PTEP

Prog. Theor. Exp. Phys. **2021** 123A01(18 pages)
DOI: 10.1093/ptep/ptab135

The Born rule and logical inference in quantum mechanics

Tsubasa Ichikawa

Mukogaoka 1-10-6-1004, Tokyo 113-0023, Japan
E-mail: ichikawa@qiqb.osaka-u.ac.jp.
Present address: Center for Quantum Information and Quantum Biology, Osaka University, Osaka 560-0043, Japan.

Received February 13, 2019; Revised August 25, 2020; Accepted October 19, 2021; Published October 21, 2021

Logical inference leads to one of the major interpretations of probability theory called logical interpretation, in which the probability is seen as a measure of the plausibility of a logical statement under incomplete information. Assuming that our usual inference procedure is working rationally for every set of logical propositions represented in terms of commuting projectors on a given Hilbert space, we extend the logical interpretation to quantum mechanics and derive the Born rule. Our result implies that, from epistemological viewpoints, we can regard quantum mechanics as a natural extension of classical probability theory.

Subject Index A60, A61, A64

1. Introduction

Inference is one of the essential building blocks in various branches of mathematical science: Jaynes' derivation of statistical mechanics [1,2] is based on the inference of the probability distribution which is most likely to reproduce given expectations of thermodynamic variables. Since the era of Laplace [3], probability has been more or less regarded as a measure of likelihood of statements being valid, although there coexists the interpretation that the probability is nothing but the relative frequency.

Inference plays an important role even in understanding fundamental aspects of probability theory. The following was shown by Cox [4–9]: Suppose there is a measure of the plausibility of statements being valid under incomplete information. If the measure follows inference rules consistent with our common sense, then the measure satisfies the product rule and the sum rule of probability, and hence it can be interpreted as the probability. In other words, without invoking the notion of relative frequency, probability theory can be derived by assuming reasonable inference rules under incomplete information. This derivation is called Cox's theorem, which allows us to see probability as a natural extension of the truth values of propositions. This interpretation is called the logical interpretation of probability [10].

Recently, several inference methods are coming to be used to reconstruct quantum mechanics. The equations of motion such as the Schrödinger equation [11], the Pauli equation [12], and the Klein–Gordon equation [13] are derived under the logical interpretation, augmented with the auxiliary requirement that the plausibility of the experimental outcomes is robust under slight changes of the experimental parameters. The wave function is also derived with other auxiliary assumptions [14,15].

In this paper we show that the measure of plausibility takes the form of the Born rule by extending the argument mentioned above to Hilbert space and the projectors thereof. Interestingly enough, the inference rule leads to the unique concrete expression of the quantum conditional probability associated with the Lüders conditionalization [16], which specifies the post-measurement quantum state. More precisely, the Lüders conditionalization requires that after a projection measurement Q on a density matrix ρ, the state changes to $\rho_Q = Q\rho Q/\mathrm{tr}(\rho Q)$, whereas we will find that the quantum conditional probability in accordance with the inference rule takes the form $\Pr(P \mid Q) = \mathrm{tr}(\rho_Q P) = \mathrm{tr}(Q\rho Q P)/\mathrm{tr}(\rho Q)$, where P is a projection operator. Thus, the quantum conditional probability is seen as the Born rule for the density matrix ρ_Q, which is obtained after the measurement of Q. We hereafter call the expression of the quantum conditional probability "Lüders rule," in order to stress the relationship between the Lüders conditionalization and the quantum conditional probability [17].

Analogously to the derivation of Cox's theorem, throughout our derivation of both rules we will assume that the plausibility measures follow our standard inference procedures for commuting projectors. Our derivation, hence, enables us to see both the rules as the measures of plausibility on which our inference makes sense for the commuting projectors.

The paper is organized as follows. In Sect. 2 we review two formulations of probability theory. In Sect. 3, we present three approaches to interpretations of probability theory. In Sect. 4, one of those interpretations, Cox's approach to the logical interpretation on the basis of inference rules, is studied in detail. In Sect. 5, we extend Cox's approach in quantum mechanics. In Sect. 6, we characterize the Lüders rule as the unique conditional probability satisfying the inference rule. In Sect. 7, we give an example where the inference rule employed in Cox's approach is no longer valid for non-commuting projectors. Section 8 is devoted to our conclusion and discussions.

2. Two formulations of probability theory

Probability and conditional probability are central notions in probability theory. Depending on which one we take as the more fundamental concept, we have two formulations of probability theory. In this section we give a brief review of them, since we need both in the following argument.

We hereafter deal with both formulations based on the finite additive class, which is defined as follows. Suppose we are given a countable infinite number of mutually different elements $\omega_1, \omega_2, \ldots$. The set of all the elements is denoted by $\Omega := \{\omega_1, \omega_2, \ldots\}$, and subsets thereof are written as $A_1, A_2, \ldots$. A set $\mathcal{F}$ of the subsets $A_1, A_2, \ldots$ is called the finitely additive class if it satisfies

$$\begin{aligned} &\emptyset \in \mathcal{F}, \\ &A_i \in \mathcal{F} \Rightarrow A_i^{\mathrm{c}} \in \mathcal{F}, \\ A_1, A_2, \ldots, A_n \in \mathcal{F} \Rightarrow &\bigcup_{i=1}^{n} A_i \in \mathcal{F}. \end{aligned} \tag{1}$$

Here, $\emptyset$ stands for the empty set, and A^{c} is the complement of A defined through $A \cap A^{\mathrm{c}} = \emptyset$ and $A \cup A^{\mathrm{c}} = \Omega$. The set $\mathcal{F}$ satisfying Eqs. (1) for infinite n is called a σ-algebra.

Let us turn to the two formulations mentioned above. In the first one, which we hereafter call the probability-based formulation, the probability is given first, and the conditional probability is defined as the ratio thereof. More precisely, the probability $\Pr(A)$ is defined as a map from $\mathcal{F}$ to $[0, 1] = \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$ which satisfies

$$\Pr(A) \geq 0 \qquad \text{for all } A \in \mathcal{F}, \tag{2a}$$

$$\Pr(\Omega) = 1, \tag{2b}$$

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \Pr(A_i) \tag{2c}$$

if $\{A_i\}_{i=1}^{n}$ is a family of mutually disjoint subsets of Ω: $A_i \cap A_j = \emptyset$ for $i \neq j$. The conditional probability $\Pr(A \mid B)$ is given as

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} \tag{3}$$

when $\Pr(B) \neq 0$.

The second formulation [18–20], in contrast, is based on the conditional probability. In this formulation, we introduce the finite additive class $\mathcal{F}$ and its non-empty subset $\mathcal{G}$, and define the conditional probability $\Pr(A \mid B)$ for $A \in \mathcal{F}$ and $B \in \mathcal{G}$ as a function which takes a value in [0, 1] and fulfills

$$\Pr(B \mid B) = 1 \qquad \text{for all } B \in \mathcal{G}, \tag{4a}$$

$$\Pr\left(\bigcup_{i=1}^{n} A_i \mid B\right) = \sum_{i=1}^{n} \Pr(A \mid B) \tag{4b}$$

for the set $\{A_i\}_{i=1}^{n}$ of mutually disjoint elements of $\mathcal{F}$ and $B \in \mathcal{G}$, and

$$\Pr(A \mid B) = \frac{\Pr(A \cap B \mid C)}{\Pr(B \mid C)} \tag{4c}$$

for $A \in \mathcal{F}$ and $B, C \in \mathcal{G}$ such that $B \subset C$ and $\Pr(B \mid C) > 0$.

The formulation based on Eqs. (4) is a natural extension of the probability-based formulation. It is shown in Ref. [19] that the conditional probability $\Pr(A \mid B)$ in the sense of Eq. (4) can be seen as the probability when B is fixed. Thus, we utilize the notation $\Pr(A)$ for the special case $B = \Omega$, that is,

$$\Pr(A) = \Pr(A \mid \Omega). \tag{5}$$

Conversely, the conditional probability in Eq. (3) satisfies Eqs. (4). Note that not all the conditional probabilities in the sense of Eqs. (4) take the form of Eq. (3). These two formulations are extended to the σ-algebra for infinite n in Eqs. (2) and (4), respectively.

3. Interpretations of probability

For interpreting probability theory, we show the existence of a mathematical quantity which is easy to understand intuitively and fulfills the conditions of Eqs. (2) or (4), depending on which formulation we take. Here we take three examples of well-known interpretations of probability: that is, frequency interpretation, subjective interpretation, and logical interpretation [21]. Note that the frequency interpretation and subjective interpretation are applicable to the finitely additive class as well as σ-algebra [22]. On the other hand, in the logical probability it is shown to be hard to implement the finitely additive class, due to Assumption 1 below, which requires an infinite number of elementary propositions [7].

3.1 *Frequency interpretation*

First, let us show that the relative frequency of the events occurring satisfies the conditions in Eq. (2). Given the finite set Ω, the relative frequency of A_i is defined by the ratio of the number of events that A_i occurs to of the total number of events. It is obvious that the relative frequency satisfies Eq. (2). Thus, we may interpret the probability as the relative frequency.

In the frequency interpretation, we need a well-defined ensemble of repeatable events to define the probability distributions. On the other hand, it has been pointed out that the frequency interpretation is not applicable to one-shot events or the measurement of physical constants [23,24]. The following two interpretations provide useful tools for the analyses of such cases.

3.2 *Subjective interpretation*

The second example is the Dutch book argument (DBA) [25]. In DBA, we consider a bet on whether a given hypothesis (it rains tomorrow, for example) is true. If the hypothesis is true, then the bettor obtains the stake S. If it is not the case, they obtain nothing. Let us now suppose that the bettor pays a wager qS, where q is called the betting quotient. Then the net payoff of the bet is $S - qS$ if the hypothesis is true, and $-qS$ otherwise.

In Ref. [25], de Finetti showed that the bookie can construct a set of bets in such a way that the bettor loses in any combination of betting outcomes (the truth values of the given hypotheses, in other words), if the set of betting quotients does not obey the conditions in Eq. (2). In other words, as far as the bettor is rational in the sense that they wish to avoid a net loss, the betting quotients have to obey the axioms of probability theory: the fair betting quotients are equal to the probability. See Refs. [21,26] for proof of the mathematical relation between the fair betting quotients and the axioms of the probability.

DBA leads to the subjective interpretation of the probability, since the betting quotients indicate how tru the bettor feels the hypotheses to be. The salient feature of DBA is that the bettor can freely determine their own betting quotients: in a bet, it is not necessary that all the participants of the bets have the same betting quotients. This implies that in the subjective interpretation we may assign several different probabilities (degrees of belief) for the plausibility of a hypothesis by our own decision.

3.3 *Logical interpretation*

The third example is the degree of plausibility of a given proposition conditioned by the prior information [4–9], which is an extension of the truth value, and the central issue of this paper. In the logical interpretation, we are given the propositions, and suppose the existence of a measure of plausibility of the proposition, on which inference works rationally. Then, there exists a function which maps the measure of plausibility to the conditional probability function satisfying Eqs. (4), suggesting that the conditional probability quantifies the plausibility of the proposition being valid.

4. Conditional probability function in the logical interpretation

In what follows, we show that the degree of plausibility is mapped to the conditional probability. This implies that the conditional probability is seen as a natural extension of the truth value of the proposition under uncertainty. Boolean operations provide a basis for this interpretation, since they characterize the relations among the (composite) propositions made through the logical operations. Given the truth values of several propositions, we can find the truth values of the composite propositions associated with them by performing Boolean operations. Since we make the argument to see conditional probability as an extension of the truth value, we provide a discussion on Boolean algebra below.

A six-tuple consisting of a set $\mathcal{S} = \{A, B, C, \ldots\}$ with the Boolean operations and ($\wedge$), or ($\vee$), and not ($\neg$), and two elements $\circledcirc, \Box \in \mathcal{S}$, is called a Boolean algebra if the following properties hold [27]:

(a) Commutativity:

$$A \wedge B = B \wedge A \in \mathcal{S}, \qquad A \vee B = B \vee A \in \mathcal{S}. \tag{6}$$

(b) Distributivity:

$$A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C) \in \mathcal{S}, \qquad A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C) \in \mathcal{S}. \tag{7}$$

(c) Identity:

$$A \wedge \Box = A \in \mathcal{S}, \qquad A \vee \circledcirc = A \in \mathcal{S}. \tag{8}$$

(d) Complements:

$$A \wedge \neg A = \circledcirc \in \mathcal{S}, \qquad A \vee \neg A = \Box \in \mathcal{S}. \tag{9}$$

Here, $\circledcirc$ and $\Box$ are the identity with respect to the or and and operations, respectively. In what follows, $\circledcirc$ plays the role of the identically false proposition, whereas $\Box$ represents the identically true proposition.

Equipped with the Boolean operations and, or, and not by

$$A \wedge B := A \cap B, \qquad A \vee B := A \cup B, \qquad \neg A := A^{\mathrm{c}}, \tag{10}$$

respectively, the finitely additive class $\mathcal{F} = \{A, B, C, \ldots\}$ mentioned earlier satisfies Eqs. (6), (7), (8), and (9) through the identification $\circledcirc := \emptyset \in \mathcal{F}$ and $\Box := \Omega \in \mathcal{F}$, showing that the finite additive class $\mathcal{F}$ with the Boolean operations is a concrete Boolean algebra. This implies that we may establish the logic on the finitely additive class, because the Boolean operation represents the relation among the logical propositions. On the basis of this observation, we may safely regard the element $\omega_i \in \Omega$ as an elementary proposition, which could be true or false. The subsets A, B, … in the finitely additive class $\mathcal{F}$, which is the concrete Boolean algebra, are therefore the set of composite propositions made by the Boolean operations, and the set Ω is seen as the set of all of these.

We now formally introduce the measure of conditional plausibility, denoted by $\varepsilon(A \mid B) \in \mathbb{R}$, whose value quantifies the degree of plausibility that A is true, given that B is true. We suppose that re exists two real numbers F and T exist such that $\mathrm{F} \leq \varepsilon(A \mid B) \leq \mathrm{T}$ for every A and B. We hereafter show that (i) there exists a function $w(\cdot)$ through which $w(\varepsilon(A \mid B))$ as a function of A and B satisfies the conditions in Eqs. (4) under reasonable assumptions given below, and (ii) we may set $w(\mathrm{F}) = 0$ and $w(\mathrm{T}) = 1$: there exists a map $w(\cdot)$ from the degree of plausibility to the conditional probability.

Before we proceed, we point out that there are variants of the sets of assumptions used to prove Cox's theorem [4,5,8,9]. This is because Cox used implicit assumptions in his original proof [6,7], and several researchers tried to construct more reasonable proofs to fill the gap.

Among these variants, we adopt the assumptions given in Refs. [4,9] for clarity and rigor of presentation.

The first assumption is the following:

Assumption 1 (van Horn). *There exists a non-empty set of real numbers* P_0 *with the following two properties:*

(a) P_0 *is a dense subset of the interval* $[\mathrm{F}, \mathrm{T}]$.
(b) *For every* $y_1, y_2, y_3 \in \mathrm{P}_0$, *there exist propositions* A_1, A_2, A_3, B *such that* $y_1 = \varepsilon(A_1 \mid B)$, $y_2 = \varepsilon(A_2 \mid A_1 \wedge B)$, *and* $y_3 = \varepsilon(A_3 \mid A_2 \wedge A_1 \wedge B)$.

Although Assumption 1 looks intricate, it is necessary to exclude models which are not the standard probability theory, but satisfy all the assumptions we will hereafter make; indeed, such a model is constructed in Ref. [6]. Moreover, it is known that Assumption 1 requires infinite numbers of the elementary propositions [7].

We make the second assumption:

Assumption 2 (van Horn). *Let* A, B, *and* C *be propositions. Then there exists a continuous function* $F : [\mathrm{F}, \mathrm{T}]^2 \to [\mathrm{F}, \mathrm{T}]$ *which is strictly increasing in both the arguments and satisfies*

$$\varepsilon(A \wedge B \mid C) = F(\varepsilon(B \mid C),\ \varepsilon(A \mid B \wedge C)). \tag{11}$$

Assumption 2 implies that the plausibility of the composite proposition $A \wedge B$ given C being true is related to two measures of plausibility of the composite propositions $\varepsilon(B \mid C)$ and $\varepsilon(A \mid B \wedge C)$. In other words, we infer the plausibility of $\varepsilon(A \wedge B \mid C)$ from successive inferences using them.

From Assumptions 1 and 2, by using the properties of Boolean algebra, it is found in Ref. [9] that the function $F(x, y)$ satisfies

$$F(x, F(y, z)) = F(F(x, y), z). \tag{12}$$

Further, the following fact [28] is useful for obtaining the concrete expression of $F(x, y)$:

Lemma 1 (Aczél). *Let* a *and* b, *with* $a < b$, *be real numbers. Suppose that* f: $[a, b]^2 \to [a, b]$ *is a continuous function, strictly increasing in both arguments, and satisfies the associativity equation*

$$f(x, f(y, z)) = f(f(x, y), z) \tag{13}$$

for all $x, y, z \in (a, b]$. *Then there exists some continuous, strictly increasing function* $g(x)$ *such that*

$$g(f(x, y)) = g(x) + g(y). \tag{14}$$

Now we define $w(x) := e^{g(x)}$ and identify f with F. By exponentiating both sides of Eq. (14) and substituting $w(x)$, we obtain

$$F(x, y) = w^{-1}\left[w(x)w(y)\right]. \tag{15}$$

Note that substitution of Eq. (15) into Eq. (11) leads to

$$w(\varepsilon(A \wedge B \mid C)) = w(\varepsilon(B \mid C))w(\varepsilon(A \mid B \wedge C)), \tag{16}$$

which is an expression of the inference rule in Eq. (11) in terms of $w(\varepsilon(A \mid B))$. For the range of $w(x)$, we have $0 \leq w(x) \leq 1$, the proof of which is given in Ref. [9]. Since $w(x)$ is a strictly increasing function, this ensures that we may set $w(\mathrm{F}) = 0$ and $w(\mathrm{T}) = 1$, as mentioned above. Note that $w(x)$ is surjective.

Further, we make the following assumption on $w(x)$:

Assumption 3 (Cox). *There exists a twice differentiable function $S(x)$ such that*

$$w(\varepsilon(A \mid B)) = S(w(\varepsilon(\neg A \mid B))). \tag{17}$$

Assumption 3 means that the plausibility of a given proposition can be inferred from that of its negation. Cox found that Eq. (17) yields

$$S(x) = (1 - x^m)^{\frac{1}{m}}, \tag{18}$$

where m is a positive finite constant. Now we set $m = 1$ to obtain

$$w(\varepsilon(A \mid B)) + w(\varepsilon(\neg A \mid B)) = 1. \tag{19}$$

Equations (16) and (19) suffice to show that the conditional plausibility $w(\varepsilon(A|B))$ satisfies Eqs. (4). To show these, we need

$$w(\varepsilon(A \vee B \mid C)) = w(\varepsilon(A \mid C)) + w(\varepsilon(B \mid C)) - w(\varepsilon(A \wedge B \mid C)), \tag{20}$$

whose proof is given in the appendix.

We first prove Eq. (4c). Suppose $B \subset C$, which leads to $B \wedge C = B$. Then it follows from Eq. (16) that

$$w(\varepsilon(A \wedge B \mid C)) = w(\varepsilon(A \mid B))w(\varepsilon(B \mid C)), \tag{21}$$

which reduces to Eq. (4c) when $w(\varepsilon(B|C)) \neq 0$.

Furthermore, by setting $A = B = C$ in Eq. (21), we find $w(\varepsilon(A \mid A))^2 = w(\varepsilon(A \mid A))$, whose solutions are either 1 or 0. Therefore we obtain Eq. (4a) by taking the solution $w(\varepsilon(A \mid A)) = 1$ and replacing A with B.

Now we proceed to prove Eq. (4b). Substituting Ω for A in Eq. (16), we find

$$w(\varepsilon(\Omega \wedge B \mid C)) = w(\varepsilon(\Omega \mid C))w(\varepsilon(B \mid \Omega \wedge C)). \tag{22}$$

Since $\Omega \wedge B = B$ and $\Omega \wedge C = C$, Eq. (22) turns into

$$w(\varepsilon(\Omega \mid C)) = 1. \tag{23}$$

Furthermore, putting $A = \Omega$ in Eq. (19), we find

$$w(\varepsilon(\emptyset \mid C)) = 0. \tag{24}$$

Hence, from Eq. (20), we obtain

$$w(\varepsilon(A \vee B \mid C)) = w(\varepsilon(A \mid C)) + w(\varepsilon(B \mid C)) \qquad \text{if } A \wedge B = \emptyset. \tag{25}$$

By induction, we find

$$w\left(\varepsilon\left(\bigvee_{i=1}^{n} A_i \mid B\right)\right) = \sum_{i=1}^{n} w(\varepsilon(A_i \mid B)) \tag{26}$$

for the set $\{A_i\}_{i=1}^n$, whose elements are mutually disjoint. Note that the right-hand side of Eq. (26) converges, since the left-hand side is bounded.

The above argument guarantees that we may put

$$\Pr(A \mid B) = w(\varepsilon(A \mid B)), \tag{27}$$

meaning that if the degree of the plausibility $\varepsilon(A \mid B)$ satisfies the assumptions mentioned above, there exists a function $w(x)$ which maps $\varepsilon(A \mid B)$ to the conditional probability. Note that we have not used the repeatability or frequency in the above argument.

5. Quantum extension

In the following we extend the preceding argument to quantum mechanics. Let $\mathcal{H}$ be a (countably) infinite-dimensional Hilbert space, and P, Q, … be the projectors acting on $\mathcal{H}$. The set of all the projectors is denoted by $\mathcal{L}(\mathcal{H})$. The range of P is denoted by $\mathrm{Im}\,P := \{P|\psi\rangle| \mid |\psi\rangle \in \mathcal{H}\}$. Note that $\mathrm{Im}\,P$ is also a Hilbert space.

Since the eigenvalues of the projectors are either 1 or 0, the eigenvalues can be seen as the truth values, and hence every projector can be thought of as a proposition.

Analogously to $\varepsilon(A \mid B)$ in the preceding section, let us now introduce $\varepsilon(P \mid Q)$ as the degree of plausibility for the proposition P, given that Q is true. We suppose that $\varepsilon(P \mid Q)$ is bounded, as $\mathrm{F} \leq \varepsilon(P \mid Q) \leq \mathrm{T}$, and well-defined for all pairs of the projectors, even for those which do not necessarily commute.

Given the propositions P and Q, the composite projectors $P \wedge Q$ and $P \vee Q$ are defined as the projectors to the meet and join of $\mathrm{Im}\,P$ and $\mathrm{Im}\,Q$, respectively. The negation $\neg P$ is the projector on the orthogonal complement of $\mathrm{Im}\,P$, which is hereafter denoted by $\mathrm{Im}\,P^{\perp}$. Note that the set of all projectors $\mathcal{L}(\mathcal{H})$ is a lattice, since the meet and join of any pair of projectors are also elements of $\mathcal{L}(\mathcal{H})$.

We shall show that the inference rule $\varepsilon(P \mid Q)$ obeys leads to the Born rule, and the conditional probability satisfying the inference rule takes the form of the Lüders rule. For this purpose, it is convenient to construct a Boolean sublattice in a set of commuting projectors. When P and Q commute with each other, we have the explicit expressions of the logical operations by associating them with corresponding projectors:

$$P \wedge Q = PQ, \qquad P \vee Q = P + Q - PQ, \qquad \neg P = \mathbb{1} - P, \tag{28}$$

where $\mathbb{1}$ is the identity operator. Note that $P \vee Q = P + Q$ if P and Q are mutually orthogonal.

Similarly to Eqs. (1), we can define a set $\mathcal{C}$ of commutative projectors such that

$$\begin{aligned} &0 \in \mathcal{C}, \\ &P \in \mathcal{C} \Rightarrow \quad \neg P \in \mathcal{C}, \end{aligned} \tag{29}$$

$$P, Q \in \mathcal{C} \quad \Rightarrow \quad P \wedge Q \in \mathcal{C}. \tag{30}$$

It is clear that the elements in $\mathcal{C}$ satisfy the properties in Eqs. (6), (7), (8), and (9). Hence, by identifying $\circledcirc$ and $\square$ with $0 \in \mathcal{L}(\mathcal{H})$ and $\mathbb{1} \in \mathcal{L}(\mathcal{H})$, respectively, the set $\mathcal{C}$ is found to be a Boolean sublattice of the non-Boolean lattice $\mathcal{L}(\mathcal{H})$.

Three remarks are in order. First, $0, \mathbb{1}$ in $\mathcal{C}$ correspond to $\emptyset$, Ω in $\mathcal{F}$, respectively. Second, the set $\mathcal{C}$ is not unique: we can find infinitely many sets satisfying Eq. (29). Indeed, given a projector P, we can construct $\mathcal{C} = \{0, P, \neg P, \mathbb{1}\}$, implying that any projector is an element of some Boolean sublattice $\mathcal{C}$. Third, the join of all $\mathcal{C}$ gives $\mathcal{L}(\mathcal{H})$.

We now extend the definitions of the probability and conditional probability given in Sect. 2 to quantum mechanics. A non-negative function $\Pr(P \mid Q)$ of the projectors $P \in \mathcal{L}(\mathcal{H})$ and $Q \in \mathcal{K}$, where $\mathcal{K}$ is a non-empty subset of $\mathcal{L}(\mathcal{H})$, is a conditional probability if it satisfies

$$\Pr(Q \mid Q) = 1 \qquad \text{for all } Q \in \mathcal{K}, \tag{31a}$$

$$\Pr\left(\bigvee_{i=1}^{n} P_i \mid Q\right) = \sum_{i=1}^{n} \Pr(P_i \mid Q) \tag{31b}$$

for $P_i \in \mathcal{L}(\mathcal{H})$ and $Q \in \mathcal{K}$, where $P_i P_j = P_i \delta_{ij}$ for any i, j, and

$$\Pr(P \wedge Q \mid Q) = \frac{\Pr(P \wedge Q \mid R)}{\Pr(Q \mid R)} \tag{31c}$$

for $P \in \mathcal{L}(\mathcal{H})$ and $Q, R \in \mathcal{K}$ such that $Q < R$ and $\Pr(Q \mid R) > 0$. Here, δ_{ij} is the Kronecker delta, and $Q < R$ means that $Q(\mathcal{H}) \subset R(\mathcal{H})$, which leads to $Q = QR = RQ$ [29]. The probability is defined through the conditional probability as

$$\Pr(P) = \Pr(P \mid \mathbb{1}), \tag{32}$$

which satisfies

$$\Pr(\mathbb{1}) = 1, \tag{33a}$$

$$\Pr\left(\bigvee_{i=1}^{n} P_i\right) = \textstyle\sum_{i=1}^{n} \Pr(P_i) \tag{33b}$$

for $P_i \in \mathcal{L}(\mathcal{H})$ such that $P_i P_j = P_i \delta_{ij}$ for any i, j.

The formal replacement of the sets A, B, C by the projectors P, Q, R in Eq. (4c) is insufficient to define the conditional probability in quantum mechanics, since such replacement overlooks the non-commutativity among the propositions. In other words, $\Pr(P \wedge Q \mid Q) = \Pr(P|Q)$ does not always hold in quantum mechanics, whereas its classical counterpart $\Pr(A \wedge B \mid B) = \Pr(A \mid B)$ always does by setting $B = C$ in Eq. (4c) [19].

The physical implication of $\Pr(P \wedge Q \mid Q) \neq \Pr(P \mid Q)$ is clear in the following example: Let us take P as the proposition that the spin of a given particle has a definite value, say p along the x-axis, and take Q as another proposition that its spin has a definite value q along the y-axis. On the one hand, then, $P \wedge Q$ is the proposition that the particle has the definite values along the x-axis and y-axis simultaneously, implying $\Pr(P \wedge Q \mid Q) = 0$ due to the uncertainty relation. On the other hand, $\Pr(P \mid Q)$ may have a non-zero value, since the eigenstate of the spin measurement along the y-axis could yield the definite value p in the spin measurement of the x-axis.

Armed with this, we hereafter make the extension of Cox's theorem. First, in parallel to section Sect. 4, we make three assumptions, which employ the analogues of F and S for commuting projectors. Second, we derive the concrete expression of the conditional probability $\Pr(P|Q)$ for the case $P \leq Q$, from which the Born rule is derived. Here, $P = Q$ stands for $P \leq Q$ and $Q \leq P$.

The first assumption is as follows:

Assumption 4. *There exists a non-empty set of real numbers* P_0 *with the following two properties:*

- *(a)* P_0 *is a dense subset of* [F, T].
- *(b) For every* $y_1, y_2, y_3 \in \mathrm{P}_0$, *there exist mutually commuting propositions* P_1, P_2, P_3, Q *such that* $y_1 = \varepsilon(P_1 \mid Q)$, $y_2 = \varepsilon(P_2 \mid P_1 \wedge Q)$, *and* $y_3 = \varepsilon(P_3 \mid P_2 \wedge P_1 \wedge Q)$.

This assumption requires an infinite number of mutually commuting projectors in a Boolean subalgebra $\mathcal{C}$, excluding the finite-dimensional Hilbert space.

To proceed, we make the following assumption:

Assumption 5. *There exists a continuous function* $G : [\mathrm{F}, \mathrm{T}]^2 \to [\mathrm{F}, \mathrm{T}]$ *which is strictly increasing in both the arguments and satisfies*

$$\varepsilon(P \wedge Q \mid R) = G(\varepsilon(Q \mid R),\ \varepsilon(P \mid Q \wedge R)) \tag{34}$$

for any mutually commuting projectors P, Q, R.

The function $G(x, y)$ ensures the validity of our inference on the composite proposition $P \wedge Q$ given R, if they commute. Conversely, if the propositions do not commute, then our inference rule in Eq. (34) does not necessarily hold, as will be shown in section Sect. 7.

By using a Assumptions 4 and 5, and employing the argument given in section Sect. 4, we can obtain

$$G(G(x, y), z) = G(x, G(y, z)), \tag{35}$$

whose solution is given by

$$G(x, y) = w^{-1}\left[w(x)w(y)\right], \tag{36}$$

where $w(x)$ is a continuous, strictly increasing non-negative function. It follows that

$$w(\varepsilon(P \wedge Q \mid R)) = w(\varepsilon(Q \mid R))w(\varepsilon(P \mid Q \wedge R)) \tag{37}$$

for any mutually commuting projectors P, Q, R. Note that $0 \leq w(\varepsilon(P \mid Q)) \leq 1$ is also derived.

By finding the function $w(x)$, we can safely assume the following:

Assumption 6. *There exists a twice differentiable function $T(x)$ such that*

$$w(\varepsilon(P \mid Q)) = T(w(\varepsilon(\neg P \mid Q))) \tag{38}$$

for any mutually commutative projectors P, Q.

Similarly to the previous section, this assumption leads to

$$T(x) = (1 - x^m)^{\frac{1}{m}}. \tag{39}$$

Setting $m = 1$, we observe

$$w(\varepsilon(P \mid Q)) + w(\varepsilon(\neg P \mid Q)) = 1 \tag{40}$$

for projectors P and Q such that $[P, Q] = 0$. Hence, in parallel to Eq. (20), it follows from Eqs. (37) and (40) that

$$w(\varepsilon(P \vee Q \mid R)) = w(\varepsilon(P \mid R)) + w(\varepsilon(Q \mid R)) - w(\varepsilon(P \wedge Q \mid R)) \tag{41}$$

for any mutually commuting projectors P, Q, R. The proof is given in the appendix.

We now prove that

$$w(\varepsilon(P \vee Q \mid R)) = w(\varepsilon(P \mid R)) + w(\varepsilon(Q \mid R)) \tag{42}$$

for any mutually commuting projectors P, Q, R satisfying $PQ = 0$. The proof is similar to that of Eq. (25). First, by setting $P = Q = R$ in Eq. (37), we find $w(\varepsilon(P \mid P))^2 = w(\varepsilon(P|P))$, whose solutions are either 1 or 0. Then we take $w(\varepsilon(P \mid P)) = 1$. Second, substituting $\mathbb{1}$ for P in Eq. (37) and using $\mathbb{1} \wedge Q = Q$ and $\mathbb{1} \wedge R = R$, we find $w(\varepsilon(\mathbb{1} \mid R)) = 1$. Furthermore, putting $P = \mathbb{1}$ in Eq. (40), we find $w(\varepsilon(0|R)) = 0$. Third, by using $w(\varepsilon(0|R)) = 0$ in to Eq. (41) for the mutually commuting projectors P, Q, R satisfying $PQ = 0$, we obtain Eq. (42), which completes the proof.

We now proceed to prove a technical lemma which that plays an essential role in determining the explicit form of $w(x)$, and shows the uniqueness of the (conditional) probability measure associated with $w(x)$ in quantum mechanics.

Lemma 2. *Suppose we are given a projector Q and two density operators ρ_1 and ρ_2 such that*

$$\mathrm{tr}(\rho_1 Q) = \mathrm{tr}(\rho_2 Q) = 1 \tag{43}$$

and

$$\mathrm{tr}(\rho_1 P) = \mathrm{tr}(\rho_2 P) \tag{44}$$

for all P such that $P \leq Q$. Then,

$$\rho_1 = \rho_2. \tag{45}$$

Proof. From Eq. (43) we obtain

$$\mathrm{tr}(\rho_1 \neg Q) = \mathrm{tr}(\rho_2 \neg Q) = 0. \tag{46}$$

Let us perform the spectral decomposition to write $\neg Q$ as

$$\neg Q = |\varphi\rangle\langle\varphi| + \cdots . \tag{47}$$

By substituting Eq. (47) into Eq. (46), we find

$$\mathrm{tr}(\rho_i \neg Q) = \mathrm{tr}(\rho_i |\varphi\rangle\langle\varphi|) + \cdots = 0 \tag{48}$$

for $i = 1, 2$. Since density operators and projectors are non-negative, we find $\mathrm{tr}(\rho_i|\varphi\rangle\langle\varphi|) = \langle\varphi|\rho_i|\varphi\rangle = 0$. This implies that

$$\rho_i|\varphi\rangle = 0 \tag{49}$$

for any vector $|\varphi\rangle \in \mathrm{Im}\, Q^{\perp}$.

Let us take a normalized vector $|\psi\rangle \in \mathcal{H}$ and decompose it as

$$|\psi\rangle = |\psi_1\rangle + |\psi_2\rangle, \tag{50}$$

where

$$|\psi_1\rangle = Q|\psi\rangle \in \mathrm{Im}\, Q, \qquad |\psi_2\rangle = (\mathbb{1} - Q)|\psi\rangle \in \mathrm{Im}\, Q^{\perp}. \tag{51}$$

Since Eq. (49) holds for $|\psi_2\rangle \in \mathrm{Im}\, Q^{\perp}$, for a projector $R = |\psi\rangle\langle\psi|$ we observe

$$\begin{aligned} \mathrm{tr}(\rho_i R) &= \langle\psi|\rho_i|\psi\rangle \\ &= \langle\psi_1|\rho_i|\psi_1\rangle + \langle\psi_2|\rho_i|\psi_2\rangle + 2\mathrm{Re}\langle\psi_1|\rho_i|\psi_2\rangle \\ &= \langle\psi_1|\rho_i|\psi_1\rangle \\ &= \|\psi_1\|^2 \mathrm{tr}(\rho_i O) \end{aligned} \tag{52}$$

for $i = 1, 2$. Here, $\|\psi_1\| = \sqrt{\langle\psi_1|\psi_1\rangle}$ is the norm of the vector $|\psi_1\rangle$ and $O = |\psi_1\rangle\langle\psi_1|/\|\psi_1\|^2$ is the projector onto the one-dimensional subspace generated by $|\psi_1\rangle$.

With the use of $|\psi_1\rangle \in \mathrm{Im}\, Q$, we find $O \leq Q$. Then, it follows from Eq. (44) that

$$\mathrm{tr}(\rho_1 O) = \mathrm{tr}(\rho_2 O), \tag{53}$$

which leads to $\mathrm{tr}(\rho_1 R) = \mathrm{tr}(\rho_2 R)$, or equivalently,

$$\mathrm{tr}(DR) = 0, \tag{54}$$

where

$$D = \rho_1 - \rho_2. \tag{55}$$

Since the operator D is normal ($DD^{\dagger} = D^{\dagger}D$, where † stands for the adjoint operation) and Eq. (54) holds for any one-dimensional projector R, which is clearly non-negative, we can employ the spectral decomposition and obtain $\rho_1 = \rho_2$, which completes the proof. □

Lemma 2 is a special case of Lemma 5, which was first shown in Refs. [31,32]. We now prove the following lemma.

Lemma 3. *If* $\dim \mathrm{Im}\, Q \geq 3$, *there exists a density operator ρ with which $w(\varepsilon(P \mid Q))$ is written as*

$$w(\varepsilon(P \mid Q)) = \frac{\mathrm{tr}(\rho P)}{\mathrm{tr}(\rho Q)} \tag{56}$$

for all P such that $P \leq Q$.

Proof. By using induction on Eq. (42), we obtain

$$w\left(\varepsilon\left(\bigvee_{i=1}^{n} P_i \mid Q\right)\right) = \sum_{i=1}^{n} w(\varepsilon(P_i \mid Q)) \tag{57}$$

for the projectors $P_1, \ldots, P_n, Q$ such that $P_i P_j = \delta_{ij} P_i$ and $[P_i, Q] = 0$. Since $[P_i, Q] = 0$ is a necessary condition of $P_i \leq Q$, Eq. (57) also holds for the set of mutually orthogonal projectors $\{P_i\}_{i=1}^n$ such that $P_i P_j = \delta_{ij} P_i$ and $P_i \leq Q$ for fixed Q. Furthermore, by noticing that Q behaves like the identity operator on $\mathrm{Im}\, Q$ due to the idempotence $Q^2 = Q$, we observe that Eq. (57) can be seen as the sufficient condition of the Gleason theorem [30], as far as we consider the set of projectors $\{P| \mid P \leq Q\}$.

On the basis of the above argument, we may employ the Gleason theorem and hence obtain

$$w(\varepsilon(P \mid Q)) = \mathrm{tr}(\rho_Q P) \tag{58}$$

for $P \leq Q$, if $\dim \mathrm{Im}\, Q \geq 3$. Here, ρ_Q is a density operator which is dependent on Q. Since $Q \leq Q$, we can set $P = Q$ to observe that

$$w(\varepsilon(Q \mid Q)) = \mathrm{tr}(\rho_Q Q) = 1, \tag{59}$$

from Eq. (31a). The density operator ρ_Q satisfying Eqs. (58) and (59) for $\{P| \mid P \leq Q\}$ is uniquely determined, due to Lemma 2.

Now it is clear from Eq. (59) that the density operator ρ_Q has non-trivial support only on the eigenspace of Q. Hence, with an appropriate density operator ρ, we can write

$$\rho_Q = \frac{Q\rho Q}{\mathrm{tr}(\rho Q)}. \tag{60}$$

Plugging Eq. (60) into Eq. (58) and using the cyclic property of trace and $PQ = QP = P$ coming from $P \leq Q$, we find Eq. (56), which completes the proof. □

Lemma 4. *$w(\varepsilon(P|Q))$ is the conditional probability for P, Q such that $P \leq Q$.*

Proof. To prove this, we show Eqs. (31) for Eq. (56). Equations (31a) and (31b) are obvious. For Eq. (31c), we find

$$\frac{w(\varepsilon(P \wedge Q \mid R))}{w(\varepsilon(Q \mid R))} = \frac{\mathrm{tr}[\rho(P \wedge Q)]}{\mathrm{tr}(\rho Q)} = w(\varepsilon(P \wedge Q \mid Q)), \tag{61}$$

which completes the proof. □

It is now straightforward to derive the Born rule:

Theorem 1. *The probability find the of proposition P being true is given by the Born rule:*

$$\Pr(P) = \mathrm{tr}(\rho P). \tag{62}$$

Proof. We set $Q = \mathbb{1}$ in Eq. (56). Then we can make use of Lemma 3 and obtain

$$w(\varepsilon(P \mid \mathbb{1})) = \mathrm{tr}(\rho P). \tag{63}$$

Since $w(\varepsilon(P \mid Q))$ is the conditional probability for $P \leq Q$, and $P \leq \mathbb{1}$ holds for any projector P, we get Eq. (62), which proves the theorem. □

Since the identity operator $\mathbb{1}$ corresponds to Ω, this corollary shows that we can regard the Born rule as the degree of plausibility of proposition P under no information.

6. Inference rule and Lüders rule

Equation (56) is seen as a special case of the Lüders rule of conditional probability [17], which takes the form

$$\mathrm{Pr}(P \mid Q) = \frac{\mathrm{tr}(Q\rho QP)}{\mathrm{tr}(\rho Q)}. \tag{64}$$

Indeed, when $P \leq Q$, the numerator of the right-hand side is written as $\mathrm{tr}(Q\rho QP) = \mathrm{tr}(Q\rho P)$ $= \mathrm{tr}(\rho PQ) = \mathrm{tr}(\rho P)$, which results in Eq. (56). We now turn to show that the Lüders rule is the concrete expression of the conditional probability which is defined for arbitrary pairs of the projectors and in accordance with logical interpretation. To this end, we recall:

Lemma 5 *(Beltrametti and Cassinelli [31,32]). For a Hilbert space $\mathcal{H}$, let $Pr(\,\cdot\,)$ be a probability measure on $\mathcal{L}(\mathcal{H})$, and let Q be any projector such that $Pr(Q) \neq 0$. Then there exists a unique probability measure on $\mathcal{L}(\mathcal{H})$, which we denote by* $\mathrm{Pr}(\,\cdot \mid Q)$*, such that, for all projectors $P \leq Q$, $Pr(P \mid Q) = Pr(P)/Pr(Q)$.*

Proof. From the Gleason theorem, we find that, for any projector Q, the probability measure $\mathrm{Pr}(\,\cdot\,|Q)$ must take the form $\mathrm{Pr}(P|Q) = \mathrm{tr}(\rho_Q P)$ for some density operator ρ_Q. If we can show the uniqueness of the density operator ρ_Q, then we complete the proof.

Now suppose that ρ_1 and ρ_2 are two density operators such that $\mathrm{tr}(\rho_1 P) = \mathrm{tr}(\rho_2 P) =$ $\mathrm{Pr}(P)/\mathrm{Pr}(Q)$ for all projectors P which satisfies that satisfy $P \leq Q$. We then find Eq. (43) by setting $P = Q$. Therefore, we can employ Lemma 2 and obtain $\rho_1 = \rho_2$, which completes the proof. □

We arrive at the following theorem.

Theorem 2. *If there exists a conditional probability satisfying*

$$\mathrm{Pr}(P \wedge Q \mid R) = \mathrm{Pr}(P \mid R)\,\mathrm{Pr}(Q \mid P \wedge R) \tag{65}$$

for any commuting projectors P, Q, R, then it is uniquely determined and takes the form of Eq. (64).

Proof. We first show that Eq. (65) leads to $\mathrm{Pr}(P \mid Q) = \mathrm{Pr}(P)/\mathrm{Pr}(Q)$ for $P \leq Q$. Since $P \leq \mathbb{1}$ and $\mathrm{Pr}(P) = \mathrm{Pr}(P \mid \mathbb{1})$, we make use of Eq. (65) to obtain

$$\mathrm{Pr}(Q)\mathrm{Pr}(P \mid Q) = \mathrm{Pr}(Q \mid \mathbb{1})\mathrm{Pr}(P \mid Q \wedge \mathbb{1}) = \mathrm{Pr}(P \wedge Q \mid \mathbb{1}) = \mathrm{Pr}(P \wedge Q) = \mathrm{Pr}(P). \tag{66}$$

Here we have used $P \leq Q$ in the last equality.

Next, we use Lemma 5. Then we find that there exists a unique density operator ρ_Q by which the conditional probability takes the form $\mathrm{Pr}(P|Q) = \mathrm{tr}(\rho_Q P)$. By setting $P = Q$, n we obtain $\mathrm{Pr}(Q \mid Q) = 1$. Thus, as with the proof of Lemma 3, we find that there exists a density operator ρ such that $\rho_Q = Q\rho Q/\mathrm{tr}(\rho Q)$, which completes the proof. □

Thus, we have arrived at the Lüders rule as the conditional probability for all the projectors, by starting from the inference rules valid at least in mutually commuting projection operators. Note that Eq. (65) is the concrete expression of the inference rule in Eq. (34).

In the rest of this section we show that the Lüders rule satisfies Eqs. (31).

Corollary 1. *The function* $\mathrm{Pr}(\cdot \mid \cdot)$ *defined by equation Eq. (64) satisfies Eq. (31b) for $P_i \in \mathcal{L}(\mathcal{H})$ and $Q \in \mathcal{K}$ such that $P_i P_j = \delta_{ij} P_i$.*

Proof. This is shown by direct calculation. □

Corollary 2. *The function* $\Pr(\cdot \mid \cdot)$ *defined by equation Eq. (64) satisfies Eq. (31c).*

Proof. From Eq. (64), we observe that

$$\frac{\Pr(P \wedge Q \mid R)}{\Pr(Q \mid R)} = \frac{\mathrm{tr}[(P \wedge Q)R\rho R]}{\mathrm{tr}(QR\rho R)}. \tag{67}$$

Since $Q < R$, we have $Q = QR = RQ$, which reduces the denominator of the right-hand side of Eq. (67) to

$$\mathrm{tr}(QR\rho R) = \mathrm{tr}(Q\rho R) = \mathrm{tr}(\rho RQ) = \mathrm{tr}(\rho Q). \tag{68}$$

Similarly, we have $P \wedge Q < Q$, since $\mathrm{Im}(P \wedge Q) \subset \mathrm{Im}\, Q$. Thus, in the same fashion as for Eq. (68), we obtain

$$\begin{aligned}
\mathrm{tr}[(P \wedge Q)R\rho R] &= \mathrm{tr}[(P \wedge Q)QR\rho R] \\
&= \mathrm{tr}[(P \wedge Q)Q\rho R] \\
&= \mathrm{tr}[Q\rho R(P \wedge Q)] \\
&= \mathrm{tr}[Q\rho RQ(P \wedge Q)] \\
&= \mathrm{tr}[Q\rho Q(P \wedge Q)].
\end{aligned} \tag{69}$$

By plugging Eqs. (68) and (69) into Eq. (67), we obtain Eq. (31c), which completes the proof. □

7. Violation of the inference rule

In the previous section we derived the Lüders rule by assuming the validity of the inference rules in Eq. (65) for projectors P, Q with the ordering relation $P \leq Q$. This suggests that the inference rules do not necessarily hold for non-commuting projectors.

However, the relation between the validity of the inference rule and the commutativity is not simple. Indeed, as suggested in the following proposition, we can show that the inference rule holds even when there exists *a pair* of non-commuting projectors. To argue this, we introduce

$$\Delta = \Pr(P \wedge Q \mid R) - \Pr(P \mid R)\Pr(Q \mid P \wedge R) \tag{70}$$

and check its behavior. If $\Delta = 0$, then the inference rule in Eq. (16) holds.

We first show the following proposition.

Proposition 1. *Suppose* $[P, Q] = [P, R] = 0$*. Then* $\Delta = 0$*.*

Proof. We show that $\Pr(P \wedge Q \mid R) = \Pr(P \mid R)\Pr(Q \mid P \wedge R)$. Since $[P, Q] = [P, R] = 0$, we have

$$\begin{aligned}
\Pr(P \wedge Q \mid R) &= \frac{\mathrm{tr}[R\rho R(P \wedge Q)]}{\mathrm{tr}(\rho R)} = \frac{\mathrm{tr}(R\rho RPQ)}{\mathrm{tr}(\rho R)}, \\
\Pr(P \mid R) &= \frac{\mathrm{tr}(\rho RP)}{\mathrm{tr}(\rho R)}, \\
\Pr(Q \mid P \wedge R) &= \frac{\mathrm{tr}[(P \wedge R)\rho(P \wedge R)Q]}{\mathrm{tr}(\rho P \wedge R)} = \frac{\mathrm{tr}(PR\rho PRQ)}{\mathrm{tr}(\rho PR)} = \frac{\mathrm{tr}(R\rho RPQ)}{\mathrm{tr}(\rho RP)}
\end{aligned} \tag{71}$$

by using the relations $P^2 = P$, $Q^2 = Q$, $R^2 = R$ and the cyclic property of the trace $\mathrm{tr}(PQ) = \mathrm{tr}(QP)$. It thus follows that $\Pr(P \wedge Q \mid R) = \Pr(P \mid R)\Pr(Q \mid P \wedge R)$, which completes the proof. □

This proposition suggests that $[Q, R] \neq 0$ is insufficient for $\Delta \neq 0$: the violation of the inference rule requires either $[P, Q] \neq 0$ or $[P, R] \neq 0$.

Second, we construct an example of the violation of the inference rule Eq. (16). Now we set $\mathcal{H} = \mathbb{C}^3$ and introduce the normalized state vectors

$$\begin{aligned} \boldsymbol{\psi} &= (\psi_1, \psi_2, \psi_3)^t \qquad (\psi_i \in \mathbb{R}), \\ \boldsymbol{\varphi} &= (\varphi_1, \varphi_2, \varphi_3)^t \qquad (\varphi_i \in \mathbb{R}), \\ \boldsymbol{z} &= (0, 0, 1)^t, \end{aligned} \tag{72}$$

where t is the transposition operation. We further define the projector onto the two-dimensional planes as

$$P = \mathbb{1}_3 - \boldsymbol{\psi}\boldsymbol{\psi}^t, \qquad Q = \mathbb{1}_3 - \boldsymbol{\varphi}\boldsymbol{\varphi}^t, \qquad R = \mathbb{1}_3 - \boldsymbol{z}\boldsymbol{z}^t, \tag{73}$$

where $\mathbb{1}_3$ stands for the identity operator on $\mathbb{C}^3$.

The meet $P \wedge Q$ is the projector onto the intersection of the two two-dimensional subspaces Im P and Im Q, which are seen as planes. Thus, if $P \neq Q$, $P \wedge Q$ is the projector onto the vector orthogonal to the normal vectors $\boldsymbol{\psi}$ and $\boldsymbol{\varphi}$:

$$P \wedge Q = \boldsymbol{\xi}(\boldsymbol{\psi}, \boldsymbol{\varphi})\boldsymbol{\xi}(\boldsymbol{\psi}, \boldsymbol{\varphi})^t, \tag{74}$$

where

$$\boldsymbol{\xi}(\boldsymbol{\psi}, \boldsymbol{\varphi}) = \frac{\boldsymbol{\psi} \times \boldsymbol{\varphi}}{|\boldsymbol{\psi} \times \boldsymbol{\varphi}|}, \tag{75}$$

and $\times$ is the exterior product. Note that

$$P\boldsymbol{\xi}(\boldsymbol{\psi}, \boldsymbol{\varphi}) = Q\boldsymbol{\xi}(\boldsymbol{\psi}, \boldsymbol{\varphi}) = \boldsymbol{\xi}(\boldsymbol{\psi}, \boldsymbol{\varphi}), \tag{76}$$

which ensures $\boldsymbol{\xi}(\boldsymbol{\psi}, \boldsymbol{\varphi}) \in \mathrm{Im}(P \wedge Q)$. By the same argument, we observe that

$$P \wedge R = \boldsymbol{\xi}(\boldsymbol{\psi}, \boldsymbol{z})\boldsymbol{\xi}(\boldsymbol{\psi}, \boldsymbol{z})^t \tag{77}$$

when $P \neq R$.

Furthermore, let us set $\rho = \mathbb{1}_3/3$. By direct calculation, we then obtain

$$\begin{aligned} \Pr(P \wedge Q \mid R) &= \frac{1}{2}\left(1 - \frac{|(\boldsymbol{\psi} \times \boldsymbol{\varphi}) \cdot \boldsymbol{z}|^2}{|\boldsymbol{\psi} \times \boldsymbol{\varphi}|^2}\right), \\ \Pr(P \mid R) &= \frac{1}{2}\left(1 + |\boldsymbol{\psi} \cdot \boldsymbol{z}|^2\right), \\ \Pr(Q \mid P \wedge R) &= 1 - \frac{|(\boldsymbol{\psi} \times \boldsymbol{z}) \cdot \boldsymbol{\varphi}|^2}{|\boldsymbol{\psi} \times \boldsymbol{z}|^2}, \end{aligned} \tag{78}$$

which leads to

$$\Delta = \frac{1}{2}\left[1 - \frac{|(\boldsymbol{\psi} \times \boldsymbol{\varphi}) \cdot \boldsymbol{z}|^2}{|\boldsymbol{\psi} \times \boldsymbol{\varphi}|^2} - \left(1 + |\boldsymbol{\psi} \cdot \boldsymbol{z}|^2\right)\left(1 - \frac{|(\boldsymbol{\psi} \times \boldsymbol{z}) \cdot \boldsymbol{\varphi}|^2}{|\boldsymbol{\psi} \times \boldsymbol{z}|^2}\right)\right] \tag{79}$$

for $P \neq Q \neq R$.

We now find the violation of the inference rule in Eq. (16): When $\boldsymbol{\psi} = (1/\sqrt{2})(1, 0, 1)^t$ and $\boldsymbol{\varphi} = (1/\sqrt{2})(0, 1, 1)^t$, we have $\Pr(P \wedge Q \mid R) = 1/3$, $\Pr(P \mid R) = 3/4$, and $\Pr(Q \mid P \wedge R) = 1/2$, resulting in $\Delta = -1/24$. Note that P, Q, R are mutually non-commuting in this case.

8. Conclusion and discussions

We obtained the Born rule and the Lüders rule through a simple and natural generalization of the logical interpretation of probability theory. We did not take the rules as postulates, but as theorems deduced from the reasonable assumption that our inference works rationally for any set of commuting projectors.

Whereas the inference rule in Eq. (34) and its concrete expression in Eq. (65) hold for any $\mathcal{C}$ by the assumptions, it does not hold for non-commuting projectors. This property of the Lüders rule is analogous to the algebraic property of the observables appearing in Mermin's magic square [33]: given a set of non-commuting observables, subsets of the commuting observables follow the algebraic relations of their eigenvalues, but this is not the case for the total set of non-commuting observables. We could say that the validity of our inference rules depends on the context. This property was overlooked in Ref. [34], where the Born rule was algebraically derived by using the associativity relation of the composite proposition $(P \vee Q) \vee R = P \vee (Q \vee R) = P \vee Q \vee R$ for mutually orthogonal projectors P, Q, R.

The results we have obtained could give new insights into quantum logic [35]. In quantum logic, we can show that (i) the projectors of the Hilbert space are seen as representations of elements of the orthomodular lattice, which governs the algebraic relations among quantum propositions [36], and (ii) the probability measure on the projectors is given by the Born rule (the Gleason theorem). Note that these outcomes are formal, and suggest nothing about the interpretations of the quantum probability.

In contrast, we have derived the Born rule based on assumptions which are easy to understand as the inference rules in the Boolean sublattice. Thus, we may say that, on the basis of quantum logic, we have arrived at an interesting and new viewpoint: the Born rule is the natural extension of the standard inference rules to the orthomodular lattice structure.

One interesting question is whether the framework presented here can be extended in such a way that the mapping $w(\varepsilon(\cdot | \mid Q))$ takes the form of Eq. (56) even for the σ-additive case (infinite n in Eq. (57)). As pointed out in Chapter 2 of [8], such extension requires "a well-defined and well-behaved limiting process from a finite set."It seems challenging to show the existence of such a limiting process for general sequences of projectors $\{P_1, P_2, \ldots, P_n\}$. Note that, for many practical purposes, it would suffice to consider the finite-set cases, since actual experimental apparatuses have finite resolution and limitations on the measurement range.

We must mention that the inference approach is one of the conceptual foundations in machine learning [37]. In particular, various non-informative prior distributions such as Jeffery's prior are proposed on the line of thought of logical interpretation. Analogously, our results may provide a conceptual basis for quantum machine learning [38], which is a recent central issue in quantum information theory [39].

Acknowledgments

We thank I. Tsutsui and S. Hanashiro for fruitful discussions. This work was supported by MEXT Quantum Leap Flagship Program (MEXT Q-LEAP) Grant Number JPMXS0120319794.

Appendix. Proofs of Eqs. (20) and (41)

We first prove Eq. (20) by following the argument given in Ref. [8]. To this end, we note De Morgan's law for elements A, B of the abstract Boolean algebra [27]:

$$\neg(A \vee B) = \neg A \wedge \neg B. \tag{A1}$$

Let us turn to the finite additive class $\mathcal{F}$, which is the concrete Boolean algebra. Three Boolean operations, and, or, and not, have been defined in Eqs. (10). By using Eqs. (A1), (16), and (17), we then obtain

$$\begin{aligned} w(\varepsilon(A \vee B \mid C)) &= 1 - w(\varepsilon(\neg(A \vee B) \mid C)) \\ &= 1 - w(\varepsilon(\neg A \wedge \neg B \mid C)) \\ &= 1 - w(\varepsilon(\neg A \mid C)) w(\varepsilon(\neg B \mid \neg A \wedge C)) \\ &= 1 - w(\varepsilon(\neg A \mid C)) [1 - w(\varepsilon(B \mid \neg A \wedge C))] \\ &= w(\varepsilon(A \mid C)) + w(\varepsilon(\neg A \mid C)) w(\varepsilon(B \mid \neg A \wedge C)) \\ &= w(\varepsilon(A \mid C)) + w(\varepsilon(\neg A \wedge B \mid C)) \\ &= w(\varepsilon(A \mid C)) + w(\varepsilon(B \mid C)) w(\varepsilon(\neg A \mid B \wedge C)) \\ &= w(\varepsilon(A \mid C)) + w(\varepsilon(B \mid C)) [1 - w(\varepsilon(A \mid B \wedge C))] \\ &= w(\varepsilon(A \mid C)) + w(\varepsilon(B \mid C)) - w(\varepsilon(A \wedge B \mid C)), \end{aligned} \tag{A2}$$

which completes the proof.

We can prove Eq. (41) by making the above argument and replacing the elements $A, B, C \in \mathcal{F}$ with $P, Q, R \in \mathcal{C}$. respectively. The Boolean operations acting on $P, Q, R \in \mathcal{C}$ are given by Eqs. (28).

References

[1] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[2] E. T. Jaynes, Phys. Rev. **108**, 171 (1957).
[3] P.-S. Laplace, A Philosophical Essay on Probabilities (John Wiley and Sons, London, 1902).
[4] R. T. Cox, Am. J. Phys. **14**, 1 (1946).
[5] R. T. Cox, The Algebra of Probable Inference (Johns Hopkins University Press, Baltimore, 1961).
[6] J. Y. Halpern, J. Artif. Intell. Res. **10**, 67 (1999).
[7] J. Y. Halpern, J. Artif. Intell. Res. **11**, 429 (1999).
[8] E. T. Jaynes, Probability Theory: The Logic of Science (Cambridge University Press, Cambridge, 2003).
[9] K. S. Van Horn, Int. J. Approx. Reasoning **34**, 3 (2003).
[10] J. M. Keynes, A Treatise on Probability (Macmillan and Co., London, 1921).
[11] H. De Raedt, M. I. Katsnelson, and K. Michielsen, Ann. Phys. (NY) **347**, 45 (2014).
[12] H. De Raedt, M. I. Katsnelson, H. C. Donker, and K. Michielsen, Ann. Phys. (NY) **359**, 166 (2015).
[13] H. C. Donker, M. I. Katsnelson, H. De Raedt, and K. Michielsen, Ann. Phys. (NY) **372**, 74 (2016).
[14] A. Caticha, Phys. Rev. A **57**, 1572 (1998).
[15] A. Caticha, J. Phys. Conf. Ser. **504**, 012009 (2014).
[16] G. Lüders, Ann. der Physik **443**, 322 (1950).
[17] P. Busch and P. Lahti, Lüders Rule, in Compendium of Quantum Physics, eds. D. Greenberger Greenberger, K. Hentschel Hentschel, and F. Weinert Weinert (Springer, Berlin, Heidelberg,2009), p. 356.
[18] A. Rényi, Acta Math. Acad. Math. Sci. Hungar. **6**, 285 (1955).
[19] A. Rényi, Teor. Veroyatnost. i Primenen. **1**, 61 (1956).
[20] A. Rényi, Foundations of Probability (Holden-Day, Inc., San Francisco, 1970).
[21] D. Gillies, Philosophical Theories of Probability (Routledge, London, 2000).
[22] J. Williamson, Br. J. Philos. Sci. **50**, 401 (1999).
[23] R. D. Cousins, Am. J. Phys. **63**, 398 (1995).
[24] L. Lyons, Contemp. Phys. **54**, 1 (2013).
[25] B. de Finetti, Theory of Probability (John Wiley and Sons, New York, 1974).
[26] C. M. Caves, C. A. Fuchs, and R. Schack, Phys. Rev. A **65**, 022305 (2002).

[27] J. E. Whitesitt, Boolean Algebra and Its Applications (Addison-Wesley, Reading, MA, 1961).
[28] J. Aczél, Lectures on Functional Equations and Their Applications (Academic Press, New York, 1966).
[29] N. I. Akhiezer and I. M. Glazman, Theory of Linear Operators in Hilbert Space (Dover, New York, 1993).
[30] A. M. Gleason, Indiana Univ. Math. J. **6**, 885 (1957).
[31] E. G. Beltrametti and G. Cassinelli, The Logic of Quantum Mechanics (Addison-Wesley, Reading, MA, 1981).
[32] J. D. Malley, Phys. Rev. A **69**, 022118 (2004).
[33] N. D. Mermin, Rev. Mod. Phys. **65**, 803 (1993).
[34] F. Holik, M. Sáentz, and A. Plastino, Ann. Phys. **340**, 293 (2014).
[35] I. Pitowsky, Quantum Probability – Quantum Logic (Springer Verlag, Berlin, Heidelberg,1989).
[36] G. Birkhoff, J. von Neumann, Ann. Math. **37**, 823 (1936).
[37] C. M. Bishop, Pattern Recognition and Machine Learning (Springer Verlag, New York, 2006).
[38] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Nature **549**, 195 (2017).
[39] J. Preskill, Quantum **2**, 79 (2018).