} p_1 \le 2p_2), \text{ then, you should pick Door (number 2)} \\ \text{if } \nu_{\text{post}}(\{\omega_1\}) \ge \nu_{\text{post}}(\{\omega_2\})(\text{i.e.,} p_1 \ge 2p_2), \text{ then, you should not pick Door (number 2).} \end{cases}$$

■

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:

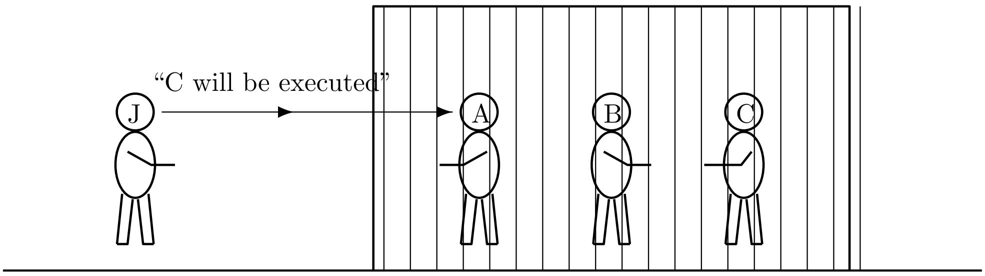
(Q) *You're on a game show and you're given the choice of three doors. Behind one door is a car, and behind the other two are goats. You choose, say, door 1, and the host, who knows where the car is, opens another door, behind which is a goat. He now gives you the choice of sticking with door 1 or switching 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. $\Big($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].$\Big)$ 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 ν_{post} ($\in \mathcal{M}_{+1}^m(\Omega)$) is given by

$$\nu_{\text{post}} = \frac{F(\{x_C\}) \times \nu_0}{\langle \nu_0, F(\{x_C\}) \rangle}. \tag{8.37}$$

That is,

$$\nu_{\text{post}}(\{\omega_a\}) = \frac{\frac{p_a^f}{2}}{\frac{p_a^f}{2} + p_b^f}, \quad \nu_{\text{post}}(\{\omega_b\}) = \frac{p_b^f}{\frac{p_a^f}{2} + p_b^f}, \quad \nu_{\text{post}}(\{\omega_c\}) = 0. \tag{8.38}$$

Thus,

- if $p_a^f = p_b^f = p_c^f = 1/3$, it holds that $\nu_{\text{post}}(\{\omega_a\}) = 1/3$, $\nu_{\text{post}}(\{\omega_b\}) = 2/3$, $\nu_{\text{post}}(\{\omega_c\}) = 0$, *and thus, the prisoner's happiness is not reasonable.* That is because $p_a^f = 1/3 = \nu_{\text{post}}(\{\omega_a\})$.

- if $p_a^f = 3/7$, $p_b^f = 1/7$, $p_c^f = 3/7$, it holds that $\nu_{\text{post}}(\{\omega_a\}) = 3/5$, $\nu_{\text{post}}(\{\omega_b\}) = 2/5$, $\nu_{\text{post}}(\{\omega_c\}) = 0$, *and thus, the prisoner's happiness is reasonable.* That is because $p_a^f = 3/7 < 3/5 = \nu_{\text{post}}(\{\omega_a\})$.

- if $p_a^f = 1/4$, $p_b^f = 1/2$, $p_c^f = 1/4$, it holds that $\nu_{\text{post}}(\{\omega_a\}) = 1/5$, $\nu_{\text{post}}(\{\omega_b\}) = 4/5$, $\nu_{\text{post}}(\{\omega_c\}) = 0$, *and thus, the prisoner's unhappiness is reasonable.* That is because $p_a^f = 1/3 > 1/5 = \nu_{\text{post}}(\{\omega_a\})$.

Also, more generally, we can say that

$$\begin{cases} \text{if } p_a^f \le \nu_{\text{post}}(\{\omega_a\})(\text{i.e.,} p_a^f + 2p_b^f \ge 1), & \text{the prisoner's happiness is reasonable} \\ \text{if } p_a^f \ge \nu_{\text{post}}(\{\omega_a\})(\text{i.e.,} p_a^f + 2p_b^f \le 1), & \text{the prisoner's unhappiness is reasonable.} \end{cases}$$

■

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:

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

and

$$\text{SMT} = \underset{\text{[Proclaim 1 (8.10)]}}{\text{statistical measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} \qquad \text{in } C^*\text{-algebra}\,,$$

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

$$\begin{aligned} \text{SMT} &= \underset{\text{[Proclaim 1 (8.10)]}}{\text{statistical measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} \\ &= \underset{\text{(Axioms 1 and 2)}}{\text{PMT}} + \underset{\text{(the probabilistic interpretation of mixed state)}}{\text{"statistical state"}} \qquad \text{in } C^*\text{-algebra}\,. \end{aligned}$$

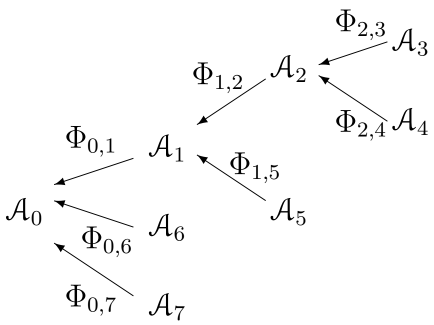
That is, Axiom 2 is common to PMT and SMT. This will be explicitly seen in the following example (= Example 8.12), which should be compared with Example 3.4. Also recalling Remark 8.3 [hybrid measurement theory (= HMT)], we say that

$$\text{HMT} = \underbrace{\text{hybrid measurement}}_{\text{[Axiom 1 (2.37) and Proclaim 1 (8.10)]}} + \underbrace{\text{the relation among systems}}_{\text{[Axiom 2 (3.26)]}} \text{ in } C^*\text{-algebra} . \tag{8.41}$$

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 $\pi(1) = \pi(6) = \pi(7) = 0$, $\pi(2) = \pi(5) = 1$, $\pi(3) = \pi(4) = 2$. $\Big($See the figure (8.42).$\Big)$ 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 $S(\rho_0^m)$.



(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",

($\sharp$) 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:

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

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):

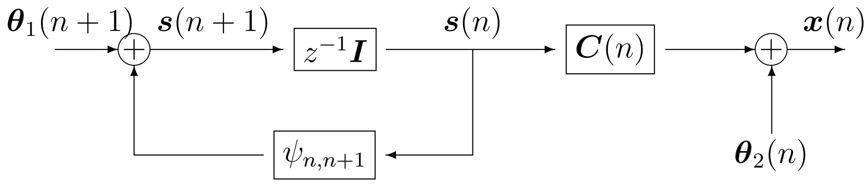
$$Tr_{+1}^m(V) \ni \rho \text{ (= pretest state)} \mapsto \text{ (posttest state =)}\frac{F(\Xi)\rho F(\Xi)}{\|F(\Xi)\rho F(\Xi)\|_{Tr(V)}} \in Tr_{+1}^m(V).$$

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,

$$\begin{cases} \boldsymbol{s}(n+1,\omega) = \psi_{n,n+1}(\boldsymbol{s}(n,\omega)) + \boldsymbol{\theta}_1(n+1,\omega) & \text{: stochastic difference state equation} \\ & (n=0,1,...,N-1). \\ \boldsymbol{x}(n,\omega) = \boldsymbol{C}(n)\boldsymbol{s}(n,\omega) + \boldsymbol{\theta}_2(n,\omega) & \text{: measurement equation} \end{cases} \tag{8.46}$$

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:

$$\text{(8.46) in DST} \leftrightarrow \text{(8.54) in SMT (or precisely, HMT, } cf. \text{ Remark 8.3).}$$

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 $[\mathbf{O}_T] \equiv [\{\mathbf{O}_t\}_{t\in T}; \{\widehat{\widehat{\Phi}}_{t_1,t_2} : C(\mathcal{S}_{t_2} \times \widehat{\Theta}_{t_2}) \to C(\mathcal{S}_{t_1} \times \widehat{\Theta}_{t_1})\}_{(t_1,t_2)\in T^2_{\leq}}]$. Then, we can construct the observable $\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, 2^{\Pi_{t\in T} X_t}, \widetilde{F}_0)$ in $C(\mathcal{S}_0 \times \widehat{\Theta}_0)$, which is the realization of the sequential observable $[\mathbf{O}_T]$, such as

$$
\begin{array}{ccccccccc}
[C(\mathcal{S}_0 \times \widehat{\Theta}_0)] & \xleftarrow{\widehat{\widehat{\Phi}}_{0,1}} & [C(\mathcal{S}_1 \times \widehat{\Theta}_1)] & \xleftarrow{\widehat{\widehat{\Phi}}_{1,2}} \cdots \xleftarrow{\widehat{\widehat{\Phi}}_{N-2,N-1}} & [C(\mathcal{S}_{N-1} \times \widehat{\Theta}_{N-1})] & \xleftarrow{\widehat{\widehat{\Phi}}_{N-1,N}} & [C(\mathcal{S}_N \times \widehat{\Theta}_N)] \\
F_0 & & F_1 & \cdots & F_{N-1} & & F_N \\
\downarrow & & \downarrow & & \downarrow & & \downarrow \\
\boxed{\begin{array}{c}(F_0 \overset{\text{qp}}{\times} \widehat{\widehat{\Phi}} \widetilde{F}_1) \\ = \widetilde{F}_0\end{array}} & \xleftarrow{\widehat{\widehat{\Phi}}_{0,1}} & \begin{array}{c}(F_1 \overset{\text{qp}}{\times} \widehat{\widehat{\Phi}} \widetilde{F}_2) \\ = \widetilde{F}_1\end{array} & \xleftarrow{\widehat{\widehat{\Phi}}_{1,2}} \cdots \xleftarrow{\widehat{\widehat{\Phi}}_{N-2,N-1}} & \begin{array}{c}(F_{N-1} \overset{\text{qp}}{\times} \widehat{\widehat{\Phi}} \widetilde{F}_N) \\ = \widetilde{F}_{N-1}\end{array} & \xleftarrow{\widehat{\widehat{\Phi}}_{N-1,N}} & \begin{array}{c}(F_N) \\ = \widetilde{F}_N\end{array}
\end{array}
\tag{8.55}
$$

(The existence of the $\widetilde{\mathbf{O}}_0 \equiv (\prod_{t\in T} X_t, 2^{\Pi_{t\in T} X_t}, \widetilde{F}_0)$ 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

(♯) 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 $\{B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t} \mid \prod_{t\in T} \Xi_t \in 2^{\Pi_{t\in T} X_t}\}$ 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

$$
\nu^{\mathcal{S}\times\widehat{\Theta}_\tau}_{\tau,\text{new}} = R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\delta_{s_0} \otimes \nu_0^{\widehat{\Theta}})
$$

where $R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t} : \mathcal{M}^m_{+1}(\mathcal{S}_0 \times \widehat{\Theta}_0) \to \mathcal{M}^m_{+1}(\mathcal{S}_\tau \times \widehat{\Theta}_\tau)$ is a normalized dual Bayes operator, i.e., $R^{(0,\tau)}_{\Pi_{t\in T}\Xi_t}(\nu) = \frac{(B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t})^*(\nu)}{\|(B^{(0,\tau)}_{\Pi_{t\in T}\Xi_t})^*(\nu)\|}$ $(\forall \nu \in \mathcal{M}^m_{+1}(\mathcal{S}_0 \times \widehat{\Theta}_0))$. 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*

$$
\begin{aligned}
&H\Big(\mathbf{M}_{C(\Omega)}(\mathbf{O}, S(\rho_0))\Big) \\
&= \sum_{n=1}^{\infty} \Big(\int_\Omega [F(\{x_n\})](\omega)\rho_0(d\omega) \int_\Omega \frac{[F(\{x_n\})](\omega)}{\int_\Omega [F(\{x_n\})](\omega)\rho_0(d\omega)} \log \frac{[F(\{x_n\})](\omega)}{\int_\Omega [F(\{x_n\})](\omega)\rho_0(d\omega)} \rho_0(d\omega) \Big) \qquad (8.56) \\
&= \sum_{n=1}^{\infty} \cdot I(\{x_n\})
\end{aligned}
$$

$$
\begin{aligned}
\text{where, } P(\{x_n\}) &= \int_\Omega [F(\{x_n\})](\omega)\rho_0(d\omega) \\
&\Big(= \textit{ the probability that a measured value } x_n \textit{ is obtained} \Big)
\end{aligned}
$$

$$
\begin{aligned}
I(\{x_n\}) &= \int_\Omega \frac{[F(\{x_n\})](\omega)}{\int_\Omega [F(\{x_n\})](\omega)\rho_0(d\omega)} \log \frac{[F(\{x_n\})](\omega)}{\int_\Omega [F(\{x_n\})](\omega)\rho_0(d\omega)} \rho_0(d\omega) \\
&= \frac{1}{P(\{x_n\})} \int_\Omega [F(\{x_n\})](\omega) \log [F(\{x_n\})](\omega)\rho_0(d\omega) - \log P(\{x_n\}) \\
&\Big(= \textit{ the information quantity when a measured value } x_n \textit{ is obtained} \Big) \qquad (8.57)
\end{aligned}
$$

$\mathbf{M}_{C(\Omega)}\big(\mathbf{O}, S(\rho_0)\big)$ *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* $H\Big(\mathbf{M}_{C(\Omega)}\ \big(\mathbf{O}, S(\rho_0)\big)\Big)$.

■

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 $\overline{\rho}_x$ ($\in \mathcal{M}_{+1}^m(\Omega)$) 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:

$$H(\mathbf{M}) = \sum_{n=1}^{\infty} \int_\Omega [F(\{x_n\})](\omega) \log[F(\{x_n\})](\omega)\rho_0(d\omega) - \sum_{n=1}^{\infty} P(\{x_n\}) \log P(\{x_n\}). \quad (8.58)$$

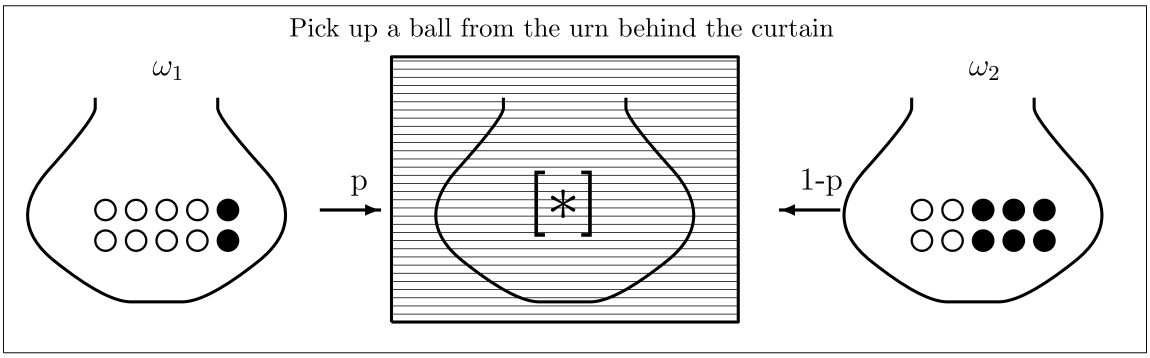
Also, if $\mathbf{O}$ is crisp, we see that $H(\mathbf{M}) = -\sum_{n=1}^{\infty} P(\{x_n\}) \log P(\{x_n\})$.

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 $\nu_0(\in \mathcal{M}_{+1}^m(\Omega))$ 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

$P(\{x\})$: the probability that a measured value x ($\in \{w, b\}$) is obtained

$I(\{x\})$: the information quantity that is acquired when we know that a measured value x ($\in \{w, b\}$) is obtained

ν_1^x : the posttest state after a measured value x ($\in \{w, b\}$) is obtained

Then,

$$P(\{w\}) = 0.8p + 0.4(1-p), \qquad P(\{b\}) = 0.2p + 0.6(1-p),$$

$$I(\{w\}) = \frac{0.8p \log 0.8 + 0.4(1-p)\log 0.4}{0.8p + 0.4(1-p)} - \log(0.8p + 0.4(1-p)),$$
$$I(\{b\}) = \frac{0.2p \log 0.2 + 0.6(1-p)\log 0.6}{0.2p + 0.6(1-p)} - \log(0.2p + 0.6(1-p)),$$

$$\nu_1^w(\{\omega_1\}) = \frac{0.8p}{0.8p + 0.4(1-p)}, \qquad \nu_1^w(\{\omega_2\}) = \frac{0.4(1-p)}{0.8p + 0.4(1-p)},$$
$$\nu_1^b(\{\omega_1\}) = \frac{0.2p}{0.2p + 0.6(1-p)}, \qquad \nu_1^b(\{\omega_2\}) = \frac{0.6(1-p)}{0.2p + 0.6(1-p)}.$$

Then, we see, by (8.58), that

$$\begin{aligned}
&H(\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0))) \\
=&[F(\{w\})](\omega_1)\log[F(\{w\})](\omega_1)p + [F(\{w\})](\omega_2)\log[F(\{w\})](\omega_2)(1-p) \\
&\quad + [F(\{b\})](\omega_1)\log[F(\{b\})](\omega_1)p + [F(\{b\})](\omega_2)\log[F(\{b\})](\omega_2)(1-p) \\
&\quad - P(\{w\})\log P(\{w\}) - P(\{b\})\log P(\{b\}) \\
=&0.8(\log 0.8)p + 0.4(\log 0.4)(1-p) + 0.2(\log 0.2)p + 0.6(\log 0.6)(1-p) \\
&\quad - (0.8p + 0.4(1-p))\log(0.8p + 0.4(1-p)) - (0.2p + 0.6(1-p))\log(0.2p + 0.6(1-p)).
\end{aligned}$$

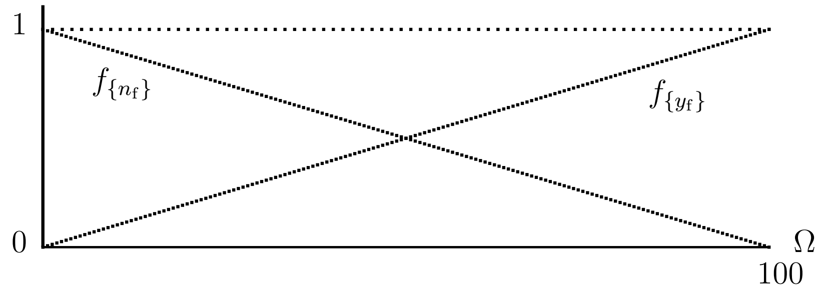
Assume that $p = 1/2$. Then, we see that

$$H(\mathbf{M}_{C(\Omega)}(\mathbf{O}, S_{[*]}(\nu_0))) = 0.6 - 0.3\log_2 3 = 0.123\cdots (\text{bit}).$$

■

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 $\mathbf{O}_\mathrm{b} \equiv (X = \{y_\mathrm{b}, n_\mathrm{b}\}, 2^X, b_{(\cdot)})$ be the crisp C^*-observable in the commutative C^*-algebra $C(\Omega)$ such that $b_{\{y_\mathrm{b}\}}(\omega_n) = 0$ (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

$$
\begin{aligned}
& H\Big(\mathbf{M}_{C(\Omega)}(\mathbf{O}_{\rm f}, S(\nu_u))\Big) \\
&= \int_\Omega f_{\{y_{\rm f}\}}(\omega) \log f_{\{y_{\rm f}\}}(\omega) \nu_u(d\omega) + \int_\Omega f_{\{n_{\rm f}\}}(\omega) \log f_{\{n_{\rm f}\}}(\omega) \nu_u(d\omega) \\
&\qquad - \|f_{\{y_{\rm f}\}}\|_{L^1(\Omega,\nu_u)} \log \|f_{\{y_{\rm f}\}}\|_{L^1(\Omega,\nu_u)} - \|f_{\{n_{\rm f}\}}\|_{L^1(\Omega,\nu_u)} \log \|f_{\{n_{\rm f}\}}\|_{L^1(\Omega,\nu_u)} \qquad (8.60) \\
&\approx 2\int_0^1 \lambda \log_2 \lambda d\lambda + 1 = -\frac{1}{2\log_e 2} + 1 = 0.278\cdots \ \text{(bit)}. \qquad (8.61)
\end{aligned}
$$

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):

$$
\text{MT}\begin{cases}
\text{PMT} = \underset{\text{[Axiom 1 (2.37)]}}{\text{measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} & \longleftrightarrow (A_1) \\
\text{SMT} = \underset{\text{(Axioms 1 and 2)}}{\text{PMT}} + \underset{\text{(the probabilistic interpretation of mixed state)}}{\text{“statistical state”}} & \longleftrightarrow (A_2) \\
\text{BMT} = \underset{\text{(Axioms 1 and 2)}}{\text{PMT}} + \underset{\text{(the principle of equal weight)}}{\text{“belief weight”}} & \longleftrightarrow (A_3)
\end{cases} \tag{8.62}
$$

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:

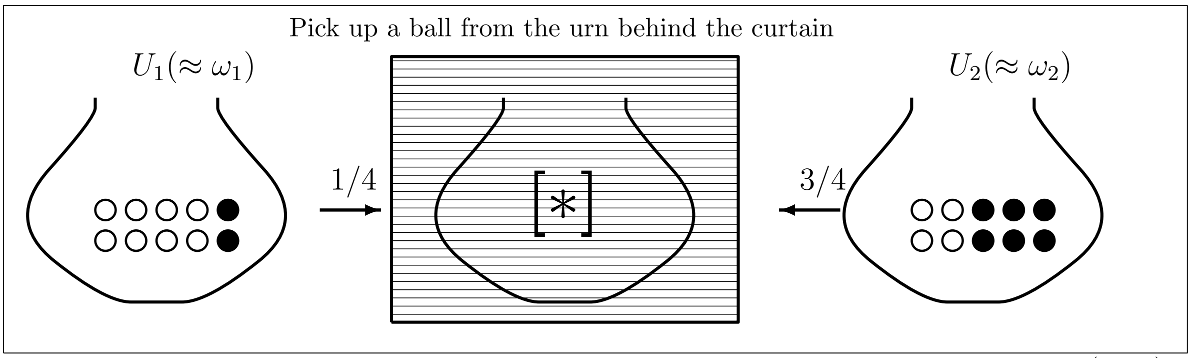
$$
\underset{\text{(mixed state)}}{\rho^m \in \mathfrak{S}^m(\mathcal{A}^*)} \cdots \begin{cases}
\underset{\text{[Proclaim 1 (8.10)]}}{\text{“probabilistic interpretation”}} & \to \underset{\text{in §8.1∼8.5}}{\text{“SMT”}} \\
\underset{\text{[the principle of equal weight (8.72)]}}{\text{“belief interpretation”}} & \to \underset{\text{in this §8.6}}{\text{“BMT”}}
\end{cases} \tag{8.63}
$$

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

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

and, $\nu_0(\{\omega_1\}) = 1/4$ and $\nu_0(\{\omega_1\}) = 3/4$. Recall that this measurement is symbolically described as follows.

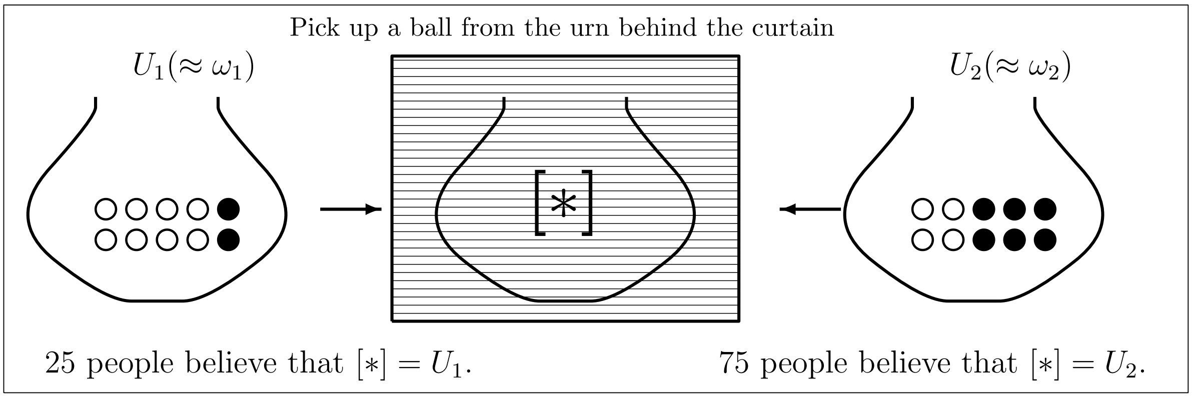


Figure(8.64)

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 $\Big(X, 2^X, {}_{\mathcal{M}(\Omega)}\Big\langle \nu, F(\cdot)\Big\rangle_{C(\Omega)}\Big)$ 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.

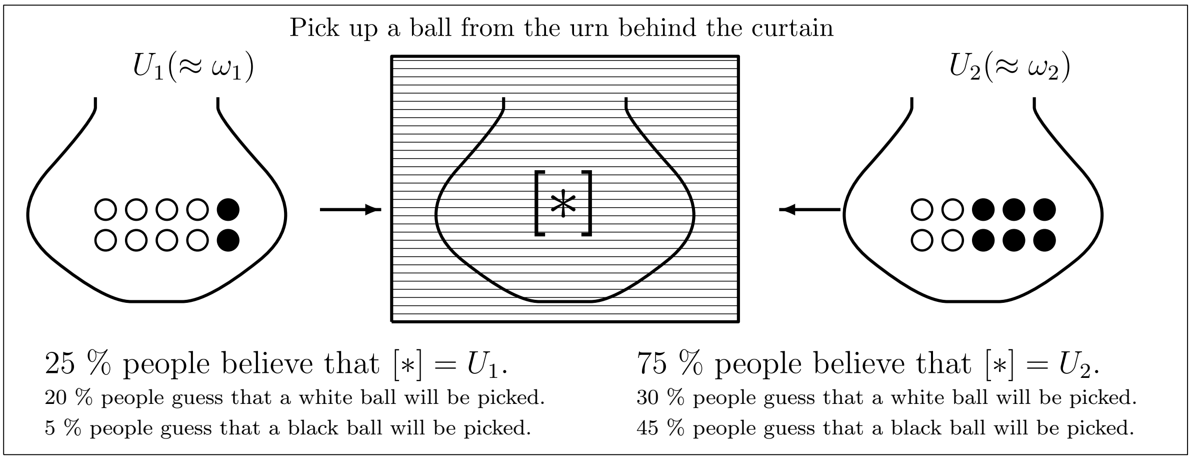
belongs to Ξ ($\in \mathcal{F}$)*. Then, we have the "Bayes theorem" such that*

$$\mathcal{M}_{+1}^m(\Omega) \ni \nu(= \textit{priori belief weight}) \mapsto \textit{(posterior belief weight =)} R_{\Xi}^{(0,0)}(\nu) \in \mathcal{M}_{+1}^m(\Omega). \tag{8.66}$$

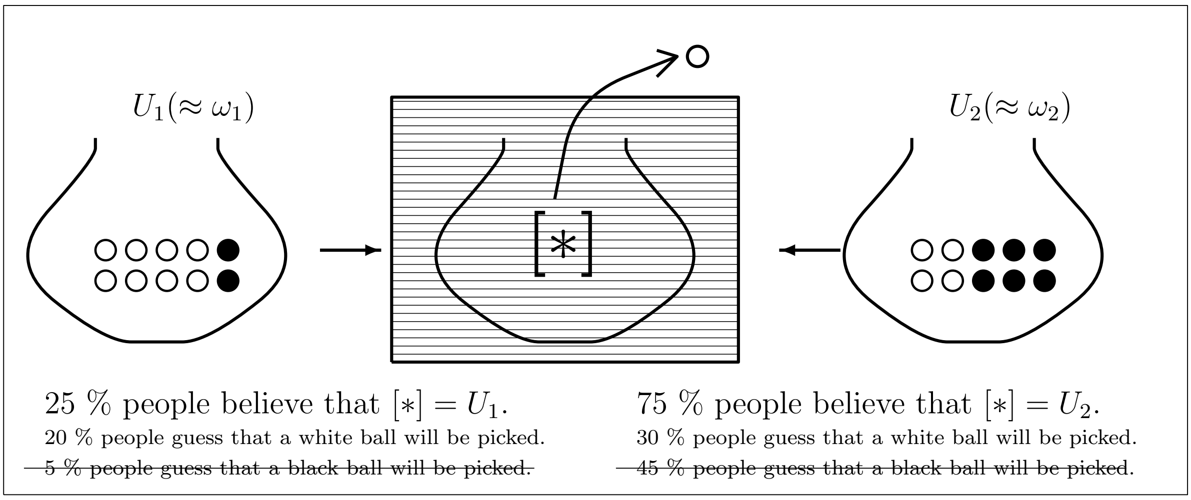
where

$$[R_{\Xi}^{(0,0)}(\nu)](D_0) = \frac{\int_{D_0}[F(\Xi)](\omega)\nu(d\omega)}{\int_{\Omega}[F(\Xi)](\omega)\nu(d\omega)} \qquad (\forall D_0 \subseteq \Omega;\ \textit{Borel set}\ ). \tag{8.67}$$

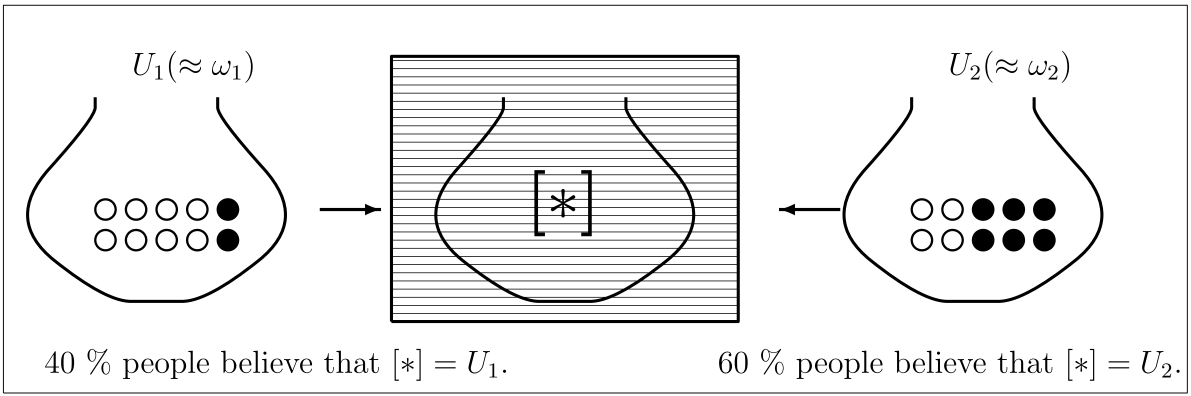
Proof. It suffices to prove a simple case since the proof of the general case is similar. For example, consider the following figure, which is essentially the same as Figure (8.65).



Assume that a "*white ball*" is picked in the above picture. Then, we see:



which is equivalent to the following figure:



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

$$H\left(\rho_0^m\right) = -\sum_{n=1}^{N} \rho_0^m(\{\omega_n\}) \log \rho_0^m(\{\omega_n\}).$$

Here, it is well known that

(i) $\sup\left\{H\left(\rho_0^m\right) : \rho_0^m \in \mathcal{M}_{+1}^m(\Omega)\ \right\} = \log N,$ (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) $\quad T_{av}(1) = 1$

(b) $\quad T_{av}(f) = T_{av}(f \circ \phi) \qquad (\forall f \in C(\Omega), \forall \text{ bijection } \phi : \Omega \to \Omega)$

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 $T_{av}(f) = \int_\Omega f(\omega)\nu_u(d\omega) \left(\equiv \frac{\sum_{n=1}^N f(\omega_n)}{N} \right)$ $(\forall f \in C(\Omega))$.

■

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_{new}^m $(\in \mathcal{M}_{+1}^m(\Omega))$ such that $\rho_{\text{new}}^m(B) = \frac{\int_B [F(\{x\})](\omega)\rho_0^m(d\omega)}{\int_\Omega [F(\{x\})](\omega)\rho_0^m(d\omega)}$ $(\forall B \in \mathcal{B}_\Omega$, Borel field).

Define the map $[R_{\{x\}}^{(0,0)}] : \mathcal{M}_{+1}^m(\Omega) \to \mathcal{M}_{+1}^m(\Omega)$ such that:

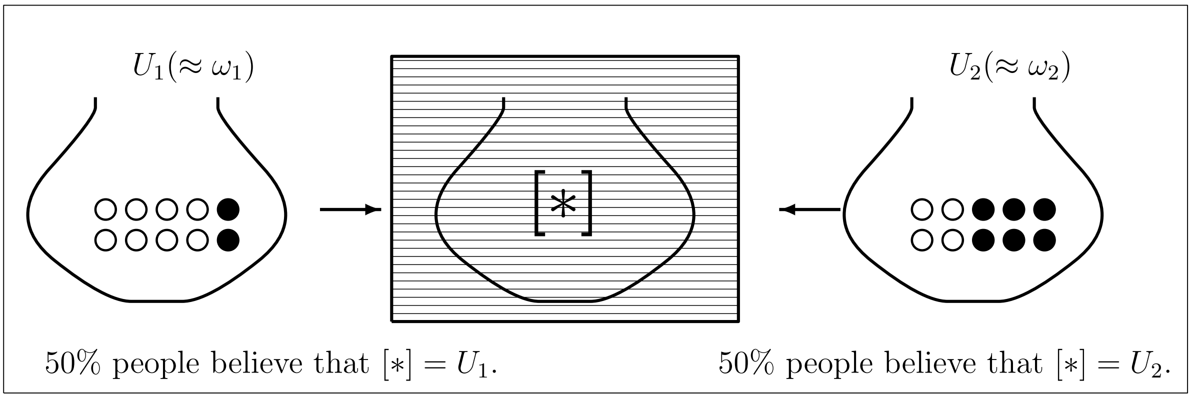
$$[R_{\{x\}}^{(0,0)}](\rho^m) = \frac{\int_{D_0}[F(\{x\})](\omega)\nu(d\omega)}{\int_\Omega [F(\{x\})](\omega)\nu(d\omega)} \qquad (\forall D_0 \subseteq \Omega;\ \text{Borel set}\ ). \tag{8.73}$$

Then, we can symbolically describe it as follows:

$$[\text{BMT}] = \begin{cases} [\text{BMT}_1] \quad \text{the loosest belief weight} \quad \longleftrightarrow \quad \nu_u(\in \mathcal{M}_{+1}^m(\Omega)) \\ \\ [\text{BMT}_2] \quad S_{[*]}((\rho))_{bw} \xrightarrow[x \text{ is obtained}]{\mathbf{M}_{C(\Omega)}(\mathbf{O},\ S_{[*]}((\rho))_{bw})} S_{[*]}(([R_{\{x\}}^{(0,0)}](\rho)))_{bw}, \end{cases} \tag{8.74}$$

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:

$$\text{MT}\begin{cases} \text{PMT} = \underset{\text{[Axiom 1 (2.37)]}}{\text{measurement}} + \underset{\text{[Axiom 2 (3.26)]}}{\text{the relation among systems}} \\ \text{SMT} = \underset{\text{(Axioms 1 and 2)}}{\text{PMT}} + \underset{\text{(the probabilistic interpretation of mixed state)}}{\text{"statistical state"}} \\ \text{BMT} = \underset{\text{(Axioms 1 and 2)}}{\text{PMT}} + \underset{\text{(the principle of equal weight)}}{\text{"belief weight"}} \end{cases} \tag{8.77}$$

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 ∼ 8.6.2). Especially, note that Proclaim 1 (8.10) is not valid in BMT.

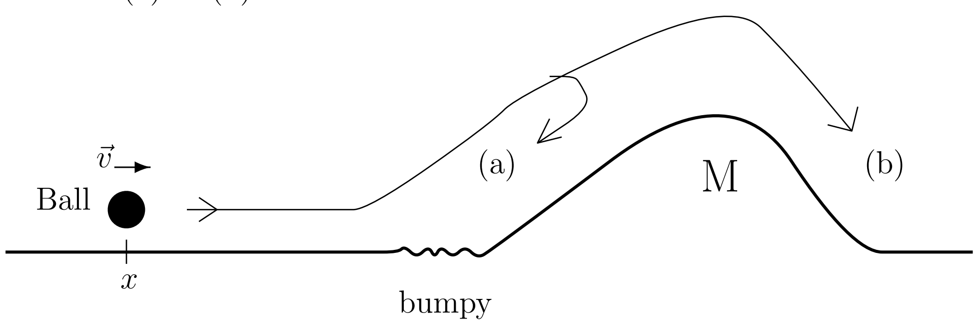
If this (R) is admitted, I agree to the above opinion (A). Thus, I recommend readers to use BMT at least until becoming accustomed to BMT. Also, it should be noted that there is a great confusion in the conventional statistics.

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 ($\in \mathcal{M}_{+1}^m(\Omega)$, 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.

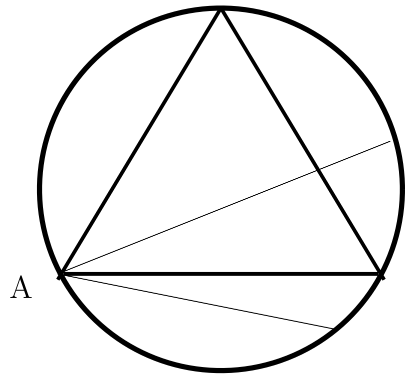
[15]The ν_u is invariant concerning any bijection ϕ on Ω, i.e., $\phi(\nu_u) = \nu_u$. In this sense, it is natural.

the following problem:

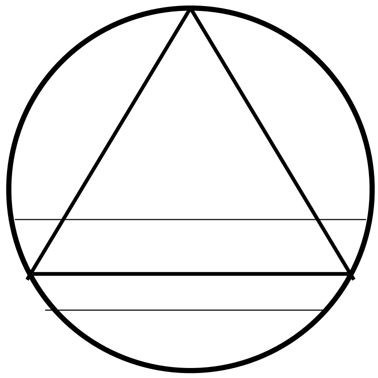
(P_1) Given a circle with the radius 1. Suppose a chord of the circle is chosen at random. What is the probability that the chord is longer than $\sqrt{3}$ (i.e., the side of an inscribed equilateral triangle)?

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

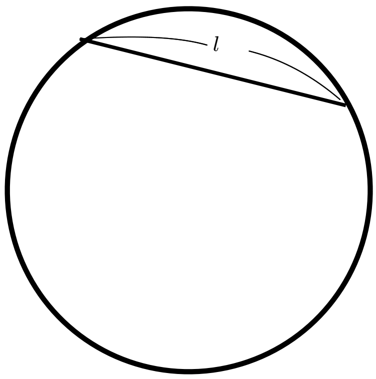
($\sharp$) If Bertrand's paradox is a paradox (i.e., if the argument in §8.7.1 is considered to be strange), it is due to the confusion between statistical states and mixed states (*cf.* (8.11)).

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 ($\in \mathcal{M}_{+1}^m(\Omega)$).

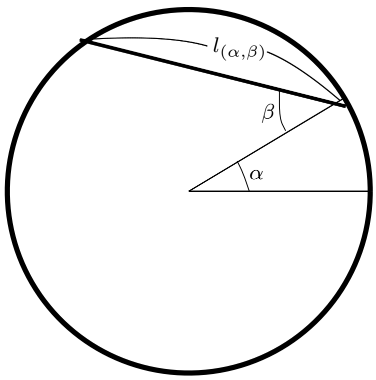
The reader will find that the (P_2) is essentially the same as the problem (P_1) in §8.7.1. Thus, the above problem has also apparently several solutions as follows:

(Fig.0)

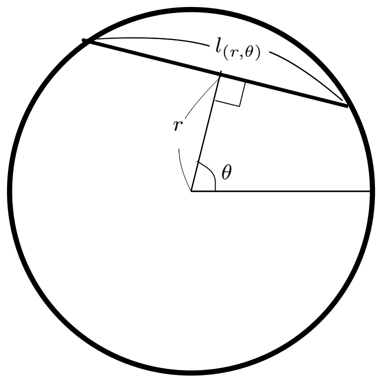


Represent a chord l
by a natural coordinate!

(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 $R_1 \equiv \{(\alpha, \beta) \mid 0 \leq \alpha \leq 2\pi,\ 0 \leq \beta \leq \pi/2(\text{radian})\}$. 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 $(\in \mathcal{M}_{+1}^m(\Omega))$ 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.