Chapter 2
Classical (Kolmogorovian) and Quantum (Born) Probability

This chapter contains short introductions to classical and quantum probabilistic models. To simplify presentation, in both cases we consider only *discrete variables*.

2.1 Kolmogorovian Probabilistic Model

We start with two notations. Let A be a set. The *characteristic function I_A* of the set A is defined as $I_A(x) = 1, \; x \in A$, and $I_A(x) = 0, \; x \notin A$. Let $A = \{a_1, \ldots, a_n\}$ be a finite set. We shall denote the number of elements n of A by the symbol $|A|$.

Sets of real and complex numbers are denoted by symbols **R** and **C**, respectively.

2.1.1 Probability Space

The modern axiomatics of probability theory was invented by Andrei Nikolaevich Kolmogorov (one of greatest mathematicians of the 20th century) in 1933, [219], see also [121, 279, 161] and it was a natural finalization of a few hundred years long development of the *set-theoretic model* for probability. A crucial point is representation of events by subsets of some basic set Ω. The collection of subsets representing events should be sufficiently rich to be able to perform set-theoretic operations such as intersection, union and difference of sets.[1] Then one assigns weights (real numbers) to these subsets:

$$A \mapsto \mathbf{P}(A) \qquad (2.1)$$

for an event A. They are chosen nonnegative and normalized by 1: $\mathbf{P}(\Omega) = 1$. The weight of a set A which is the disjoint union of sets A_1 and A_2 is equal to the sum of weights of these subsets. The latter property is called *finite additivity*. The map given by (2.1) with mentioned properties is measure-theoretic probability. If the

---

[1] However, it should not be unreasonably large. If too extended a system of subsets is selected, then it may represent "events" that cannot be interpreted in a reasonable way.

basic set Ω is finite[2] one can proceed with this simple definition. However, if the Ω is countable, i.e., it is infinite and its points can be enumerated or "continuous" – e.g., a segment of the real line **R** – then finite additivity is not sufficient for creating a fruitful mathematical model. It is extended to so-called σ-additivity (countable additivity). A rich mathematical model is created. However, by proceeding with σ-additivity one should not forget Kolmogorov's remark [219] that σ-*additivity is a purely mathematical and totally nonphysical notion.* It is impossible to perform a real experiment an infinite number of times. In principle, the model based on σ-additivity might produce probabilistic artifacts that have no real interpretation.[3]

We now start rigorous presentation of probability theory. But, in principle, one can read practically the whole book by considering the model based on a finite set Ω, a collection of events represented by all its subsets and finite-additive probability given by assigning weights to points of Ω : $\omega \to \mathbf{P}(\omega)$. For example, the uniform probability is given by equal weights. For $\Omega = \{\omega_1, \ldots, \omega_N\}$, $\mathbf{P}(\omega_j) = 1/N$. Here, for $A \subset \Omega$, $\mathbf{P}(A) = |A|/N$.

Let Ω be a set. We recall that a σ-*algebra* is a system of subsets of Ω that is closed with respect to operations of countable intersection, union and difference of sets and containing Ω and the empty set $\emptyset$.[4]

The simplest example of a σ-algebra is the system consisting of just two sets: Ω and $\emptyset$. However, it is too small to do anything interesting. Another example is given by the family of all subsets of Ω. As was mentioned, such a σ-algebra is useful if the set Ω is finite or even countable. However, if Ω is "continuous", then consideration of all possible subsets as representing events induces visible probabilistic pathologies, see [161] for details. So, the σ-algebra consisting of all subsets (of, e.g., a segment $[a, b]$ of the real line) is too large. One chooses a smaller σ-algebra, the so-called *Borel σ-algebra*. For example, for $\Omega = \mathbf{R}$, it is generated by all half-open intervals: $[\alpha, \beta), \alpha < \beta$. However, in this book we will practically never operate with continuous Ω.

Let Ω be a set and let $\mathcal{F}$ be a σ-algebra of its subsets. A probability measure **P** is a map from $\mathcal{F}$ to the segment [0, 1], which is normalized $\mathbf{P}(\Omega) = 1$ and σ-additive

$$\mathbf{P}(A_1 \cup \ldots \cup A_n \cup \ldots) = \mathbf{P}(A_1) + \ldots + \mathbf{P}(A_n) + \ldots,$$

for disjoint sets belonging to $\mathcal{F}$.

By the Kolmogorov axiomatics [219], see also [280], the *probability space* is a triple

$$\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P}).$$

Points ω of Ω are said to be *elementary events*, elements of $\mathcal{F}$ are *events*, **P** is *probability*.

---

[2] Since Ω can contain billions of points, this model is useful in a huge class of applications.

[3] Unfortunately, this point made by Kolmogorov has been totally forgotten.

[4] In some books on probability theory the terminology σ-*field* is used, instead of σ-*algebra*.

Discrete random variables[5] on the Kolmogorov space $\mathcal{P}$ are by definition functions $a : \Omega \to X_a$, where $X_a = \{\alpha_1, \ldots, \alpha_n, \ldots\}$ is a countable set (the *range of values*), such that sets

$$C_\alpha^a = \{\omega \in \Omega : a(\omega) = \alpha\}, \quad \alpha \in X_a, \tag{2.2}$$

belong to $\mathcal{F}$.

It is typically assumed that the range of values X_a is a subset of the real line. We will proceed under this assumption practically everywhere, but sometimes, e.g., in cognitive and psychological modeling, it will be more convenient to consider Boolean labels, e.g. α = yes, no.

We shall use the symbol $RVD(\mathcal{P})$ to denote the space of discrete random variables for the probability space $\mathcal{P}$. The probability distribution of $a \in RVD(\mathcal{P})$ is defined by $\mathbf{P}(a = \alpha) = \mathbf{P}(C_\alpha^a), \alpha \in X_a$, where the set C_α^a is given by (2.2). It is convenient to proceed with a shorter symbol

$$p^a(\alpha) \equiv \mathbf{P}(\omega \in \Omega : a(\omega) = \alpha).$$

We remark that:

$$p^a(\alpha_1) + \ldots + p^a(\alpha_n) + \ldots = 1, \quad p^a(\alpha_n) \geq 0. \tag{2.3}$$

The *average* (mathematical expectation) of a random variable a is defined as

$$\bar{a} \equiv Ea = \alpha_1 \ p^a(\alpha_1) + \ldots + \alpha_n \ p^a(\alpha_n) + \ldots. \tag{2.4}$$

For a family of random variables $a_1, \ldots, a_m$ taking values $\alpha_j^1, \ldots, \alpha_j^m, j = 1, 2, \ldots$, respectively, their joint probability distribution is defined as

$$p^{a_1 \ldots a_m}(\alpha_{j_1}^1, \ldots, \alpha_{j_m}^m) = \mathbf{P}(\omega \in \Omega : a_1(\omega) = \alpha_{j_1}^1, \ldots, a_m(\omega) = \alpha_{j_m}^m). \tag{2.5}$$

We remark that joint probability is symmetric with respect to permutations; e.g., for two random variables a and b, we have

$$p^{ab}(\alpha, \beta) = \mathbf{P}(\omega \in \Omega : a(\omega) = \alpha, b(\omega) = \beta) = p^{ba}(\beta, \alpha). \tag{2.6}$$

It is an important feature of the Kolmogorov model.

For two random variables a and b *covariance* is defined as

$$\text{cov}(a, b) = E(a - \bar{a})(b - \bar{b}) = \sum_{\alpha\beta}(\alpha - \bar{a})(\beta - \bar{b}) \ p^{ab}(\alpha, \beta). \tag{2.7}$$

---

[5] In Chaps. 1–9 we consider only discrete random variables. In Chaps. 10 and 11 random variables having continuous ranges of values will be used.

It is easy to see that

$$\text{cov}(a, b) = Eab - \bar{a}\bar{b}. \tag{2.8}$$

We remark that covariance is symmetric

$$\text{cov}(a, b) = \text{cov}(b, a). \tag{2.9}$$

2.1.2 Conditional Probability

Kolmogorov's probability model is based on a probability space equipped with the operation of conditioning. In this model *conditional probability* is defined by the well-known *Bayes' formula*

$$\mathbf{P}(B|C) = \mathbf{P}(B \cap C)/\mathbf{P}(C), \quad \mathbf{P}(C) > 0. \tag{2.10}$$

By Kolmogorov's interpretation it is the *probability that an event B occurs under the condition that an event C occurred.* We remark that this formula is a definition, it cannot be derived. The use of this definition of conditional probability is one of fundamental constraints induced by the Kolmogorov model.

We remark that $\mathbf{P}_C(B) \equiv \mathbf{P}(B|C)$ is again a probability measure on $\mathcal{F}$. For a set $C \in \mathcal{F}, \mathbf{P}(C) > 0$, and a (discrete) random variable a, the conditional probability distribution is defined as

$$p_C^a(\alpha) \equiv \mathbf{P}(a = \alpha|C), \quad \alpha \in X_a.$$

We naturally have

$$p_C^a(\alpha_1) + \ldots + p_C^a(\alpha_n) + \ldots = 1, \quad p_C^a(\alpha_n) \geq 0. \tag{2.11}$$

The conditional expectation of a random variable a is defined by

$$E(a|C) = \alpha_1 \ p_C^a(\alpha_1) + \ldots + \alpha_n \ p_C^a(\alpha_n) + \ldots . \tag{2.12}$$

For two random variables a and b, consider conditional probabilities

$$p_{\beta|\alpha} \equiv \mathbf{P}(b = \beta|a = \alpha), \quad p_{\alpha|\beta} \equiv \mathbf{P}(a = \alpha|b = \beta).$$

Following tradition, we will call these probabilities *transition probabilities,* although this terminology might be misleading for our further considerations, see Remark 2.1.

These conditional probabilities can also be written in the form

$$p_{\beta|\alpha} = \mathbf{P}(b = \beta|C_\alpha^a), \ p_{\alpha|\beta} = \mathbf{P}(a = \alpha|C_\beta^b), \tag{2.13}$$

where, e.g., C_α^a is defined by (2.2). It is, of course, assumed that in the first case $p^a(\alpha) > 0$ and in the second case $p^b(\beta) > 0$.

Remark 2.1 The terminology "transition probabilities" may be rather misleading for this book. Typically $p_{\beta|\alpha}$ is considered as the probability of transition from the state α of some system to another state β of the same system. That is why the symbol $p_{\alpha\beta}$ is typically used, instead of our $p_{\beta|\alpha}$. To come to the standard notation, one should change $p_{\beta|\alpha} \to p_{\alpha\beta}$ and vice versa. However, we will not consider states of systems. For us, $p_{\beta|\alpha}$ is probability of obtaining the value $b = \beta$ of the observable b under the condition that the result $a = \alpha$ was observed in the previous measurement of the observable a. Nevertheless, we will also use the standard terminology – transition probabilities.

It is convenient to use the following definition. A random variable a is said to be *nondegenerate* if

$$p^a(\alpha) > 0 \tag{2.14}$$

for any $\alpha \in X_a$. In future considerations we shall use the matrices of conditional probabilities for successive measurements – transition probabilities

$$\mathbf{P}^{b|a} = (p_{\beta|\alpha}), \quad \mathbf{P}^{a|b} = (p_{\alpha|\beta}). \tag{2.15}$$

The first matrix is well defined if a is nondegenerate and the second if b is nondegenerate. We remark that these matrices are always *left stochastic*. A left stochastic matrix is a square matrix whose columns consist of nonnegative real numbers whose sum is 1. For example, for $\mathbf{P}^{b|a}$, we have that

$$\sum_\beta p_{\beta|\alpha} = \sum_\beta \mathbf{P}(b = \beta|a = \alpha) = \sum_\beta \mathbf{P}_{C_\alpha^a}(b = \beta) = 1 \tag{2.16}$$

for any fixed $a = \alpha$. It is a consequence of the fact that, for any set C of strictly positive probability, $\mathbf{P}_C$ is also a probability measure. In (2.16) we select $C = C_\alpha^a$.

Coming back to Remark 2.1, we notice that in standard notation a matrix of "transition probabilities" is not left, but right stochastic, i.e., all rows sum to 1. We point out the following equality connecting the joint probability distribution of two random variables a and b with their transition probabilities:

$$p^{ab}(\alpha, \beta) = p^a(\alpha)p_{\beta|\alpha} = p^b(\beta)p_{\alpha|\beta} = p^{ba}(\beta, \alpha). \tag{2.17}$$

Conditional probabilities are basic in considerations on *independent random variables,* see Sect. 12.1.1.

In our further considerations one special class of matrices of transition probabilities will play a fundamental role. These are so-called doubly stochastic matrices. We recall that in a *doubly stochastic matrix* all entries are nonnegative and all rows

and all columns sum to 1. Of course, in general $\mathbf{P}^{b|a}$ is only left stochastic, not doubly stochastic. The following proposition characterizes random variables inducing doubly stochastic matrices.

Proposition 2.1 *Let a and b be nondegenerate random variables. Then the following conditions are equivalent:*

DS-DS *Both matrices $\mathbf{P}^{a|b}$ and $\mathbf{P}^{b|a}$ are doubly stochastic.*
UD *Random variables are uniformly distributed: $p^a(\alpha) = p^b(\beta) = 1/2$.*
SC *Random variables are "symmetrically conditioned" in the sense*

$$p_{\beta|\alpha} = p_{\alpha|\beta}. \tag{2.18}$$

In the Kolmogorovian model one can guarantee double stochasticity for both $b|a$- and $a|b$-conditioning only for uniformly distributed random variables. This is not the case in non-Kolmogorovian models, e.g., for the quantum probabilistic model, see Sect. 2.4. Here equivalence of conditions **DS-DS** and **SC** plays a crucial role. In fact, the latter is coupled to the symmetry of the scalar product.

Consider now a pair of dichotomous random variables $a = \alpha_1, \alpha_2$ and $b = \beta_1, \beta_2$. The matrix of transition probabilities $\mathbf{P}^{b|a}$ has the form

$$\mathbf{P}^{b|a} = \begin{pmatrix} p_{\beta_1|\alpha_1} & p_{\beta_1|\alpha_2} \\ p_{\beta_2|\alpha_1} & p_{\beta_2|\alpha_2} \end{pmatrix} \tag{2.19}$$

It is doubly stochastic iff $p_{1|1} = p_{2|2}$ and $p_{1|2} = p_{2|1}$, i.e.,

$$\mathbf{P}^{b|a} = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix} \tag{2.20}$$

In particular, it is automatically symmetric. In this case **SC** is equivalent to the condition $\mathbf{P}^{b|a} = \mathbf{P}^{a|b}$.

2.1.3 Formula of Total Probability

In our further considerations an important role will be played by the *formula of total probability* (FTP). It is a theorem of the Kolmogorov model. Let us consider a countable family of disjoint sets A_k belonging to $\mathcal{F}$ such that their union is equal to Ω and $\mathbf{P}(A_k) > 0$, $k = 1, \ldots$. Such a family is called a *partition* of the space Ω.

Theorem 2.1 *Let $\{A_k\}$ be a partition. Then, for every set $B \in \mathcal{F}$, the following formula of total probability holds:*

$$\mathbf{P}(B) = \mathbf{P}(A_1)\mathbf{P}(B|A_1) + \ldots + \mathbf{P}(A_k)\mathbf{P}(B|A_k) + \ldots. \tag{2.21}$$

Proof We have

$$\mathbf{P}(B) = \mathbf{P}(B \cap \cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mathbf{P}(B \cap A_k) = \sum_{k=1}^{\infty} \mathbf{P}(A_k) \frac{\mathbf{P}(B \cap A_k)}{\mathbf{P}(A_k)}.$$

Especially interesting for us is the case such that a partition is induced by a discrete random variable a taking values $\{\alpha_k\}$. Here

$$A_k = C_{\alpha_k}^a = \{\omega \in \Omega : a(\omega) = \alpha_k\}. \tag{2.22}$$

Let b be another random variable. It takes values $\{\beta_j\}$. For any $\beta \in X_b$, we have

$$\mathbf{P}(b = \beta) = \mathbf{P}(a = \alpha_1)\mathbf{P}(b = \beta | a = \alpha_1) + \ldots + \mathbf{P}(a = \alpha_k)\mathbf{P}(b = \beta | a = \alpha_k) + \ldots \tag{2.23}$$

or in compact notation

$$p^b(\beta) = p^a(\alpha_1)p_{\beta|\alpha_1} + \ldots + p^a(\alpha_k)p_{\beta|\alpha_k} + \ldots \quad . \tag{2.24}$$

2.2 Probabilistic Incompatibility: Bell–Boole Inequalities

If the reader has not yet been excited by *Bell's inequality* and such mysterious consequences of its violation as *quantum nonlocality* and *death of realism* in QM, then I strongly recommend him or her to omit this section as well as Sect. 2.7. Bell's inequality will not play a fundamental role in this book (nevertheless, it will appear in Sect. 9.6).

Bell's inequality is really the central point of modern QM. Therefore the reader may be surprised to find it not in Sect. 2.3, devoted to QM, but in the section devoted to classical probability theory (Kolmogorov's model). However, I think that it is the right place for the appearance of Bell's inequality, i.e., before saying anything about QM. My personal opinion is that this inequality is the standard subject of classical probability theory. Moreover, we will see that Bell-type inequalities appeared in probability theory long before not only Bell's invention, but even the discovery of QM. My main message to the reader is that attempts (which are very popular in modern QM, especially in the quantum information community) to associate Bell-type inequalities with *quantum nonlocality* or *death of realism* are not sufficiently justified. In classical probability theory such inequalities were used for one hundred years (!) without any reference to the mentioned fundamental problems or to QM in general.

2.2.1 Views of Boole, Kolmogorov, and Vorob'ev

In his book [219] Andrei Nikolaevich Kolmogorov emphasized that each experimental arrangement (context) generates its *own probability space.* For him it was totally clear that it is very naive to expect that all experimental contexts can be described by a single (perhaps huge) probability space. In particular, the following problem arises. Suppose that a family of observables, say $\mathcal{O} = \{a_1, a_2, a_3, \ldots\}$, is given. However, it is impossible to measure them all simultaneously. Thus the joint probability distribution is not given. Nevertheless, it is possible to measure some groups of these observables and joint probability distributions for such groups are given.

Is it possible to construct a single probability space serving for the whole family $\mathcal{O}$?

Thus we are interested in the possibility of embedding the family of observables $\mathcal{O}$ into the space of random variables on a single probability space. If the answer is yes, then such observables exhibit *probabilistic compatibility* (PC), and in the opposite case, *probabilistic incompatibility* (PI), see [204] for details.

It seems that G. Boole (the inventor of *Boolean logic* and *Boolean algebra*) was the first to study this problem. He formulated a necessary condition for PC of a family of three dichotomous observables, $a_1, a_2, a_3 = \pm 1$, such that they can be measured pairwise, but not all simultaneously. This condition coincides with the famous Bell's inequality [31], which plays a fundamental role in modern QM![6]. Later the most general problem of PC (i.e., for an arbitrary family of observables) was solved by Soviet mathematician Vorob'ev [302], who applied these results to problems of optimal control and game theory. Unfortunately, Vorob'ev's results were also practically forgotten.[7] Of course, practically complete disregard of the PC problem in the probabilistic community played an extremely negative role in the development of science. In particular, if Bell or at least someone from the quantum community had been aware of the results of Boole or Vorob'ev or at least of Kolmogorov's message, *"context induces a probability space",* discovery of Bell-type inequalities need not have induced coupling to such mysterious (and nowadays extremely popular) things such as quantum nonlocality or death of realism in QM.

A pragmatic guy [161] could be completely satisfied with recognition that probabilistic data collected in a few incompatible experiments (and this is the case in

---

[6] Boole's results were totally forgotten. Itamar Pitowsky found these results and compared them with Bell's inequality, see [259, 260] and also the preface in [167].

[7] Walter Philipp discovered Vorob'ev's article [302] and together with Karl Hess advertised it a lot [143], in particular during the Växjö series of conferences on quantum foundations, see, e.g., [165, 167, 5, 6]. The main problem of the classical probabilistic community was concentration on mathematical problems related to a single Kolmogorov space, especially various limit theorems. In such a situation even the idea that something could not be embedded in such a space was not especially welcome. Vorob'ev's works were not highly estimated by the Soviet probabilistic community (which was one of the strongest in the world) and, as a result, not by the international community either.

application of Bell's inequality in QM) cannot be described by a single probability space, or in other words, observables are not of the PC-type. Compare the views of Accardi, Aerts, Fine, Fuchs, Garola and Solombrino, Hess and Philipp, Khrennikov, Kupczynski, Larsson, Pitowsky, Rastal, Sozzo, Adenier (see [204] for the corresponding bibliography and this author's book [161] for mathematical details). Then one can try to find sources of PI that are different from quantum nonlocality or death of realism, see [204].

We now discuss the problem in mathematical notation. In principle, it is a repetition of previous considerations, but with mathematical symbols.

Consider a system of three observables $a_i, i = 1, 2, 3$. Suppose for simplicity that they take discrete values and moreover they are dichotomous: $a_i = \pm 1$. Suppose that these observables as well as their pairs can be measured and hence joint probabilities for pairs are well defined: $p^{a_i a_j}(\alpha_i, \alpha_j) \geq 0$ and $\sum_{\alpha_i, \alpha_j = \pm 1} p^{a_i a_j}(\alpha_i, \alpha_j) = 1$.

Question *Is it possible to construct the joint probability distribution, $p^{a_1 a_2 a_3}(\alpha_1, \alpha_2, \alpha_3)$, for any triple of pairwise measurable observables?*

This is the simplest case of the general problem – to find conditions for existence of probability distribution with given *marginal probabilities*. First of all, it is easy to give numerous examples of nonexistence.

Example 2.1 (see [302]) Suppose that

$$\mathbf{P}(a_1 = +1, a_2 = +1) = \mathbf{P}(a_1 = -1, a_2 = -1) = 1/2;$$
$$\mathbf{P}(a_1 = +1, a_3 = +1) = \mathbf{P}(a_1 = -1, a_3 = -1) = 1/2;$$
$$\mathbf{P}(a_2 = +1, a_3 = -1) = \mathbf{P}(a_2 = -1, a_3 = +1) = 1/2.$$

Hence, $\mathbf{P}(a_1 = +1, a_2 = -1) = \mathbf{P}(a_1 = -1, a_2 = +1) = 0$; $\mathbf{P}(a_1 = +1, a_3 = -1) = \mathbf{P}(a_1 = -1, a_3 = +1) = 0$, $\mathbf{P}(a_2 = +1, a_3 = +1) = \mathbf{P}(a_2 = -1, a_3 = -1) = 0$. It is impossible to construct a probability measure which would produce these marginal distributions. We can show this directly [302]. Suppose that one can find a family of real constants $\mathbf{P}(\varepsilon_1, \varepsilon_2, \varepsilon_3)$, $\varepsilon_j = \pm 1$, such that

$$\mathbf{P}(\varepsilon_1, \varepsilon_2, +1) + \mathbf{P}(\varepsilon_1, \varepsilon_2, -1) = \mathbf{P}(a_1 = \varepsilon_1, a_2 = \varepsilon_2), \dots,$$
$$\mathbf{P}(+1, \varepsilon_2, \varepsilon_3) + \mathbf{P}(-1, \varepsilon_2, \varepsilon_3) = \mathbf{P}(a_2 = \varepsilon_2, a_3 = \varepsilon_3).$$

Then it is easy to see that some of these numbers should be negative. In a more fashionable way one can apply Bell's inequality, e.g., for correlations (Sect. 2.2.2) and see that it is violated.

We emphasize that *for mathematicians consideration of Bell-type inequalities did not induce revolutionary reconsideration of the laws of nature. The joint probability distribution does not exist simply because those observables could not be measured simultaneously.*

2.2.2 Bell's and Wigner's Inequalities

Let $\mathcal{P} = (\Omega, \mathcal{F}, P)$ be a Kolmogorov probability space. We recall that *covariance* of two random variables is given by (2.7).

Theorem 2.2 (Bell inequality for covariances) *Let $a_1, a_2, a_3 = \pm 1$ be random variables on $\mathcal{P}$. Then Bell's inequality*

$$|\langle a_1, a_2 \rangle - \langle a_2, a_3 \rangle| \leq 1 - \langle a_3, a_1 \rangle \tag{2.25}$$

holds.

The proof of this inequality (in such a rigorous mathematical formulation) can be found, e.g., in [161]; see also the original work of Bell [31] for a proof in the physical framework.

We now turn to Example 2.1. If a_1, a_2, a_3 can be realized on the same probability space, then (2.25) would hold. On the other hand, we have

$$\langle a_1, a_2 \rangle = 1; \quad \langle a_1, a_3 \rangle = 1; \langle a_2, a_3 \rangle = -1.$$

Bell's inequality should imply: $1 - (-1) = 2 \leq 1 - 1 = 0$. We remark that in accordance with Boole we consider Bell's inequality just as a necessary condition for probabilistic compatibility.

We also recall the following simple mathematical result, see Wigner [304]:

Theorem 2.3 (Wigner inequality) *Let $a_1, a_2, a_3 = \pm 1$ be arbitrary random variables on a Kolmogorov space $\mathcal{P}$. Then the following inequality holds:*

$$\mathbf{P}(a_1 = +1, a_2 = +1) + \mathbf{P}(a_2 = -1, a_3 = +1) \tag{2.26}$$

$$\geq \mathbf{P}(a_1 = +1, a_3 = +1).$$

Its proof is very simple, see Sect. 12.2. The crucial point is the use of a single probability measure $\mathbf{P}$.

2.2.3 Bell-type Inequalities for Conditional Probabilities

The original *Boole–Bell inequality* served to solve the problem of PC. In its simplest version this problem is based on the assumption that pairwise probability distributions are well defined – observables can be measured pairwise. However, even such an assumption is not always satisfied. Sometimes even for pairs of observables joint measurements are impossible, but it is possible to perform conditional measurements. For example, first the observable a_1 is measured and the result $a_1 = \alpha_1$ is obtained. Then under this condition the observable a_2 is measured. Conditional probability $\mathbf{P}(a_2 = \alpha_2 | a_1 = \alpha_1)$ can be found. The simplest test of PC – the possibility of realizing three observables a_1, a_2, a_3 on a single Kolmogorov probability

space – is based on conditional probabilities. As the author of this book noticed, by using Bayes' formula (2.10), i.e., assuming the validity of the Kolmogorovian definition of conditional probability, Wigner's inequality can be easily rewritten as an inequality for conditional probabilities:

Theorem 2.4 (Wigner–Khrennikov inequality) *Let $a_1, a_2, a_3 = \pm 1$ be arbitrary random variables on a Kolmogorov space $\mathcal{P}$. Then the following inequality holds:*

$$\mathbf{P}(a_1 = +1)\mathbf{P}(a_2 = +1|a_1 = +1) + \mathbf{P}(a_2 = -1)\mathbf{P}(a_3 = +1|a_2 = -1) \quad (2.27)$$

$$\geq \mathbf{P}(a_3 = +1)\mathbf{P}(a_1 = +1|a_3 = +1).$$

Thus if conditional probabilities for a triple of dichotomous observables violate this inequality, they exhibit PI; see Sect. 9.6 for application to game theory.

2.3 Quantum Probabilistic Model

The mathematical formalism of quantum mechanics is the theory of self-adjoint operators on complex Hilbert spaces. The symbols $\mathcal{H}$ and $\langle \cdot, \cdot \rangle$ denote separable complex Hilbert space and the scalar product on it; $\|\psi\| = \sqrt{\langle \psi, \psi \rangle}$ the norm of $\psi \in \mathcal{H}$;

$$S = \{\psi \in \mathcal{H} : \|\psi\| = 1\}$$

is the unit sphere in $\mathcal{H}$. We also consider the set of equivalence classes in the unit sphere S with respect to the equivalence relation: $\psi_1 \sim \psi_2$ iff $\psi_1 = c\psi_2$, where $c \in \mathbf{C}$ and $|c| = 1$. Denote this set by the symbol $\tilde{S}$.

Although real quantum physics is described by infinite-dimensional Hilbert space (of square integrable complex valued functions), quantum information is totally fine with finite dimensional spaces:

$$\mathcal{H}_n = \mathbf{C}^n = \mathbf{C} \times \ldots \times \mathbf{C}. \quad (2.28)$$

Since our considerations relate merely to informational features of the quantum model, we can proceed (practically everywhere) in the same way as in quantum information. The space $\mathcal{H}_n$ is endowed with the scalar product

$$\langle \psi, \phi \rangle = \sum_{j=1}^{n} \psi_j \overline{\phi_j}, \ \psi = (\psi_1, \ldots, \psi_n), \phi = (\phi_1, \ldots, \phi_n) \in \mathcal{H}_n. \quad (2.29)$$

Self-adjoint operators can be represented by Hermitian matrices, $\widehat{a} = (a_{ij})$, such that $a_{km} = \overline{a_{mk}}$, where $z = x + iy \rightarrow \bar{z} = x - iy$ is the operation of complex conjugation. The spectrum, $\text{Spec}(\widehat{a})$, is nothing else than the set of eigenvalues:

$\widehat{a}\psi = \alpha\psi$. We remark that all eigenvalues are real. Eigenvectors corresponding to the same eigenvalue α form a linear subspace. Its dimension gives the degree of degeneration of α. The orthogonal projector on this subspace is denoted by the symbol P_α^a. It acts similarly to the orthogonal projector to a plane or line in $\mathbf{R}^3$. Of course, the use of complex spaces makes direct geometric illustration impossible even for the space $\mathcal{H}_2$ – it is the four-dimensional real space.

2.3.1 Postulates

The probabilistic model of quantum theory can be formulated as the following series of postulates:

Postulate 1 (The mathematical description of quantum states.) *Quantum (pure) states (wave functions) are represented by normalized vectors ψ (i.e., $\|\psi\|^2 = \langle\psi, \psi\rangle = 1$) of a complex Hilbert space $\mathcal{H}$. Every normalized vector $\psi \in \mathcal{H}$ may represent a quantum state. If a vector ψ corresponding to a state is multiplied by any complex number c, $|c| = 1$, the resulting vector will correspond to the same state.*[8]

The physical meaning of "a quantum state" is not defined by this postulate. It must be provided by a separate postulate; see Postulates 6, 6a.

Postulate 2 (The mathematical description of physical observables.) *A physical observable a is represented by a self-adjoint operator $\widehat{a}$ in complex Hilbert space $\mathcal{H}$. Different observables are represented by different operators.*

Postulate 3 (Spectral) *For a physical observable a which is represented by the self-adjoint operator $\widehat{a}$ we can predict (together with some probabilities) values $\lambda \in$ Spec$(\widehat{a})$ (the spectrum of $\widehat{a}$).*

We restrict our considerations to the simplest self-adjoint operators, which are analogous to discrete random variables. We recall that a self-adjoint operator $\widehat{a}$ has *a purely discrete spectrum* if it can be represented as

$$\widehat{a} = \alpha_1 P_{\alpha_1}^a + \ldots + \alpha_m P_{\alpha_m}^a + \ldots, \quad \alpha_m \in \mathbf{R}, \tag{2.30}$$

where $P_{\alpha_m}^a$ are orthogonal projection operators related to the orthonormal eigenvectors $\{e_{km}^a\}_k$ of $\widehat{a}$ corresponding to the eigenvalues α_m by

$$P_{\alpha_m}^a \psi = \sum_k \langle\psi, e_{km}^a\rangle e_{km}^a, \quad \psi \in \mathcal{H}. \tag{2.31}$$

Here k labels the eigenvectors e_{km}^a which belong to the same eigenvalue α_m of $\widehat{a}$. Thus Spec$(\widehat{a}) = \{\alpha_1, \ldots, \alpha_m, \ldots\}$.

---

[8] Thus states are given by elements of $\tilde{S}$.

Postulate 4 (Born's rule – in formalization of Dirac and von Neumann) *Let a physical observable a be represented by a self-adjoint operator $\widehat{a}$ with purely discrete spectrum. The probability $\mathbf{P}_\psi(a = \alpha_m)$ of obtaining the eigenvalue α_m of $\widehat{a}$ for measurement of a in a state ψ is given by*

$$\mathbf{P}_\psi(a = \alpha_m) = \| P_m^a \psi \|^2. \tag{2.32}$$

If the operator $\widehat{a}$ has *nondegenerate* (purely discrete) spectrum, then each α_m is associated with a one-dimensional subspace. The latter can be fixed by selecting any normalized vector, say e_m^a. In this case orthogonal projectors act simply as

$$P_{\alpha_m}^a \psi = \langle \psi, e_m^a \rangle e_m^a. \tag{2.33}$$

The formula (2.32) takes a very simple form

$$\mathbf{P}_\psi(a = \alpha_m) = |\langle \psi, e_m^a \rangle|^2. \tag{2.34}$$

This is Born's rule in the Hilbert space formalism.

To obtain original Born's rule, one should choose $\mathcal{H}$ as the L_2-space of square integrable functions, $\psi : \mathbf{R} \mapsto \mathbf{C}$. (We consider a one-dimensional particle.) The position observable x is represented by the multiplication operator $\widehat{x}$

$$\widehat{x}(\psi)(x) = x\psi(x). \tag{2.35}$$

This operator has a continuous spectrum. It coincides with the whole real line. So, this operator is unbounded. Its eigenvectors do not belong to the L_2-space. They are given by Dirac's δ-functions, i.e., these are generalized eigenvalues, see Dirac [90]

$$\widehat{x}(e_\alpha)(x) = \alpha e_\alpha(x), \quad \alpha \in \mathbf{R}, \tag{2.36}$$

where $e_\alpha(x) = \delta(x - \alpha)$. One can reasonably define paring[9]

$$\langle \psi, e_\alpha \rangle = \psi(\alpha). \tag{2.37}$$

Then the rule (2.34) gives

$$\mathbf{P}_\psi(x = \alpha) = |\psi(\alpha)|^2. \tag{2.38}$$

---

[9] In fact, the situation is little bit more complicated from the mathematical viewpoint. In the rigorous mathematical framework, elements of the L_2-space are given by equivalent classes of functions. Two functions belong to the same class if the measure of points where they are distinct is equal to zero. To proceed rigorously, one should select a subspace in the L_2-space and consider Dirac's delta function and its shifts $e_\alpha(x) = \delta(x - \alpha)$ as continuous linear functionals on this subspace. This can be done in the framework of distribution theory. Here paring (2.37) is nothing else than action of the functional e_α to the test function ψ. However, physicists typically do not pay attention to such mathematical problems.

*Remark 2.2 (*Origin of Born's rule.*)* This rule was invented in the following way. Originally Schrödinger considered the ψ-function as a classical field – similar to the electromagnetic field. The quantity $E(\alpha) = |\psi(\alpha)|^2$ is the energy density of this field. Born invented the rule (2.38) by criticizing Schrödinger's interpretation. Instead of the energy density, he considered this quantity as the probability density. The latter induces automatically the normalization condition

$$1 = \int_{-\infty}^{+\infty} |\psi(\alpha)|^2 = \langle \psi, \psi \rangle$$

which was absent in Schrödinger's model. After a few years of struggle, Schrödinger gave up and kept to Born's interpretation.

In the same way one can consider momentum measurement. Schrödinger defined the momentum operator as

$$\widehat{p}(\psi)(x) = -i\frac{\mathrm{d}}{\mathrm{d}x}\psi(x). \tag{2.39}$$

(We eliminate the Planck constant from consideration by choosing the appropriate system of units.) It is easy to see that its spectrum is also continuous and it coincides with **R**. Its generalized eigenfunctions can be easily found from the equation

$$-i\frac{\mathrm{d}}{\mathrm{d}x}e_\beta^p(x) = \beta e_\beta^p(x), \quad \beta \in \mathbf{R}.$$

Thus $e_\beta^p(x) = \mathrm{e}^{i\beta x}$. Thus by (2.34)

$$\mathbf{P}_\psi(p = \beta) = |\langle \psi, e_\beta^b \rangle|^2. \tag{2.40}$$

By taking into account that

$$\langle \psi, e_\beta^b \rangle = \int_{-\infty}^{+\infty} \psi(x)\mathrm{e}^{-i\beta x}\mathrm{d}x = \tilde{\psi}(\beta)$$

is the Fourier transform of ψ, we write Born's rule for the momentum measurement as

$$\mathbf{P}_\psi(p = \beta) = |\tilde{\psi}(\beta)|^2, \tag{2.41}$$

cf. (2.38).

*Remark 2.3 (*Classical description of quantum measurements.*)* For any state ψ, each quantum observable $\widehat{a}$ can be represented as a classical random variable. In the discrete case we take $\Omega = \{\alpha_1, \ldots, \alpha_m, \ldots\} \equiv \mathrm{Spec}(\widehat{a})$, the σ-algebra consists of all subsets of Ω, and the probability measure is defined as $\mathbf{P}(A) = \sum_{\alpha_m \in A} \mathbf{P}_\psi$

$(a = \alpha_m)$, where $\mathbf{P}_\psi(a = \alpha_m)$ is given by Born's rule. Thus each concrete quantum measurement can be described classically. Problems arise only when one tries to describe classically data collected for a few incompatible observables. We remark that such attempts contradict Kolmogorov's ideology [219]. Kolmogorov emphasized that each probability space is determined by the corresponding complex of experimental conditions (context). The same message came from Bohr, who pointed out that the whole experimental arrangement should be taken into account and whose *principle of complementarity* supports Kolmogorovian ideology. For example, the impossibility of embedding the collection of probabilities for the position and momentum measurements (for all possible quantum states) into a single probability space is often considered as a new astonishing probabilistic situation. However, Kolmogorov's ideology implies that attempts at such an embedding have no justification – since the position and momentum measurements for a quantum system cannot be performed in a single experimental setting.

By using Born's rule (2.32) and the classical probabilistic definition of average (2.4), it is easy to see that the average value of an observable a in a state ψ belonging to the domain of definition of the operator $\widehat{a}$ is given by

$$\langle a \rangle_\psi = \langle \widehat{a}\, \psi, \psi \rangle. \tag{2.42}$$

Postulate 5 (Time evolution of wave function.) *Let $\widehat{H}$ be the Hamiltonian of a quantum system, i.e., the self-adjoint operator corresponding to the energy observable. The time evolution of the wave function $\psi \in \mathcal{H}$ is described by the Schrödinger equation*

$$i\frac{d}{dt}\psi(t) = \widehat{H}\psi(t) \tag{2.43}$$

with the initial condition $\psi(0) = \psi$.

2.3.2 Quantization

We remark that the operators of position and momentum, $\widehat{x}$ and $\widehat{p}$, see (2.35) and (2.39), do not commute and they satisfy Heisenberg's *canonical commutation relation*

$$[\widehat{x}, \widehat{p}] = \mathrm{i}. \tag{2.44}$$

Consider any real-valued function on the classical phase space, i.e., a function of classical coordinate and momentum, $f(x, p)$. The quantization procedure is the map

$$f \mapsto \widehat{f} = f(\widehat{x}, \widehat{p}). \tag{2.45}$$

In general, it is a tricky mathematical problem to define a function of two noncommuting operators. It is typically done by using the calculus of pseudo-differential operators.[10]

However, in the simplest case the operator of energy $\widehat{H}$ can be easily defined. Consider a classical particle with the mass m moving in the potential $V(x)$. Its Hamiltonian function (representing classical energy of this particle) is given by

$$H(x, p) = \frac{p^2}{2m} + V(x). \tag{2.46}$$

Quantization gives us the operator

$$\widehat{H} = H(\widehat{x}, \widehat{p}) = \frac{\widehat{p}^2}{2m} + V(x). \tag{2.47}$$

2.3.3 Interpretations of Wave Function

Now we are going to discuss one of the most important and complicated notions of quantum mechanics: the notion of a quantum state. There are two main points of view, which are formulated in the following postulates.

Postulate 6 (The ensemble interpretation.) *A wave function provides a description of certain statistical properties of an ensemble of similarly prepared quantum systems.*

This interpretation is upheld, for example by Einstein, Popper, Blokhintsev, Margenau, Ballentine, Klyshko, and in recent years by, e.g., de Muynck, De Baere, Holevo, Santos, Khrennikov, Nieuwenhuizen, Adenier and many others.

Postulate 6a (The Copenhagen interpretation.) *A wave function provides a complete description of an individual quantum system.*

This interpretation was supported by a great variety of scientists, from Schrödinger, in his original attempt to identify the electron with a wave function solution of his equation, to the proponents of the several versions of the Copenhagen interpretation (for example, Heisenberg, Bohr, Pauli, Dirac, von Neumann, Landau, Fock and, in recent years, e.g., Greenberger, Mermin, Lahti, Peres, Summhammer[11]). Nowadays

---

[10] See [160] for the most general presentation of quantization procedure on the mathematical level of rigorousness, including both bosons and fermions as well as supersymmetric systems, quantum field theory, strings and superstrings and corresponding string field theories; see [158, 159] for operator quantization over non-Archimedean (in particular, p-adic) number fields.

[11] There is an interesting story about the correspondence between Bohr and Fock on the individual interpretation. This story was told to me by a former student of Fock, who pointed out that one of the strongest supporters of this interpretation was Vladimir A. Fock, and that even though Bohr himself had doubts about its consistency, Fock demonstrated to Bohr inconsistency in the Einsteinian ensemble interpretation. Thus interpretation, which is commonly known as the *Copenhagen interpretation*, might just as well be called the "*Leningrad interpretation*."

the individual interpretation is extremely popular, especially in quantum information and computing.

Instead of Einstein's terminology "*ensemble interpretation*", Ballentine [25, 26] used the terminology *"statistical interpretation."* However, Ballentine's terminology is rather misleading, because the term "statistical interpretation" was also used by von Neumann for individual randomness! For him "statistical interpretation" had a meaning that is totally different from Ballentine's "ensemble-statistical interpretation." John von Neumann wanted to emphasize the difference between deterministic (Newtonian) classical mechanics, in which the state of a system is determined by values of two observables (position and momentum), and quantum mechanics, in which the state is determined not by values of observables, but by probabilities. We shall follow Albert Einstein and use the terminology *"ensemble interpretation."*

Finally, we point out recent papers concerning the foundations and, in particular, various interpretations of quantum mechanics: [3–6, 13, 21, 106, 116, 130, 4, 17, 21, 27, 45, 47, 66, 77, 79, 83–85, 106, 116, 117, 130, 119, 127, 142–144, 148, 161–214, 215, 225, 226, 261–264, 272, 284–292].

2.4 Quantum Conditional Probability

As in the classical Kolmogorov probabilistic model, Born's postulate should be completed by a definition of conditional probability. We present the contemporary definition that is conventional in quantum logic [32] and quantum information theory.

Definition 2.1 Let physical observables a and b be represented by self-adjoint operators with purely discrete (possibly degenerate) spectra:

$$\widehat{a} = \sum_m \alpha_m P^a_{\alpha_m}, \ \widehat{b} = \sum_m \beta_m P^b_{\beta_m} \tag{2.48}$$

Let ψ be a pure state and let $P^a_{\alpha_k} \psi \neq 0$. Then the probability of obtaining the value $b = \beta_m$ under the condition that the value $a = \alpha_k$ was observed in the preceding measurement of the observable a on the state ψ is given by

$$\mathbf{P}_\psi(b = \beta_m | a = \alpha_k) \equiv \frac{\| P^b_{\beta_m} P^a_{\alpha_k} \psi \|^2}{\| P^a_{\alpha_k} \psi \|^2} \tag{2.49}$$

Let the operator $\widehat{a}$ have a nondegenerate spectrum, i.e., for any eigenvalue α the corresponding eigenspace (i.e., generated by eigenvectors with $\widehat{a}\psi = \alpha\psi$) is one dimensional. We can write

$$\mathbf{P}_\psi(b = \beta_m | a = \alpha_k) = \| P^b_{\beta_m} e^a_k \|^2 \tag{2.50}$$

(here $\widehat{a}e_k^a = \alpha_k e_k^a$). Thus the conditional probability in this case does not depend on the original state ψ. We can say that the memory of the original state has been destroyed. If also the operator $\widehat{b}$ has a nondegenerate spectrum then we have $\mathbf{P}_\psi(b = \beta_m | a = \alpha_k) = |\langle e_m^b, e_k^a \rangle|^2$ and $\mathbf{P}_\psi(a = \alpha_k | b = \beta_m) = |\langle e_k^a, e_m^b \rangle|^2$. By using symmetry of the scalar product we obtain:

Proposition 2.2 *Let both operators $\widehat{a}$ and $\widehat{b}$ have purely discrete nondegenerate spectra and let $P_k^a \psi \neq 0$ and $P_m^b \psi \neq 0$. Then conditional probability is symmetric and it does not depend on the original state ψ :*

$$\mathbf{P}_\psi(b = \beta_m | a = \alpha_k) = \mathbf{P}_\psi(a = \alpha_k | b = \beta_m) = |\langle e_m^b, e_k^a \rangle|^2.$$

We remark that classical (Kolmogorov–Bayes) conditional probability is not symmetric, except in very special situations; the same is valid for my general contextual probabilistic model, see Chapt. 3. Thus *QM is described by a very specific probabilistic model.*

Consider two nondegenerate observables. Set $p_{\beta|\alpha} = \mathbf{P}(b = \beta | a = \alpha)$. The matrix of transition probabilities $\mathbf{P}^{b|a}$, see (2.15) for the definition (but do not forget that transition probabilities are no longer defined by Bayes' rule!), is not only *stochastic* but *doubly stochastic.* It is easy to see that

$$\sum_\alpha p_{\beta|\alpha} = \sum_\alpha |\langle e_\beta^b, e_\alpha^a \rangle|^2 = \langle e_\beta^b, e_\beta^b \rangle = 1.$$

Double stochasticity is also a very specific property of quantum probability, cf. the Kolmogorovian model and my model Chap. 3. In fact, condition **DS-DS** holds: both matrices of transition probabilities $\mathbf{P}^{a|b}$ and $\mathbf{P}^{b|a}$ are doubly stochastic. Moreover, any pair of quantum observables (with nondegenerate spectra) satisfies to condition **SC**; they are "symmetrically conditioned", see (2.18).

In the quantum framework *independent observables* are considered in Sect. 12.1.2.

2.5 Interference of Probabilities in Quantum Mechanics

We will show that quantum probabilistic calculus violates the conventional FTP, see Sect. 2.1.3.

Let $\mathcal{H}_2 = \mathbf{C} \times \mathbf{C}$ be the two-dimensional complex Hilbert space and let $\psi \in \mathcal{H}_2$ be a quantum state. Let us consider two dichotomous observables $b = \beta_1, \beta_2$ and $a = \alpha_1, \alpha_2$ represented by self-adjoint operators $\widehat{b}$ and $\widehat{a}$, respectively (one may consider simply Hermitian matrices). Let $e^b = \{e_\beta^b\}$ and $e^a = \{e_\alpha^a\}$ be two orthonormal bases consisting of eigenvectors of the operators. The state ψ can be represented in the two ways

$$\psi = c_1 e_1^a + c_2 e_2^a, \ c_\alpha = \langle \psi, e_\alpha^a \rangle; \tag{2.51}$$

$$\psi = d_1 e_1^b + d_2 e_2^b, \ d_\beta = \langle \psi, e_\beta^b \rangle. \tag{2.52}$$

By Postulate 4 we have

$$\mathbf{P}(a = \alpha) \equiv \mathbf{P}_\psi(a = \alpha) = |c_\alpha|^2; \tag{2.53}$$

$$\mathbf{P}(b = \beta) \equiv \mathbf{P}_\psi(b = \beta) = |d_\beta|^2. \tag{2.54}$$

The possibility of expanding one basis with respect to another basis induces connection between the probabilities $\mathbf{P}(a = \alpha)$ and $\mathbf{P}(b = \beta)$. Let us expand the vectors e_α^a with respect to the basis e^b

$$e_1^a = u_{11} e_1^b + u_{12} e_2^b; \tag{2.55}$$

$$e_2^a = u_{21} e_1^b + u_{22} e_2^b, \tag{2.56}$$

where $u_{\alpha\beta} = \langle e_\alpha^a, e_\beta^b \rangle$. Thus $d_1 = c_1 u_{11} + c_2 u_{21}, \ d_2 = c_1 u_{12} + c_1 u_{22}$. We obtain the *quantum rule* for transformation of probabilities

$$\mathbf{P}(b = \beta) = |c_1 u_{1\beta} + c_2 u_{2\beta}|^2. \tag{2.57}$$

On the other hand, by the definition of quantum conditional probability, see (2.49), we obtain

$$\mathbf{P}(b = \beta | a = \alpha) \equiv \mathbf{P}_\psi(b = \beta | a = \alpha) = |\langle e_\alpha^a, e_\beta^b \rangle|^2. \tag{2.58}$$

By combining (2.53), (2.54) and (2.57), (2.58) we obtain the *quantum formula of total probability – the formula of interference of probabilities*:

$$\mathbf{P}(b = \beta) = \sum_\alpha \mathbf{P}(a = \alpha)\mathbf{P}(b = \beta | A = \alpha)$$
$$+2\cos\theta \sqrt{\mathbf{P}(a = \alpha_1)\mathbf{P}(b = \beta | a = \alpha_1)\mathbf{P}(a = \alpha_2)\mathbf{P}(b = \beta | a = \alpha_2)} \tag{2.59}$$

In general $\cos\theta \neq 0$. Thus the quantum FTP does not coincide with the classical FTP (2.23) which is based on Bayes' formula

$$\mathbf{P}(b = \beta) = \sum_\alpha \mathbf{P}(a = \alpha)\mathbf{P}(b = \beta | a = \alpha). \tag{2.60}$$

2.6 Contextual Point of View of Interference

The difference between the quantum rule (2.59) and the classical rule (2.60) is not surprising. As was pointed out in Remark 2.2, there are no reasons to expect that data obtained for observables a and b which could not be jointly measured can be described by a single Kolmogorov probability space. However, the classical FTP, see Sect. 2.1.3, was derived under the assumption that both observables can be represented by random variables belong to the same Kolmogorov space.[12]

The crucial point is that one cannot use the same symbol $\mathbf{P}$ to denote all probabilities in (2.59). In one formula, (2.59), one combines probabilistic data obtained in four different experiments (experimental contexts):

a) measurement of the observable a under the complex of physical conditions (context) C which is represented by the initial state ψ;
b) measurement of the observable b under the same context C;

After performing the a-measurement one can create through selection procedures C_{α_1} and C_{α_2} (selections of systems with respect to the values $a = \alpha_1$ and $a = \alpha_2$) two new ensembles of systems S_{α_1} and S_{α_2}. In quantum mechanics (with the ensemble interpretation) these ensembles are represented by the eigenvectors e_1^a, e_2^a of the operator $\widehat{a}$. Therefore we can perform the b-measurement for two new contexts:

a1) measurement of the observable b under the complex of physical conditions (context) C_{α_1} which is represented by the state e_1^a;
a2) measurement of the observable b under the complex of physical conditions (context) C_{α_2} which is represented by the state e_2^a.

The a)-experiment gives probabilities $\mathbf{P}_\psi(a = \alpha)$; the b)-experiment – $\mathbf{P}_\psi(b = \beta)$; the a1)-experiment – $\mathbf{P}_{e_1^a}(b = \beta)$; the a2)-experiment – $\mathbf{P}_{e_2^a}(b = \beta)$.

What could be the reason to assume that we can use a single probability measure $\mathbf{P}$ in all these experiments?

2.7 Bell's Inequality in Quantum Physics

As was pointed out in Sect. 2.2, inequalities of Boole–Bell type provide necessary conditions for probabilistic compatibility (PC) of families of observables and, hence, their violations provide sufficient conditions for probabilistic incompatibility (PI). As was first pointed out by Bell, see [31] for details, quantum formalism

---

[12] We remark that Feynman [105] considered violation of FTP in the two-slit experiment as violation of the laws of classical probability. For him it was an exhibition of special, even mystical, properties of quantum systems. A similar comment by d'Espagnat on violation of FTP can be found in [87].

predicts the existence of such quantum states[13] that inequality (2.25) is violated for a special choice of a family of pairwise measurable observables.[14] Thus these observables are of the PI-type.

In any domain of science, one should look for special roots of PI. In particular, in physics Bell found two possible roots: quantum nonlocality and death of realism. Moreover, he was sure that one can still proceed in QM by using the realistic description in its strongest (Einsteinian) form: assigning values of observables to the state of a quantum system before measurement.[15] In principle, one cannot exclude that he found the right possible roots.

My approach is essentially more general. By considering the problem from the PI viewpoint, we can look for other roots of PI, which need not coincide with those proposed by Bell. One can still keep to realism and locality. PI can arise from, e.g., taking into account parameters of measurement devices (so considering values of observables as depending on internal states not only of systems, but also of measurement devices[16]), or from unfair sampling; details can be found in [204, 7].

Moreover, Bell-type inequalities for probability distributions (or covariances) of pairwise measurements are not the simplest tests of PC. As was mentioned in Sect. 2.2.3, PC can be tested by conditional measurements of three observables by using the Wigner–Khrennikov inequality (2.27). It is easy to see [214] that this inequality is violated for specially selected projections of spin or polarization. Conditional measurements, e.g., spin projections to one direction and then to another, can be performed on a single particle. Unlike Bell's original scheme, we need not consider pairs of entangled particles. Hence, PI of spin or polarization projections take place even for a single particle. It is completely clear that the source of PI is the impossibility of measuring these observables simultaneously. It would be surprising if PI for spin or polarization projections derived by using Bell's original inequality for entangled pairs has another explanation, e.g., nonlocality. By operating with the Wigner–Khrennikov inequality for conditional probabilities one can see how artificial Bell's appeal to nonlocality was.

---

[13] These are so called EPR-type states, see Einstein, Podolsky, Rosen [99] for details.

[14] For example, spin or polarization projections to specially chosen directions.

[15] Here by state we understood "prequantum state", hidden variable, λ. Thus, first of all, J. Bell was sure that QM does not provide the complete description of phenomena. As well as Einstein, he was sure that one can finally find a better description of physical reality than given by QM. The reason of Bell's belief in naive Einsteinian realism were precise correlations (anti-correlations) exhibited by measurements for EPR-type states. Thus Bell was sure that violation of inequality (2.25) implies nonlocality. For him, the best model of prequantum reality was given by Bohmian mechanics. Later, as is often happen in science, majority of people combine nonlocality with rejection of realism. The monster of mysterious "quantum nonlocality" was born. It is clear that Bell would not be happy with such an interpretation of his studies. However, it is clear as well that Einstein would not be happy with nonlocal realism. His reaction to creation of Bohmian mechanics was negative.

[16] Such sort of realism differs from naive *Einsteinian realism* and it is closer to Bohr's views; cf. also with Accardi's chameleon effect [1, 4] and Ohya's adaptive dynamics [245, 246].

2.8 Växjö Interpretation of Quantum Mechanics

The Växjö interpretation [177] is a variant of the ensemble interpretation, Postulate 6:

A wave function provides a description of certain statistical properties of an ensemble of similarly prepared quantum systems.

However, "properties" are not Einsteinian properties, which can be assigned to a system before measurement. Properties should be understood in Bohr's sense: as results of interaction of systems with measurement devices. However, unlike Bohr, I do not claim that QM is complete and it is in principle impossible to provide a finer description of reality, e.g., by taking into account internal states of measurement devices, see [214, 184, 191].