Interpretations of Probability

Andrei Khrennikov

Walter de Gruyter

Khrennikov · Interpretations of Probability

Andrei Khrennikov

Interpretations of Probability

Second revised and extended edition

Walter de Gruyter
Berlin · New York

Author

Andrei Khrennikov
International Center for Mathematical Modeling
Växjö University
Vejdes plats 7
35195 Växjö, Sweden
E-mail: andrei.khrennikov@vxu.se

Mathematics Subject Classification 2000: 60-02, 60A05

Keywords: Non-Kolmogorov probability models, negative probabilities, quantum probabilities, *p*-adic probability theory, contextual probability, interference of probabilities, complex probability amplitudes

The first edition was published by VSP (Utrecht, The Netherlands) in 1999.

⊗ Printed on acid-free paper which falls within the guidelines of the ANSI to ensure permanence and durability.

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

ISBN 978-3-11-020748-4

Printed in Germany.
Cover design: Martin Zech, Bremen.
Typeset using the author's LATEX files: Kay Dimler, Müncheberg.
Printing and binding: Hubert & Co. GmbH & Co. KG, Göttingen.

I dedicate this book to Olga

Preface: edition 2008

The main aim of this book is to show that probability theory as well as geometry is characterized by a huge diversity of mathematical models. Various probabilistic models are induced by various interpretations of probability. Unfortunately, during last 70 years one special probabilistic model, namely, the measure-theoretic model [133] (Kolmogorov, 1933) became totally dominating in science – in mathematics, physics, biology, psychology, economics. The aim of this book is not to criticize the Kolmogorovean model. We recognize that it is an important probabilistic model which demonstrated tremendous success in many domains of science. The main problem is not that this model is bad. The problem is totality of the use of this model. Such a totality may induce improper applications. It would be too naive for a modern scientist to hope that any model might describe all possible natural and social phenomena. Any model has its own domain of application. Hence, the Kolmogorovean model also has its domain of validity. An attempt to apply it outside such a domain might induce paradoxical consequences for the corresponding science (to which this model is applied).[1]

We would like to show that unsatisfactory situation in quantum foundations is a consequence of improper application of the Kolmogorovean model.[2] The main problem is an attempt to embed quantum probabilistic data collected in a few experiments into a single Kolmogorov probability space – to manipulate with the absolute probability **P** without taking care about its coupling to special experimental contexts. Roughly speaking for any fixed experimental context a Kolmogorov space can be used. Probabilistic data collected in a fixed experiment in quantum (as well as classical) physics can be described by the conventional measure-theoretic model, but not data collected for a few incompatible observables. If one understood this, then such mystical things as e.g. interference of probabilities (which Feynman considered as as exhibition of violation of laws of classical probability, namely, its additivity) or violation of Bell's inequality can be easily explained in the classical approach, but based on intelligent taking into account context-dependence of probabilities.

---

[1] In the same way the Euclidean model plays the fundamental role in geometry. However, nowadays everybody understands well terrible consequences of an attempt to reduce the use of geometry in physics (especially relativity theory) to the Euclidean model. Since discovery of the first of non-Euclidean model by Nikolay Lobachevsky, nobody tries to find the 'best geometric model' which might be used in all applications.

[2] Total absence of natural realistic grounds for quantum mechanics, the common use of Copenhagen interpretation by which it is simply forbidden to try to create such grounds, the idea that quantum probability is totally different from classical probability, consideration of interference of probabilities (e.g. in the famous two slit experiment) and violation of Bell's inequality as violations of rules of classical probability theory. Recently (in connection with Bell's inequality) the 'state of health of the quantum patient' became even worse. Nowadays people (especially in quantum information theory) openly speak about such things as quantum nonlocality or quantum teleportation.

We remark that Andrei Nikolaevich Kolmogorov understood well this contextual story. He emphasized that each complex of experimental physical conditions is described by its own probability space [133]. Unfortunately, he was not able to explain this crucial point to simple users of his axiomatics. If one takes randomly a book on probability theory, a single probability space presentation would be found with probability one. We also remark that Niels Bohr pointed out that all experimental arrangement should be taken into account [19]. In particular, in this way he explained for called Einstein–Podolsky–Rosen paradox [58] in quantum foundations. Unfortunately, Bohr never tried to present his arguments in the probabilistic framework.

One of the basic classical probabilistic models which is different from the measure-theoretic one is the frequency model. Here one starts not directly with probabilities, but with random sequences, collectives. Probabilities are defined as limits of relative frequencies. This model was developed by Richard von Mises [169–171] in 1919 (see even Venn, 1866 – [165]). Kolmogorov used in [133] the Misesian model as the intuitive basis of his model. The main distinguishing feature of the Misesian model is its fundamental contextuality expressed in the form of dependence of probabilities on corresponding collectives. Therefore it would be natural to use in quantum theory not the Kolmogorovean model, but the Misesian model.[3] We shall do this in this book. It will be shown that one can proceed without mysteries and paradoxes if quantum probability is considered as classical, but frequency probability.

We are well aware about difficulties in definition of collective (random sequence).[4] This is an extremely complicated problem. Nowadays it is commonly accepted that random sequences can be defined either via Kolmogorov's algorithmic complexity or via Martin-Löf's theory of recursive statistical tests. However, as was pointed out by Van Lambalgen [163] such a viewpoint is not totally justified. It should recognized that we still do not have a satisfactory mathematically rigorous definition of random sequence.

We would like to escape difficulties related to the notion of random sequence in application of the frequency probability model in quantum theory. At least at the moment quantum physicists are not interested in study of randomness of sequences of results of measurements. It is assumed that quantum nature is intrinsically random. Hence it automatically produces 'random sequences' (even if mathematicians were not able to define them rigorously). The main aim is to find relative frequencies for results of observations. Thus we can proceed by using, instead of von Mises collectives, so called S-sequences: sequences of results of observations in which relative frequencies have

[3] As was already mentioned, one can proceed even in contextual Kolmogorovean framework. But dependence on an experimental context is not present in the Kolmogorov axiomatics. It is an implicit assumption which was clear for Kolmogorov, but it was practically forgotten. In the Misesian model contextual dependence of probabilities is really intrinsic.

[4] Von Mises – [169–171], see also Ville – [166], Wald – [173], Tornier – [162], Church – [39] as well as latter attempts Kolmogorov – [135], [134], [159], Chaitin [38], Solomonoff – [160], Martin-Löf – [142–144], Schnorr – [158], Zvonkin and Levin – [162], Van Lambalgen – [163], Khrennikov and Yamada [130], see Chapter 1, Section 8, of this book for a short introduction.

limits when the number of trials goes to infinity. We are not interested in randomness of such sequences. It is sufficient for our purposes that we can define frequency probabilities and identify them with probabilities provided by the mathematical formalism of quantum mechanics.

More details about book's content can be found in the preface for 1999-edition (in particular, about 'exotic probabilities' such as negative and p-adic probabilities). We just point to main changes which have been done comparing with previous editions. The first part of Chapter 2 was essentially rewritten. Probabilistic interpretations of the wave function are presented in clearer way: ensemble realist (Einsteinian, which is also known as Ballentine's 'statistical interpretation' [20]), ensemble empiricist (Bohrian), ensemble realist observational (Växjö interpretation), individual realist and empiricist (versions of the Copenhagen interpretation). The chapter devoted to information and probability was cancelled, because it did not match completely to the main stream of presentation. The book was completed by two new chapters, Chapters 6, 7, devoted to the detailed presentation of contextual realistic probabilistic model (Växjö model) and its representation in complex Hilbert space. The main aim of these chapters is to show that quantum-like representations of probability in complex (and more general) linear spaces can be created by starting with a classical probability model. We present an algorithm – quantum-like representation algorithm (QLRA) – which transfers probabilistic data of any origin into complex probability amplitudes (normalized vectors in complex Hilbert space).

By applying QLRA one can transform data obtained in any domain of science, e.g. in psychology or economics, into probabilistic vectors and then operate with this data by using the mathematical formalism of quantum mechanics (e.g. to apply quantum computational algorithms).

Växjö–Copenhagen, May 2008.

Preface: edition 1999

The modern axiomatics of probability theory was created by Andrei Nikolaevich Kolmogorov in 1933. This axiomatics is based on the measure-theoretical approach to probability. The main advantage of the Kolmogorov probabilistic formalism is the high level of abstractness. The use of abstract probability spaces gives the possibility to develop general probabilistic calculus in that structures of concrete spaces of elementary events do not play any role. The Kolmogorov probabilistic formalism was successfully developed in many directions. This formalism is now the mathematical basis for numerous probabilistic models in physics, technique, biology, finance. Everything is perfect in the landscape of Kolmogorov probability theory except for one dark cloud obscuring the horizon.

This cloud is the probability foundation of quantum mechanics. This cloud was generated by A. Einstein, B. Podolsky and N. Rosen (EPR) in 1935 (just two years after creation of the axiomatics of probability theory). EPR started the discussion on *completeness* of quantum mechanics.[5] In fact, the problem of completeness has the close connection with foundations of probability theory. EPR proposed some arguments that can be interpreted as the evidence of incompleteness of quantum mechanics. The following discussion on the EPR arguments demonstrated that quantum mechanics has quite marshy foundations. Often the EPR arguments are even considered as the paradox in foundations of quantum mechanics. During following thirty years dark quantum cloud obscuring the landscape of the Kolmogorov probability theory was rather small. The probabilistic roots of the EPR paradox were not so evident. Nobody tried to connect the paradox in foundations of quantum mechanics with foundations of probability theory. The first attempt to provide the probabilistic representation of the EPR considerations was done by J. Bell who found in 1964 famous Bell's inequality for covariations of physical observables involved in the EPR experiment. The black quantum cloud became quite large. Even in 1964 it is not only blotted out the sun of the Kolmogorov landscape, but it was gathering to obscure the beautiful idea of unique and general probability theory. However, nothing occurred in 1964. Moreover, nothing occurred in the following thirty years. And it seems to be that nothing is gathering to occur with unique and general Kolmogorov probability theory.

The great Kolmogorov probability community is still working in the standard measure-theoretical formalism. They do not pay attention to quantum clouds. On the other hand, the majority of the physical community observes this cloud. However, physicists do not understand the hidden probabilistic structure of this cloud. Some of

[5] N. Bohr and W. Heisenberg as well as Pauli, Landau, Fock were sure that quantum mechanics is complete, i.e., the wave function provides the complete description of the state of a quantum system. For them it was totally impossible to go beyond quantum mechanics, i.e., to provide a finer description of the state than given by quantum mechanics.

them support the idea of the *death of reality.* They think that it is impossible to use realism in quantum considerations. And if there is no reality at all, they do not afraid this non-real cloud. Other physicists support the idea of *nonlocality.* They think that physical reality is nonlocal. Thus by doing some measurement for a quantum system in Moscow we change the quantum state of the correlated quantum system which is located in Vladivostok. The adherents of nonlocality also do not observe the black cloud: this cloud is distributed everywhere and, hence, such a cloud could not induce storm.

There are many reasons for this strange situation. One of them is purely psychological. Mathematicians are not interested in quantum physics (mainly because they still do not know quantum theory). Physicists are not interested in foundations of probability theory (mainly because they know not so much even about the standard Kolmogorov measure-theoretical approach). In principle, even J. Bell in 1964 could pay attention that Bell's inequality is connected not only with such properties of physical observables as realism and locality, but also with the way of the probabilistic description. However, this was not done[6]. Bell's inequality was not considered as a sign for reconsideration of the foundations of probability theory. In the opposite to geometry probability theory was not transformed in an elastic formalism containing numerous probabilistic models which can be used for descriptions of different physical phenomena. Probability theory is still a rigid structure. This structure can be compared with the rigid Euclidean cub. Attempts to use the unique Kolmogorov model for describing all physical phenomena can be compared with attempts to represent all geometrical models by Euclidean cubs. However, geometric reality is not restricted to reality of cubs as well as probabilistic reality is not restricted to reality of Kolmogorov probability spaces.

In this book we demonstrate that 'pathological behaviour' of 'quantum probabilities is a consequence of the use of Kolmogorov's approach. The high level of abstractness does not give the possibility to control connection between probabilities and *statistical ensembles* or *random sequences* (collectives). Formal manipulations with abstract Kolmogorov probabilities produce such monsters as Bell's inequality (in fact, this idea was already discussed by L. de Broglie and later improved by G. Lochak).

First attempts to reduce the EPR paradox to the use of one concrete probability model, namely, the Kolmogorov model, were the works of L. Accardi and I. Pitowsky. L. Accardi proposed rather formal non-Kolmogorovean model which did *not contain Bayes' formula.* I. Pitowsky proposed to consider events which are described by *non-measurable sets.* The latter formalism was strongly improved by S. Gudder who developed the theory of *probability manifolds.* The main disadvantage of all these models is that they have even higher level of abstractness than Kolmogorov's model. Therefore they provide merely a new description of quantum phenomena. They could not explain probabilistic roots of quantum behaviour. The same can be said about so called *quantum probabilities.* Of course, quantum probability calculus gives the useful and

[6]I had numerous discussions with scientists worked with J. Bell. Unfortunately the general opinion is that J. Bell had never been interested in probability theory.

convenient description of quantum phenomena. However, quantum probability has no direct connection with probability. This is just rather speculative use of the word ‘probability’ in some formal mathematical constructions.

In this book we *explain* quantum probabilistic behaviour by using two basic interpretations of probability: *ensemble* and *frequency*. We demonstrate that (despite of the common opinion) the ensemble and frequency probability models are not in general equivalent to Kolmogorov’s model.

The frequency model is von Mises’ theory of collectives (random sequences), 1919. In this model probabilities are defined as limits of relative frequencies, $\nu_N = n/N$, in a collective x. This model was intensively studied in probability theory to find a reasonable definition of a random sequence (Kolmogorov algorithmic complexity and Martin-Löf theory of tests for randomness). It is the common opinion that frequency probabilities can be reduced to Kolmogorov probabilities. We explain (using the original arguments of R. von Mises) that such a viewpoint is totally wrong. In particular, the law of large numbers does not describe the statistical stabilization of relative frequencies, $\nu_N = n/N$, in a collective x. It will be shown that the frequency probability model has many features which extremely differ from features of Kolmogorov’s model. The most important feature of frequency probability is dependence of probabilities on a collectives x. In particular, the careful control of such a dependence gives the possibility to eliminate Bell’s inequality from considerations. The frequency model also differs from Kolmogorov’s model in the approach to conditional probabilities and independence. Such a difference also plays the important role in quantum considerations.

The ensemble model, as well as the frequency model, is one of basic ‘pre-Kolmogorov’ models. For example, the well-known Bernoulli theorem is, in fact, a theorem for ensemble probabilities. It is commonly supposed that the definition of ensemble probability (as a proportion in an ensemble) is just a particular case of Kolmogorov’s measure-theoretical definition. It is not right. The ensemble probability model cannot be reduced to Kolmogorov’s one. The most important feature of ensemble probabilities is dependence on an ensemble. In particular, the careful control of such a dependence gives the possibility to eliminate Bell’s inequality from considerations. It is impossible to use Kolmogorov’s measure-theoretical approach for describing ensemble probabilities for infinite ensembles. In fact, Kolmogorov’s measures are not proportional distributions of properties of ensembles of physical systems. These are measures on ensembles of all possible sequences of results of measurements. To obtain the adequate mathematical description of ensemble (proportional) probabilities for infinite ensembles (and quantum states describe such ideal ensembles), we have to leave the domain of Kolmogorov’s probability model and, moreover, the domain of real analysis. We have to use number systems which contain *actual infinities.* In this book we use systems of so called p-adic numbers $\mathbb{Q}_p$ (where $p > 1$ are prime numbers) for the description of some infinite statistical ensembles ($\mathbb{Q}_p$ contains actual infinities).

The origin of p-adic ensemble probabilities can be illustrated by the following example. Let S be an infinite ensemble of balls. Each ball has some colour $c \in C = \{0, 1, 2, \ldots, k, \ldots\}$ (countable system of colours). The S has the following colour structure: there are $n_k = 2^k$ balls with the colour $k \in C$ in S. The 'volume' $N = |S|$ of S can be easily found:

$$N = \sum_{k=0}^{\infty} n_k = \sum_{k=0}^{\infty} 2^k.$$

Of course, this series diverges in the field of real numbers $\mathbb{R}$. But it converges in the field of 2-adic numbers $\mathbb{Q}_2$. The sum of this series $\mathbb{Q}_2$ can be found by using ordinary formula for the sum of infinite geometric progression (because $2^k \to 0$, $k \to \infty$, in $\mathbb{Q}_2$):

$$N = \sum_{k=0}^{\infty} 2^k = \frac{1}{1-2} = -1.$$

We can now find the proportion of balls with the colour $k \in C$ in the ensemble S:

$$\mathbf{P}_S(k) = \frac{n_k}{N} = -2^k.$$

We remark that, as $n_k = 2^k$ is a finite number and $N = -1$ is an infinite number, the probability $\mathbf{P}_S(k)$ is *infinitely small probability.* And such a probability is represented by a negative number. This approach induces the rigorous mathematical theory of *negative probabilities.*

In fact, negative probabilities is other cloud obscuring the Kolmogorov probability landscape. Negative probabilities (which could not be justified by Kolmogorov's model) arise with the strange regularity in practically all quantum models. The most famous are Wigner distribution on the phase space and Dirac's negative probability distributions in the formalism of relativistic quantization. Such 'probability distributions' are considered as monsters of quantum theory. For example, physicists always underline that Wigner distribution is not really a probability distribution. At the same time they continue to use it for describing probabilistic phenomena. In this book negative probabilities (in particular, the Wigner distribution) are realized as probabilities with respect to infinitely large statistical ensembles. In many physical models these probability have the interpretation of infinitely small probabilities. Negative probabilities can also be obtained in the frequency approach as limits of relative frequencies, $\nu_N = n/N$, with respect to some topology on the set rational numbers $\mathbb{Q}$ which differs from the standard real topology (and frequencies $\nu_N = n/N$ always belong to $\mathbb{Q}$). For example, in the p-adic topology the probability $\mathbf{P} = -1$ can be obtained as the limit of frequencies:

$$\mathbf{P} = -1 = \lim \nu_N.$$

Typically in the frequency approach the presence of negative probabilities is the exhibition of the violation of the principle of the statistical stabilization for relative frequencies with respect to the real topology. Negative frequency probabilities can also appear via, for example, p-adic splitting of conventional probability $\mathbf{P} = 0$. In the latter case negative probabilities can be again interpreted as infinitely small probabilities.

The last chapter of the book has purely mathematical character. Here we develop a p-adic analogue of the Martin-Löf theory of tests for randomness (to find a p-adic analogue of a random sequence). p-adic theory strongly differs from the real one. There is no universal test for randomness (at least for the uniform p-adic probability distribution). In this sense the p-adic theory of randomness is similar to Schnorr's theory for Kolmogorov probabilities. We also obtained a large class of limit theorems for p-adic probabilities. In particular, these limit theorems can be applied to negative probabilities. We remark that the first limit theorem for negative probabilities was proved by M. Barnett in 1944.

Main consequences of the book:

1. Kolmogorov's probability theory (measure-theoretical approach) is just one of many probability models.
2. Two fundamental interpretations of probability, namely, the *ensemble and frequency interpretations,* can be used as the basis for numerous non-Kolmogorovean models.
3. Negative probabilities are well defined on the mathematical level of rigorousness.
4. Pathological (or nonclassical) behaviour of 'quantum probabilities' (and, in particular, Bell's inequality) is a consequence of the formal use of Kolmogorov's probability model.
5. Bell's inequality could not be used as an argument for *nonlocality or nonreality.* In may be that physical reality nonlocal or nonobjective. However, Bell's inequality has nothing to do with these problems.
6. The Wigner distribution is well defined both in the ensemble and frequency frameworks.
7. From the frequency viewpoint non-Kolmogorovean probabilistic behaviour is (typically) the exhibition of the violation of the law of large numbers.
8. From the ensemble viewpoint non-Kolmogorovean probabilistic behaviour is a consequence of the use of ensembles of infinitely large 'volume'. There is no statistical *reproducibility* of properties for finite approximations of these infinite ensembles.
9. Quantum states (wave functions) describe such infinite (ideal) ensembles with statistical nonreproducibility of properties.

A large number of mathematicians and physicists took part in the discussion of results exposed in this book. I want to use the opportunity to express my deepest gra-

titude to all of them. I feel myself especially indebted to L. Accardi, S. Albeverio, H. Atmanspacher, Z. Hradil, W. de Muynck, H. Rauch, J. Summhammer for fruitful discussions.

Växjö–Clermont-Ferrand–Tokyo, 1998–99.

Contents

1 Foundations of probability theory

There is no 'general probability theory'. There exist an incredible number of different mathematical probabilistic formalisms [132–136], [33], [34], [63], [44], [62], [75], [169–171], [2], [66], [148], [80], and, moreover, each of these formalisms has a few different interpretations. We shall discuss some of these theories which will be useful in further physical considerations.

1 A few words about measures

We recall some notions of measure theory. A system F of subsets of a set Ω is called an *algebra* if the sets $\varnothing, \Omega$ belong to F and the union, intersection and difference of two sets of F also belong to F. In particular, for any $A \in F$, a *complement* $\bar{A} = \Omega \setminus A$ of A belongs to F. Denote by F_Ω the family of all subsets of Ω. This is the simplest example of an algebra.

Let F be an algebra. A map $\mu : F \to \mathbb{R}_+$ is said to be a *measure* if $\mu(A \cup B) = \mu(A) + \mu(B)$ for $A, B \in F$, $A \cap B = \varnothing$. A measure μ is called σ-additive if, for every sequence $\{A_n\}_{n=1}^{\infty}$ of sets $A_n \in F$ such that their union $A = \bigcup_{n=1}^{\infty} A_n$ belongs to F, $\mu(A) = \sum_{n=1}^{\infty} \mu(A_n)$.

An algebra $\mathcal{F}$ is said to be a σ-algebra if, for every sequence $\{A_n\}_{n=1}^{\infty}$ of sets $A_n \in \mathcal{F}$, their union $A = \bigcup_{n=1}^{\infty} A_n$ belongs to $\mathcal{F}$.

Let Ω_1, Ω_2 be arbitrary sets and let G_1, G_2 be some systems of subsets of Ω_1 and Ω_2, respectively. A map $\xi : \Omega_1 \to \Omega_2$ is called *measurable* (or more precisely $((\Omega_1, G_1), (\Omega_2, G_2))$-measurable) if, for any set $A \in G_2$, the set $\xi^{-1}(A) \in G_1$. We shall use the notation $\xi : (\Omega_1, G_1) \to (\Omega_2, G_2)$ to indicate the dependence on G_1, G_2. Typically we shall consider measurability of maps in the case in that G_j, $j = 1, 2$, are algebras or σ-algebras.

Let A be a set. A *characteristic function* I_A of the set A is defined as $I_A(x) = 1$, $x \in A$, and $I_A(x) = 0$, $x \in \bar{A}$.

Let $A = \{a_1, \dots, a_n\}$ be a finite set. We shall denote the *cardinality* n of A by the symbol $|A|$.

2 Classical and ensemble definitions of probability

2.1 Classical definition of probability

The theory of probability originated from the study of problems connected with ordinary games of chance. In all these games the results that are *a priori* possible may be arranged in a finite number of cases assumed to be perfectly symmetrical, such as the cases represented by the six sides of a dice, the 52 cards in an ordinary pack of cards, and so on. This fact seemed to provide a basis for a rational explanation of the observed stability of statistical frequencies, and the 18th century mathematicians were thus led to the introduction of the famous *principle of equally possible cases*. According to this principle, a division into equally possible cases is possible in all random experiments, and the probability of an event is defined as the ratio between the number of cases favorable to the event, and the total number of possible cases. The main disadvantage of this probability theory is that the idea of symmetry cannot be applied to all random phenomena. For example, the classical definition of probability describes only a symmetric coin or dice. This definition cannot be used in the case of a violation of symmetry (see von Mises [88] for an extended critique of the classical definition). Denote by C the set of all possible cases. The classical theory operated on finite sets $C = \{c_1, \ldots, c_N\}$. For example, if a dice is considered, then $C = \{1, \ldots, 6\}$. Let E belong to the algebra F_C of all subsets of the set C. Then classical probability is defined by the equality

$$\mathbf{P}(E) = |E|/|C|. \tag{2.1}$$

The map $\mathbf{P} : F_C \to T_C \subset \mathbb{R}_+$, where $T_C = \{x = k/N : k = 0, 1, \ldots, N\}, N = |C|$, is a measure and $\mathbf{P}(C) = 1$. This measure is *uniform:* $\mathbf{P}(\{c_j\}) = 1/N$ and $\mathbf{P}(E) = \frac{1}{N} \sum_{c_j \in E} 1$.

We could not use (2.1) for infinite sets C in the framework of real analysis (there are no actual infinities in $\mathbb{R}$). This problem seems to be solved on the basis of the Kolmogorov measure-theoretic approach. But the classical definition (2.1) is not preserved in that approach. There are other possibilities to extend the classical definition of probability to infinite sets C. In principle we need not identify the set T_C of values of the classical probability with a subset of the set $\mathbb{R}$ of real numbers. It can be considered as just a subset of the set $\mathbb{Q}$ of rational numbers. It would be possible to extend the classical definition of probability by identifying T_C with a subset of other number system X such that $\mathbb{Q} \subset X$, see Chapter 4.

2.2 Ensemble (proportional) definition of probability

We start with the following classical example. There is an urn which contains balls of two colours, black and white. Let N_b and N_w be respectively the numbers of black and white balls; $N = N_b + N_w$ is the total number of balls in the urn. By definition a probability is the coefficient of the proportion between the number of balls of the concrete colour and the total number of balls: $\mathbf{P}(b) = \frac{N_b}{N}$ and $\mathbf{P}(w) = \frac{N_w}{N}$. In the general case we have a finite set S (an ensemble). Elements s of S have some properties. Denote the set of these properties by π_S. Each property $\xi \in \pi_S$ can be described as a map $\xi : S \to K_\xi$, where $K_\xi = \{1, 2, \ldots, k_\xi\}$ is a finite set (a numerical cod of the property ξ). We set $S(\xi = j) = \{s \in S : \xi(s) = j\}$; denote by $F(\pi_S)$ the collection of all these sets. By definition these are *events* and their probability is defined by

$$\mathbf{P}(S(\xi = j)) = \frac{|S(\xi = j)|}{|S|}. \tag{2.2}$$

If we assume that $F(\pi_S)$ is an algebra of sets then the map $\mathbf{P} : F(\pi_S) \to T_S \subset \mathbb{R}_+$, where $T_S = \{x = k/N : k = 0, 1, \dots, N\}$ and $N = |S|$, is a measure and $\mathbf{P}(S) = 1$. If all one point sets s belong to the algebra $F(\pi_S)$, then $F(\pi_S)$ is the algebra of all subsets of S (i.e., $F(\pi_S) = F_S$) and $\mathbf{P}$ is the uniform distribution: $\mathbf{P}(\{s\}) = 1/N$. In this case we can connect the ensemble (proportional) definition with the classical definition: *the elements of the ensemble S can be interpreted as equally possible cases.*

The conditional probabilities will play an essential role in further quantum considerations. Now we demonstrate how these probabilities are introduced in the ensemble approach. Let $B = S(\xi = l)$, $A = S(\eta = k)$, $\xi, \eta \in \pi_S$. Let the set $C = A \cap B \in F(\pi_S)$. This means that there exists a property $\theta \in \pi_S$ such that $C = S(\theta = m)$. Conditional probability of the event B under the condition A is defined as

$$\mathbf{P}_S(B/A) \equiv \mathbf{P}_A(B) = |B \cap A|/|A|$$

(we must extract from the ensemble S the sub-ensemble A and find the proportion of elements $s \in A$ which has the property $\xi(s) = l$). Thus we can easily obtain that

$$\mathbf{P}_S(B/A) = \mathbf{P}_S(B \cap A)/\mathbf{P}_S(A), \ \mathbf{P}_S(A) > 0. \tag{2.3}$$

This is the well-known Bayes' formula. We note that in the ensemble framework it is a theorem. In standard textbooks the ensemble index is omitted:

$$\mathbf{P}(B/A) = \mathbf{P}(B \cap A)/\mathbf{P}(A), \quad \mathbf{P}(A) > 0. \tag{2.4}$$

Remark 2.1. If $F(\pi_S)$ is not an algebra, then $A, B \in F(\pi_S)$ need not imply that $C = A \cap B \in F(\pi_S)$. In this case we could not use Bayes' formula (2.4). Moreover, in such a case it is insensible to speak about conditional probabilities. There is no property θ of elements s of S such that $C = S(\theta = m)$. Thus the set $C = \{s \in S : \xi(s) = l\} \cap \{s \in S : \eta = k\}$ cannot be described by properties of S. From the physical viewpoint it means that we could not verify two properties ξ and η simultaneously. If we try to extract the sub-ensemble A from S by verifying the property η, then we change the property ξ of $s \in S$.

As a simple consequence of (2.4) we obtain another important formula:

$$\mathbf{P}(A \cap B) = \mathbf{P}(B/A)\mathbf{P}(A). \tag{2.5}$$

By symmetry we find

$$\mathbf{P}(A \cap B) = \mathbf{P}(A/B)\mathbf{P}(B). \tag{2.6}$$

Thus we have:

$$\mathbf{P}(A/B) = \frac{\mathbf{P}(B/A)\mathbf{P}(A)}{\mathbf{P}(B)}. \tag{2.7}$$

To be more careful, we have to indicate the dependence of probabilities on corresponding ensembles: $\mathbf{P}_B(A) = \frac{\mathbf{P}_A(B)\mathbf{P}_S(A)}{\mathbf{P}_S(B)}$.

In further quantum considerations we shall often use the following consequence of Bayes' formula. Let $A_k \in F(\pi_S)$, $k = 1, \ldots, m$, $\bigcup_{k=1}^m A_k = S$ and $A_k \cap A_l = \emptyset$, $k \neq l$. Then, for every $C \in F(\pi_S)$ such that $C \cap A_k \in F(\pi_S)$, we have:

$$\mathbf{P}_S(C) = \sum_{k=1}^{m} \mathbf{P}_S(A_k)\mathbf{P}_{A_k}(C).$$

It is the well-known formula of *total probability.* In standard textbooks this formula is written as

$$\mathbf{P}(C) = \sum_{k=1}^{m} \mathbf{P}(A_k)\mathbf{P}(C/A_k).$$

Thus concrete ensembles which are used to define left and right hand sides probabilities are not taken into account. We shall see that in quantum formalism this manipulation with the ensemble index will imply such unexpected consequences as *non-locality of space-time and super-luminal signals and death of reality.*

The direct generalization of proportional formula (2.2) for ensemble probabilities to infinite ensembles S is impossible in the framework of real analysis, because there are no actual infinities (infinitely large numbers) in the field of real numbers $\mathbb{R}$. A measure-theoretical approach (see Section 4) provides some indirect generalization. However, this measure-theoretical approach is not the unique possibility to extend the proportional definition of probability to infinite ensembles. In Chapter 4 we shall consider ensembles which have structures of trees with an infinite number of vertexes (with p branches leaving each vertex; there $p > 1$ is a prime number). For such ensembles we can directly use (2.2) to define ensemble probabilities (there $N = |S|$ can be an infinite large number belonging to the field of so called p-adic numbers). Another possibility for extending (2.2) to infinite ensembles S is to use nonstandard analysis (see [10]).

3 Frequency theory of probability

This theory was the first where the principle of the stabilization of statistical frequencies was realized on a mathematical level. In fact, this principle was used as the definition of probability. Let us recall the main notions of a frequency theory of probability [169–171] of Richard von Mises (1919).[1] This theory is based on the notion of a collective. Consider a random experiment $\mathcal{S}$ and denote by $L = \{s_1, \ldots, s_m\}$ the set of all possible results of this experiment. The set L is said to be the label set, or the set of attributes. We consider only finite sets L. Let us consider N realizations of $\mathcal{S}$ and

[1] In fact, already in 1866 John Venn, see [165], tried to define a probability explicitly in terms of relative frequencies.

write a result x_j after each realization. Then we obtain the finite sample:

$$x = (x_1, \dots, x_N), \quad x_j \in L. \tag{3.1}$$

A *collective* is an infinite idealization of this finite sample:

$$x = (x_1, \dots, x_N, \dots), \quad x_j \in L, \tag{3.2}$$

for which the following two von Mises' principles are valid.

The first is the *statistical stabilization of relative frequencies* of each attribute $\alpha \in S$ in the sequence (3.2). Let us compute frequencies $\nu_N(\alpha; x) = n_N(\alpha; x)/N$ where $n_N(\alpha; x)$ is the number of realizations of the attribute α in the first N tests. The principle of the statistical stabilization of relative frequencies says: *the frequency $\nu_N(\alpha; x)$ approaches a limit as N approaches infinity for every label $\alpha \in L$.* This limit $\mathbf{P}(\alpha) = \lim \nu_N(\alpha; x)$ is said to be the probability of the label α in the frequency theory of probability. Sometimes this probability will be denoted by $\mathbf{P}_x(\alpha)$ (to show a dependence on the collective x).

"We will say that a collective is a mass phenomenon or a repetitive event, or simply a long sequence of observations for which there are sufficient reasons to believe that the relative frequency of the observed attribute would tend to a fixed limit if the observations were infinitely continued. This limit will be called the probability of the attribute considered within the given collective" [170].

The second principle is the so-called principle of *randomness.* Heuristically it is evident that we cannot consider, for example, the sequence $z = (0, 1, 0, 1, \dots, 0, 1, \dots)$ as a random object (generated by a statistical experiment). However, the principle of the statistical stabilization holds for z and $\mathbf{P}(0) = \mathbf{P}(1) = 1/2$. Thus, we need an additional restriction for sequences (3.2). This condition was proposed by von Mises: *The limits of relative frequencies have to be stable with respect to a place selection (a choice of a subsequence) in (3.2).*

In particular, z does not satisfy this principle. For example, if we choose only even places, then we obtain the zero sequence $z_0 = (0, 0, \dots)$ where $\mathbf{P}(0) = 1, \mathbf{P}(1) = 0$.

However, this very natural notion was the hidden bomb in the foundations of von Mises' theory. The main problem was to define a class of place selections which induces a fruitful theory. The main and very natural restriction is that a place selection in (3.2) cannot be based on the use of attributes of elements. For example, we cannot consider a subsequence of (3.2) constructed by choosing elements with the fixed label $\alpha_k \in L$. Von Mises proposed the following definition of a place selection:

(PS) "a subsequence has been derived by a place selection if the decision to retain or reject the nth element of the original sequence depends on the number n and on label values $x_1, \dots, x_{n-1}$ of the $(n-1)$ preceding elements, and not on the label value of the nth element or any following element",

see [170], p. 9. Thus a place selection can be defined by a set of functions f_1, $f_2(x_1)$, $f_3(x_1, x_2)$, $f_4(x_1, x_2, x_3)$, $\dots$, each function yielding the values 0 (rejecting the nth element) or 1 (retaining the nth element).

Here are some examples of place selections: (1) choose those x_n for which n is prime; (2) choose those x_n which follow the word 01; (3) toss a (different) coin; choose x_n if the nth toss yields heads. The first two selection procedures may be called *lawlike*, the third random. It is more or less obvious that all of these procedures are place selections: the value of x_n is not used in determining whether to choose x_n.

The principle of randomness ensures that no strategy using a place selection rule can select a subsequence that allows different odds for gambling than a sequence that is selected by flipping a fair coin. This principle can be called the *law of excluded gambling strategy*.

The definition (PS) induced some mathematical problems. If a class of place selections is too extended then the notion of the collective is too restricted (in fact, there are no sequences where probabilities are invariant with respect to all place selections). This was the main point of criticism of von Mises' theory. This problem has been investigated since the 1930s and solved only in the 1970s on the basis of Kolmogorov's notion of algorithmic complexity [135].

However, von Mises himself was satisfied by the following operational solution of this problem. He proposed [171] to fix for any collective a class of place selections which depends on the physical problem described by this collective. Thus he removed this problem outside the mathematical framework.

The frequency theory of probability is not, in fact, the calculus of probabilities, but it is the calculus of collectives which generates the corresponding calculus of probabilities. We briefly discuss some of the basic operations for collectives (see [171] for the details).

As probability is defined on the basis of the principle of the statistical stabilization of relative frequencies, it is possible to develop quite fruitful probabilistic calculus by using only this principle. Sequence (3.2) which satisfies the principle of the statistical stabilization of relative frequencies is said to be a *S-sequence*. Thus limits of relative frequencies in a S-sequence x need not be invariant with respect to some class of place selections[2].

(a) *Mixing and additivity.* Let x be a collective with the (finite) label space L_x and let $E = \{\alpha_{i_1}, \dots, \alpha_{i_l}\}$ be a subset of L_x. The sequence (3.2) of x is transformed into a new sequence y_E by the following rule. If $x_j \in E$ then we write 1; if $x_j \notin E$ then we write 0. Thus the label set $L_{y_E} = \{0, 1\}$. It is easy to show that this sequence has the property of statistical stabilization for its labels. For example,

$$\mathbf{P}_{y_E}(1) = \lim \nu_N(E; x) = \lim \sum_{k=1}^{l} \nu_N(\alpha_{i_k}; x) = \sum_{k=1}^{l} \mathbf{P}_x(\alpha_{i_k}), \tag{3.3}$$

[2] Of course, the use of S-sequences contradicts to the philosophy of the modern probability theory which is based on generalizations of Mises' principle of randomness (such as Kolmogorov complexity [135] and Martin-Löf [142] theory of statistical tests). However, it seems that all this machinery of randomness is not used in quantum physics. Experimentalists are only interested in the statistical stabilization of relative frequencies.

where $\nu_N(E; x) \equiv \nu_N(1; y_E) = n_N(1; y_E)/N$ is the relative frequency of 1 in y_E. To obtain (3.3) we have only used the fact that the addition is a continuous operation on the field of real numbers $\mathbb{R}$. We can show that the sequence y_E also satisfies the principle of randomness, see [171]. Hence this is a new collective. By this operation any collective x generates a probability distribution on the algebra F_{L_x} of all subsets of L_x : $\mathbf{P}(E) = \mathbf{P}_{y_E}(1)$. Sometimes it will be convenient also to denote this probability distribution by $\mathbf{P}_x(E)$ to distinguish probabilities corresponding to different collectives. Now we find the properties of this probability. As $\mathbf{P}(E) = \lim \nu_N(E; x)$ and $0 \le \nu_N(E) \le 1$, then (by the elementary theorem of real analysis) $0 \le \mathbf{P}(E) \le 1$. Hence the probability must yield values in the segment [0, 1]. Further, as the collective y_{L_x} corresponding to the whole label set L_x does not contain zeros, we obtain that $\nu_N(L_x; x) \equiv \nu_N(1; y_{L_x}) \equiv 1$ and, consequently, $\mathbf{P}(L_x) = 1$. Finally by (3.3) we find that the set function $\mathbf{P} : F_{L_x} \to [0, 1]$ is additive. Thus $\mathbf{P}$ is a normalized measure on the algebra F_{L_x} which yields values in [0, 1]. We remark that all these considerations can be repeated for S-sequences.

(b) *Partition and conditional probabilities.* Let x be a collective and let $A \in F_{L_x}$ and $\mathbf{P}(A) \neq 0$. We derive a new sequence $z(A)$ by retaining only those elements of x which belong to A and discarding all other elements (thus the label set $L_{z(A)} = A$). This operation is obviously not a place selection, since the decision to retain or reject an element of x depends on the label of just this element. The sequence $z(A)$ is again a collective, see [171]. Suppose that $\alpha_j \in A$ and let y_A be the collective generated by x with the aid of the mixing operation. Then $\mathbf{P}_{z(A)}(\alpha_j) = \lim_{N\to\infty} \nu_N(\alpha_j; z(A)) = \lim_{k\to\infty} \nu_{N_k}(\alpha_j; z(A))$, where $N_k \to \infty$ is an arbitrary sequence. As $\mathbf{P}(A) \neq 0$ then $M_k = n_k(1; y_A) \to \infty$ (this is the number of labels belonging to A among the first k elements of x). Thu

$$\begin{aligned}
\mathbf{P}_{z(A)}(\alpha_j) &= \lim_{k\to\infty} \nu_{M_k}(\alpha_j; z(A)) = \lim_{k\to\infty} n_{M_k}(\alpha_j; z(A))/M_k \\
&= \lim_{k\to\infty} [n_{M_k}(\alpha_j; z(A))/k] : [M_k/k] \\
&= \mathbf{P}_x(\alpha_j)/\mathbf{P}_x(A).
\end{aligned}$$

We have used the property that $n_{M_k}(\alpha_j; z(A))$, the number of α_j among first M_k elements of $z(A)$, is equal to $n_k(\alpha_j; x)$, the number of α_j among first k elements of x. The probability $\mathbf{P}_{z(A)}(\alpha_j)$ is the conditional probability of the label α_j if we know that a label belongs to A. It is denoted by $\mathbf{P}(\alpha_j/A) \equiv \mathbf{P}_x(\alpha_j/A)$. As a consequence of this formula we obtain Bayes' formula:

$$\begin{aligned}
\mathbf{P}_{z(A)}(B) &= \sum_{\alpha_j \in B\cap A} \mathbf{P}_x(\alpha_j/A) \\
&= \sum_{\alpha_j \in B\cap A} \mathbf{P}_x(\alpha_j)/\mathbf{P}_x(A) = \mathbf{P}_x(B \cap A)/\mathbf{P}_x(A). \qquad (3.4)
\end{aligned}$$

In fact, this formula connects probabilities defined with respect to different collectives. The left hand side probability is $\mathbf{P}_{z(A)}$ and the right hand side probabilities are $\mathbf{P}_x$. As in the case of the ensemble probability, sometimes we shall use the symbol $\mathbf{P}_A(B)$ instead of $\mathbf{P}(B/A)$. It useful to remark that $\mathbf{P}_A : F_{L_x} \to [0, 1]$ is a measure normalized by 1. In particular, the probability $\mathbf{P}$ may be written as the conditional probability $\mathbf{P}_{L_x}$.

As in the ensemble framework, here we can also obtain the formula of total probability (2.8). Formula (2.8) is often applied in the wrong way: probabilities $\mathbf{P}(A_k)$ are found with respect to one collective and conditional probabilities $\mathbf{P}(C/A_k)$ with respect to other collective. To apply this formula in the right way we have to use the index of a collective:

$$\mathbf{P}_x(C) = \sum_{k=1}^{m} \mathbf{P}_x(A_k)\mathbf{P}_x(C/A_k). \tag{3.5}$$

Formulas (2.5)–(2.7) can be also easily obtained in the frequency framework.

Remark 3.1. The Bayes formula in the frequency framework is a consequence of the possibility of using the operation of partition for collectives. It should be noticed that from the physical point of view the operation of partition is a physical condition, which means that by extracting the collective $z(A)$ from the original collective x we do not change the property of belonging to B or not. If the physical system does not satisfy this condition, we cannot use the Bayes formula (3.4). This does not mean that we cannot define the conditional probability $\mathbf{P}_A(B)$. But we cannot use (3.4) to find this probability.

It is important to remark that the conditional probabilities in (2.7) are defined with respect to different collectives, $z(A)$ and $z(B)$. From the physical point of view the connection (2.7) between these probabilities is possible only for physical systems which satisfy conditions discussed in Remark 3.1.

It is evident that we can also consider countable sets of attributes $L_x = \{\alpha_1, \alpha_2, \dots, \alpha_m, \dots\}$. If we use the additional condition $\sum_{j=1}^{\infty} \mathbf{P}(\alpha_j) < \infty$ for the probabilities of labels then $\mathbf{P}$ is a (discrete) measure on F_{L_x}. Moreover, this measure is σ-additive. However, the generalization of the frequency theory of probability to 'continuous' sets of attributes is a nontrivial mathematical problem, see [171], [162].

4 Kolmogorov's measure-theoretical theory

The axiomatics of the modern probability theory was proposed by Andrei Nikolaevich Kolmogorov [133] in 1933 to provide a reasonable mathematical description of this theory. The basis of Kolmogorov axiomatics was prepared at the beginning of this century in France by investigations of Borel [33, 34] and Fréchet [63] on the measure-theoretic approach to probability. At the same time Kolmogorov used ideas of von Mises [169] about the frequency definition of probability (see remarks in [133]).

By the Kolmogorov axiomatics the *probability space* is defined as the triple $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$, where Ω is an arbitrary set (points ω of Ω are said to be *elementary events*), $\mathcal{F}$ is an arbitrary σ-algebra of subsets of Ω (elements of $\mathcal{F}$ are said to be *events*), $\mathbf{P}$ is a σ-additive measure on $\mathcal{F}$ which yields values in the segment [0, 1] of the real line and normalized by the condition $\mathbf{P}(\Omega) = 1$.

Random variables on $\mathcal{P}$ are defined as measurable functions $\xi : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra on the real line[3]. We shall use the symbol $\mathrm{RV}(\mathcal{P})$ to denote the space of random variables over $\mathcal{P}$. Probability distribution of $\xi \in \mathrm{RV}(\mathcal{P})$ is defined as $\mathbf{P}_\xi(B) = \mathbf{P}(\xi^{-1}(B))$ for $B \in \mathcal{B}$. This is a σ-additive measure on the Borel σ-algebra.

A. N. Kolmogorov motivated additivity of probability by additivity of frequency probability (see formula (3.3)); he also used frequency reasons to take the segment [0, 1] as the range of values of a probabilistic measure. On the other hand, the condition of σ-additivity was considered by Kolmogorov as an additional mathematical (technical) condition to provide a fruitful integration theory based on the Lebesgue integral. In fact, Kolmogorov started with finite additive probabilities defined on algebras of sets. The spaces with σ-additive probabilities defined on σ-algebras were called generalized probability space.

The Kolmogorov theory also contains the additional axiomatic definition of conditional probabilities. By definition $\mathbf{P}(B/A)$ is defined by formula (2.4). Kolmogorov did not give any motivation for this definition in his book [133]. However, as he gave a clear motivation of all other properties of $\mathbf{P}$ on the basis of the von Mises frequency theory, it seems to be that he used the same frequency reasons for (2.4). In Kolmogorov's model two events A and B are said to be independent if

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) P(B) \tag{4.1}$$

or

$$\mathbf{P}(B/A) = P(B), \quad \mathbf{P}(A) > 0. \tag{4.2}$$

In the standard framework of Lebesgue integration we start with a σ-additive measure μ defined on some algebra F and then μ is extended over the σ-algebra $\mathcal{F}$ generated by F (Borel σ-algebra). This extension procedure, which is well defined from the mathematical point of view, is not so innocent from the probabilistic point of view. Kolmogorov remarked: "Even if the sets (events) A of F can be interpreted as actual and (perhaps only approximately) observed events, it does not, of course, follow from this that the sets of $\mathcal{F}$ reasonably admit of such an interpretation. Thus there is the possibility that while a field of probability $(F, \mathbf{P})$ may be regarded as the image (idealized, however) of actual random events, the extended field of probability $(\mathcal{F}, \mathbf{P})$ will still remain merely a mathematical structure. Thus sets of $\mathcal{F}$ are merely ideal events to which nothing corresponds in the outside world. However, if reasoning which utilizes probabilities of such ideal events leads us to a determination of the probability

[3] Thus $\xi^{-1}(B) \in \mathcal{F}$ for every $B \in \mathcal{B}$.

of an actual event of F, then, from an empirical point of view also, this determination will automatically fail to be contradictory”, see [133], p. 17. It should be noticed that the adherents of Kolmogorov’s measure-theoretical approach to probability theory did not pay large attention to these ideas of Kolmogorov. This implied that manipulations with abstract probabilities belonging to $\mathcal{F}$ were considered as real probabilistic investigations. Moreover, if we need not pay attention to the difference between real and abstract probabilities, we could in principle omit the concrete probabilistic model from our considerations and operate with ‘events’ belonging to abstract σ-algebras. This is the main problem of worldwide use of Kolmogorov’s measure-theoretical approach.

Remark 4.1. For example, Cramer, who used the Kolmogorov axiomatics to create the mathematical theory of statistics, had another point of view on the problem of verification: “any probability assigned to a specific event must, in principle, be liable to verification” [44]. The question of verification was the cornerstone of the von Mises theory for the continuous label set S. He showed that in the case $L_x = \mathbb{R}$ (or $\mathbb{R}^n$) a probability measure of an event E has the frequency interpretation iff the measure of the boundary of E is equal to 0, [171].

On the other hand, Kolmogorov himself developed actively the viewpoint that probability theory is a purely mathematical theory. Therefore the concrete structure of set algebra (or σ-algebra) does not play any role in probabilistic considerations. In his manifest “General Measure Theory and Probability Calculus”, 1929 (see [159]), he wrote: “To outline the context of theory, it suffices to single out from probability theory those elements that bring out its intrinsic logical structure, but have nothing to do with the specific meaning of theory.”

Finally we remark that in Kolmogorov’s approach Bayes’ formula (2.4) is just the definition of a conditional probability. I like to underline this fact. I have the experience that many scientists working in applications of probability are sure that Bayes’ formula is a theorem. But this is right only for ensemble and frequency approaches. On the other hand, the formula of total probability (2.8) is a theorem of the Kolmogorov’s theory. Here it holds true for a countable family of sets $A_k \in \mathcal{F}$, $\mathbf{P}(A_k) > 0$, $k = 1, \dots,$ such that $\bigcup_{k=1}^{\infty} A_k = \Omega$ and $A_k \cap A_l = \varnothing$, $k \neq l$: for every $C \in \mathcal{F}$, $\mathbf{P}(C) = \sum_{k=1}^{\infty} \mathbf{P}(A_k)\mathbf{P}(C/A_k)$. To obtain this formula, we need to use the σ-additivity of probability and the definition (Bayes’ formula) of conditional probabilities.

5 Kolmogorov’s ideas on probability

It should be noticed that before to create the system of axioms of probability theory, A. N. Kolmogorov discussed ([159], 1929) some examples of ‘generalized probabilities’ which could not be described by his axiomatics. Moreover, probably we need not call these objects ‘generalized probabilities’. It seems more natural to call ordinary probabilities (described by Kolmogorov’s axiomatics) ‘restricted probabilities’.

There is other side of the common use of Kolmogorov's approach which is not so visible as the disappearance of concrete probabilistic spaces. This is the idea that only Lebesgue measurable sets could play some role in probabilistic considerations. Of course, this is a consequence of the fact that Kolmogorov discussed merely the Lebesgue extension [159] (or the Borel extension [133]). However, in principle some sets which are not Lebesgue measurable may appear in probabilistic models connected with some natural phenomena. We shall discuss such a model in Chapter 2. On the other hand, Kolmogorov discussed in [159] non-Lebesgue extension of the linear Lebesgue measure μ on the segment [0, 1], namely, the result of Banach that μ can be extended to a measure $\bar{\mu}$ defined on the (σ-)algebra $F_{[0,1]}$ all subsets of [0, 1]. It seems to be that Kolmogorov considered this measure as a good candidate to be probability. He also considered multidimensional case and pointed out that an extension $\bar{\mu}$ on $F_{[0,1]^n}$ of the Lebesgue measure μ on $[0,1]^n$ can be obtained by using the metric equivalence of a cube $[0,1]^n$, $n = 2, 3, \ldots$, and the segment [0, 1]. Then he mentioned that in the case $n > 2$ such a measure does not satisfy the principle of equality of the measure of congruent sets. This is a consequence of example on the decomposition of a sphere into three sets being congruent to the sum of two others to within a countable set (see, for example, [73] for the proof):

Theorem 5.1. *A sphere S can be decomposed into disjoint sets $S = A \cup B \cup C \cup Q$ such that: (i) the sets A, B, C are congruent to each other; (ii) the set $B \cup C$ is congruent to each of the sets A, B, C; (iii) Q is countable.*

We continue to study the question on a domain of definition of probability. As we have seen, the ensemble approach does not imply automatically that the system of sets (events) $F(\pi_S)$ (corresponding to properties π_S of the ensemble S) must be an algebra. On the other hand, if Kolmogorov's axiomatics is used, then we have to start with (at least) an algebra. However, there may be random phenomena which do not possess the structure of an algebra. Why the union $A \cup B$ of two events A, B must always be an event? Why the complement $D = \Omega \setminus C$ of an event C must always be an event? It is interesting that, before to propose the general axiomatics of probability theory [133] (1933), Kolmogorov discussed the problem of a domain of definition of probability [159] (1929). At that time he had the viewpoint which coincided with our viewpoint: "It is also doubtful if a measure connected with some problem in probability calculus need be closed" (i.e., defined on an algebra). In [159] Kolmogorov pointed out that "one should not assume, however, that the existence of measures of two intersecting sets implies the existence of measure for their sum or difference: there are certain important measures without this property." In particular, he discussed the following example.

Example 5.1 (density of natural numbers; see, for example, [69], [147] for the details). For a subset $A \subset \mathbb{N}$ the quantity

$$\delta(A) = \lim_{N\to\infty} \frac{|A \cap \{1, \ldots, N\}|}{N},$$

is called the *density* of A if the limit exists. Let $\mathcal{G}_d$ denote the collection of all subsets of $\mathbb{N}$ which admit density. It is evident that each finite $A \subset \mathbb{N}$ belongs to $\mathcal{G}_d$ and $\delta(A) = 0$. It is also evident that each subset $B = \mathbb{N} \setminus A$, where A is finite, belongs to $\mathcal{G}_d$ and $\delta(B) = 1$ (in particular, $\mathbf{P}(\mathbb{N}) = 1$). The reader can easily find examples of sets $A \in \mathcal{G}_d$ such that $0 < \delta(A) < 1$.

Proposition 5.1. *Let $A_1, A_2 \in \mathcal{G}_d$ and $A_1 \cap A_2 = \varnothing$. Then $A_1 \cup A_2 \in \mathcal{G}_d$ and*

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2). \tag{5.1}$$

Proof. As $A_1 \cap A_2 = \varnothing$, then $|(A_1 \cup A_2) \cap \{1, \dots, N\}| = |A_1 \cap \{1, \dots, N\}| + |A_2 \cap \{1, \dots, N\}|$. □

Proposition 5.2. *Let $A_1, A_2 \in \mathcal{G}_d$. The following conditions are equivalent:*

(1) $A_1 \cup A_2 \in \mathcal{G}_d$; (2) $A_1 \cap A_2 \in \mathcal{G}_d$;

(3) $A_1 \setminus A_2 \in \mathcal{G}_d$; (4) $A_2 \setminus A_1 \in \mathcal{G}_d$.

There are standard formulas:

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 \cap A_2); \tag{5.2}$$

$$\mathbf{P}(A_1 \setminus A_2) = \mathbf{P}(A_1) - \mathbf{P}(A_1 \cap A_2). \tag{5.3}$$

Proof. We have

$$|(A_1 \cup A_2) \cap \{1, \dots, N\}| = |A_1 \cap \{1, \dots, N\}| + |A_2 \cap \{1, \dots, N\}| \\ - |(A_1 \cap A_2) \cap \{1, \dots, N\}|.$$

Therefore, if, for example, $A_1 \cap A_2 \in \mathcal{G}_d$ then there exists a limit of the right hand side. It implies $A_1 \cup A_2 \in \mathcal{G}_d$ and (5.2) holds. Other implications are proved in the same way. □

It is possible to find sets $A, B \in \mathcal{G}_d$ such that, for example, $A \cap B \notin \mathcal{G}_d$. Let A be the set of even numbers. Take any subset $C \subset A$ which has no density. In fact, you can find C such that

$$\frac{1}{N}|C \cap \{1, 2, \dots, N\}|$$

is oscillating. There happen two cases: $C \cap \{2n\} = \{2n\}$ or $= \varnothing$. Set

$$B = C \cup \{2n - 1 : C \cap \{2n\} = \varnothing\}$$

Then, both A and B have densities one half. But $A \cap B = C$ has no density. Thus $\mathcal{G}_d$ is not a set algebra.

In 1929 A. N. Kolmogorov wrote [159]: "It is not known whether every measure is closable. If closure is possible, then it is not necessarily closable in only one way. It would seem that it is very difficult to find a measure that closes the measure given by the density of natural numbers."

We can prove that the density of natural numbers can be closed (extended on the algebra $F_{\mathbb{N}}$ of all subsets of $\mathbb{N}$), see Theorem 5.4. To formalize our considerations on the density of natural numbers, we propose the following definition.

Definition 5.1. A system of subsets $\mathcal{G}$ of a set Ω, which has the properties described by Proposition 5.2 and contains $\varnothing$ and Ω, is called *semi-algebra*.

Definition 5.2. A function $\mathbf{P} : \mathcal{G} \to [0, 1]$, where $\mathcal{G}$ is semi-algebra, is said to be a *probability semi-measure* if it satisfies the additivity condition (5.1) and $\mathbf{P}(\Omega) = 1$.

Definition 5.3. The system $\mathcal{P} = (\Omega, \mathcal{G}, \mathbf{P})$, where $\mathbf{P}$ is a probability semi-measure on semi-algebra $\mathcal{G}$, is called a *semi-probability space*.

Unfortunately we could not say anything more about such a generalization of a probability space, because the theory of integration with respect to probability semi-measures is not well developed.

We present the simplest construction of an extension of a measure μ on the algebra of all subsets. This construction is based on a representation of μ by a continuous linear functional on some space of functions and the application of the Hahn–Banach theorem.

Theorem 5.2. *Let μ be a (finite additive) measure on an algebra F of subsets of Ω. Then there exists a finite-additive extension $\bar{\mu}$ of μ on the algebra F_{Ω} of all subsets of Ω.*

To prove this theorem, it is sufficient to apply the following well-known theorem of functional analysis (but we would like to escape such a functional analytic consideration):

Theorem 5.3 (Hahn–Banach). *Let E be a normed linear space and let U be a linear subspace. Every continuous linear functional $l : U \to \mathbb{R}$ can be extended to a continuous linear functional $L : E \to \mathbb{R}$ in such a way that norms of the functionals l and L coincide:*

$$\|L\| = \|l\|. \tag{5.4}$$

On the other hand, the $\bar{\mu}$ may be not σ-additive even if μ is σ-additive. It seems that an answer to the question

"Is it possible in the general case to construct a σ-additive extension $\bar{\mu}$ on the algebra F_{Ω} of a σ-additive measure μ?"

is unknown.

Another difficulty is that the proof of the Hahn–Banach theorem is based on the *axiom of choice*. Therefore we also have to use this axiom to obtain an extension of probability. However, the place of the axiom of choice in quantum physics is not clear. Thus it is not easy to find the range of possible applications of probabilities extended on F_Ω with the aid of the Hahn–Banach theorem.

It seems that in general case it is impossible to obtain the existence of an extension $\bar{\mu}$ of μ without the axiom of choice.

However, the main problem is *non-uniqueness* of an extension $\bar{\mu}$. By our construction $\bar{\mu}$ is determined by an extension L_μ of the functional l_μ. In general such an extension is not unique.

Corollary 5.1. *Let* $\mathbf{P}$ *be a probabilistic measure on an algebra* F *of subsets of* Ω. *Then there exists a finite-additive extension* $\bar{\mathbf{P}}$ *of* $\mathbf{P}$ *on the algebra* F_Ω *of all subsets of* Ω.

In some physical models we may use 'probabilities' defined on the algebra F_Ω of all subsets of Ω which are obtained via the Hahn–Banach theorem. As it has been noticed, in general these probabilities are not σ-additive. However, finite-additivity is merely a mathematical problem. The real problem is non-uniqueness of an extension $\bar{\mathbf{P}}$.

For instance, we start with a σ-additive probability $\mathbf{P}$ defined on a σ-algebra $\mathcal{F}$. Let us assume that some events $A \in F_\Omega \setminus \mathcal{F}$ have a physical meaning. Let $\bar{\mathbf{P}}_1$ and $\bar{\mathbf{P}}_2$ be different extensions of $\mathbf{P}$ to the algebra F_Ω. In principle, $\bar{\mathbf{P}}_1(A) \neq \bar{\mathbf{P}}_2(A)$. As mathematical arguments are not sufficient to fix a 'probability', we need to use some additional physical arguments to obtain the 'right extension'.

It seems that the situation with nonuniqueness is even more complicated. As in the above considerations, let us start with a σ-additive probability $\mathbf{P}$ defined on a σ-algebra $\mathcal{F}$. Let $\mathbf{P}_\mathrm{L}$ be the Lebesgue extension of $\mathbf{P}$ on the σ-algebra $\mathcal{F}_L$ of Lebesgue measurable sets. The $\mathbf{P}_\mathrm{L}$ is the unique σ-additive extension of $\mathbf{P}$ on $\mathcal{F}_L$. On the other hand, there may exist finite-additive extensions $\bar{\mathbf{P}}$ of $\mathbf{P}$ on $\mathcal{F}_L$ which do not coincide with $\mathbf{P}_\mathrm{L}$. As we have discussed many times, the condition of σ-additivity is a purely mathematical condition. Therefore from the physical viewpoint there are no reasons to choose only the σ-additive extension. Thus the standard choice, $\mathbf{P}_\mathrm{L}(A)$, of probability for events $A \in \mathcal{F}_L$ does not seem so natural from the physical viewpoint. We think that some paradoxes in quantum formalism are a consequence of the common opinion that only the Lebesgue extension $\mathbf{P}_L$ gives 'right physical probability'. In particular, the proof of famous Bell's inequality is based on such an assumption. Thus the Einstein–Podolsky–Rosen paradox (see Chapter 2) might be a consequence of the conventional (but probably non-physical) choice of an extension of probability.

We have discussed norm-preserving extensions of probability obtained via the Hahn–Banach theorem. In principle there may exist extensions of linear functionals which increase the norm. If we define an extension of a measure $\mu : F \to \mathbb{R}_+$ with the aid of such a functional extension, then we could not be sure that $\bar{\mu}$ is non-negative. In this way starting with probability $\mathbf{P} : F \to [0, 1]$ we may obtain generalized probabilities $\bar{\mathbf{P}} : F_\Omega \to \mathbb{R}$ with negative values as well as with values which extend one. We shall see in Chapter 3 that such generalized probabilities may have physical meaning. We note that if $\mathbf{P} : F \to [0, 1]$ is a σ-additive probability, then it may be that a (norm-increasing) extension $\bar{\mathbf{P}} : F_\Omega \to \mathbb{R}$ is also σ-additive. In such a case we obtain a signed probability measure (a charge), see Chapter 3.

We turn back to the density of natural numbers.

Theorem 5.4. *The density of natural numbers* $\delta : \mathcal{G}_d \to [0,1]$ *can be extended to a finite-additive measure* $\bar{\delta} : F_{\mathbb{N}} \to [0,1]$.

Proof. We apply again the Hahn–Banach theorem. Denote by LIM($\mathbb{N}$) the subspace of the normed space $B(\mathbb{N})$ (of all bounded functions $f : \mathbb{N} \to \mathbb{R}$) consisting of all functions f for which there exists the mean value:

$$l_\delta(f) = \lim_{n\to\infty} \frac{1}{n}(f(1) + \cdots + f(n)).$$

The map

$$l_\delta : \mathrm{LIM}(\mathbb{N}) \to \mathbb{R}$$

is a continuous linear functional and $l_\delta(I_A) = \delta(A)$ for each $A \in \mathcal{G}_d$. By the Hahn–Banach theorem l_δ can be extended to a continuous linear functional $L_\delta : B(\mathbb{N}) \to \mathbb{R}$ and $1 = \delta(\mathbb{N}) = \|l_\delta\| = \|L_\delta\|$. We set $\bar{\delta}(A) = L_\delta(I_A)$ for $A \in F_{\mathbb{N}}$. The linearity of L_δ implies that $\bar{\delta} : F_{\mathbb{N}} \to \mathbb{R}$ is additive. One can show that $\bar{\delta}$ is non-negative. □

If $A \in F_{\mathbb{N}} \setminus \mathcal{G}_d$, then

$$\bar{\delta}(A) \neq \lim_{n\to\infty} \frac{|A \cap \{1, 2, \ldots, n\}|}{n}.$$

Thus the frequency verification of the event A is impossible (the principle of the statistical stabilization is violated; compare with Chapters 2 and 4).

An extension of δ from semi-algebra $\mathcal{G}_d$ on the σ-algebra $F_{\mathbb{N}}$ given by Theorem 5.3 is not unique. If in some physical model some sets $A \in F_{\mathbb{N}} \setminus \mathcal{G}_d$ are considered as physical events, then there must be special physical reasons to choose one or another extension of δ. And we have to remember that the principle of the statistical stabilization is violated for events $A \in F_{\mathbb{N}} \setminus \mathcal{G}_d$. As in the case of measures defined on algebras, there may exist extensions L_δ of l_δ which do not preserve the norm: $\|L_\delta\| > 1$. In such a case an extension $\bar{\delta}$ corresponding to L_δ can take negative values.

In principle we might consider Theorem 5.4 as the answer to the question of A. N. Kolmogorov on a possibility to close the density of natural numbers δ. However, Kolmogorov wanted **to find** a measure which closes δ. Theorem 5.4 is not constructive (it is based on the axiom of choice). Really it does not give the answer to Kolmogorov's question. The construction which has been used in Theorem 5.4 could not be applied to an arbitrary semi-measure μ. Thus we do not know the answer to the question: "Is it possible to close an arbitrary semi-measure?" (compare with Kolmogorov, 1929).

6 Measure-theoretical approach and interpretations of probability

Now we are going to discuss possible probability interpretations of the Kolmogorov measure-theoretical approach (the mathematical theory of a special class of measures). As we have seen, the probabilistic measures can be associated with all probability models (classical, ensemble, and frequency). Therefore it is in principle possible to use the measure-theoretical formalism and classical, ensemble or frequency interpretation. However, A. N. Kolmogorov proposed not only a mathematical formalism but also an interpretation of this formalism. We shall start with this interpretation.

6.1 Ensemble-frequency interpretation

Kolmogorov interpreted a probability in the following way: "[...] we may assume that to an event A which may or may not occur under conditions Σ is assigned a real number $\mathbf{P}(A)$ which has the following characteristics: (a) one can be practically certain that if the complex of conditions Σ is repeated a large number of times, N, then if n be the number of occurrences of event A, the ratio n/N will differ very slightly from $\mathbf{P}(A)$; (b) if $\mathbf{P}(A)$ is very small, one can be practically certain that when conditions Σ are realized only once the event A would not occur at all". This interpretation is a mixture of the frequency and ensemble interpretations. In fact, (a) is the frequency interpretation and (b) is the ensemble interpretation. However, we cannot identify Kolmogorov's interpretation with any of these interpretations (for example, we may not assume (see [171], p. 5) that *each* infinite repetition of Σ will generate a collective). This mixture of interpretations generated some problems and played a negative role in applications of probability theory. Kolmogorov did not separate the proportion (measure) in ensemble and the frequency of realizations. Moreover, it seems to be that he often reduced the proportion in an ensemble to the proportion (2.1) for possible cases[4]. For example, he considered the experiment of tossing a coin twice and obtained a finite space of elementary events $\Omega = \{HH, HT, TH, TT\}$, where the labels H, T are used for the sides of a coin. I think that Kolmogorov understood very well the weakness of his interpretation. For this reason he considered this problem again 30 years later and proposed the theory of algorithmic complexity of random sequences [135]. However, the latter theory is nothing other than the attempt to justify the frequency probability theory of von Mises.

Remark 6.1. As the ensemble-frequency interpretation is based on both frequency and proportional arguments, the range of applications of Bayes' formula (2.4) is restricted by Remarks 2.1 and 3.1. In fact the Bayes formula is the additional postulate of the Kolmogorov axiomatics. In principle we can use the Kolmogorov theory (probability spaces) without Bayes' formula (2.4). This theory will describe the physical systems

[4]There Kolmogorov followed the historical tradition.

with a violation of (2.4). This framework was developed by Accardi [2]; we shall discuss it in the connection with quantum theory.

As we have seen, there are two (essentially different) contributions of Kolmogorov to probability theory. The first is the measure-theoretical approach and the second is the ensemble-frequency interpretation. The first is purely mathematical and the second is phemenological. Of course, it is possible to combine Kolmogorov's measure-theoretical formalism with other interpretations of probability. However, we have to pay attention to the problem that the use of a specific interpretation induces some restrictions to Kolmogorov's measure-theoretical approach. We start with the ensemble probability.

6.2 Measure-theoretical approach and the ensemble interpretation

Let S be an arbitrary (probably infinite) ensemble. Let $\mathcal{P}_S = (S, \mathcal{F}, \mathbf{P}_S)$ be Kolmogorov's probability space based on S. This space can be used for probabilistic analysis on S. However, we have to remember that the set π_S of properties can differ from the set of random variables $\mathrm{RV}(\mathcal{P}_S)$. There can exist random variables $\xi \in \mathrm{RV}(\mathcal{P}_S)$ (and, in particular, sets $A \in \mathcal{F}$) which are not properties of $s \in S$. As it was mentioned by Kolmogorov, analysis based on $\xi \in \mathrm{RV}(\mathcal{P}_S)$ can give some results for elements $\eta \in \pi_S$ which have no real physical meaning. On the other hand, there can exist properties $\eta \in \pi_S$ which are not random variables (these are non-measurable maps η on $\mathcal{P}_S$). Other important thing is that all probability distributions depend on the ensemble S.

Let us consider the following example. Let $\mathcal{P}_j = (\Omega_j, \mathcal{F}_j, \mathbf{P}_j)$, $j = 1, 2$, be Kolmogorov's probability spaces and let $\xi_j : \Omega_j \to \mathbb{R}$ be random variables with probability distributions $\mathbf{P}_{\xi_j}$. Then in Kolmogorov's formalism it is **always** possible to construct a probability space $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$ such that there are well defined random variables $\bar{\xi}_j \in \mathrm{RV}(\mathcal{P})$, $j = 1, 2$, such that $\mathbf{P}_{\bar{\xi}_j} = \mathbf{P}_{\xi_j}$. We can simply set $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$, $\mathbf{P} = \mathbf{P}_1 \otimes \mathbf{P}_2$ and $\bar{\xi}_j(\omega_1, \omega_2) = \xi_j(\omega_j)$. However, it does not sound reasonable that we can do the same thing in the ensemble framework. Let $\Omega_1 = \Omega_2 = S$ and $\xi_j \in \pi_S$, $j = 1, 2$. In general it does not sensible to use the ensemble $\Omega = S \times S$ for representing properties of the original ensemble S.

6.3 Measure-theoretical approach and frequency probability

The original viewpoint of R. von Mises was that Kolmogorov's probability measure is nothing other than the probability distribution $\mathbf{P}_x$ (on the label set L_x) of a collective x. The Kolmogorov probability space in Mises' theory is chosen as $\mathcal{P}_x = (L_x, \mathcal{F}_{L_x}, \mathbf{P}_x)$ where in general case $\mathcal{F}_{L_x}$ is some σ-algebra of subsets of L_x. As we have already pointed out, in the continuous case not all sets $A \in \mathcal{F}_{L_x}$ have the frequency meaning. In particular, if the measure-theoretical approach is used for the description of the frequency phenomena, then the possibility of the frequency verification for events $A \in$

$\mathcal{F}_{L_x}$ must be controlled. However, it is even more important to control continuously the dependence of a probability space on a collective.

Let us consider the following example (which is similar to the example considered in the ensemble framework). Let x^j, $j = 1, 2$, be two collectives with the label sets L_j and probability distributions $\mathbf{P}_{x^j}$. Let $\mathcal{P}_j = (L_j, \mathcal{F}_{L_j}, \mathbf{P}_{x^j})$, $j = 1, 2$, be the corresponding Kolmogorov's probability spaces. Let $A_j \in \mathcal{F}_{L_j}$. Suppose that somewhere we need to use conditional probability $\mathbf{P}(A_1/A_2)$. What is the meaning of Bayes' formula (2.4) in this case?

6.4 Measure-theoretical approach and ensemble-frequency interpretation

As we have already mentioned, typically Kolmogorov's measure-theoretical formalism on abstract probability spaces is used together with the ensemble-frequency interpretation of probability. However, as in the cases of the ensemble and frequency theories, we must be careful with applications of the abstract measure-theoretical formalism. We study the question of a choice of a probability space for the concrete probability experiment.

The part (a) of the ensemble-frequency interpretation of probability implies that the space Ω must describe occurrences of events in very long sequences of repetitions of some condition Σ (in the mathematical formalism sequences can have infinite length). It seems that collectives can be used for the description of such a phenomenon. However, the part (b) is related to occurrences of events under a single realization of conditions Σ. Probability of a single realization is nonsense for collectives. Let us try to solve the contradiction between probability in a long sequence of repetitions of Σ and a single realization of Σ. We may consider the space C of all possible collectives which can be induced by repetitions of Σ. Then we may introduce on C a probability measure $\mathbf{P}$ (which seems to have the meaning of an ensemble probability for the ensemble C) that would provide a mathematical description of the part (b). The latter would mean that if $\mathbf{P}(A)$ is very small, then a single realization of A (in one of collectives $x \in C$) is practically impossible (from the ensemble viewpoint). However, in the standard formalism the space C of all collectives is not used as a space of elementary events Ω.[5] Instead of C, there is used the space $\Omega = L^\infty$ of all infinite sequences of labels $\alpha \in L$. Such a choice gives measure-theoretical advantages. However, this implies the consideration of sequences which have no probabilistic meaning.

We construct now Kolmogorov's probability measure $\mathbf{P}^{\text{Kol}}$ on the space of sequences Ω which gives (as it is commonly accepted) the mathematical realization for (a) and (b). We start with the consideration of a symmetric coin (with sides denoted by symbols 0 and 1), $L = \{0, 1\}$. Here we can use the classical definition of probabilities as the starting point for the construction of $\mathbf{P}^{\text{Kol}}$. As there are two equally possible cases,

[5] The constructive probability theory (see, for example, [137]) can be considered as an attempt to realize on the mathematical level of rigorousness the idea to use C as a space of elementary events.

the classical probabilities $\mathbf{P}^{\mathrm{cl}}(0) = \mathbf{P}^{\mathrm{cl}}(1) = 1/2$. Now consider m trials of the coin and write all possible samples (3.2). At this point it seems that the formalism is developed in the same way as in the von Mises theory. However, the next step demonstrates the crucial difference between two the approaches. Denote by $S_m = L^m$ the set of all vectors of length m with coordinates $0, 1$. This set is considered as a statistical ensemble. Thus, for $\mathbf{i} = (i_1, \dots, i_m) \in S_m$, $\mathbf{P}^{\mathrm{ens}}(\mathbf{i}) \equiv \mathbf{P}_{S_m}(\mathbf{i}) = 1/|S_m| = 1/2^m$. Bernoulli proved the following mathematical result for these ensemble probabilities:

Theorem 6.1 (Bernoulli). *The larger m is, the larger is the proportion of those vectors in S_m in which the relative number of zeros (or of ones) deviates from 1/2 by less than a given ϵ.*

Obviously this is the result for proportional probability. But Bernoulli and most authors state this result as the result for the frequency probability: if one throws a 'true' coin long enough it is almost certain that the relative number of heads will deviate by less than ϵ from 1/2.

The Kolmogorov probability measure $\mathbf{P}^{\mathrm{Kol}}$ on the space of elementary events

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n, \dots) : \omega_j \in L\},$$

where $L = \{0, 1\}$, will be defined with the aid of the ensemble probabilities $\mathbf{P}_{S_m}$. For $\mathbf{i} = (i_1, \dots, i_m) \in S_m$, a *cylindrical subset* of Ω with the base $\mathbf{i}$ is defined as $B_{\mathbf{i}} = \{\omega \in \Omega : \omega_1 = i_1, \dots, \omega_m = i_m\}$. We set $\mathbf{P}^{\mathrm{Kol}}(B_{\mathbf{i}}) = \mathbf{P}_{S_m}(\mathbf{i}) = 1/2^m$. Denote the σ-algebra generated by all cylindrical subsets by $\mathcal{F}$ (i.e., this is the minimal σ-algebra which contains all cylindrical subsets of Ω). $\mathbf{P}^{\mathrm{Kol}}$ is extended as a σ-additive measure on the σ-algebra $\mathcal{F}$.

It is typically assumed that the frequency part (a) of the interpretation can be described by following mathematical result for the measure $\mathbf{P}^{\mathrm{Kol}}$.

Theorem 6.2 (Law of large numbers). *For any $\epsilon > 0$,*

$$\mathbf{P}^{\mathrm{Kol}}(\{\omega \in \Omega : |\nu_m(1;\omega) - 1/2| > \epsilon\}) \to 0, \quad m \to \infty,$$

where $\nu_m(1;\omega) = n_m(1;\omega)/m$ and $n_m(1;\omega) = \sum_{j=1}^m \omega_j$.

However, like the classical Bernoulli theorem, the law of large numbers is not connected with the frequency approximation of probabilities. This is the statement on the approximation of classical probabilities $\mathbf{P}^{\mathrm{cl}}(0) = \mathbf{P}^{\mathrm{cl}}(1) = 1/2$ by ensemble probabilities. On the other hand, we could use the so called strong law of large numbers.

Theorem 6.3 (Strong law of large numbers). *There exists a subset $\Omega' \in \mathcal{F}$ such that $\mathbf{P}^{\mathrm{Kol}}(\Omega') = 1$ and $\nu_m(1,\omega) \to 1/2$, $m \to \infty$, for all sequences $\omega \in \Omega'$.*

But on the basis of this statement we could not say anything about the statistical stabilization of $\nu_m(1;\omega)$ for any concrete sequence $\omega \in \Omega$. The strong law of large numbers do not say anything about a frequency approximation of ensemble probabilities; this is the statement about the frequency approximation of the classical probabilities $\mathbf{P}^{\mathrm{cl}}(0) = \mathbf{P}^{\mathrm{cl}}(1) = 1/2$ in the sense of the ensemble probabilities.

Conclusion. The laws of large numbers cannot be applied for describing the statistical stabilization of frequencies in sampling experiments.

We construct now Kolmogorov's measure $\mathbf{P}^{\mathrm{Kol}}$ in the general case. The classical definition of probability cannot be used for nonsymmetrical coin. We use the frequency definition. The statistical experiments for coin's tossing produce collectives x with the label sets $L = \{0, 1\}$. Let us assume that all these collectives have the same probability distribution: $q_0 = \mathbf{P}^{\mathrm{fr}}(0)$ and $q_1 = 1 - q_0 = \mathbf{P}^{\mathrm{fr}}(1)$. For cylindric set $B_{\mathbf{i}} = \{\omega \in \Omega : \omega_1 = i_1, \ldots, \omega_k = i_k\}$, $\mathbf{i} = (i_1, \ldots, i_k)$, $i_l \in L = \{0, 1\}$, the probability is defined as

$$\mathbf{P}^{\mathrm{Kol}}(B_{\mathbf{i}}) = q_1^{|\mathbf{i}|} q_0^{k-|\mathbf{i}|}, \tag{6.1}$$

where $|\mathbf{i}| = i_1 + \cdots + i_k$.

In the symmetric case ($q_0 = q_1 = 1/2$) the origin of formula (6.1) has been explained in the ensemble framework. In the general case we could not apply the ensemble framework. Here we can apply frequency arguments. Let $x^{(j)} = (x_t^{(j)})_{t=1}^{\infty}$, $j = 1, 2, \ldots, k$, be collectives having the same label space $L = \{0, 1\}$ and probability distribution $\mathbf{P}_{x^{(j)}}(0) = q_0$, $\mathbf{P}_{x^{(j)}}(1) = q_1$. We form a new collective $x = (x_t)_{t=1}^{\infty}$ with the label space $L^k = L \times \cdots \times L$ by setting $x_t = (x_t^{(j)})_{j=1}^{k}$, $t = 1, 2, \ldots$. We assume that collectives $x^{(j)}$ are independent (see sections 9,10 for the details). In particular, this imply the factorization of the probability distribution $\mathbf{P}_x$ in a product of probability distributions $\mathbf{P}_{x^{(j)}}$. Thus, for each $\mathbf{i} = (i_1, \ldots, i_k)$, $i_l = 0, 1$, there exists

$$\mathbf{P}_x(\mathbf{i}) = \lim_{M \to \infty} \nu_M(\mathbf{i}; x) = \prod_{l=1}^{k} \lim_{M \to \infty} \nu_M(i_l; x^{(l)}) = \prod_{l=1}^{k} \mathbf{P}_{x^{(l)}}(i_l), \tag{6.2}$$

where $\nu_M(\mathbf{i}; x) = n_M(\mathbf{i}; x)/M$ and $\nu_M(i_l; x^{(l)}) = n_M(i_l; x^{(l)})/M$ are relative frequencies for labels $\mathbf{i} \in L^k$ and $i_l \in L$ (in collectives x and $x^{(l)}$, respectively). Formula (6.2) can be used as the motivation for definition (6.1) of probability of a cylindric subset $B_{\mathbf{i}}$ of Ω.

The $\mathbf{P}^{\mathrm{Kol}}$ defined on cylindric subsets by (6.2) can be extended to a probability measure on the σ-algebra $\mathcal{F}$ of Ω generated by cylindric subsets.

We analyse now how the $\mathbf{P}^{\mathrm{Kol}}$ serves to purposes (a) and (b). Here we can also use the strong law of large numbers:

Theorem 6.4 (Strong law of large numbers for nonsymmetrical distributions). *There exists a subset $\Omega' \in \mathcal{F}$ such that $\mathbf{P}^{\mathrm{Kol}}(\Omega') = 1$ and $\nu_M(\alpha; \omega) = n_M(\alpha; \omega)/M \to q_\alpha$, $M \to \infty$, $\alpha = 0, 1$, for all sequences $\omega \in \Omega'$.*

It seems that (with the same remarks as in the symmetric case) this is the mathematical realization of (a). The part (b) can be interpreted in the following way. If, for example, $q_0 \ll 1$ then, for each j, $\mathbf{P}^{\mathrm{Kol}}(\omega : \omega_j = 0) = q_0 \ll 1$. Thus 'probability to obtain 0 in the jth test is practically zero.' In fact, the problem is more complicated. In the nonsymmetrical case we could not interpret sequences $\omega \in \Omega$ (even some of them) as collectives generated by the statistical experiment[6]. The construction of Kolmogorov's measure $\mathbf{P}^{\mathrm{Kol}}$ demonstrates that $\omega \in \Omega$ have the meaning of (infinite) 'multi-labels' for the collective $x = (x_l)_{l=1}^{\infty}$, where $x_l = (x_l^{(j)})_{j=1}^{\infty} \in \Omega$, which is obtained on the basis of a sequence $\{x^{(j)}\}$, $j = 1, 2, \ldots$, of independent collectives having the same probability distribution $q_\alpha, \alpha = 0, 1$ (i.e., a sequence (for $j = 1, 2, \ldots$) of parallel running sequences (for $l = 1, 2, \ldots$) of coins' tossings). Thus the strong law of large numbers says that 'practically all' these 'multi-labels' have the property of the statistical stabilization and limits of relative frequencies (accidentally!) coincide with probabilities q_α corresponding to collectives. Thus the probability measure $\mathbf{P}^{\mathrm{Kol}}$ describes the frequency approximation of probabilities only indirectly.

The $\mathbf{P}^{\mathrm{Kol}}$ describes only random phenomena which have the property of *ergodicity*. The ergodicity has the following meaning. First we consider the statistical experiment in that one person makes a long run of coin's tossings, $u = (u_1, \ldots, u_M, \ldots)$, and obtains the relative frequencies $\nu_M(\alpha; u) = n_M(\alpha; u)/M$, $\alpha = 0, 1$. Then we consider another statistical experiment in that all persons belonging to a large statistical ensemble S (population) make simultaneously just one coin's tossing. As the result of the latter experiment we obtain the proportions (in S), $\nu_S(\alpha) = |S(\alpha)|/|S|, \alpha = 0, 1$, of persons who have obtained the label α. Then $\nu_M(\alpha; u) \approx \nu_S(\alpha)$ for large M and $|S|$. Of course, we could not assume that all random phenomena have the property of ergodicity. Thus in general the ensemble and frequency interpretations of probability must be separated.

Remark 6.2. Let collectives $x^{(j)}$ be independent, but not in general equally distributed: $\mathbf{P}_{x^{(j)}}(\alpha) = q_{\alpha j}$, $\alpha = 0, 1$. Then we obtain that $\mathbf{P}_x(\mathbf{i}) = \prod_{l=1}^{N} q_{i_l l}$. This can be used as the motivation to define a probability of a cylindric subset of Ω by $\mathbf{P}^{\mathrm{Kol}}(B_{\mathbf{i}}) = \prod_{l=1}^{N} q_{i_l l}$. We underline again that there we could not use ensemble arguments to define probabilities of cylindric subsets.

Conclusion. Kolmogorov's ensemble-frequency interpretation can be used only for ergodic random phenomena.

7 Subjective (Bayesian) probability theory

According to the subjective interpretation of probability, it is the *degree of belief* in the occurrence of an event attributed by a given person at a given instant and with given set of information that is important. It is very important for our further quantum mechanical considerations that changing information changes probabilities. We illustrate this by an example.

[6]Thus in the nonsymmetrical case the strong law of large numbers could not be interpreted in the same way as in the symmetric case. Only in the symmetric case we can interpret some of 'elementary events' $\omega \in \Omega$ as collectives generated by coin's tossing.

Example 7.1. I have forgotten something: Have I sent a letter to my friend or not? I can propose my subjective probabilities q_1 (the letter was sent), q_2 (it was not sent), $q_1 + q_2 = 1, q_j \in [0,1]$. Suppose that we have an ideal postal system, i.e., a letter could not disappear in the postal service. If I telephone to my friend and he tells me that he has received the letter, then at that moment the probabilities will immediately change: $q_1 \to 1$ and $q_2 \to 0$, in the opposite case: $q_1 \to 0$ and $q_2 \to 1$.[7]

In fact the subjective theory of probability is a sufficiently good theory from the operational point view. The main problem of this approach is how to choose the subjective probabilities in a concrete case. In this theory it is postulated that the probability depends on the *status of information* which is available to whoever evaluates probability. Thus the evaluation of probability is conditioned by some *a priori* ('theoretical') prejudices and by some facts ('experimental data'). However, in applications all this information is nothing other than information about frequency or proportional probabilities.

It must be noted that the subjective probability theory is described mathematically by the Kolmogorov probability space $(\Omega, \mathcal{F}, \mathbf{P})$. The Bayes formula (2.4) is the cornerstone of this theory (therefore, it is also called Bayesian theory). As we have discussed, in principle we can exclude (2.4) from the Kolmogorov theory and consider a more general formalism which describes violations of (2.4). Such an approach is impossible in the subjective framework. The subjective probability theory is applied in the following form. There is a fixed set of hypotheses (events) $H_i \in \mathcal{F}$: $\bigcup_i H_i = \Omega, H_i \cap H_j = \varnothing, i \neq j$. Let $E \in \mathcal{F}$ be an event. Suppose that we know conditional probabilities $\mathbf{P}(E/H_i)$. Then we find $\mathbf{P}(H_i/E)$ by (3.8) and the formula of total probability: $\mathbf{P}(E) = \sum_i \mathbf{P}(E/H_i)\mathbf{P}(H_i)$, i.e.,

$$\mathbf{P}(H_i/E) = \frac{\mathbf{P}(E/H_i)\mathbf{P}(H_i)}{\sum_j \mathbf{P}(E/H_j)\mathbf{P}(H_j)}. \tag{7.1}$$

This is the standard form of *Bayes' theorem*.

Remark 7.1. Of course, Bayes' formula plays a great role in probability theory. However, as we have seen, there are restrictions for using this formula. These also are restrictions for using Bayesian probability theory. According to Bayesian theory $\mathbf{P}_H(E) = \mathbf{P}(E/H)$ is a subjective probability (a measure of an individual belief) on the basis of the known set of conditions H; in particular, $\mathbf{P}(E) = \mathbf{P}(E/\Omega)$ correspond to the set of all conditions. Therefore it is assumed that we can always extract the information H from the total amount of information Ω.

The main positive consequence of the subjective approach to probability theory is the connection between probability and information. The idea that probability is a measure of information on a random phenomenon (for example, a statistical ensemble)

[7] In quantum formalism such a reduction of subjective probabilities is nothing other than so called *collapse* of a wave function ($\phi = \sqrt{q_1}\phi_1 + \sqrt{q_2}\phi_2$).

looks attractive. Typically such an information is reduced to our subjective knowledge about a random phenomenon. This information probability is coded by real number $q_{\text{inf/pr}} \in [0, 1]$. Intuitively it is identified with the classical probability (based on the proportion of equally possible cases) or with the ensemble probability. However, the relation between subjective probability and classical or ensemble probabilities is indirect. The subjective probability approach claims that $q_{\text{inf/pr}}$ is chosen on the basis of 'subjective reasons' of an individual.

The subjective probability approach can be strongly improved if we assume that 'subjective reasons' are nothing other than the calculation of probability with respect to an ensemble of ideas S (in the brain of an individual) which are connected with the concrete random phenomenon. Thus $q_{\text{inf/pr}}(\alpha) = |S(\alpha)|/|S|$, where $S(\alpha)$ is the sub-ensemble of ideas which imply the property α.

8 Foundations of randomness

We study some special questions of the frequency probability theory connected with the principle of randomness (see, for example, [137], [163], [18] for the details).

8.1 Existence of collectives; Kamke's objection

As we have already remarked, the principle of randomness based on the invariance of limits of relative frequencies with respect to the set of *all possible place selections* is too general. In fact, there are no sequences which satisfy this principle. To show this, we follow arguments of E. Kamke [74] (see also [137]).

Let $L = \{0, 1\}$ and $x = (x_j)_{j=1}^{\infty}$, $x_j \in L$, be a collective which induces the probability distribution $\mathbf{P}(\alpha) = 1/2$, $\alpha = 0, 1$. Consider the set SI of all strictly increasing sequences of natural numbers. This set can be formed independently of x; but, among elements of SI, we have the strictly increasing sequence $\{n : x_n = 1\}$. This sequence define a place selection which selects the subsequence $(11 \dots 1 \dots)$ from x. Hence x is not a collective after all!

The reader may well feel uncomfortable with the mathematical structure of the argument. Kamke claims to have shown that for every putative collection x there exists a place selection ϕ that disturbs the statistical stabilization of frequencies to probability $1/2$. The use of the existential quantifier here classical (Platonistic). Indeed, it seems impossible to exhibit explicitly a procedure which satisfies von Mises' criterion (independence on value x_n) and at the same time selects the subsequence $(11 \dots 1 \dots)$ from x. The interesting analysis of this problem can be found in the review of M. van Lambalgen [163]. He is convinced that a satisfactory treatment of random sequences is possible only in set theories lacking the set power axiom, in which random sequences "are not already there." However, even we uncritical accept classical mathematics, Kamke's argument is somewhat beside the mark in that it fails to appreciate the purpose of von Mises' axiomatization. It refers to what **could** happen, whereas Mises' axioms are rooted in experience and refer to what **does** happen.

Remark 8.1. In various places von Mises likens the principle of randomness to the first law of thermodynamics. Both are statements of impossibility: the principle of randomness is the principle of the excluded gambling strategy, while the first law (conservation of energy) is equivalent to the impossibility of a *perpetuum mobile* of the first kind. I think that such an analogy is not so natural. It would be more natural to connect the first law of thermodynamics with the first von Mises principle, the principle of the statistical stabilization of relative frequencies. The impossibility to perform precise measurements implies that the law of conservation of energy is only a statistical law. Thus it is just one of exhibitions of the principle of the statistical stabilization of relative frequencies. M. van Lambalgen compared the principle of randomness with the second law of thermodynamics, the law of increase of entropy or the impossibility of a perpetuum mobile of the second kind. Indeed, Kamke's objection is reminiscent of Maxwell's celebrating demon, that "very observant and neat-fingered being", invented to show that entropy decreasing evolutions may occur. Maxwell's argument of course in no way detracts from the validity of the second law, but serves to highlight the fact that statistical mechanics cannot provide an absolute foundation for entropy increase, since it does not talk about what **actually** happens (see [163] for further mathematical details).

The early attempts to formalize Mises' principle of randomness were based on considerations of different classes of lawlike place selection. The idea was to fix some class of lawlike place selections and then construct a set of collectives with respect to that class. Various authors (e.g. Popper, Reichenbach, Copeland) independently arrived at the so called *Bernoulli* selections. To discuss this class of place selections, it is convenient to formalize the definition of place selection.

Denote by L^* the set of all finite words $x = (x_1, \ldots, x_m)$, $x_j \in L$, $m = 1, 2, \ldots$ in the alphabet (label set) $L = \{\alpha_0, \alpha_1, \ldots, \alpha_l\}$, $l > 1$; as usual the symbol L^∞ is used to denote the set of all infinite sequences $x = (x_1, \ldots, x_m, \ldots)$, $x_j \in L$. Set $x_{1:n} = (x_1, \ldots, x_n)$ for $x \in L^\infty$ (this is the initial segment of the length n of the sequence x). A place selection ϕ is defined on the basis of a function $f : L^* \mapsto \{0, 1\}$. The domain of definition of a place selection ϕ corresponding to f is the set

$$\operatorname{dom} \phi = \{x \in L^\infty : \forall n \; \exists k \geq n : \; f(x_{1:k}) = 1\} \subset L^\infty.$$

For $x \in \operatorname{dom} \phi$, we set $\phi(x) = \bigcap_n \bar{\phi}(x_{1:n})$, where the map $\bar{\phi} : L^* \mapsto L^*$ is defined as $\bar{\phi}(u\alpha) = \bar{\phi}(u)\alpha$ if $f(u) = 1$ and $\bar{\phi}(u\alpha) = \bar{\phi}(u)$ if $f(u) = 0$ (here $u = (u_1, \ldots, u_m)$ and $u\alpha = (u_1, \ldots, u_m, \alpha)$, $u_j, \alpha \in L$). Thus a place selection ϕ is a partial function $\phi : L^\infty \mapsto L^\infty$.

Example 8.1 (Bernoulli sequences). Let $w = (w_1, \ldots, w_s)$, $w_j \in L$, be a fixed word. For a sequence $x \in L^\infty$, we choose all x_n such that w is a final segment of $x_{1:n}$.

The domain of this place selection, ϕ_w, is the set of all sequences $x \in L^\infty$ which contain infinitely many occurrences of the word w. Formally ϕ_w is defined on the basis of the function $f_w : L^* \mapsto \{0,1\}$, $f_w(u) = 1$, if w is a final segment of u, $f_w(u) = 0$, if not. A *Bernoulli sequence* (with respect to a probability distribution $P(\alpha_j) = p_j$, $j = 1, \ldots, l$, $L = \{\alpha_1, \ldots, \alpha_l\}$) is a sequence $x \in L^\infty$ such that $\lim_{n\to\infty} \nu_n(\alpha_j; x) = p_j$, $j = 1, \ldots, l$, and for all words $w \in L^*$

$$\lim_{n\to\infty} \nu_n(\alpha_j; \phi_w x) = p_j, \quad j = 1, \ldots, l, \tag{8.1}$$

where $\nu_n(\alpha_j; \phi_w x)$ is the relative frequency of occurrence of the label α_j in the initial segment of length n of the sequence $\phi_w x$. If $L = \{0,1\}$ is the binary alphabet and $\mathbf{P}_x(1) = p, \mathbf{P}_x(0) = 1 - p$, then (8.1) has the form

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} (\phi_w x)_j = p$$

for all words $w \in \{0,1\}^*$.

The sets of Bernoulli place selections and sequences are denoted by symbols $\mathcal{U}_B$ and X_B, respectively.

A. Church [39] suggested to consider the set $\mathcal{U}_{\mathrm{Ch}}$ of place selections which are generated by total recursive functions $f : L^* \to \{0,1\}$ (functions which can be computed by using algorithms). Church's collectives (random sequences) are sequences $x \in L^\infty$ which satisfy the principle of the statistical stabilization and the principle of randomness for the set of place selections $\mathcal{U}_{\mathrm{Ch}}$. Denote the set of Church's collectives by the symbol X_{Ch}.

Both the sets $\mathcal{U}_B$ and $\mathcal{U}_{\mathrm{Ch}}$ are countable. The existence of Bernoulli sequences and Church's collectives is a consequence of the general result of A. Wald [173].

Let $p = (p_j)$: $p_j = \mathbf{P}(\alpha_j)$, $j = 1, \ldots, l$, $L = \{\alpha_1, \ldots, \alpha_l\}$, be a probability distribution on the label set L. Let $\mathcal{U}$ be a set of place selections. We set

$$X(\mathcal{U}, p) = \{x \in L^\infty : \forall \phi \in \mathcal{U} \lim_{n\to\infty} \nu_n(\alpha_j; \phi x) = p_j, \ j = 1, \ldots, l\}.$$

Theorem 8.1 (Wald). *For any countable set $\mathcal{U}$ of place selections and any probability distribution p on the label set L, the set of sequences $X(\mathcal{U}, p)$ has the cardinality of the continuum.*

Thus at least for countable sets of place selections $\mathcal{U}$ Mises' frequency theory of probability can be developed on the mathematical level of rigorousness. R. von Mises was completely satisfied by this situation (see [171]). However, he was strongly against the idea to fix once and for all a set of place selections. By Mises the concrete set of place selections is determined by a physical problem. But mathematicians prefer to consider fixed classes of place selections. In particular, the large part of mathematical community consider Church's choice as the most reasonable. The author does not think that the choice of total recursive functions as place selections can be justified by some physical arguments. The idea that reality which can be studied by human mind can be reduced to reality produced by Turing machines looks rather primitive in the light of modern investigations of the processes of thinking. It seems that the brain uses transformations $\bar{\phi} : L^* \to L^*$ which based on non-recursive functions, [88], [90].

8.2 Geometric and frequency spaces

According to the modern ideology of geometry, geometric model is a pair (X, G), where X is a set of points and G is a group of transformations of X. Such an approach is closely connected with von Mises' approach to probability theory. Here we have a system of place selections $\mathcal{U}$ (which plays the role of a group of transformations G) and the space $X(\mathcal{U}, p)$ of 'probabilistic points'. The pair $(X(\mathcal{U}, p), \mathcal{U})$ can be called a *frequency probability model.* Moreover, as in geometry, we have to consider some algebraic structure on the system of transformations $\mathcal{U}$. We shall demonstrate that we have to use semigroups (with unit) of transformations $\mathcal{U}$.

Let $\mathcal{U}$ be a system of place selections containing the identity transformation. If $x \in X(\mathcal{U}, p)$, it is natural to assume that, for each $\phi \in \mathcal{U}$, $y = \phi x \in X(\mathcal{U}, p)$: each element ϕ of $\mathcal{U}$ transforms an $\mathcal{U}$-collective x in a new $\mathcal{U}$-collective. Thus, for each $\psi \in \mathcal{U}$, the sequence $z = \psi y = \psi \circ \phi x$ satisfies the principle of the statistical stabilization. Let $f = \psi \circ \phi \notin \mathcal{U}$. Then we can extend the system of transformations $\mathcal{U}$ by setting $\mathcal{U}' = \mathcal{U} \cup \{f\}$. It is evident that (under our assumption) the set of points $X(\mathcal{U}', p)$ coincides with the set $X(\mathcal{U}, p)$. Therefore it would be natural to assume from the beginning that $\mathcal{U}$ is a semigroup.

One of nice examples of frequency probability spaces is the space $(X_{\mathrm{Ch}}, \mathcal{U}_{\mathrm{Ch}})$ based on the system of totally recursive functions.

8.3 Ville's objection

Although Wald's reformulation of von Mises' ideas solved the problem of consistency, it lead to an objection of entirely different kind.

Theorem 8.2 (Ville, [166]). *Let $L = \{0, 1\}$ and let $\mathcal{U} = \{\phi_n\}_{n=1}^{\infty}$ be a countable set of place selections. Then there exists $x \in L^{\infty}$ such that*

(a) for all n, $$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} (\phi_n x)_j = \frac{1}{2};$$

(b) for all N, $$\frac{1}{N} \sum_{j=1}^{N} (\phi_n x)_j \geq \frac{1}{2}.$$

Such an x is a collective with respect to $\mathcal{U}$ ($x \in X(\mathcal{U}, 1/2)$), but seems to be far too regular to be called random. Formally, x's with property (b) form a set of Lebesgue measure 0 (this is a consequence of the law of iterated logarithm).

8.4 Ensemble probability approach to randomness

Ville and Fréchet used Theorem 8.2 to argue that collectives in the sense of von Mises and Wald do not necessarily satisfy all intuitively required properties of randomness. Ville introduced a new way of characterizing random sequences, based on the following idea: *a random sequence should satisfy all properties of probability one.* Strictly speaking, this is of course impossible: we have to choose countably many from among those properties. It must be underlined that Ville's idea is really completely foreign to von Mises. For von Mises, a collective $x \in L^\infty$ induces a probability on the set of labels L, not on the set of all sequences L^∞. Hence there is no connection at all between properties of probability one in L^∞ and properties of individual collectives.

Per Martin-Löf [142, 144] proposed to consider 'recursive properties of probability one' (i.e., properties which can be tested with the aid of algorithms). Such an approach induces the fruitful theory of recursive tests for randomness (see, for example, [137], [175]). Similar approach was developed by Schnorr [158]. We underline that approaches of Martin-Löf and Schnorr (as well as Ville and Fréchet) have nothing to do with the justification of Mises' frequency probability theory.

8.5 Kolmogorov complexity

A. N. Kolmogorov tried to find foundations of randomness by reducing this notion to the notion of complexity. Let $L = \{0, 1\}$ and $x \in L^*$.

Definition 8.1 (Kolmogorov). Let $\mathcal{A}$ be an arbitrary algorithm. The *complexity of a word* x with respect to $\mathcal{A}$ is $K_{\mathcal{A}}(x) = \min l(\pi)$, where $\{\pi\}$ are the programs which are able to realize the word x with the aid of $\mathcal{A}$.

Here $l(\pi)$ denotes the length of a program π. This definition depends on the structure of an algorithm $\mathcal{A}$. Later Kolmogorov proved the following theorem:

Theorem 8.3. *There exists an algorithm $\mathcal{A}_0$ (optimal algorithm) such that*

$$K_{\mathcal{A}_0}(x) \preceq K_{\mathcal{A}}(x) \tag{8.2}$$

for every algorithm $\mathcal{A}$.

As usual, (8.2) means that there exists a constant C such that $K_{\mathcal{A}_0}(x) \leq K_{\mathcal{A}}(x)+C$ for all words x. An optimal algorithm $\mathcal{A}_0$ is not unique.

Definition 8.2. The *complexity $K(x)$ of the word x* is equal to the complexity $K_{\mathcal{A}_0}$ with respect to one fixed (for all considerations) optimal algorithm $\mathcal{A}_0$.

The original idea of Kolmogorov [135], [134] was that complexity $K(x_{1:n})$ of initial segments $x_{1:n}$ of a random sequence x has to have the asymptotic $K(x_{1:n}) \sim n$, $n \to \infty$, i.e., we might not find a short code for $x_{1:n}$. However, this nice idea was rejected due to an objection of Per Martin-Löf [144]. To discuss this objection and connection of Kolmogorov complexity with Martin-Löf randomness, it is better to use conditional Kolmogorov complexity $K(x;n)$ instead of complexity $K(x)$. Complexity $K_{\mathcal{A}}(x;n)$ is defined as the length of a minimal program π which produces the output x on the basis of information that the length of the output x is equal to n.

Theorem 8.4. *Let f be a total recursive function such that $\sum_{n=1}^{\infty} 2^{-f(n)} = \infty$. Then, for every sequence x, $K(x_{1:n};n) < n - f(n)$ for infinitely many n.*

In particular, we can choose $f(n) = \log_2 n$. Thus, for any binary sequence x, $K(x_{1:n};n) < n - \log_2 n$ for infinitely many n. Hence 'Kolmogorov random sequences' do not exit.

P. Martin-Löf obtained also an estimate of $K(x_{1:n};n)$ from below:

Theorem 8.5. *Let f be such that $\sum_{n=1}^{\infty} 2^{-f(n)} < \infty$. Then, with probability one, $K(x_{1:n};n) \geq n - f(n)$ for all but finitely many n.*

In particular, we can choose $f(n) = 2\log_2 n$. Thus, for almost all binary sequences x, $K(x_{1:n};n) \geq n - 2\log_2 n$ for all but finitely many n. Therefore for almost all binary sequences Kolmogorov complexity $\phi(n) = K(x_{1:n};n)$ oscillates between graphs of the functions $g_{\max}(n) = n$ and $g_{\min}(n) = n - 2\log_2 n$ (with finitely many intersections with $g_{\min}(n)$). The graph of the function $\phi(n)$ has infinitely many intersections with the graph of the function $f_{\mathrm{mid}}(n) = \log_2 n$.

The following two theorems [144] give the connection between high Kolmogorov complexity (for infinitely many initial segments) and Martin-Löf randomness:

Theorem 8.6. *Let f be a total recursive function such that $\sum_{n=1}^{\infty} 2^{-f(n)}$ is recursively convergent. Then, if x is random in the sense of Martin-Löf, then $K(x_{1:n};n) \geq n - f(n)$ for all but finitely many n.*

That $\sum_{n=1}^{\infty} 2^{-f(n)}$ is recursively convergent means that there is a recursive sequence $n_1, n_2, \ldots, n_k, \ldots$ such that

$$\sum_{n_m+1}^{\infty} 2^{-f(n)} \leq 2^{-m}, \quad m = 1, 2, \ldots .$$

Theorem 8.7. *If there exists a constant c such that $K(x_{1:n};n) \geq n - c$ for infinitely many n, then the sequence x is random in the sense of Martin-Löf.*

Critical remarks

1) Despite of the great success of Kolmogorov and Martin-Löf approaches, it is doubtful that these approaches provide the adequate description of randomness in physical reality. The main objection is against the use of recursive functions (algorithms). On one hand, there are no reasons to suppose that random sequences produced by physical

phenomena must pass all recursive tests for randomness (even the law of large numbers). On the other hand, 'randomness' of such sequences may be characterized by some systems of non-recursive transformations.

2) It seems impossible to reduce Martin-Löf randomness to Mises' randomness. Denote the class of Martin-Löf random sequences (with respect to the uniform distribution) by the symbol RM. The reduction of Martin-Löf randomness to Mises' randomness must be given by the equality $\mathrm{RM} = X(\mathcal{U}, 1/2)$ for some class $\mathcal{U}$ of place selections. However, it seems impossible to find such a class $\mathcal{U}$. For example, let $\mathcal{U} = \mathcal{U}_{\mathrm{Ch}}$ be the class (semigroup) of Church place selections. Then, as each $\phi \in \mathcal{U}_{\mathrm{Ch}}$ gives a recursive property of probability one, we have $\mathrm{RM} \subset X(\mathcal{U}, 1/2)$. Ville's result, combined with the observation that the Martin-Löf random sequences satisfy the law of the iterated logarithm, shows that the inclusion is strict. Moreover, it can be shown (see M. van Lambalgen [163]) that the set of sequences $X(\mathcal{U}, 1/2) \setminus \mathrm{RM}$ is rather large.

Therefore approaches of Martin-Löf–Ville–Fréchet and von Mises give totally different viewpoints to the notion of randomness. The first approach is based on the ensemble interpretation of probability and the second approach is based on the frequency interpretation of probability. As we have already noticed, these interpretations could not be unified in one (mixed) ensemble-frequency interpretation.

9 Operation of combining of collectives

In the three basic operations discussed in Section 3, one single collective x served each time as point of departure for the construction of a new collective. We consider the problem of combining of two or more given collectives. We start with S-sequences (sequences which satisfy the principle of the statistical stabilization).

Let $x = (x_j)$ and $y = (y_j)$ be two S-sequences with label sets L_x and L_y, respectively. We define a new sequence

$$z = (z_j), \quad z_j = (x_j, y_j). \tag{9.1}$$

(in general z is not an S-sequence with respect to the label set $L_z = L_x \times L_y$). Let $a \in L_x$ and $b \in L_y$. Among the first N elements of z there are $n_N(a; z)$ elements with the first component equal to a. As $n_N(a; z) = n_N(a; x)$ is a number of $x_j = a$ among the first N elements of x, we obtain that $\lim_{N\to\infty} \frac{n_N(a;z)}{N} = \mathbf{P}_x(a)$. Among these $n_N(a; z)$ elements, there are a number, say $n_N(b/a; z)$ whose second component is equal to b. The frequency $\nu_N(a, b; z)$ of elements of the sequence z labeled (a, b) will then be

$$\frac{n_N(b/a; z)}{N} = \frac{n_N(b/a; z)}{n_N(a; z)} \frac{n_N(a; z)}{N}.$$

We set $\nu_N(b/a; z) = \frac{n_N(b/a;z)}{n_N(a;z)}$. Let us assume that, for each $a \in L_x$, the subsequence $y(a)$ of y which is obtained by choosing y_j such that $x_j = a$ is an S-sequence[8]. Then,

[8] In general such a choice of the subsequence $y(a)$ of y is not a place selection.

for each $a \in L_x, b \in L_y$, there exists

$$\mathbf{P}_z(b/a) = \lim_{N\to\infty} \nu_N(b/a; z) = \lim_{N\to\infty} \nu_N(b; y(a)) = \mathbf{P}_{y(a)}(b).$$

We have

$$\sum_{b\in L_2} \mathbf{P}_z(b/a) = 1. \tag{9.2}$$

The existence of $\mathbf{P}_z(b/a)$ implies the existence of $\mathbf{P}_z(a,b) = \lim_{N\to\infty} \nu_N(a,b;z)$. Moreover, we have

$$\mathbf{P}_z(a,b) = \mathbf{P}_x(a)\mathbf{P}_z(b/a) \tag{9.3}$$

and $\mathbf{P}_z(b/a) = \mathbf{P}_z(a,b)/\mathbf{P}_x(a)$, if $\mathbf{P}_x(a) \neq 0$. By (9.2) and (9.3) we obtain

$$\sum_{a\in L_a} \sum_{b\in L_2} \mathbf{P}_z(a,b) = 1.$$

Thus in this case the sequence z is an S-sequence with the probability distribution $\mathbf{P}_z(a,b)$ defined by (9.3). The S-sequence y is said to be *combinable* with the S-sequence X. This relation is denoted by $\overleftarrow{xy}$. The relation of combining is a symmetric relation on the set of pears of S-sequences with strictly positive probability distributions. To show this, we write

$$\nu_N(a/b; z) = \frac{\nu_N(a,b;z)}{\nu_N(b;z)},$$

$a \in L_x, b \in L_y$. If $\overleftarrow{xy}$ and $\mathbf{P}_x(a) > 0$, $\mathbf{P}_y(b) > 0$, $a \in L_x, b \in L_y$, then, for each $b \in L_y, a \in L_x$, there exists

$$\mathbf{P}_z(a/b) = \lim_{N\to\infty} \nu_N(a/b; z) = \frac{\mathbf{P}_z(a,b)}{\mathbf{P}_y(b)} = \frac{\mathbf{P}_z(b/a)\mathbf{P}_x(a)}{\mathbf{P}_y(b)}.$$

Thus we obtain that $\overleftarrow{yx}$. On the other hand, if, for example, $\mathbf{P}_y(b) = 0$ and $\mathbf{P}_z(a,b) = 0$, then in principle $\nu_N(a/b; z)$ *may fluctuate.* In that case x is not combinable with y. The previous considerations can be summarized as the following proposition.

Proposition 9.1. *Let x and y be two S-sequences with strictly positive probability distributions. Then the following conditions are equivalent: (1) $\overleftarrow{xy}$; (2) $\overleftarrow{yx}$; (3) the sequence z defined by (9.1) is an S-sequence.*

If $\overleftarrow{yx}$ and $\overrightarrow{xy}$, then x and y are said to be combinable. This relation is denoted by $\overline{xy}$. If $\overleftarrow{xy}$, then the conditional probabilities $\mathbf{P}_z(b/a)$ are well defined even if $\mathbf{P}_x(a) = 0$. Typically such probabilities do not play any role in probabilistic considerations. We say that y is (mod $\mathbf{P}_x$)–combinable with x, $\overleftarrow{xy}$ (mod $\mathbf{P}_x$), if, for each $a \in L_x$, $\mathbf{P}_x(a) > 0$, the sequence $y(a)$ is an S-sequence. By Proposition 9.1 we have $\overleftarrow{yx}$ (mod $\mathbf{P}_x$) $\Leftrightarrow$ z is an S-sequence $\Leftrightarrow$ $\overleftarrow{xy}$ (mod $\mathbf{P}_y$). Thus we need not use the arrow to denote this relation of combining. This relation will be denoted as $\overline{xy}$ (mod $\mathbf{P}_z$).

We introduce the operation of combining for collectives. We start with some preliminary considerations on place selections. Let $x = (x_j)$, $x_j \in L_x$, and $y = (y_j)$, $y_j \in L_y$, be two arbitrary sequences and let $z = (z_j)$, $z_j = (x_j, y_j)$. Let Φ_1, Φ_2 and G be some systems of place selections operated in x, y and z, respectively. Let ϕ belongs to G: $\phi z = (z_{n_1}, z_{n_2}, \ldots, z_{n_k}, \ldots)$. We set $\phi^{(1)} x = (x_{n_j})$ and $\phi^{(2)} y = (y_{n_j})$. It should be noticed that in general $\phi^{(1)}$ and $\phi^{(1)}$ are not place selections in x and y, respectively[9]. We set $G_1 = \{f = \phi^{(1)} : \phi \in G\}$, $G_2 = \{g = \phi^{(2)} : \phi \in G\}$. Let $x = (x_j)$ and $y = (y_j)$ be Φ_1 and Φ_2 collectives, respectively. Let G be a system of place selections operated in $z = (z_j)$, $z_j = (x_j, y_j)$, such that, $\Phi_1 \subset G_1$ and $\Phi_2 \subset G_2$. If x and y are mod $\mathbf{P}$-combinable as S-sequences, then they are said to be (mod $\mathbf{P}, G$)-combinable collectives if: (1) the limits $\mathbf{P}_x(a)$, $a \in L_x$, and $\mathbf{P}_y(b)$, $b \in L_y$, are insensitive to transformations belonging to G_1 and G_2, respectively; (2) the limits $\mathbf{P}_z(b/a)$, $\mathbf{P}_x(a) > 0$, and $\mathbf{P}_z(a/b)$, $\mathbf{P}_y(b) > 0$, are insensitive to place selections belonging to G. We can easily prove that x and y are (mod $\mathbf{P}, G$)-combinable collectives iff z is the G-collective.

Proof. (1) Let x and y be (mod $\mathbf{P}, G$)-combinable. Let $\phi \in G$. For $\mathbf{P}_x(a) > 0$, we have:

$$\begin{aligned}\mathbf{P}_{\phi z}(a,b) &= \lim_{N\to\infty} \nu_N(a,b;\phi z) = \lim_{N\to\infty} \nu_N(b/a;\phi z)\nu_N(a;\phi z)\\ &= \lim_{N\to\infty} \nu_N(b/a;\phi z)\nu_N(a;\phi^{(1)}x) = \mathbf{P}_z(b/a)\mathbf{P}_x(a) = \mathbf{P}_z(a,b).\end{aligned}$$

For $\mathbf{P}_x(a) = 0$ we have: $\nu_N(a,b;\phi z) \le \nu_N(a;\phi z) = \nu_N(a;\phi^{(1)}x)$. But

$$\lim_{N\to\infty} \nu_N(a;\phi^{(1)}x) = \lim_{N\to\infty} \nu_N(a;x) = \mathbf{P}_x(a) = 0.$$

Thus $\mathbf{P}_{\phi z}(a,b) = 0 = \mathbf{P}_z(a,b)$.

(2) Let z be the G-collective. Then we obtain

$$\mathbf{P}_{\phi^{(1)}x}(a) = \sum_{b\in L_y} \mathbf{P}_{\phi z}(a,b) = \sum_{b\in L_y} \mathbf{P}_z(a,b) = \mathbf{P}_x(a).$$

In the same way we obtain that $\mathbf{P}_{\phi^{(2)}y}(b) = \mathbf{P}_y(b)$. Finally, we have

$$\mathbf{P}_{\phi z}(a/b) = \mathbf{P}_{\phi z}(a,b)/\mathbf{P}_{\phi^{(1)}x}(a) = \mathbf{P}_z(a,b)/\mathbf{P}_x(a)) = \mathbf{P}_z(b/a),$$

for $\mathbf{P}_x(a) > 0$. □

[9] Let ϕ be defined by a function $f = (f_1, f_2(z_1), f_3(z_1, z_2), \ldots)$, where $f_j = 0, 1$. If we have $f_n(z_1, z_2, \ldots, z_{n-1}) = 1$, then the element z_n is chosen for a new sequence. Let $x = (x_1, x_2, \ldots, x_n, \ldots)$ be a sequence and let $y = (x_m, x_{m+1}, \ldots)$, $m > 1$. Here $z = (z_j)$ has the form: $z_1 = (x_1, x_m), z_2 = (x_2, x_{m+1}), \ldots$. Thus, in particular, $f_2(z_1)$ depends (in general) not only on x_1 but also on x_m. Therefore $\phi^{(1)}$ is not a place selection for x.

The reader can easily define for the collectives x and y the relations $\overleftarrow{yx}$, $\overleftarrow{xy}$ and $\overline{xy}$ with respect to the G which are not based, respectively, on mod $\mathbf{P}_x$, mod $\mathbf{P}_y$ and mod $\mathbf{P}$ factorizations. In general $\overleftarrow{yx}$ does not imply $\overleftarrow{xy}$ or that the sequence z is the G-collective (and vice versa).

Probabilities $\mathbf{P}_z(a/b)$, $\mathbf{P}_z(b/a)$, $a \in L_x$, $b \in L_y$, have the meaning of conditional probabilities. If $\mathcal{S}_x$ and $\mathcal{S}_y$ are statistical experiments which generate x and y, respectively, then, for example, $\mathbf{P}_z(b/a)$ is nothing other than the conditional probability of the result b for $\mathcal{S}_y$ if we knew that a was the result of $\mathcal{S}_x$. It is easy to show that probabilities $\mathbf{P}_z(a,b)$ (or $\mathbf{P}_z(a/b)$) can be obtained on the basis of the general definition of conditional probabilities based on the operation of partition. For each $a \in L_x$, we consider the set $A_a = \{u = (a,b),\ b \in L_y\} \subset L_x \times L_y$ and the point set $B_{a,b} = (a,b) \subset A_a$. Let z be an S-sequence (in particular, a collective). It easy to see that the conditional probability $\mathbf{P}(B_{a,b}/A_a)$ (for $\mathbf{P}(A_a) > 0$) defined on the basis of the operation of partition for z coincides with the probability $\mathbf{P}_z(b/a)$. However, the approach based on the operation of combining seems more attractive than the approach based on the operation of partition. In the first case the conditional probabilities have the natural interpretation as a measure of dependence between collectives x and y.

10 Independence of collectives

Let x and y be S-sequences and let $\overleftarrow{yx}$. The y is said to be independent from x if all S-sequences $y(a)$, $a \in L_x$, have the same probability distribution which coincides with the probability distribution $\mathbf{P}_y$ of y. This implies that

$$\mathbf{P}_z(b/a) = \lim_{N\to\infty} \nu_N(b/a; z) = \lim_{N\to\infty} \nu_N(b; y(a)).$$

Hence

$$\mathbf{P}_z(a,b) = \mathbf{P}_x(a)\mathbf{P}_y(b), \quad a \in L_x, \quad b \in L_y. \tag{10.1}$$

Thus the independence implies the factorization of the two dimensional probability $\mathbf{P}_z(a,b)$. However, in general the multiplication rule (10.1) does not imply independence. If (10.1) holds, but $\mathbf{P}_x(a) = 0$, then in principle $\mathbf{P}_z(b/a)$ may depend on a (or it may be that $\mathbf{P}_z(b/a) = \text{Const} \neq \mathbf{P}_y(b)$). By similar reasons the condition "y is independent from x" does not imply that x is independent from y. Dependence on a such that $\mathbf{P}_x(a) = 0$ (or b, $\mathbf{P}_y(b) = 0$) does not play any role in probabilistic considerations[10]. Therefore it is natural to consider (mod $\mathbf{P}$)-independence.

Let x and y be two (mod $\mathbf{P}$)-combinable S-sequences (or collectives). They are said to be (mod $\mathbf{P}$)-independent if (a) $\mathbf{P}_{y(a)} \equiv \mathbf{P}_y$ for all $a \in L_x$, $\mathbf{P}_x(a) > 0$ and (b) $\mathbf{P}_{x(b)} \equiv \mathbf{P}_x$ for all $b \in L_y$, $\mathbf{P}_y(b) > 0$. In fact, (a) implies (b) and vice versa. For

[10]Of course, we could not completely exclude the possibility that there may exist physical phenomena in that the dependence on labels having zero probabilities plays some rule.

instance, let (a) take place. Then for $\mathbf{P}_y(b) > 0$, we have

$$\mathbf{P}_z(a/b) = \mathbf{P}_z(a,b)/\mathbf{P}_y(b) = \mathbf{P}_z(b/a)\mathbf{P}_x(a)/\mathbf{P}_y(b)$$
$$= \mathbf{P}_{y(a)}(b)\mathbf{P}_x(a)/\mathbf{P}_y(b) = \mathbf{P}_x(a).$$

It is evident that the multiplication rule (10.1) holds for (mod $\mathbf{P}$)-independent sequences. On the other hand, if $\overline{xy}$(mod $\mathbf{P}$) and the multiplication rule (10.1) hold, then x and y are (mod $\mathbf{P}$)-independent.

Remark 10.1. The reader can easily generalize the frequency approach to conditional probabilities and independence to countable sets of labels. Non-countable sets of labels were considered in [162].

Example 10.1. Assume that two coins are tossed simultaneously, the corresponding sequences being x and y (with $L_x = L_y = \{0,1\}$). Our experience says that in mathematical models we can assume that x and y are collectives with probability distributions $\mathbf{P}_x(a)$, $\mathbf{P}_y(b)$, $a,b = 0,1$. We choose two subsequences of y: (1) $y(0) = (y_{j_l})$, where, for the first coin, $x_{j_l} = 0$; (2) $y(1) = (y_{j_l})$, where, for the first coin, $x_{j_l} = 1$. Our experience says that in the mathematical model (for ordinary coins) we can assume that there exist $\mathbf{P}(b/0) = \lim_{N\to+\infty} \nu_N(b; y(0))$ and $\mathbf{P}(b/1) = \lim_{N\to\infty} \nu_N(b; y(1))$, $b = 0,1$. If tossing of the second coin does not depend in any way on the tossing of the first coin, then the relative frequencies in $y(0)$ and $y(1)$ have the same behaviour as relative frequencies in y (this is again the experimental fact). Thus we can assume that collectives x and y are independent.

Example 10.2. Assume that an urn contains balls each marked with a number a, where a belongs to the set $S = \{a_1, \ldots, a_n\}$. The sequence x is induced by the experiment $\mathcal{S}_x$: we draw a ball from the urn, write its label and return it into the urn. The sequence y is induced by the experiment $\mathcal{S}_y$: after drawing the first ball and before returning it, a second ball is drawn from the urn and its label is written. As usual, we define subsequences $y(a_j)$, $j = 1, \ldots, n$, of y. Our experience says that in the mathematical model we can assume that x, y and $y(a_j)$ are collectives and x and y are combinable. Thus the conditional probabilities $\mathbf{P}(b_j/a_i) = \mathbf{P}_{y(a_i)}(b_j)$ are well defined. However, if the distribution of balls in the urn is not symmetric, then $\mathbf{P}(b_j/a)$ depends on a. Thus the collectives x and y are not independent.

11 Frequency and measure – theoretical viewpoints on independence

If we use the frequency approach and take combining as our starting point, then the mathematical and physical conditions for independence concern the interconnection of the two one-dimensional collectives x and y (two statistical experiments $\mathcal{S}_x$ and

$\mathcal{S}_y$) or in terms of $\mathbf{P}_z(a, b)$, the type of this two-dimensional distribution: factorization (10.1); they are not concerned with properties on each single collective (statistical experiment). On the other hand, measure-theoretical definition (4.1), (4.2) of independence does not relate in any way to two-dimensional distribution. Of course, definition (4.1), (4.2) can be considered as a generalization of the factorization rule (10.1). However, (4.1), (4.2) extends (10.1) too much. In general (4.1), (4.2) has no relation with original physical motivations of independence. We wish to consider this problem carefully. Consider the following example [171]: The label space S consists of six points $1, \ldots, 6$ with distribution p_i, $i = 1, \ldots, 6$; the event (or set) A consists of the three points $2, 3, 4$, the event C of the two points $1, 2$; the intersection $A \cap C$ is the point 2, and $\mathbf{P}(C/A) = p_2/(p_2 + p_3 + p_4)$ (due to measure-theoretical definition (2.4) or frequency definition (3.4) which is based on the operation of partition). Now the following question is asked: Under what conditions is $\mathbf{P}(C/A)$ equal to $\mathbf{P}(C)$ (or $\mathbf{P}(C \cap A) = \mathbf{P}(A)\mathbf{P}(C)$)? In our example $p_2/(p_2 + p_3 + p_4) = p_1 + p_2$? The example is so chosen that this is true for $p_i = 1/6, , i = 1, \ldots, 6$. The statement is then made that, in this case, the events A and C are independent. Let us analyze this statement.

Let us consider a set A consisting of the points $2, 3, 4$ and a set C of point 2; here $C \subset A$. Then $\mathbf{P}(2/A) = p_2/(p_2 + p_3 + p_4)$. Here $\mathbf{P}(2/A)$ certainly does not remain unchanged if we vary the set A, and certainty for no A, $\mathbf{P}(A) \neq 1$, is $\mathbf{P}(2/A)$ equal to p_2. Now, however, in order to make such an equality possible, one consider other sets, C, such that $A \cap C = \{2\}$ but $C \supset A \cap C$. Such subsets of S are, for example, $C_1 = \{1, 2\}$, $C_2 = \{2, 5\}$, $C_3 = \{1, 2, 5, 6\}$, $C_4 = \{1, 2, 5\}$. Then, for each of these $C_i, \mathbf{P}(C_i/A) = p_2/(p_2 + p_3 + p_4)$. Thus, having the choice of sets C_i one may ask whether for one or more of them, and with some given distribution, $\mathbf{P}(C/A) = \mathbf{P}(C)$. If all $p_i = 1/6$, this holds true for $C_1 = \{1, 2\}$ or for $C_2 = \{2, 5\}$ but not for C_3 or C_4. If we take $p_1 = p_5 = 1/12$, $p_2 = p_3 = p_4 = 1/6$, $p_6 = 1/3$, then the above equality holds for $C_4 = \{1, 2, 5\}$ but no longer for C_1 and C_2, and so on. It seems that the measure-theoretical definition allows the possibility to purely numerical accidents. From the physical point of view, it is not clear: What is the meaning of the statement that, for a given distribution "the events $A = \{2, 3, 4\}$ and $C = \{2, 5\}$ are independent" while "the events $\{2, 3, 4\}$ and $\{1, 2, 5\}$ are dependent" or "events $\{1, 6\}$ and $\{2, 3, 4\}$ are dependent"?

One may say that the intersection of two sets A and C has the 'property' of belonging to A and the 'property' of belonging to C (and many others). Nevertheless, the label "2" – the result of the ordinary tossing of one die – is not a two-dimensional label like "blond hair, blue eyes" or "first die 3, second die 5". Therefore a concept of independence of two 'properties' which may or may not influence each other is meaningful. However, this concept must be discussed on the basis of the procedure of combining of collectives corresponding to measurements of these properties.

Conclusion. Independence should be defined for collectives rather than for isolated events.

12 Generalization of the operation of combining

In fact, to consider the relation of combining $\overleftarrow{xy}$ we need not start with two collectives (or S-sequences) x and y. It suffices to have one collective x and a family $\{y(a)\}_{a \in L_x}$ of collectives having the same label set L_y. We denote the system $(x, \{y(a)\}_{a \in L_x})$ by the symbol U_{xy}. In this framework we can also define conditional probabilities $\mathbf{P}_{U_{xy}}(b/a) = \mathbf{P}_{y(a)}(b)$, $a \in L_x$, $b \in L_y$, and two-dimensional probability distribution

$$\mathbf{P}_{U_{xy}}(a,b) = \mathbf{P}_{U_{xy}}(b/a)\mathbf{P}_x(a), \quad a \in L_x, \quad b \in L_y.$$

In fact, a sequence $z = (z_j)$ corresponding to measurements of pears $z_j = (x_j, y_j)$ may be not defined. Such a situation is common for measurements of so called incompatible observables in quantum mechanics (i.e., observables represented by noncommuting operators), see Chapter 2. In that case it is impossible to perform a simultaneous measurement of two observables x and y (i.e., we could not form the collective z). Nevertheless, we could speak about two properties A and B of the physical system. The conditional probability $\mathbf{P}_{U_{xy}}(b/a)$ has the following meaning: if the result of a measurement of the property A is equal a, then the probability to obtain the value b of the property B is equal $\mathbf{P}_{U_{xy}}(b/a)$.

We suppose now that it is also possible to perform a measurement of the property B and, for each $B = b$, to perform a measurement of the property A. Mathematically such measurements are described by a collective y (corresponding to a measurement of B) and a system $\{x(b)\}_{b \in L_y}$ of collectives (corresponding to measurements of A under the condition $B = b$). Thus we have the system $U_{yx} = (y, \{x(b)\}_{b \in L_y})$. Here we can also define the conditional probabilities $\mathbf{P}_{U_{yx}}(b/a)$ and two-dimensional probability distribution

$$\mathbf{P}_{U_{yx}}(b,a) = \mathbf{P}_{U_{yx}}(a/b)\mathbf{P}_y(b), \quad a \in L_x, \quad b \in L_y.$$

It may be that $\mathbf{P}_{U_{xy}}(a,b) \neq \mathbf{P}_{U_{yx}}(b,a)$. In such a case the two-dimensional probability distribution $\mathbf{P}(a,b)$ corresponding to pears $(A = a, B = b)$ does not exist.

13 Comparative probability

All probability models discussed in the previous sections are called quantitative probability models. Terse the quantitative statement "$P(A) = p$" read "the probability of A is p" is the basis of these theories[11]. On the other hand, the modal or classificatory statement "A is probable" or "A is likely" seems to be most common in ordinary discourse. To formalize such an approach, we can consider, for example, a binary relation P_2 in the set $D \times D$, where D is the set of events. This relation can be read as follows: If $(A, B) \in P_2$, then A is at least as probable as B, $A \geq B$. Such a formalization gives so called *comparative probability* formalism (see, for example, T. Fine [62]).

[11]There arises natural question Why do we consider only real numbers p as quantities?

Comparative probability induces more extended class of probability models (with larger domains of application) than quantitative probability.

For example, having observed that 10 tosses of a strange coin resulted in 7 heads, we are more justified in asserting that "heads are more probable than tails" then asserting that "the probability of heads is 0.7". There exist relatively simple mathematical models in that we consider to be valid comparative probability statements that, are incompatible with any representation in quantitative theory[12].

However, my opinion is that comparative probability models have to be considered as "derivatives" of the three fundamental models (classical, frequency and ensemble). To define the binary relation P_2, we need to use one of fundamental models (or their generalizations).

Typically it is assumed that the binary relation P_2 satisfies the following axioms (see T. Fine [62], p. 17):

C0. (Nontriviality) $\Omega \succ \varnothing$, where $\varnothing$ is the null or empty set.

C1. (Comparability) $A \gtrsim B$ or $B \gtrsim A$.

C2. (Transitivity) $A \gtrsim B, B \gtrsim C \Rightarrow A \gtrsim C$.

C3. (Improbability of impossibility) $A \gtrsim \varnothing$.

C4. (Disjoint unions) $A \cap (B \cup C) = \varnothing \Rightarrow (B \gtrsim C \Leftrightarrow A \cup B \gtrsim A \cup C)$.

Axiom C1 and C2 establish that the relation $\gtrsim$ is a linear complete order. The requirement that all events be comparable is not insignificant and as been denied by some authors [62]. To illustrate the latter possibility, we consider the following example. There is an ensemble S, $|S| = N$, of coins having different centers of mass. The first coin tossing experiment (for all coins $s \in S$) gave N_1 heads and the second experiment give N_2 heads. If $N_1 > M_1 = N - N_1$, but $N_2 < M_2 = N - N_2$, then we cannot assert neither "heads are least as probable as tails" ($A \gtrsim B$) nor "tails are at least as probable as heads".

In Chapter 4 we construct a quantitative probability model (with the field of p-adic numbers as quantitative space) that induces a comparative probability model in that there exist noncomparable events. In this model the axiom (C4) is also violated.

Remark 13.1 (Subjective probability as comparative probability). It seems that the comparative interpretation is the one of possible interpretation of subjective probability. We remark that is does not sound reasonable to use the fixed ordered set (the segment $[0, 1]$) for quantitative representation of subjective probabilities. The use of $[0, 1]$ is the root of misunderstandings related to subjective probability. This implies that numbers $p \in [0, 1]$ are often interpreted as frequency of ensemble probabilities.

[12] At least if $\mathbf{R}$ is used as a "quantitative space".

2 Quantum probabilities

In this chapter various interpretations of quantum mechanics will be discussed. Comparative analysis of statistics of results of measurements for classical and quantum physical systems will be performed. We shall see that (at least formally) quantum statistics differs crucially from classical one. We present classical (but in general non-Kolmogorovean) probabilistic analysis of this situation

1 Classical and quantum probability rules

1.1 Properties of physical systems

The notion of a *property* of a physical system will play an important role in our analysis. Therefore we shall start with a discussion on this notion. We shall use not only physical, but also philosophic arguments. Those who are not interested in such discussions may start directly with probability interpretations of the wave function (see Subsection 2).

Before to start, we consider following simple examples of physical properties.

Example 1.1. Let S be an ensemble of bodies. Suppose that these bodies have one of two colors, black or white, and one of two forms, ball or cube. The color and form are properties of $s \in S$. Numerically these properties, A and B, can be described by quantities $A = 0, 1$ for black and white bodies, respectively, and $B = 0, 1$ for ball and cube, respectively.

Example 1.2. Let a be a particle (classical or quantum). The position q and momentum p of a are properties of a. Numerically these properties are described by continuous spectrum of values (by the field of real numbers $\mathbb{R}$). In what follows we shall mainly study properties which are described by discrete spectra of values. In the case of the position and momentum we can make the following discretization. Let D_q and D_p be domains in $\mathbb{R}^3$. We set $A = 1$ if $q \in D_q$ and $A = 0$ if $q \notin D_q$; $B = 1$ if $p \in D_p$ and $B = 0$ if $p \notin D_p$. The quantities A and B are properties of a.

The physical community is characterized by huge diversity of views on the notion of a property of a quantum physical system (see, for example, [58], [67], [172], [20], [21], [26–28], [31], [30], [49], [48], [54] [71], [45], [70], [50], [51], [139], [1, 4–9, 12–15, 17, 19, 22, 23, 25, 36, 37, 57, 60, 65, 97, 100, 102–105, 107, 108, 114–126] for the details).

1.2 Realism

Some scientists keep to *realism.* They assume that a property is an *objective* characteristic of a quantum system. Thus any property is a property of an object (as well as it is assumed in classical physics). Such a property has no direct relation to acts of measurement. In particular, adherents of realism (de Broglie [49], [48], Einstein [58], Bohm [31], [30], Bell [26], [28], ...) assume that both classical and quantum particles have well-defined positions and momentums or vectors of spin or polarizations.

Adherents of realism can be split in two subgroups. This splitting is based on two different answers to the following question:

Does the quantum formalism operate with *initial* values of properties (i.e., values before acts of measurements) or *final* values of properties (i.e., values after acts of measurements)?

This question is very important in quantum mechanics, because here a measurement can change crucially values of properties of physical systems. Roughly speaking a quantum systems is extremely small and a measurement device is huge (comparing to a system).

We shall call adherents of the initial values hypothesis *i-realists* and adherents of the final values hypothesis *f-realists*[1]. For example, Einstein, Podolsky and Rosen were i-realists. In their famous work [58] on the EPR-paradox they proved (at least they were sure that they did this) that a quantum particle has definite position and momentum[2], see also [57].

1.3 Empiricism

Other part of the physical community supports (following Bohr) the ideas of *empiricism,* see e.g. De Muynck's book [53] – the Bible of modern empiricism. They assume that the quantum formalism does not describe microreality such as it is.[3] Properties obtained via quantum measurements are not properties of quantum systems (not properties of objects). They are merely properties of measurement phenomena (properties

[1]For example, any measurement of the position of a quantum particle should change the localization of this particle. I-realists suppose that a quantum measurement gives the initial value of the position, $q = q_i$. Thus, although the post-measurement value of position q_f may differ essentially from q_i, nevertheless, a detector gives the pre-measurement value q_i. On the other hand, f-realists suppose that a quantum measurement gives the final value of the position, $q = q_f$.

[2]Since in quantum mechanics position and momentum are incompatible observables and hence they could not be measured simultaneously, Einstein, Podolsky and Rosen concluded that quantum mechanics is not complete. They also pointed [58] that the only alternative to incompleteness is quantum nonlocality: measurement on one quantum particle changes the state of not only this particle, but also of any particle which is entangled with it (and which can be located far away). However, Einstein, Podolsky and Rosen considered [58] quantum nonlocality as a totally absurd alternative to their i-realism. Surprisingly nowadays quantum nonlocality became well established in quantum mechanics, especially in quantum information theory.

[3]In philosophy people speak about *ontic reality:* reality as it is when nobody performs measurement, see [17] for discussion on the ontic and epistemological levels of descriptions of reality. The famous question of Einstein to Pauli: 'Does Moon exists when nobody looks at it?' - is related to this problem. Empiricists do not consider quantum mechanics as a story about ontic reality. It is an epistemological story.

of instruments and physical circumstances in that these instruments are used). In particular, adherents of empiricism (N. Bohr, P. Dirac, J. von Neumann, ...) claim that positions and momentums of quantum particles are not objective. For example, an electron has no definite position before the act of a measurement.

Some adherents of empiricism think that a property depends not only on a *measurement procedure* $\mathcal{M}$, but also on so called *preparation procedure* $\mathcal{E}$ (see Section 2 for the details) which is used to prepare quantum systems for acts of measurements. We call the adherents of the latter viewpoint *Bohrian empiricists.*

N. Bohr was father of the latter form of empiricism: "all experimental arrangement should be taken into account" as we have learned from him.

Example 1.3 (Bohrian property). Let us keep to Bohrian empiricism. As cats cannot fly, the velocity of flying v cannot be considered as an objective property of a cat. We consider the following procedure $\mathcal{E}$ preparing cats to fly: each cat is placed in front of the pilot-desk of an airplane which is equipped with a system of autopilot. By manipulations by buttons a cat can change the velocity v of the airplane. A large statistical ensemble S of cats in airplanes is prepared. The measurement procedure $\mathcal{M}$ is a measurement of average (for e.g. one hour of flying) of the velocity v of a randomly chosen airplane with a cat. The v is not a property of the cat (on the other hand, it is neither a property of the airplane). It is a property of the preparation and measurement procedures. Nevertheless, nobody would deny reality of cats and airplanes. If cats can choose only a finite set of speeds $v_1, \dots, v_k$, then the measurement $\mathcal{M}$ will produce discrete probability distribution $\mathbf{P}(v = v_i)$, $i = 1, 2, \dots, k$.

1.4 Idealism

Empiricism is sometimes identified with *idealism.* By idealists viewpoint quantum systems have no objective properties at all. This approach immediately implies a *death of realism* (not only reality of the microworld, but also reality of macroworld which is composed of microsystems). However, in principle empiricism need not imply idealism. It is very well possible to believe in the objective existence of atoms and electrons without being committed to the thesis that this reality is described by the quantum mechanical formalism. It was Bohr's position.

Of course, one should not forget that Bohr's position was time-dependent (views of everybody may change crucially during his life). Early Bohr's wrote about positions and momentums of quantum particles. He could be considered as f-realist: measurement devices change values of positions and momentums of quantum particles. However, from the very beginning he claimed that quantum mechanics is *complete.* One could not expect creation of new (more fundamental) theory which would provide an access to e.g. positions and momentums (at least simultaneously). This is a result of uncontrollable exchange of momentum between a particle and a measurement device. Late Bohr did not write more about properties of quantum systems, but solely about results of measurements. Thus he moved from the camp of f-realists and he became pure empiricist.

1.5 Comparing realism and empiricism

The realist philosophy is very attractive for scientists working in classical physics. However, we shall see that the realist viewpoint induces some problems, e.g. Einstein–Podolsky–Rosen paradox [58] in the foundations of quantum physics.

The empiricists approach seems to be free of such problems. However, empiricism is not so attractive as the philosophic basis for the investigation of reality. If we even do not keep to idealism – not deny existence of objective reality (which is independent to our observations), then by the empiricists ideology we still have to assume that the quantum formalism describes not objective reality of microworld, but reality of equipment in our laboratories.

In principle, the empiricists ideology were not be so bad, if it would not deny (following Bohr) a possibility to go *beyond quantum mechanics,* i.e., to create of a new more fundamental theory for which quantum mechanics would play the role of an approximation. However, such thoughts were totally forbidden in Bohr's kingdom. Nowadays some leading empiricists, e.g. already mentioned De Muynck as well as De Baere, see [52,53], and Ballentine [21], do not exclude a possibility of construction of a subquantum model with mentioned features. We point out that in his early works, e.g. the fundamental paper [20], Ballentine presented the realist position, but later he kept to the empiricist one [21].

1.6 Probability interpretation of a quantum state

We discuss now a probability interpretation of quantum mechanics. We may restrict our considerations to two-dimensional quantum systems. Already such quantum systems demonstrate all delicate features of this problem. Let us consider a large statistical ensemble S of quantum systems. Suppose that each system s has two properties, A and B. Let $\mathcal{H} = \mathbb{C} \times \mathbb{C}$ be the two-dimensional complex linear space with the inner product $(\cdot,\cdot) : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$. In the quantum formalism a statistical ensemble of identically prepared systems S is described by a normalized vector $\phi \in \mathcal{H}$ (i.e., $\|\phi\|^2 = (\phi,\phi) = 1$). This vector is called a (pure) *quantum state.* The properties (physical observables) A and B are described by symmetric operators $\widehat{A}$ and $\widehat{B}$, respectively. Let $e_A = (\phi_0, \phi_1)$ and $e_B = (\psi_0, \psi_1)$ be two orthonormal bases in $\mathcal{H}$ consisting of eigenvectors of the operators $\widehat{A}$ and $\widehat{B}$, respectively. The quantum state ϕ can be represented in two ways:

$$\phi = c_0\phi_0 + c_1\phi_1, \qquad \text{where} \qquad c_0, c_1 \in \mathbb{C}, \qquad |c_0|^2 + |c_1|^2 = 1; \tag{1.1}$$

$$\phi = d_0\psi_0 + d_1\psi_1, \qquad \text{where} \qquad d_0, d_1 \in \mathbb{C}, \qquad |d_0|^2 + |d_1|^2 = 1. \tag{1.2}$$

By the probability interpretation of expansion (1.1) of the quantum state ϕ the probability $\mathbf{P}(A = \alpha)(\equiv \mathbf{P}_\phi(A = \alpha))$ that $s \in S$ has the property $A = \alpha$ is equal to $|c_\alpha|^2$ (Born's rule):

$$\mathbf{P}_\phi(A = \alpha)) = |(\phi, \phi_\alpha)|^2.$$

In the same way expansion (1.2) gives that $\mathbf{P}(B = \beta)(\equiv \mathbf{P}_\phi(B = \beta)) = |d_\beta|^2$. The possibility to expand one basis with respect to other basis induces connection between the probabilities $\mathbf{P}(A = \alpha)$ and $\mathbf{P}(B = \beta)$. Let us expend the vectors ϕ_0 and ϕ_1 with respect to the basis e_B:

$$\phi_0 = u_{00}\psi_0 + u_{01}\psi_1, \quad \text{where} \quad u_{0\alpha} \in \mathbb{C}, \quad |u_{00}|^2 + |u_{01}|^2 = 1; \tag{1.3}$$

$$\phi_1 = u_{10}\psi_0 + u_{11}\psi_1, \quad \text{where} \quad u_{1\alpha} \in \mathbb{C}, \quad |u_{10}|^2 + |u_{11}|^2 = 1. \tag{1.4}$$

Thus $d_0 = c_0 u_{00} + c_1 u_{10}$, $d_1 = c_0 u_{01} + c_1 u_{11}$ and we obtain the *quantum rule* for transformation of probabilities due to transition from one orthogonal basis in complex Hilbert space to another (from the a-representation to the b-representation):

$$\mathbf{P}(B = \beta) = |c_0 u_{0\beta} + c_1 u_{1\beta}|^2, \quad \beta = 0, 1. \tag{1.5}$$

This rule is often interpreted as an exhibition of *quantum interference.* In the next section we shall compare (1.5) with the classical formula of total probability, see Chapter 1. We shall see that (1.5) can be considered as a quantum generalization of this classical probabilistic law.

1.7 Contradiction between classical and quantum formulas of total probability

On the other hand, by the probability interpretation of expansions (1.3), (1.4) we obtain that

$$\mathbf{P}(B = \beta / A = \alpha) = |u_{\alpha\beta}|^2.$$

Indeed, in (1.3), (1.4) the quantum states ϕ_α, $\alpha = 0, 1$, describe statistical ensembles $\bar{S}(A = \alpha)$ of physical systems which have the property $A = \alpha$. Therefore the expansion of the ϕ_α with respect to the basis e_B gives corresponding probabilities for $B = \beta$ (under the condition that $A = \alpha$). Hence, the formula of total probability, see Chapter 1, implies:

$$\begin{aligned} \mathbf{P}(B = \beta) &= \sum_{\alpha=0,1} \mathbf{P}(A = \alpha)\mathbf{P}(B = \beta / A = \alpha) \\ &= |c_0|^2 |u_{0\beta}|^2 + |c_1|^2 |u_{1\beta}|^2. \end{aligned} \tag{1.6}$$

Thus in general 'quantum rule' (1.5) – the quantum formula of total probability – differs from 'classical rule' (1.6) – the classical formula of total probability. The standard viewpoint to the contradiction between (1.5) and (1.6) is that it is the exhibition of ‘violation of laws of classical probability’. Such a position stimulated mystification of quantum mechanics. Typically people think: ‘If laws of quantum probability are so different from laws of classical probability, then even quantum physical laws should be very different from classical physical laws,’ see Feynman [60] for presentation of such a sort of interpretation of violation of the formula of total probability (he considered it as violation of the law of additivity of probability). Finally, people are looking for solutions such as quantum nonlocality and death of realism. However, careful analysis will show that this contradiction is a consequence of formal manipulations with Kolmogorov probabilities.

Remark 1.1. The reader can easily understand that (1.5) differs from (1.6) only if the operators $\widehat{A}$ and $\widehat{B}$ do not commute. Observables (properties) A, B which are represented by noncommuting operators $\widehat{A}$, $\widehat{B}$ are said to be *incompatible.* By the quantum formalism incompatible observables cannot be measured simultaneously. Hence, for incompatible, observables A, B the two-dimensional probability distribution $\mathbf{P}(\alpha, \beta) = \mathbf{P}(A = \alpha, B = \beta)$ cannot be defined on the basis of real physical measurements.

2 Interpretations of wave function

The wave function has many different physical interpretations.[4] We discuss the most important of them.

2.1 Ensemble realist interpretation – Einsteinian interpretation

This interpretation is often called a *statistical interpretation* (following to L. Ballentine, [20]).[5] It is assumed that ϕ describes a statistical ensemble S of identically prepared systems (a system is denoted by the symbol s, may be with indexes). Properties of any s are its objective properties. On the basis of this interpretation, it is possible to keep both to i-realism and f-realism. However, the main part of investigations is based on i-realism. As was mentioned, A. Einstein was one of creators of the statistical interpretation (in the i-realists framework). We call this interpretation Einsteinian interpretation.

Here the wave function ϕ (given by a normalized vector of complex Hilbert space) describes probabilistic distributions of properties of elements s of a statistical ensemble S. If we keep to i-realism, then these are distributions of initial properties of these elements; if we keep to f-realism, then these are distributions of final properties

[4] Of course, such a proliferation of paradigms is characteristic of a crisis in the development of quantum theory, see [51], [54] for the details.

[5] However, such a terminology might be misleading. Since in all interpretations the experimental verification is performed via statistical verification of Born’s rule, some adherents of other interpretations, e.g., the orthodox Copenhagen interpretation are sure that they use ‘statistical interpretation’.

(obtained via measurements) of these elements. In any event these are objective properties. Einsteinian interpretation is an attempt to generalize directly classical statistical mechanics to quantum systems.

2.2 Individual realists interpretation – 'Leningrad interpretation'

It is assumed that ϕ describes not the probabilistic features of a statistical ensemble S, but the *state of an individual physical system* s. Properties of s are considered as being objective. This is one of forms of the orthodox Copenhagen interpretation. Nowadays this interpretation is widely used in *quantum information theory.* By taking into account violation of Bell's inequality (Section 6), we shall see that this interpretation can be consistent only in combination with quantum nonlocality. Although assigning of the wave function to an individual quantum system (e.g. an electron) is associated with Copenhagen, Bohr had never told anything about the individual interpretation of the wave function. It seems that this interpretation was elaborated by Vladimir Fock, so it would be more natural to call it the *Leningrad interpretation.*[6]

2.3 Ensemble empiricists interpretation – Bohrian interpretation

Here it is assumed that ϕ describes the probabilistic features of a statistical ensemble of quantum systems produced by a preparation procedure. 'Properties' of quantum systems are simply results of measurements. Probabilistic distributions which are related to this statistical ensemble are merely (see Example 1.3) probabilities associated with preparation and measurement procedures.

The problem of objectivity of measured properties is either not discussed or it is claimed that they are not objective – in the sense that they cannot be assigned solely to quantum systems. The latter is Bohrian (at least late Bohrian) interpretation of quantum mechanics. We emphasize that Bohrian interpretation is very different from the interpretation which is called the orthodox Copenhagen interpretation!

As was already pointed out, early Bohr kept to the statistical f-realist interpretation. Thus early Bohr's interpretation was not so much different from the Einsteinian: one kept f-realism, another i-realism.[7]

[6] Among modern supporters of this interpretation I can mention Paul Busch, Marian Grabowski, Pekka Lahti [65], [37]. Majority of experimenters (especially working in quantum computing and quantum cryptography) use this interpretation. It is not easy for them to abandon reality of parameters that they measure. As the price of realism, they get quantum nonlocality. Of course, they are sure that this version of the Copenhagen interpretation is the original Bohr's interpretation. They are extremely surprised to hear that Bohr would never support such ideas.

[7] The main reason for their debates was Bohr's claim that quantum mechanics could not be improved: uncontrollable perturbations induced by measurements could not provide pre-measurement values of quantum variables, e.g. position and momentum. Since Bohr by himself was at the position of f-realism, it was not easy to explain why one could not determine (via development of measurement technology) pre-measurement values, in particular, to perform simultaneous measurements. I think it was one of the reasons why later Bohr took the purely empiricist position.

2.4 Växjö interpretation

As was pointed out, Bohr emphasized the completeness of quantum mechanics – impossibility to create a finer description of micro-processes. In fact, this Bohr's claim is merely of the philosophic value. There are no reasons which forbid to combine statistical (ensemble) interpretation of the wave function and interpretation of 'properties' of quantum systems as results of observations with possible existence of 'subquantum models', models with 'hidden variables.' Last years such an interpretation became known as the *Växjö interpretation of quantum mechanics*: subquantum realist statistical observational interpretation of quantum mechanics, see e.g. [6, 7, 97, 100, 102–105, 107, 108]:

(a) A wave function (as well as a density matrix) describes probabilistic features of an ensemble prepared by a preparation procedure;

(b) Self-adjoint operators describe observables. Results of observations could not be assigned solely to prepared systems: parameters of measurement devices play the crucial role.

(c) The Bohr principle of complementarity takes place.

(d) There might exist subquantum models with 'hidden variables' which induce probabilities given by the quantum formalism (but need not the values of quantum observables).

Concerning (c) we remark that it implies that one could not expect creation of a theory with hidden variables, say λ, such that incompatible quantum observables described by noncommuting operators $\widehat{A}$, $\widehat{B}$ would be represented by random variables $A(\lambda)$, $B(\lambda)$ defined on the same Kolmogorov probability space. In particular, Växjönese is not looking for the joint probability distribution for the position and momentum. It does not exist, because observations of them are based on incompatible experimental arrangements.

Concerning (d) we remark that Växjönese is looking for subquantum variables, say λ, such that so called quantum randomness can be reduced to classical randomness of such variables. However, such a reduction has essentially more complicated structure than the representation by random variables $\lambda \to A(\lambda)$, $B(\lambda)$, see e.g. [114–120, 122–126].

However, this book is not devoted to the problem of creation of subquantum models, see e.g. [114–120, 122–126] for an attempt. In this book we shall merely keep to the *statistical empiricist interpretation.* Thus we shall not take care about (c) and (d).

Finally, we remark that known so called 'no-go' theorems against hidden variables such as Bell's theorem are directed against Einsteinian interpretation, i.e., statistical realist interpretation, or the orthodox Copenhagen interpretation in the form of individual realists interpretation. The empiricist interpretation in general and the Växjö interpretation in particular do not contradict to these theorems [121].

2.5 Individual empiricists interpretation – Dirac–von Neumann interpretation

Here it is assumed that ϕ describes the state of an individual quantum system s. Many adherents of this interpretation keep to idealism and suppose that s has no objective properties at all. 'Properties' of s are solely based on specific acts of measurements. This is also a form of the orthodox Copenhagen interpretation. I can mention Dirac and von Neumann as main supporters of this interpretation. The latter developed [172] theory of *irreducible quantum randomness*, i.e. randomness which could not be reduced to ensemble randomness.

2.6 Individual interpretation and subjective probability – Fuchsian interpretation

In this book we shall not consider individual interpretations of ϕ. In particular, we shall not deal with various forms of the orthodox Copenhagen interpretation.

It seems that the use of subjective probability is the only reasonable way to give the probabilistic foundation to these interpretations. Recently Christopher Fuchs formalized this idea in the form of so called *Fuchsian interpretation of quantum mechanics.* It is clear that the Fuchsian interpretation is totally opposite to the Växjö interpretation. Therefore sometimes the Fuchsian interpretation is called anti-Växjö interpretation, see [121].

However, it would be rather strange to use such an argument as a 'measure of the personal belief' as the cornerstone of the fundamental physical theory. As it has been mentioned in Section 7, Chapter 1, at the moment the only real possibility to justify the use of subjective probabilities is to reduce them to ensemble or frequency probabilities.

Let the wave function ϕ describe the measure of our belief that, for example, the position q of an (individual) electron would be observed in a domain D. But how can this measure of our belief be found? The only way is to use our ensemble or frequency experience[8].

We recall that Jaynes [29] strongly criticized both the Copenhagen and ensemble interpretation for using objective probabilities. He was sure that all problems, including violation of Bell's inequality could be immediately eliminated by using subjective interpretation of quantum probabilities.

In this book we shall concentrate our studies on ensemble (statistical) interpretations. These interpretations can be used on the basis of ensemble and frequency probability theories. Ensemble probability theory provides the basis of the statistical interpretation in the framework of i-realism. Frequency probability theory must be used as the basis of the statistical interpretation in the f-realists framework and the ensemble empiricists interpretation.

[8] Another possibility to deal with subjective probabilities is to consider them as measures of information. However, even if probability is considered as subjective information, then it can be again reduced to ensemble or frequency probability for distribution of ideas in the brain.

2.7 Preparation and measurement procedures

Many physicists imagine a quantum measurement process as it is split in two procedures:

(1) a *preparation procedure* $\mathcal{E}$,

(2) a *measurement procedure* $\mathcal{M}$.

A *preparation procedure* $\mathcal{E}$ produces an ensemble $S \equiv S_{\mathcal{E}}$ of 'identically prepared' systems: $s \in S$.

The next step is a *measurement procedure* $\mathcal{M}$. The $\mathcal{M}$ is used for a measurement of some property B of systems prepared via $\mathcal{E}$.[9]

The values of B obtained via a measurement procedure $\mathcal{M}$ are considered (depending on a viewpoint) as B_i (objective initial values), B_f (objective final values) or Bohrian (values determined by the preparation procedure $\mathcal{E}$ and the measurement $\mathcal{M}$).

The main feature of the quantum mechanical formalism is that theoretically the probability distribution of B can be found on the basis of the purely algebraic computations via representation (1.2), Born's rule.

The statement that quantum systems are 'identically prepared' by a preparation procedure is not verifiable experimentally. It is a metaphysical proposition. What can be experimentally verified?

The fact that a preparation procedure determines the definite probability distributions for any property of a system which can be described by quantum mechanics (for any quantum observable). Thus by definition a preparation procedure induces stabilization of relative frequencies of results of observations for any quantum observable.[10]

Depending on interpretation any property B can be considered either as an objective property of a quantum system $s \in S$ – realism or as merely a property of preparation and measurement procedures – empiricism. Probabilities $\mathbf{P}(B = \beta)$ have different meanings in different approaches to quantum mechanics.

By the statistical (ensemble realists) interpretation in the i-realists framework

$$\mathbf{P}(B = \beta) = \mathbf{P}_S(B_i = \beta)$$

is the distribution of initial values B_i of the property B in the ensemble S. By the same statistical interpretation, but in the f-realists framework

$$\mathbf{P}(B = \beta) = \mathbf{P}_b(B_f = \beta)$$

[9] It is more or less commonly accepted that it is meaningless to perform a measurement without a preparation procedure. First one should determine conditions of ensemble preparation and only after this a measurement procedure can be run.

[10] Typically stabilization of relative frequencies for observations is considered as a law of nature. However, in practice we simply select preparation procedures and measurement procedures in a consistent way. If experimenters find that the principle of statistical stabilization (the law of large numbers in the Kolmogorov measure-theoretic model) is violated, they would simply consider such an experiment as badly performed (instability of parameters, nonreproducibility of results of measurements and so on).

is the distribution of final values B_f of the property B in the collective

$$b = (\beta_1, \beta_2, \dots, \beta_k, \dots)$$

induced by measurements $\mathcal{M}$ of B for systems $s \in S$. If we keep to the ensemble empiricists interpretation, then probabilities $\mathbf{P}(B = \beta)$ have also the frequency meaning.

2.8 Filters – preparation via selection

A preparation procedure $\mathcal{E}$ often consists of a source and a *filter* (a selection procedure) with respect to values of some property of emitted systems, say A (observable in the empiricist approach). It is common to consider filters which determine completely probabilistic features of the output ensemble. Thus for two different types of sources, say I_1 and I_2, probabilistic features of ensembles obtained via such a filtration, say S_1 and S_2, are indistinguishable. In the quantum formalism such a property (observable) used for filtration is represented by an operator $\widehat{A}$ having nondegenerate spectrum.

Let $A = 0, 1$ be a property of systems. Consider filters F_α corresponding to fixed values of $A : F_\alpha$ selects only systems for which $A = \alpha$. Here it is assumed that A is represented by operator in two-dimensional Hilbert space. Statistical ensembles $\bar{S}(A = \alpha), \alpha = 0, 1$, results of selection with respect to corresponding values of A are represented by pure states ϕ_α, eigenvectors of $\widehat{A}$:

$$\widehat{A}\phi_\alpha = \alpha\phi_\alpha.$$

2.9 Preparation and measurement procedures in quantum formalism

In the simplest case the formalism of quantum mechanics works in the following way. A preparation procedure $\mathcal{E}$ is represented by a pure state ϕ (in general preparation procedures could correspond to so called mixed states). For any quantum observable B represented by a self-adjoint operator $\widehat{B}$ with purely discrete nondegenerate spectrum (the latter we assume for simplicity) the probability to obtain the value β is given by Born's rule:

$$\mathbf{P}_\phi(B = \beta) = |(\phi, \psi_\beta)|^2,$$

where ψ_β is an eigenvector of $\widehat{B}$ corresponding to the eigenvalue β. In other words one should use the expansion (1.2) of ϕ with respect to the orthonormal basis consisting of eigenvectors of $\widehat{B}$. We remind that one may use various interpretations for the probability in the left-hand side of Born's rule. The well-known conflict between 'classical probability' and 'quantum probability' arises only if one tries to proceed in the Kolmogorov framework. By using the frequency (von Mises) framework we escape any problem.

Let us consider now a property (observable) $A = \alpha_1, \alpha_2$ represented by an operator $\widehat{A}$ acting in two-dimensional Hilbert space. Consider two preparation procedures based on corresponding filtration, $F_{\alpha_i}, i = 1, 2$. They produce ensembles $\bar{S}(A = \alpha_i)$.

We now consider another property ('measure another observable') $B = \beta_1, \beta_2$ for elements of the ensemble $\bar{S}(A = \alpha_i)$ (for the fixed i). We obtain probabilities $\mathbf{P}_{\bar{S}(A=\alpha_i)}(B = \beta_j)$ for $j = 1, 2$.

The ensembles $\bar{S}(A = \alpha_i)$ are represented by eigenvectors ϕ_{α_i} of $\widehat{A}$. The property (observable) B is represented by an operator $\widehat{B}$ having eigenvectors ψ_{β_j}. Born's rule gives prediction for probabilities:

$$\mathbf{P}_{\bar{S}(A=\alpha_i)}(B = \beta_j) \equiv \mathbf{P}_{\phi_{\alpha_i}}(B = \beta_j) = |(\phi_{\alpha_i}, \psi_{\beta_j})|^2.$$

As was pointed out, the problem arises if one interpret these probabilities in the Kolmogorov framework as conditional probabilities $\mathbf{P}(B = \beta_j / A = \alpha_i)$ (given by Bayes' rule).

3 'Contradiction' between quantum and classical probability calculi

We now come back to our comparative analysis of the conventional formula of total probability, see Chapter 1, with quantum formula of total probability, see Section 1.6, formula (1.5). In this section we demonstrate that:

R1). One of possible roots of the contradiction between quantum rule (1.5) and classical rule (1.6) is the identification of conditional probabilities $\mathbf{P}(B = \beta / A = \alpha)$ for an ensemble S of quantum systems represented by the quantum state ϕ, see (1.2), with probabilities $\mathbf{P}_{\phi_\alpha}(B = \beta)$ for ensembles $\bar{S}(A = \alpha_i)$ represented by the quantum states ϕ_α, $\alpha = 0, 1$, see (1.4).

R2). Another possible root is a possibility that conditional probabilities $\mathbf{P}(B = \beta / A = \alpha)$ do not exist at all.

In the R1-case one simply put probabilities related to three different preparations (of ensembles S and $\bar{S}(A = \alpha_i)$) into a single probability space. And what is the reason for this?

In the R2-case one considers the possibility that

(a) conditional probabilities given by any filtration are well defined;

(b) but micro-parameters may fluctuate in such a way that conditional probabilities $\mathbf{P}(B = \beta / A = \alpha)$ for an ensemble S do not exist.[11]

As we have already discussed, A. N. Kolmogorov eliminated the concrete structures of probabilistic spaces from his model.[12] People manipulate with rather mystical symbol $\mathbf{P}$ of abstract probability which was not related to any concrete statistical ensemble

[11] These are ontic probabilities. We are never able to find them on the basis of observational results.

[12] In his book [133] he emphasized the role of experimental context in determination of an appropriative probability space. Unfortunately, we did not make this point sufficiently clear for ordinary users of his axiomatics.

S or collective x. It seems that in quantum physics Kolmogorov’s abstract probabilities are often used formally. This implies identification of (conditional) probabilities which are related to different ensembles or collectives. However, such probabilities need not be equal. On the other hand, Kolmogorov’s definition of conditional probabilities via Bayes’ formula induces the opinion that (at least in ‘classical probability theory’) existence of probabilities $\mathbf{P}(E_j),\ j = 1, 2$, with $\mathbf{P}(E_2) > 0$ must automatically imply existence of conditional probability $\mathbf{P}(E_1/E_2)$. However, such an assumption need not be true for all statistical phenomena. In the ensemble probability framework (see Section 2, Chapter 1) we need not assume that the family of all events $F(\pi_S)$ (determined by the family π_S of properties of elements $s \in S$) is an algebra. Thus $E_1, E_2 \in F(\pi_S)$ need not imply $E_1 \cap E_2 \in F(\pi_S)$. Here conditional probability $\mathbf{P}(E_1/E_2)$ could not be defined by Bayes’ formula.

3.1 Ensemble approach: disturbance effects

Let us try to keep even to i-realism. Here both properties A and B are objective properties of elements of the statistical ensemble S represented by the quantum state ϕ. Measurements give initial values of these properties, $A \equiv A_i$, $B \equiv B_i$. We can consider sub-ensembles $S(A = \alpha)$ and $S(B = \beta)$, $\alpha, \beta = 0, 1$, of S which consist of elements s having the properties $A = \alpha$ and $B = \beta$, respectively.

By the ensemble definition of probability

$$\mathbf{P}_S(A = \alpha) = |S(A = \alpha)|/|S|$$

and

$$\mathbf{P}_S(B = \beta) = |S(B = \beta)|/|S|$$

and by the ensemble definition of the conditional probability

$$\mathbf{P}_S(B = \beta/A = \alpha) = \mathbf{P}_{S(A=\alpha)}(B = \beta) = \frac{|S(A = \alpha) \cap S(B = \beta)|}{|S(A = \alpha)|}\,. \tag{3.1}$$

We can use Bayes’ formula (and the formula of total probability) for these probabilities. It seems that we should obtain the above contradiction. However, there is one delicate point.

In general we cannot assume that the conditional probabilities $\mathbf{P}_S(B = \beta/A = \alpha) = \mathbf{P}_{S(A=\alpha)}(B = \beta),\ \alpha, \beta = 0, 1$, can be obtained from expansions (1.3), (1.4). We cannot identify the sub-ensembles $S(A = \alpha)$, $S(B = \beta)$ of S with ensembles $\bar{S}(A = \alpha)$, $\bar{S}(B = \beta)$ which are described by the quantum states ϕ_α and ϕ_β, respectively.

There are different preparation procedures $\mathcal{E}$, $\mathcal{E}(A = \alpha)$, $\mathcal{E}(B = \beta)$, $\alpha, \beta = 0, 1$. They produce ensembles S, $\bar{S}(A = \alpha)$, $\bar{S}(B = \beta)$, respectively, which are represented by quantum states ϕ, ϕ_α, ψ_β, respectively. We cannot identify the sub-ensembles $S(A = \alpha)$, $S(B = \beta)$ of S with ensembles $\bar{S}(A = \alpha)$, $\bar{S}(B = \beta)$. For instance, the preparation procedures $\mathcal{E}(A = \alpha), \alpha = 0, 1$, can be realized as filters F_α: only systems with property property $A = \alpha$ can pass F_α. However, such a filtration changes the value of the property B for s.

Thus in general we have:

$$\mathbf{P}_S(B = \beta/A = \alpha) = \mathbf{P}_{S(A=\alpha)}(B = \beta) \neq \mathbf{P}_{\bar{S}(A=\alpha)}(B = \beta).$$

Moreover, $\mathbf{P}_S(B = \beta/A = \alpha)$ may be not well defined. Despite of the fact that $A, B \in \pi_S$, it may be that the set $\{A = \alpha\} \cap \{B = \beta\}$ is not described by any property $C \in \pi_S$.

Such a phenomenon is not essentially nonclassical. An example of the selection of a sub-ensemble on the basis of one fixed property which can change the probability distribution of other property can be easily found for classical systems. We can illustrate this problem by Example 1.1. Let S be an ensemble of bodies having different colours, $A = 0, 1$, and forms, $B = 0, 1$. There are sub-ensembles $S(A = \alpha)$, $\alpha = 0, 1$, corresponding to fixed colours and $S(B = \beta)$, $\beta = 0, 1$, corresponding to different forms. To extract elements of the S having the fixed colour α, we use a device D_α which changes randomly the form of a body (some bodies of the form $B = 0$ are transformed in bodies of the form $B = 1$ and vice versa). By this procedure we obtain new ensembles $\bar{S}(A = \alpha)$, $\alpha = 0, 1$. Of course, the distributions of B in $\bar{S}(A = \alpha)$, $\alpha = 0, 1$, may differ from the initial distributions of B in the ensembles $S(A = \alpha)$, $\alpha = 0, 1$.

Conclusion. The contradiction between ‘quantum and classical probabilistic rules’ (1.5), (1.6) need not be regarded to the specific (‘nonclassical’) behaviours of statistical ensembles of quantum systems. The possible root of this contradiction is the formal use of Kolmogorov’s measure-theoretical approach in that we do not control the relation between probabilities and statistical ensembles. The identification of probabilities corresponding to different statistical ensembles implies (in general) the use of wrong values, $|u_{\alpha\beta}|^2$, for conditional probabilities $\mathbf{P}(B = \beta/A = \alpha)$ (which, in fact, must be calculated on the basis of (3.1)). This induces the illusion of the violation of Bayes’ formula (and the formula of total probability) in the quantum formalism.

3.2 Ensemble approach: no conditional probabilities

It must be pointed out that any quantum state ϕ represents not a finite statistical ensemble consisting of N quantum systems, but an infinite ideal statistical ensemble S. For any property C, probabilities $\mathbf{P}_\phi(C = \gamma)$ are, in fact, probabilities $\mathbf{P}_S(C = \gamma)$ with respect to this infinite ensemble S. Of course, in each concrete run $R = \{1, 2, \ldots, N\}$ of experiments we can obtain only a finite statistical ensemble $S^{(R)}$ and ‘experimental ensemble probabilities’ are probabilities (relative frequencies) with respect to $S^{(R)}$:

$$\mathbf{P}_\phi^{\exp}(C = \gamma) = \mathbf{P}_{S^{(R)}}(C = \gamma) = \frac{|\{s \in S^{(R)} : C = \gamma\}|}{|S^{(R)}|}\ .$$

For different runs R and R', these probabilities are rather different. The main feature of quantum systems (as many other physical systems) is that these probabilities have the property of the statistical stabilization, namely, $\lim_{|R|\to\infty} \mathbf{P}_{S(R)}(C = \gamma) = |d_\gamma|^2$, where d_γ are coefficients in the expansion of ϕ with respect to the system of eigenvectors of the operator $\widehat{C}$. These limiting probabilities are probabilities with respect to the infinite ideal ensemble S:

$$|d_\gamma|^2 = \mathbf{P}_\phi(C = \gamma) = \frac{|\{s \in S : C = \gamma\}|}{|S|} . \tag{3.2}$$

However, as the field of real numbers $\mathbb{R}$ does not contain actual infinities, formula (3.2) has no meaning in the framework of real analysis. Instead of (3.2), mathematicians (and, as a consequence, physicists) use the measure-theoretical approach. However, (despite of the common opinion) this approach cannot be used as a justification of the ensemble probability theory even in the case of a countable ensemble $S = \{s_1, s_2, \ldots, s_k, \ldots\}$. For example, let us try to define the uniform σ-additive probability on S : $\mathbf{P}(\{s_1\}) = \mathbf{P}(\{s_2\}) = \cdots = \mathbf{P}(\{s_k\}) = \cdots \neq 0$. Then $\mathbf{P}(S) = \sum_{j=1}^{\infty} \mathbf{P}(\{s_j\}) = \infty$.

Such 'pathological' properties of the field of real numbers (the absence of actual infinities) is one the reasons to use the ensemble-frequency interpretation instead of the purely ensemble interpretation. By the ensemble-frequency interpretation a Kolmogorov probability space is based not on the ensemble S of quantum systems, but (roughly speaking) on the ensemble Ω of all possible (ideal infinite) runs of experiments. However, this approach to the definition of a probability space was, in fact, never used by physicists. They typically assume that a Kolmogorov probability space gives the mathematical representation of an ensemble of quantum systems. One of the main reasons to do so and to reject the ensemble-frequency approach is the impossibility to construct a probability measure on the space of all runs for measurements corresponding incompatible properties.

One of problems of the Kolmogorov axiomatics is that probability $\mathbf{P}$ must be closed (defined on the σ-algebra or at least algebra of sets). Thus if probabilities of events $\{A = \alpha\}$ and $\{B = \beta\}$ are well defined, then automatically probability of the event $\{A = \alpha\} \cap \{B = \beta\}$ must be well defined. Hence the conditional probability $\mathbf{P}(B = \beta/A = \alpha)$ must be well defined in Kolmogorov's framework. However, such conditional probabilities are not observed. In principle, there are no reasons to assume that they are even well defined for each quantum state ϕ. For example, why we cannot assume that the ensemble $S = \mathbb{N}$ and 'probability' $\mathbf{P} = \delta$, where δ is the density of natural numbers? In such situations there is no Bayes' formula at all and the problem of difference between 'quantum and classical probability rules' is meaningless. Of course, these are non-Boolean models. However, this non-Boolean structure of probabilities has no special nonclassical features.

Detailed analysis of the problem of existence of conditional probabilities will be presented in Section 4.

3.3 Frequency probability viewpoint to quantum probabilistic rule

The frequency approach to probability gives more freedom than the ensemble approach. Here we need not assume that properties of quantum systems are objective[13]. Thus in principle we can consider various combinations of objective and nonobjective properties of quantum systems.

We start with the general scheme in that we do not suppose that any of properties $A(=0,1)$ and $B(=0,1)$ has an objective character. Let $\phi \in \mathcal{H}$ be a quantum state. As in Section 1, we consider the two-dimensional Hilbert space $\mathcal{H}$. Properties A and B are represented by symmetric operators $\widehat{A}$ and $\widehat{B}$; $e_A = (\phi_0, \phi_1)$ and $e_B = (\psi_0, \psi_1)$ are orthonormal bases in $\mathcal{H}$ consisting of eigenvectors of operators $\widehat{A}$ and $\widehat{B}$, respectively. Thus we have: $\phi = c_0\phi_0 + c_1\phi_1 = d_0\psi_0 + d_1\psi_1$, where $c_0, c_1, d_0, d_1 \in \mathbb{C}$, and $|c_0|^2 + |c_1|^2 = 1$, $|d_0|^2 + |d_1|^2 = 1$.

A quantum state ϕ represents an *ideal infinite ensemble* S of quantum systems. This ensemble is characterized in the following way: frequency probability distribution of any property $C(=0,1)$ is given by squares of absolute values of coefficients in the expansion of ϕ with respect to the system of eigenvalues of the operator $\widehat{C}$ representing C. In particular, we have:

(1) Any series of measurements $\mathcal{N}$ of the property A for elements $s_j \in S$, $j = 1, 2, \ldots,$ induces a collective

$$a_{\mathcal{N}} = (\alpha_1, \alpha_2, \ldots, \alpha_k, \ldots), \quad \alpha_j = 0, 1,$$

such that frequency probabilities $\mathbf{P}_{a_{\mathcal{N}}}(\alpha) = \lim_{k\to\infty} \nu_k(\alpha; a_{\mathcal{N}})$ are equal to $|c_\alpha|^2$, $\alpha = 0, 1$. Here, as usual, $\nu_k(\alpha; a_{\mathcal{N}}) = n_k(\alpha; a_{\mathcal{N}})/k$ is the relative frequency of realizations of the value $A = \alpha$.

(2) Any series of measurements $\mathcal{M}$ of the property B for elements $s_j \in S$, $j = 1, 2, \ldots,$ induces a collective

$$b_{\mathcal{M}} = (\beta_1, \beta_2, \ldots, \beta_k, \ldots), \quad \beta_j = 0, 1, \tag{3.3}$$

such that frequency probabilities $\mathbf{P}_{b_{\mathcal{M}}}(\beta) = \lim_{k\to\infty} \nu_k(\beta; b_{\mathcal{M}})$ are equal to $|d_\alpha|^2$, $\alpha = 0, 1$. Here, as usual, $\nu_k(\beta; b_{\mathcal{M}}) = n_k(\beta; b_{\mathcal{M}})/k$ is the relative frequency of realizations of the value $B = \beta$.

Remark 3.1. As we have already discussed, infinite statistical ensembles could not arise in any real physical experiment. We always operate with finite statistical ensembles (samples of finite lengths) S_N which are prepared by some preparation procedure $\mathcal{E}$ after N steps. However, a quantum state $\phi = \phi_{\mathcal{E}}$ (corresponding to this preparation procedure) cannot be considered as a representation of any of these finite ensembles S_N. A measurement for elements of S_N gives only a relative frequency, but not a

[13] We recall (see Section 2) that nonobjective character of some properties (creation of these properties in the process of a measurement) does not imply 'essentially quantum features' of systems.

probability. These frequencies may fluctuate when N is changed. Only asymptotically frequencies in S_N approach probabilities in S. Different properties may have different behaviour of fluctuations of frequencies before stabilization[14].

In the same way quantum states $\phi_\alpha = u_{\alpha 0}\psi_0 + u_{\alpha 1}\psi_1$, $\alpha = 0, 1$ describe some ideal infinite statistical ensembles $\bar{S}(A = \alpha)$ of quantum systems. In particular, these ensembles have the following frequency properties:

(1α) Any series of measurements $\mathcal{N}$ of the property A for elements $s_j \in \bar{S}(A = \alpha)$, $j = 1, 2, \ldots$, induces a collective

$$w_{\alpha\mathcal{N}} = (\theta_1, \theta_2, \ldots, \theta_k, \ldots), \quad \theta_j = 0, 1,$$

such that frequency probabilities $\mathbf{P}_{w_{\alpha\mathcal{N}}}(\alpha) = \lim_{k\to\infty} \nu_k(\alpha; w_{\alpha\mathcal{N}}) = 1$ and $\mathbf{P}_{w_{\alpha\mathcal{N}}}(1-\alpha) = \lim_{k\to\infty} \nu_k(1-\alpha; w_{\alpha\mathcal{N}}) = 0$, $\alpha = 0, 1$.

(2α) Any series of measurements $\mathcal{M}$ of the property B for elements $s_j \in \bar{S}(A = \alpha)$, $j = 1, 2, \ldots$, induces a collective

$$b_{\alpha\mathcal{M}} = (\lambda_1, \lambda_2, \ldots, \lambda_k, \ldots), \quad \lambda_j = 0, 1, \tag{3.4}$$

such that frequency probabilities $\mathbf{P}_{b_{\alpha\mathcal{M}}}(\beta) = \lim_{k\to\infty} \nu_k(\beta; b_{\alpha\mathcal{M}})$ are equal to $|u_{\alpha\beta}|^2$, $\beta = 0, 1$.

As we have already seen, by quantum calculus $\mathbf{P}_b(\beta) = |d_\beta|^2 = |c_0 u_{0\beta} + c_1 u_{1\beta}|^2$, $\beta = 0, 1$. As in the case of ensemble probabilities, if we forget about dependence of probabilities on collectives and identify in the formula of total probability conditional probabilities $\mathbf{P}(\beta/\alpha) \equiv \mathbf{P}(B = \beta/A = \alpha)$, $\alpha, \beta = 0, 1$, with probabilities $\mathbf{P}_{b_{\alpha\mathcal{M}}}(\beta)$ (of the property $B = \beta$ in the collectives $b_{\alpha\mathcal{M}}, \alpha = 0, 1$), then we arrive to the contradiction (between classical and quantum probability calculi). However, in the frequency approach there are even less reasons for such identification of probabilities than in the ensemble approach. Moreover, there immediately arises the problem of the correct frequency definition of conditional probabilities $\mathbf{P}(\beta/\alpha)$.

These conditional probabilities can be defined on the basis of the operation of combining of collectives, see Section 9, Chapter 1. However, it is not clear which collectives

$$\bar{a} = (\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots), \qquad A = \alpha_j = 0, 1,$$

and

$$\bar{b} = (\beta_1, \beta_2, \ldots, \beta_n, \ldots), \qquad B = \beta_j = 0, 1,$$

[14]I do not agree with the viewpoint of A. N. Kolmogorov: "The frequency concept based on the notion of *limiting frequency* as the number of trials increases to infinity does not contribute anything to substantiate the application of the results of probability theory to real practical problems where we always have to deal with a finite number of trials."

we have to combine to obtain conditional probabilities $\mathbf{P}(\beta/\alpha)$ (which must be equal to probabilities $\mathbf{P}_{b_{\alpha\mathcal{M}}}(\beta)$ for observed collectives (3.4)). In any case the direct combination of observed collectives $a_{\mathcal{N}}$ and $b_{\mathcal{M}}$ would not produce such conditional probabilities, because these collectives are independent and here $\mathbf{P}(\beta/\alpha) = \mathbf{P}_{b_{\mathcal{M}}}(\beta) \neq \mathbf{P}_{b_{\alpha\mathcal{M}}}(\beta)$.

Moreover, we have shown (see Section 9, Chapter 1), that (in the case of strictly positive probabilities) the condition of combining of $\bar{a}$ and $\bar{b}$ is equivalent to the existence of a two-dimensional collective

$$z = (z_1, z_2, \dots, z_n, \dots), \quad z_j = (\alpha_j, \beta_j),$$

where α_j and β_j are elements of $\bar{a}$ and $\bar{b}$, respectively. Hence the two-dimensional probability distribution

$$\mathbf{P}_z(\alpha, \beta) \equiv \mathbf{P}_z(A = \alpha, B = \beta) = \lim_{k\to\infty} \nu_k((\alpha, \beta); z)$$

must be well defined. Here $\nu_k((\alpha, \beta); z) = n_k((\alpha, \beta); z)/k$, $\alpha, \beta = 0, 1$, are relative frequencies of the realization of two-dimensional labels $\gamma = (\alpha, \beta)$ in the collective z. We recall that the two-dimensional probability distribution $\mathbf{P}_z(\alpha, \beta)$ and conditional probabilities are connected as $\mathbf{P}_z(\alpha, \beta) = \mathbf{P}_z(\beta/\alpha)\mathbf{P}_{\bar{a}}(\alpha)$. Thus if we assume that probabilities $\mathbf{P}_{\bar{a}}(\alpha)$ and conditional probabilities $\mathbf{P}_z(\beta/\alpha)$ are obtained on the basis of observed quantities, then we must assume that the two-dimensional probability distribution $\mathbf{P}_z(\alpha, \beta)$ can be also obtained on the basis of observed quantities. Hence we have to assume the possibility of the simultaneous measurement of the pair $(A = \alpha, B = \beta)$. However, if the properties A and B are incompatible (represented by noncommuting operators $\widehat{A}, \widehat{B}$), then the existence of the simultaneous distribution for $(A = \alpha, B = \beta)$ contradicts to the quantum formalism.

Conclusion. In the frequency probability framework it seems to be impossible to define conditional probabilities $\mathbf{P}_z(B = \beta/A = \alpha)$ on the basis of combining of collectives generated by some observations of incompatible properties A and B. Therefore the formula of total probability cannot be used for frequency physical probabilities.

3.4 Kolmogorov formalism and quantum measurements

We consider now Kolmogorov's (ensemble-frequency) interpretation of 'quantum probabilities'. If we use the abstract measure-theoretical formalism, then we might identify some probabilities related to different probability spaces. This imply the contradiction between 'classical' and 'quantum' probabilities. In fact, different preparation procedures $\mathcal{E}$ are described by different probability spaces $(\Omega_{\mathcal{E}}, \mathcal{F}_{\mathcal{E}}, \mathbf{P}_{\mathcal{E}})$.

The quantum state $\phi = c_0\phi_0 + c_1\phi_1$, see (1.1), which describes the ensemble S (consisting of a statistical mixture of quantum systems with property $A = \alpha = 0, 1$ with probabilities $\mathbf{P}_\phi(A = \alpha) = |c_\alpha|^2$) is prepared via a preparation procedure $\mathcal{E}$. It is described by a Kolmogorov probability space $\mathcal{P} = (\Omega_{\mathcal{E}}, \mathcal{F}_{\mathcal{E}}, \mathbf{P}_{\mathcal{E}})$. The states ϕ_α,

$\alpha = 0, 1$, which describe ensembles $\bar{S}(A = \alpha)$ (consisting of quantum systems with the definite value α of the property A) are prepared via other preparation procedures $\mathcal{E}_\alpha$ (filters with respect to $A = \alpha$). These states must be described by other Kolmogorov probability spaces $(\Omega_{\mathcal{E}_\alpha}, \mathcal{F}_{\mathcal{E}_\alpha}, \mathbf{P}_{\mathcal{E}_\alpha})$.

Typically physicists apply the formula of the total probability by mixing conditional probabilities $\mathbf{P}_{\mathcal{E}}(\beta/\alpha)$ with respect to the probability space $\mathcal{P}$ with probabilities $\mathbf{P}_{\mathcal{E}_\alpha}(\beta)$, $\alpha = 0, 1$. Such a manipulation induces the contradiction between ‘classical’ and ‘quantum’ probabilities.

3.5 Interference

Here we do not present statistical models which might explain the interference phenomena on the basis of the corpuscular picture (see [91], [89], [92]). We want just to illustrate our analysis of the notion of probability in quantum formalism by the example of the two slit experiment. This is the simplest experiment for demonstrating interference of light. There is a point source of light O and two screens L and L'. The screen L contains two slits h_0 and h_1. Light passes through S (through slits) and finally reaches the screen L' where the interference fringes are observed. The wave explanation of the existence of interference fringes is well known: the light reaching L' can travel by one of two routes–either through h_0 or through h_1; but the distances travelled by lights waves following these two paths are not equal and the light waves do not generally arrive at the screen ‘in step’ with each other.

On the other hand, O is a source of quantum particles, photons. To exclude the interaction between photons in a beam, we perform the experiment with very weak light, so that at any time there is only one photon in the region between O and L. The screen L' is replaced by a photographic plate or film (also denoted by L'). Individual spots appear on L' more or less chaotically. However, there appear standard interference fringes for a sufficiently long exposure. By this experiment we can compute the probability distribution of spots on L'. Here the property $A = 0, 1$, is given by a slit which is passed by a photon. The property B is obtained by the discretization of a measurement of the position on the screen L'. Let D be a domain on the plane L'. We set $B = 1$ if a photon is observed in D and $B = 0$ if a photon is observed in $L' \setminus D$. The formal application of the formula of total probability gives that $\mathbf{P}(B = \beta) = \mathbf{P}(A = 0)\mathbf{P}(B = \beta/A = \alpha) + \mathbf{P}(A = 1)\mathbf{P}(B = \beta/A = 1)$. However, experimental data demonstrates the violation of this equality. Mainly physicists interpret this violation as ‘nonclassical behaviour’ of photons. They claim that a photon does not pass one fixed slit.

Our ensemble analysis of the quantum formalism implies that we have to consider three ensembles:

(1) S consists of all particles that pass through the screen L when both slits are open;

(2) $\bar{S}_0$ consists of all particles that pass through the screen L when the slit h_0 is open and the slit h_1 is closed;

(3) $\bar{S}_1$ consists of all particles that pass through the screen L when the slit h_1 is open and the slit h_0 is closed.

These ensembles of particles are represented by quantum states ϕ and ϕ_0, ϕ_1, respectively. In fact, probabilities which physicists use in the formula of the total probability for the two slit experiment are related to different ensembles of particles:

$$\mathbf{P}(A = \alpha) = \mathbf{P}_S(A = \alpha)$$

and

$$\mathbf{P}(B = \beta/A = \alpha) = \mathbf{P}_{\bar{S}_\alpha}(B = \beta).$$

Of course, the formula of total probability must hold true if instead of probabilities $\mathbf{P}_{\bar{S}_\alpha}(B = \beta)$ we would use probabilities $\mathbf{P}_S(B = \beta/A = \alpha)$. However, to find latter probabilities, we have to use sub-ensembles S_α, $\alpha = 0, 1$, of the ensemble S consisting of particles that pass slits h_α. These sub-ensembles could not be found without to disturb the property B (because to find a slit, we have to perform the additional measurement, which, of course, change the distribution of B). Therefore it is insensible to discuss experimental verification of the formula of total probability for the two slit experiment.

In the following frequency analysis we shall use the framework of preparation/measurement. We consider preparation procedures $\mathcal{E}$, $\mathcal{E}_0$, $\mathcal{E}_1$ corresponding to the following configurations of open slits: h_0 and h_1 are open, h_0 is open and h_1 is closed, h_1 is open and h_0 is closed. The measurement procedure $\mathcal{M}$ is a measurement of the position B on the screen L'. Our frequency analysis of the quantum formalism implies that we have to consider three different collectives $b_\mathcal{M}$, $b_{0\mathcal{M}}$, $b_{1\mathcal{M}}$ which are obtained by the measurement $\mathcal{M}$ for particles prepared by $\mathcal{E}$, $\mathcal{E}_0$, $\mathcal{E}_1$, respectively. There are no reasons to identify probabilities $\mathbf{P}(B = \beta/A = \alpha)$ with frequency probabilities in the collectives $b_{\alpha\mathcal{M}}$, $\alpha = 0, 1$ (but typically they are identified). Moreover, probabilities $\mathbf{P}(B = \beta/A = \alpha)$ may be not well defined in the standard frequency framework.

3.6 Non-ergodic interpretation of quantum mechanics

We discuss now another delicate problem in the probabilistic foundations of quantum mechanics. As it has been pointed out, Kolmogorov's (ensemble-frequency) interpretation of probability implies identification of the ensemble and frequency approaches to probability. As a consequence, it is always assumed that frequency probabilities in the collective $b_\mathcal{M}$, see (3.3), can be identified with probabilities with respect to the ensemble $\mathcal{S}$ of pairs (a, u), where a is a quantum particle and u is an equipment which is used to measure the property B of a.[15] However, this postulate has never been tested experimentally[16]. Brazilian physicist N. Buonumano proposed a *non-ergodic inter-*

[15] Despite of the fact that in all real experiments this collective is generated by the long chain of successive experiments with the same equipment u_{fix}.

[16] The group of H. Rauch at Atominstitute in Vienna did some indirect experiments in this direction, see [156] (the most interesting experiment was performed by J. Summhammer, [161]).

pretation of quantum mechanics, [35]. By this interpretation frequency probabilities need not coincide with $\mathcal{S}$-ensemble probabilities. This imply that in principle trials need not be independent (see [91], [89]). Thus there may be correlations between x_i and x_j, $i \neq j$, in x (of course, in this case x would not be Mises' collective).

4 Probabilities with respect to objective conditions

The formalism of quantum mechanics implies that it is impossible to perform experiments for a simultaneous measurement $(A = \alpha, B = \beta)$ of two incompatible properties A and B of a quantum system[17]. Therefore two-dimensional probabilities $\mathbf{P}(A = \alpha, B = \beta)$ (or equivalently conditional probabilities $\mathbf{P}(B = \beta/A = \alpha)$ for $\mathbf{P}(A = \alpha) > 0$) could not be found on the experimental basis.

However, it is still sensible to discuss the existence of such probabilities if it is supposed that properties A, B are objective. Of course, the simultaneous existence of two objective properties does not imply automatically the possibility to perform a simultaneous measurement of these properties.

We shall study the most general situation. It is supposed that B is some observed property (it may be objective or created by the act of a measurement) and A is some objective property which cannot be observed simultaneously with B. We shall study the problem of existence of conditional probabilities $\mathbf{P}(B = \beta/A = \alpha)$ in the frequency and ensemble frameworks. The present scheme covers different approaches to properties of quantum systems:

(i) We may keep to i-realism. Here the observed value of B coincides with its initial value. Thus we study probabilities $\mathbf{P}(B_i = \beta/A_i = \alpha)$ where A_i and B_i are initial values of properties.

(f) We may keep to f-realism. Here the observed value B_f of the property B can differ from its initial value B_i. Thus we study probabilities $\mathbf{P}(B_f = \beta/A_i = \alpha)$.

(e) We may keep to empiricism (or even idealism). Here the property B is created by the act of a measurement. It is meaningless to speak about values of B before a measurement.

In fact, f-realism seems to be the most attractive. Here we can use the preparation/measurement approach. Some preparation procedure $\mathcal{E}$ produces a statistical ensemble S of particles with the definite probability distribution for the property A (which is typically supposed to be objective). Then a measurement $\mathcal{M}$ of other property B is performed for particles $s \in S$. This measurement gives final values B_f of the property B.

[17] In fact, on the physical level incompatible properties are defined as properties for which it is impossible to perform a simultaneous measurements. The representation of such properties by noncommuting operators in a Hilbert space of quantum states is only a consequence of such impossibility of simultaneous measurements.

The main result of our considerations is that conditional probabilities $\mathbf{P}(B = \beta/A = \alpha)$ and two-dimensional probabilities $\mathbf{P}(A = \alpha, B = \beta)$ may be not exist (both in frequency and ensemble approaches). This result seems to be rather strange from the Kolmogorov probability viewpoint.

Our models in that $\mathbf{P}(B = \beta/A = \alpha)$ do not exist give examples of (non-Kolmogorovean) probabilistic models without conditional probabilities (probabilities are well defined, but conditional probabilities could not be defined). Here probability is not closed. It is defined on the system of events which do not form a set algebra.

4.1 Frequency probabilities

To define frequency conditional probabilities $\mathbf{P}(B = \beta/A = \alpha)$, we must combine two collectives a and b corresponding to values of A and B. We choose a collective $b = b_{\mathcal{M}}$, see (3.3), induced by a measurement $\mathcal{M}$ of the property B as one of collectives for combining. As the property A is objective, then each element $s_j \in S$ has this property. Hence parallel to the constructing of the collective $b_{\mathcal{M}}$ we can imagine the process of construction of a 'hidden sequence'

$$a_h = (\alpha_1, \alpha_2, \dots, \alpha_n, \dots), \quad \alpha_j = 0, 1,$$

where $\alpha_j = \alpha$ if the property A has the value α for a quantum system s_j. Suppose that the sequence a_h is a collective. We choose $a = a_h$ as another collective for combining. Suppose that probability distributions $\mathbf{P}_b(\beta)$ and $\mathbf{P}_a(\alpha)$ are strictly positive. Finally we suppose that collectives a and b are combinable. Therefore the two-dimensional sequence

$$z = (z_1, z_2, \dots, z_j, \dots), \quad z_j = (\alpha_j, \beta_j),$$

corresponding to these collectives is also a collective. Hence frequency conditional probabilities $\mathbf{P}_z(\beta/\alpha) \equiv \mathbf{P}_z(B = \beta/A = \alpha)$ are defined via the standard scheme:

Suppose that there are $M_k(\alpha; z)$ elements with the first coordinate α among the first k elements of z, and there are $n_k(\beta/\alpha; z)$ elements with the first coordinate β among these $M_k(\alpha; z)$ elements. We introduce the relative frequencies:

$$\nu_k(\alpha; z) = \frac{M_k(\alpha; z)}{k} \quad \text{and} \quad \nu_k(\beta/\alpha; z) = \frac{n_k(\beta/\alpha; z)}{M_k(\alpha; z)}.$$

Conditional probability is defined as

$$\mathbf{P}_z(\beta/\alpha) = \lim_{k\to\infty} \nu_k(\beta/\alpha; z).$$

This definition can be reformulated in the following way. For each fixed $\alpha = 0, 1$, we choose a subsequence

$$b_\alpha = (\beta_1, \beta_2, \dots, \beta_n, \dots), \quad \beta_j = 0, 1,$$

of the sequence z consisting of second coordinates of $z_j = (\alpha_j, \beta_j)$ with $\alpha_j = \alpha$. Then $\mathbf{P}_z(\beta/\alpha) = \lim_{k\to\infty} \nu_k(\beta; b_\alpha)$.

The conditional probability $\mathbf{P}_z(\beta/\alpha)$ has the following meaning: *it is probability to observe the value β of the property B under the condition that the hidden (but objectively existing) value of the property A is equal to α.* The two-dimensional probability distribution $\mathbf{P}_z(\alpha, \beta) = \mathbf{P}_a(\alpha)\mathbf{P}_z(\beta/\alpha)$ is also well defined. This probability has the following physical meaning: it is probability that the hidden property $A = \alpha$ and the observed property $B = \beta$.

4.2 No conditional probabilities, no Bayes' formula

In principle there may be some frequency 'pathologies'. It may be that the hidden sequence $a = a_h$ is not a collective or it is a collective, but the collectives a and b are not combinable (physical experience implies that the sequence $b = b_{\mathcal{M}}$ is always a collective). Let us analyze such a situation more carefully. To simplify our considerations, we start with the case in that $a = a_h$ is a collective. Here frequency probabilities $\mathbf{P}_a(\alpha) = \lim_{k\to\infty} \nu_k(\alpha; a)$ are well defined. However, we do not more assume that the collectives a and b are combinable.

We have $n_k(\beta; b) = n_k(\beta/0; z) + n_k(\beta/1; z)$. Thus we obtain

$$\nu_k(\beta; b) = \frac{n_k(\beta/0; z)}{M_k(0; z)} \frac{M_k(0; z)}{k} + \frac{n_k(\beta/1; z)}{M_k(1; z)} \frac{M_k(1; z)}{k} .$$

Thus we have

$$\nu_k(\beta; b) = \nu_k(\beta/0; z)\nu_k(0; a) + \nu_k(\beta/1; z)\nu_k(1; a) . \tag{4.1}$$

We also note that by our assumptions there exist $\mathbf{P}_b(\beta) = \lim_{k\to\infty} \nu_k(\beta; b)$ and $\mathbf{P}_a(\alpha) = \lim_{k\to\infty} \nu_k(\alpha; a)$. We ask the following question:

Is it possible that (despite of the existence of the above limits and despite of equality (4.1)) $\lim_{k\to\infty} \nu_k(\beta/\alpha; z)$ *does not exist?*

Yes, it is surely possible!

Example 4.1. Let $\mathbf{P}_a(\alpha) = \lim_{k\to\infty} \nu_k(\alpha; a) = 1/2$ for $\alpha = 0, 1$. As always $\nu_k(0/\alpha; z) + \nu_k(1/\alpha; z) = 1$ for $\alpha = 0, 1$, it is possible to represent conditional frequencies in the form

$$\nu_k(0/\alpha; z) = \sin^2 \phi_{\alpha,k}, \quad \nu_k(1/\alpha; z) = \cos^2 \phi_{\alpha,k},$$

where the *phase* $\phi_{\alpha,k} = \arcsin \sqrt{\nu_k(0/\alpha; z)}$. In the case of regular conditional behaviour angles $\phi_{\alpha,k}$ stabilize (mod 2π) to some values ϕ_α when $k \to \infty$. Here conditional probabilities are well defined: $\mathbf{P}_z(0/\alpha) = \sin^2 \phi_\alpha$ and $\mathbf{P}_z(1/\alpha) = \cos^2 \phi_\alpha$. Equality (4.1) implies the formula of total probability:

$$\mathbf{P}_b(0) = \frac{1}{2}(\sin^2 \phi_0 + \sin^2 \phi_1), \quad \mathbf{P}_b(1) = \frac{1}{2}(\cos^2 \phi_0 + \cos^2 \phi_1).$$

Let us consider now the case of irregular conditional behaviour. Here angles $\phi_{\alpha,k}$ do not stabilize (mod 2π) when $k \to \infty$. But by (4.1) we have that limits

$$\mathbf{P}_b(0) = \frac{1}{2}\lim_{k\to\infty}(\sin^2\phi_{0,k} + \sin^2\phi_{1,k}), \quad \mathbf{P}_b(1) = \frac{1}{2}\lim_{k\to\infty}(\cos^2\phi_{0,k} + \cos^2\phi_{1,k})$$

must exist. For example, these conditions can be satisfied if we choose $\phi_{1,k} \approx \frac{\pi}{2} - \phi_{0,k}, k \to \infty$. Thus there is no contradiction between nonexistence of frequency conditional probabilities and formula (4.1).

What is a physical meaning of fluctuations of conditional relative frequencies $\nu_k(\beta/\alpha)$ (nonexistence of conditional probabilities $\mathbf{P}_z(\beta/\alpha)$)?

A quantum state ϕ contains only information on asymptotic behavior of frequencies for observations of each fixed property. However, ϕ does not contain information on statistical relations between different properties. This relation is given by conditional frequencies which are not determined by the quantum formalism. Therefore in principle behavior of relative frequencies in the statistical ensemble S (represented by ϕ) may be extremely irregular. But these fluctuations of conditional frequencies may compensate one another and give well-defined frequency probabilities for observed properties.

A real preparation procedure $\mathcal{E}$ can produce (after N steps) only a finite approximation S_N of the (ideal infinite) ensemble S represented by ϕ. Fluctuations of conditional frequencies imply that the statistical relation between two properties A and B (or more precisely the reaction of a quantum system s with the fixed (hidden) value α of the property A to a measurement of the property B) may strongly depend on the number N of experiments N. Let us consider again Example 4.1 and let $\phi_{0,k} \approx \frac{\pi k}{2m}, \phi_{1,k} \approx \frac{\pi}{2} - \frac{\pi k}{2m}, k \to \infty$, where $m > 1$ is the fixed natural number. Here 'conditional probabilities'

$$\mathbf{P}^k(0/0) \equiv \nu_k(0/0; z) \approx \sin^2\frac{\pi k}{2m}, \qquad \mathbf{P}^k(1/0) \equiv \nu_k(1/0; z) \approx \cos^2\frac{\pi k}{2m},$$
$$\mathbf{P}^k(0/1) \equiv \nu_k(0/1; z) \approx \cos^2\frac{\pi k}{2m}, \qquad \mathbf{P}^k(1/1) \equiv \nu_k(1/1; z) \approx \sin^2\frac{\pi k}{2m}$$

oscillate with the period $T = 2m$, when $k \to \infty$. Let m be very large. Then, for $k = 2mj + 1$, $\mathbf{P}^k(0/0) = \mathbf{P}^k(1/1) \approx 0$ and $\mathbf{P}^k(1/0) = \mathbf{P}^k(0/1) \approx 1$. Therefore in the ensemble S_k practically every quantum system s having the property $A = 0$ will exhibit the property $B = 1$ and practically every quantum system s having the property $A = 1$ will exhibit the property $B = 0$ (in the measurement $\mathcal{M}$ of B.) However, after $(m-1)$ steps statistical conditional behavior changes crucially. For $k' = 2mj + m, \mathbf{P}^k(0/0) = \mathbf{P}^k(1/1) \approx 1$ and $\mathbf{P}^k(1/0) = \mathbf{P}^k(0/1) \approx 0$. Therefore in the ensemble $S_{k'}$ practically every quantum system s having the property $A = 0$ will exhibit the property $B = 0$ and practically every quantum system s having the property $A = 1$ will exhibit the property $B = 1$. At the same time observed 'probabilities' $\mathbf{P}^k(B = \beta) = \nu_k(\beta; b)$ do not depend on these oscillations of conditional probabilities.

Remark 4.1 (On fluctuations of ensemble conditional probabilities). The above arguments can be used for ensemble conditional probabilities. A quantum state ϕ represents an *infinite ideal ensemble* S of quantum systems. As we have already discussed in Section 3, real analysis does not give a possibility to use the proportional definition of probability with respect to S. Typically probabilities $\mathbf{P}_S(B = \beta)$ with respect to S are considered as limits of probabilities $\mathbf{P}_{S_N}(B = \beta)$ with respect to finite approximations S_N of S. Such an approach to probabilities $\mathbf{P}_S(B = \beta)$ is justified by the incredible number of quantum experiments. However, it is often supposed that conditional probabilities $\mathbf{P}_S(B = \beta/A = \alpha)$ can be also defined as limits of probabilities $\mathbf{P}_{S_N}(B = \beta/A = \alpha)$ with respect to finite approximations S_N of S. Such an assumption has not been (and probably it will never be) verified experimentally. Example 4.1 (which can be used in the ensemble framework) demonstrates that in principle conditional probabilities $\mathbf{P}_{S_N}(B = \beta/A = \alpha)$ may oscillate with the increasing of N. In such a case conditional probabilities $\mathbf{P}_S(B = \beta/A = \alpha)$ cannot be defined. Therefore Bayes' formula and the formula of total probability cannot be used for such quantum states.

Finally we remark that it may be that the sequence $a = a_h$ is not collective. For example, if we keep to f-realism, then the statistical stabilization of frequencies $\nu_k(A_f = \alpha)$ need not imply the statistical stabilization of frequencies $\nu_k(A_i = \alpha)$.

Example 4.2 (Fluctuating probabilities and stabilized conditional probabilities). Suppose that $\nu_k(0, a) \approx \sin^2 \phi_k$ and $\nu_k(1, a) \approx \cos^2 \phi_k, k \to \infty$. If phases ϕ_k do not stabilize (mod 2π) when $k \to \infty$, then frequencies $\nu_k(\alpha; a)$ fluctuate when $k \to \infty$. Thus frequency probabilities $\mathbf{P}_a(\alpha)$ do not exist. Suppose that, however, frequency conditional probabilities $\mathbf{P}_z(\beta/\alpha) = \lim_{k\to\infty} \nu_k(\beta/\alpha; z)$ exist and they are equal to 1/2. Therefore sizes of populations $S_{N,\alpha}$ with the fixed value $A = \alpha$ fluctuate, but reactions of quantum systems $s \in S_{N,\alpha}$ to the measurement $\mathcal{M}$ of the B are stable. In this case we find that the limit in (4.1) exists

$$\mathbf{P}_b(\beta) = \lim_{k\to\infty} (\nu_k(0; a)\nu_k(\beta/0; z) + \nu_k(1; a)\nu_k(\beta/1; z)) = 1/2, \quad \beta = 0, 1.$$

Thus there is no contradiction between nonexistence of frequency probabilities $\mathbf{P}_a(\alpha)$ and formula (4.1).

5 Einstein–Podolsky–Rosen paradox: probability, reality and locality

In the previous sections we have shown that there is no principal difference between 'quantum' and 'classical' probabilities and as consequence between classical and quantum systems. We can use ensemble or frequency definitions of probability. However, we have to control the relation between probabilities and ensembles or collectives. On the other hand, we cannot use the conventional Kolmogorov formalism in that the structure of an ensemble or a collective does not play any role. In principle, it is possible to consider 'quantum properties' as objective properties (by using both i-realism and f-realism). Of course, probabilistic distributions of these properties depend on ensembles or collectives.

However, there are some quantum experiments which seem to demonstrate that there is a large difference between classical and quantum systems. All such experiments are based on the idea to eliminate disturbance effects by separating quantum systems in space-time (and to use correlations between these separated quantum systems). The starting point was the famous Einstein–Podolsky–Rosen (EPR) experiment [58]. We present a brief description of this experiment. We start with the definition of *Einsteinian separability:*

Two space-time regions U and V are said to be spatially separated, if the real factual situation within V is independent of what is done in U.

We recall that A. Einstein was an adherent of i-realism. Thus values of physical properties which will be discussed later (namely, positions q and momentums p) are initial values of this properties.

It should be noticed that the study of distinguishing features of 'quantum probabilities' was not the original aim of EPR's considerations. EPR wanted to show that quantum mechanics is not a *complete physical theory.* The completeness means that quantum mechanics provides a complete description of the atomic and subatomic phenomena. The opinion that quantum mechanics is complete (and hence we need not more detailed description of reality than quantum mechanics) was, already at that time (1933), so much engrained in the mind of physicists that the EPR arguments against the completeness was soon referred as a paradox. EPR wanted to demonstrate that there exist elements of reality which could not be described by a quantum state. Of course, in this framework the question on the meaning of an *element of reality* arose immediately. EPR thought that it would be impossible to propose the exact definition of an element of reality. However, they proposed the following criterion of reality:

"If, without in any way disturbing a system we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity."

EPR proposed the following arguments based on this criterion for elements of reality and the notion of separability for two space-time regions U and V.

Let us consider a statistical ensemble S of pairs (a^1, a^2) of correlated particles. For example, these are pairs of particles which are emitted by excited atoms. We consider the one-dimensional model with particles moving in the opposite directions. For each pair the correlation implies the conservation of the momentum, $p^{(1)} + p^{(2)} = 0$, and the relative position, $q^{(1)} - q^{(2)} = 0$ (correlations between properties of particles). For any pair (a^1, a^2) of correlated particles, we can measure the position $q^{(1)}$ of a^1 in U which (due to the correlation) gives the position $q^{(2)}$ of a^2 (without to disturb a^2). Thus the position $q^{(2)}$ of a^2 is an element of reality. By the similar considerations we obtain that the momentum $p^{(2)}$ of a^2 is an element of reality. On the other hand, the quantum formalism implies the Heisenberg uncertainty relation

$$\Delta q \Delta p \geq h/2$$

for any quantum state ϕ. Thus the definite values of the position and momentum of a quantum particle cannot be simultaneously elements of reality for the same quantum state. EPR interpreted this as the evidence of the incompleteness of quantum mechanics: in the EPR experiment two elements of reality (the position and momentum) exist simultaneously, but they could not be described by any quantum state (thus the formalism of quantum mechanics does not provide the description of the whole physical reality).

The EPR considerations (which are often regarded as the paradox in the foundations of quantum mechanics) induced great debates (which were initiated by A. Einstein and N. Bohr). Numerous arguments were used by both sides. It is interesting to remark that at the first stage of these debates probability reasons were not used. There was no analysis of the probability basis of the EPR considerations. In particular, nobody tried to study the problem of difference between 'classical and quantum probabilities' to disprove the simultaneous reality of the position and momentum. However, later such analysis has been done and one of the results of this analysis was famous Bell's inequality (see Section 6).

I support the viewpoint that quantum mechanics is not complete. The incompleteness of quantum mechanics is rather a consequence of all physical experience which demonstrated that no physical theory (at least before quantum mechanics) turned out to be universally valid. Every single theory was valid only if applied to a restricted part of reality, its domain of application. Do we have any reason to believe that quantum mechanics is different, and will hold true for whatever future experiments we may be able to think of? But at the same time I think that EPR arguments do not imply the conclusion that quantum mechanics is not complete. I am not satisfied by EPR's criterion of reality. There are strong probabilistic arguments against this criterion. The meaning of 'unit probability' in this criterion is unclear. In fact, this 'unit probability' must depend on an ensemble or a collective. We shall see that it is impossible to find the same ensemble or collective for the positions and momentums of particles a^2 (or particles a^1).

To save completeness of quantum mechanics, some physicists accept *nonlocality of space-time.* They claim that, for example, a measurement of the position of the particle a^1 located in Moscow changes properties of the particle a^2 located in Vladivostok. Some of them assume the possibility of the *faster-than-light-influences* (of course, such an assumption contradicts to theory of relativity). Other adherents of nonlocality consider this nonlocality as only *information nonlocality.* They think that, for example, a measurement of the position $q^{(1)}$ of the particle a^1 does not change objective properties of the particle a^2, but such a measurement changes our information about the particle a^2. Hence they need not use the *faster-than-light-influences*.

Another group of physicists thinks that the root of the problem is the realist viewpoint of EPR. If we reject realism and keep to empiricism (or even idealism), then we could not assign any physical meaning to values of the position $q^{(2)}$ and momentum $p^{(2)}$ of the particle a^2 which are predicted on the basis of measurements for the particle a^1. Adherents of empiricism have also some differences in views. One part of them think that the root of the problem is the impossibility to perform a simultaneous measurement of the position $q^{(1)}$ and momentum $p^{(1)}$ for the same particle a^1 (and thus obtain the 'simultaneous prediction'). Other keep to the rigid empiricists line. They think that the root of the problem is the impossibility to perform a simultaneous measurement of the position $q^{(2)}$ and momentum $p^{(2)}$ for the same particle a^2.

As I have already mentioned, it seems that EPR arguments do not imply incompleteness, nonlocality or impossibility to keep to realism. EPR considerations imply only that we could not manipulate with formal (abstract) probabilities which are not related to concrete ensembles or collectives.

By fixing a value $\alpha \in \mathbb{R}$ of $q^{(1)}$ we can construct an ensemble $\bar{S}_\alpha$ of particles a_j^2 for that $q^{(2)} = \alpha$ and the distribution of the momentums is the same as in the original ensemble. Of course, probability $\mathbf{P}_{\bar{S}_\alpha}(q^{(2)} = \alpha) = 1$. Thus $q^{(2)} = \alpha$ is an element of reality for the ensemble $\bar{S}_\alpha$. However, $\mathbf{P}_{\bar{S}_\alpha}(p^{(2)} = \beta) \neq 1$ for any fixed value $\beta \in \mathbb{R}$. Thus $p^{(2)} = \beta$ is not an element of reality for this ensemble.

In the same way by fixing the value $\beta \in \mathbb{R}$ of $p^{(1)}$ we can construct an ensemble $\bar{R}_\beta$ of particles a_j^2 for that $p^{(2)} = \beta$ and the distribution of the positions is the same as in the original ensemble. Probability $\mathbf{P}_{\bar{R}_\beta}(p^{(2)} = \beta) = 1$. Thus $p^{(2)} = \beta$ is an element of reality for the ensemble $\bar{R}_\beta$. However, $\mathbf{P}_{\bar{R}_\beta}(q^{(2)} = \alpha) \neq 1$ for any fixed value $\alpha \in \mathbb{R}$. Thus $q^{(2)} = \alpha$ is not an element of reality for this ensemble. EPR did not present any idea how we could construct an ensemble $W_{\alpha\beta}$ of particles a_j^2 such that $\mathbf{P}_{W_{\alpha\beta}}(q^{(2)} = \alpha) = 1$ and $\mathbf{P}_{W_{\alpha\beta}}(p^{(2)} = \beta) = 1$. Therefore the EPR arguments give no reason to conclude that quantum mechanics is not complete. There are no reasons to use nonlocality or to reject realism to explain the EPR arguments. It must be pointed out that EPR arguments could not be used as a 'proof' that i-realism can (or even must) be used to describe quantum phenomena[18]. In fact, from the mathematical point of view the only 'argument' of EPR was that the notion of the unit probability can be used without connection to concrete ensembles.

Our previous considerations can be repeated in the frequency framework. Here we can keep not only to i-realism, but also to f-realism in the EPR scheme[19]. Let us perform a measurement $\mathcal{N}$ of $q^{(1)}$ and a measurement $\mathcal{M}$ of $p^{(2)}$ and save only the results for that $q^{(1)} = \alpha$, where $\alpha \in \mathbb{R}$, is some fixed value. We obtain the two-

[18]EPR obtained incompleteness of quantum mechanics by presenting the experiment which demonstrates that both the position and momentum of a quantum particle can be elements of reality for the same quantum state. This is often interpreted as a proof of the possibility to use i-realists approach in quantum mechanics.

[19]In the latter case $q^{(l)} \equiv q_f^{(l)}$ and $p^{(l)} \equiv p_f^{(l)}$, $l = 1, 2$, and $q_f^{(1)} - q_f^{(2)} = 0$, $p_f^{(1)} + p_f^{(2)} = 0$.

dimensional collective:

$$x_\alpha = (x_1, x_2, \dots, x_k, \dots), \quad x_j = (q_j^{(1)}, p_j^{(2)}),$$

where $q_j^{(1)} \equiv \alpha$.

As $q_j^{(2)} = q_j^{(1)} \equiv \alpha$, we can (parallel to the construction of x_α) construct another two-dimensional collective

$$x'_\alpha = (x'_1, x'_2, \dots, x'_k, \dots), \quad x'_j = (q_j^{(2)}, p_j^{(2)}),$$

where $q_j^{(2)} \equiv \alpha$.

In the same way, for each fixed value $\beta \in \mathbb{R}$ of the momentum, we construct two-dimensional collectives

$$y_\alpha = (y_1, y_2, \dots, y_k, \dots), \quad y_j = (q_j^{(2)}, p_j^{(1)}),$$

where $p_j^{(1)} \equiv \beta$, and

$$y'_\alpha = (y'_1, y'_2, \dots, y'_k, \dots), \quad y'_j = (q_j^{(2)}, p_j^{(2)}),$$

where $p_j^{(2)} \equiv \beta$. Of course,

$$\mathbf{P}_{x'_\alpha}(q^{(2)} = \alpha) = 1$$

and $q^{(2)} = \alpha$ is the element of reality for the collective x'_α and

$$\mathbf{P}_{y'_\beta}(p^{(2)} = \beta) = 1$$

and $p^{(2)} = \beta$ is the element of reality for the collective y'_β. However, EPR did not present any idea how we can construct a collective $z_{\alpha\beta}$ such that

$$\mathbf{P}_{z_{\alpha\beta}}(q^{(2)} = \alpha) = 1, \quad \text{and} \quad \mathbf{P}_{z_{\alpha\beta}}(p^{(2)} = \beta) = 1.$$

In fact, the presented argument against the EPR-paper is the probabilistic version of Bohr's argument in his reply [72] to Einstein. Unfortunately, Bohr did not use probabilities in his response. He proceeded with complementarity principle. But the absence of a collective $z_{\alpha\beta}$ is simply von Misesian version of Bohr's complementarity principle.

6 Bell's inequality

The EPR idea to consider statistical ensembles of correlated spatially separated particles was developed by D. Bohm. He proposed a simpler example in that it is possible to use discrete variables.

6.1 EPR experiment for measurement of spin projections: Bohm's experiment

Instead of the position and momentum of a quantum particle, Bohm considered its spin[20] components. Let $\mathbf{s} \in \mathbb{R}^3$ be the spin of a quantum particle. For any axis $n \in \mathbb{R}^3$ we denote the projection of $\mathbf{s}$ to this axis by the symbol $\mathbf{s}_n$:

$$\mathbf{s}_n = \frac{(\mathbf{s}, n)n}{\|\mathbf{s}\|},$$

where $(\cdot,\cdot)$ and $\|\cdot\|$ are, respectively, the inner product and norm on $\mathbb{R}^3$. Consider a measurement device M_n for measuring the spin projection $\mathbf{s}_n$. It is Stern–Gerlach magnet with the orientation n.

By keeping to realism we can say: such a measurement disturbs a quantum particle and changes its spin. However, by keeping empiricism one cannot speak about vector of spin before measurement. Our aim is to test realism.

A measurement device $M_{n,n'}$ which can measure two components $\{\mathbf{s}_n, \mathbf{s}_{n'}\}$, $n \neq n'$, simultaneously does not exist. However, for correlated particles (a^1, a^2) (with spins $\mathbf{s}^1, \mathbf{s}^2$) we can use the conservation law for the spins of these particles:

$$\mathbf{s}^1 + \mathbf{s}^2 = 0.$$

Hence the measurement M_n^1 for a^1 gives automatically the value $\mathbf{s}_n^2$ of the spin of a^2. As usual, it is assumed that particles a^1 and a^2 satisfy the condition of Einsteinian separability. By the EPR reality criterion we obtain that there exists an element of reality corresponding to the spin component $\mathbf{s}_n^2$ (and by symmetry for the $\mathbf{s}_n^1$) for any axis $n \in \mathbb{R}^3$. Therefore the spin $\mathbf{s}$ is an element of reality. However, by probabilistic reasons (discussed in the previous section) we do not want to apply the EPR reality criterion. We (following Bell) can study the following problem:

Is it possible to keep to realism to describe spin measurements for correlated particles?

We restrict our consideration to two-dimensional model. Here each direction n can be characterized by an angle $\phi : n = n_\phi$. We set

$$\mathbf{s}_\phi = \mathrm{sign}(\mathbf{s}, n_\phi).$$

In the real physical model we have to use probabilities of simultaneous measurements of $\mathbf{s}_{\phi_j}^1$ and $\mathbf{s}_{\phi_l}^2$ for three angles ϕ_j, $j = 1, 2, 3$.[21] It is possible to obtain some inequality for these probabilities, namely, Bell's inequality in the probabilistic form – Wigner's inequality. As there are two particles a^1 and a^2, to describe the model, we must use

[20] The scientists whose interests are far from quantum mechanics may imagine spin as an arrow $\mathbf{s} \in \mathbb{R}^3$ which is associated with each quantum particle indicating the 'internal rotation' of the particle. Of course, it is a realistic model of spin. It would be tested.

[21] In fact, in experiments we need to use even four angles.

the four-dimensional Hilbert space. However, we can obtain the same results on the basis of the two-dimensional Hilbert space by using the following toy-model.

Let $\mathcal{H}$ be the two-dimensional Hilbert space and let $e_\phi = (e_{\phi,+}, e_{\phi,-}), \phi \in [0, 2\pi)$, be orthonormal bases in $\mathcal{H}$ which are connected by the following unitary transformation:

$$e_{\phi,+} = \cos(\theta - \phi)e_{\theta,+} + i \sin(\theta - \phi)e_{\theta,-}, \tag{6.1}$$

$$e_{\phi,-} = i \sin(\theta - \phi)e_{\theta,+} + \cos(\theta - \phi)e_{\theta,-}. \tag{6.2}$$

We introduce the quantum state

$$\Psi = \frac{e^{i\gamma}}{\sqrt{2}}e_{\gamma,+} + \frac{e^{i\gamma}}{\sqrt{2}}e_{\gamma,-} \ .$$

We consider observables (properties) $\mathbf{s}_\phi$ corresponding to bases $e_\phi : \widehat{\mathbf{s}}_\phi e_\phi = \pm e_\phi$. By the probability interpretation of the quantum state Ψ we have $\mathbf{P}_\Psi(\mathbf{s}_\phi = \pm 1) = 1/2$.

Since we want to keep to realism, we consider a possibility to represent quantum observables for spin-projections by classical random variables $\mathbf{s}_\theta(\omega), \omega \in \Omega$, where Ω is the space of chance parameters (hidden variables). Thus we identify 'quantum probabilities' with classical ones:

$$\mathbf{P}_\Psi(\mathbf{s}_\phi = \pm 1) = \mathbf{P}(\omega \in \Omega : \mathbf{s}_\phi(\omega) = \pm 1) = 1/2.$$

If conditional probabilities $\mathbf{P}(\omega \in \Omega : \mathbf{s}_\theta(\omega) = \epsilon/\mathbf{s}_\phi(\omega) = \delta), \epsilon, \delta = \pm 1$, are identified with quantum probabilities given by expansions (6.1), (6.2), then we obtain:

$$\mathbf{P}(\omega \in \Omega : \mathbf{s}_\theta(\omega) = +1/\mathbf{s}_\phi(\omega) = -1) = \sin^2(\theta - \phi)$$

and

$$\mathbf{P}(\omega \in \Omega : \mathbf{s}_\theta(\omega) = +1/\mathbf{s}_\phi(\omega) = +1) = \cos^2(\theta - \phi).$$

Now by using Bayes' formula for classical probabilities we obtain:

$$\begin{aligned}\mathbf{P}(\omega \in \Omega &: \mathbf{s}_\theta(\omega) = +1, \mathbf{s}_\phi(\omega) = -1) \\ &= \mathbf{P}(\omega \in \Omega : \mathbf{s}_\phi(\omega) = -1)\mathbf{P}(\omega \in \Omega : \mathbf{s}_\theta(\omega) = +1/\mathbf{s}_\phi(\omega) = -1) \\ &= \frac{1}{2}\sin^2(\theta - \phi)\end{aligned} \tag{6.3}$$

and

$$\begin{aligned}\mathbf{P}(\omega \in \Omega &: \mathbf{s}_\theta(\omega) = +1, \mathbf{s}_\phi(\omega) = +1) \\ &= \mathbf{P}(\omega \in \Omega : \mathbf{s}_\phi(\omega) = +1)\mathbf{P}(\omega \in \Omega : \mathbf{s}_\theta(\omega) = +1/\mathbf{s}_\phi(\omega) = +1) \\ &= \frac{1}{2}\cos^2(\theta - \phi).\end{aligned} \tag{6.4}$$

6.2 Bell's inequality for probabilities (Wigner's inequality)

We prove now some inequality for events defined by three variables $\mathbf{s}_\gamma(\omega)$, $\gamma = 0, \phi, \theta$. In fact, this inequality does not depend on the form of the probability distributions of random variables $\mathbf{s}_\gamma(\omega)$. We shall use only the fact that there exists the Kolmogorov probability space $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$ on which these random variables are defined:

$$\begin{aligned} &\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1) \\ &\quad = \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1, \mathbf{s}_\theta(\omega) = +1) \\ &\qquad + \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1, \mathbf{s}_\theta(\omega) = -1), \end{aligned} \tag{6.5}$$

$$\begin{aligned} &\mathbf{P}(\omega \in \Omega : \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1) \\ &\quad = \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1) \\ &\qquad + \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = -1, \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1), \end{aligned} \tag{6.6}$$

and

$$\begin{aligned} &\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\theta(\omega) = +1) \\ &\quad = \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1, \mathbf{s}_\theta(\omega) = +1) \\ &\qquad + \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1). \end{aligned} \tag{6.7}$$

If we add together the equations (6.5) and (6.6) we obtain

$$\begin{aligned} &\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1) + \mathbf{P}(\omega \in \Omega : \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1) \\ &\quad = \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1, \mathbf{s}_\theta(\omega) = +1) \\ &\qquad + \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1, \mathbf{s}_\theta(\omega) = -1) \\ &\qquad + \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1) \\ &\qquad + \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = -1, \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1). \end{aligned} \tag{6.8}$$

But the first and the third terms on the right hand side of this equation are just those which when added together make up the term $\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\theta(\omega) = +1)$ (Kolmogorov probability is additive). It therefore follows that:

$$\begin{aligned} &\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1) + \mathbf{P}(\omega \in \Omega : \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1) \\ &\quad = \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\theta(\omega) = +1) \\ &\qquad + \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1, \mathbf{s}_\theta(\omega) = -1) \\ &\qquad + \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = -1, \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1) \end{aligned} \tag{6.9}$$

By using nonnegativity of probability we obtain the inequality:

$$\begin{aligned} &\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1) \\ &\qquad + \mathbf{P}(\omega \in \Omega : \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1) \\ &\geq \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\theta(\omega) = +1) \end{aligned} \tag{6.10}$$

which is a variant of Bell's inequality (for probabilities).

We turn back to physics and apply the inequality (6.10) to the 'quantum probabilities' $\mathbf{P}_q$, see (6.3), (6.4), which were computed in the framework of quantum mechanics. We obtain: $\cos^2 \phi + \sin^2(\theta - \phi) \geq \cos^2 \theta$. Now set $\phi = 3\theta$. We obtain: $g(\theta) = \cos^2 3\theta + \sin^2 2\theta - \cos^2 \theta \geq 0$. However, the latter inequality holds only for *sufficiently large* angles $\theta : \theta \geq \pi/6$. Thus for $\theta < \pi/6$ the inequality (6.10) is violated.

7 Bell's mystification

First of all we remark that the violation of Bell's inequality for 'quantum probabilities' may be in principle interpreted as an evidence of violations of quantum mechanical laws for the spin model. However, numerous experiments were performed in the connection with this problem, see, for example, [16], [40–42], [71]. All these experiments demonstrated that quantum mechanical laws hold true: experimental probabilities coincide (of course, with some precision) with quantum probabilities $\mathbf{P}_q(\mathbf{s}_\theta = \epsilon, \mathbf{s}_\phi = \delta)$, $\epsilon, \delta = \pm 1$, computed via (6.3), (6.4). Bell's inequality for experimental probabilities is violated.

7.1 Probability and reality

It is widely accepted by a part of physical community that the violation of Bell's inequality has demonstrated that the realists philosophy cannot be used for the description of quantum phenomena: the spin is not an objective property of a quantum particle.

Remark 7.1. Other part of the physical community connects Bell's inequality and nonlocality of space-time: in the real physical experiments observables $\mathbf{s}_\theta = \mathbf{s}_\theta^1$ and $\mathbf{s}_\phi = \mathbf{s}_\phi^2$ correspond to two particles which are separated in space. However, our probabilistic analysis will demonstrate that there are no traces of nonlocality in Bell's framework. Therefore we will mainly concentrate our considerations on connection between Bell's inequality and realism.

The problem of existence (reality) of spin is often mixed with the problem of existence of random variables $\mathbf{s}_\phi(\omega)$, $\phi \in [0, 2\pi]$, defined on some Kolmogorov probability space. However, these are two different problems. Kolmogorov's model is just one of possible models of reality. Besides Kolmogorov's model, there exist frequency and ensemble models. We shall demonstrate that Bell's inequality does not present in the latter models. Thus there is no problem with experimental violations of this inequality. The spin can be in principle considered as an objective property of a quantum system. We shall show that we can even keep to 'real realism', namely, i-realism.

7.2 Realism and Bell's inequality

Let s_ϕ, $\phi \in [0, 2\pi]$, be initial values. We use the ensemble approach to probability. The main distinguishing feature of this approach is that all probabilities in (6.5)–(6.7) depend on corresponding ensembles. There are three ensembles $S^N_{0\phi}$, $S^N_{\phi\theta}$, $S^N_{0\theta}$ (of cardinality N) which are used to obtain observed probabilities $\mathbf{P}_{S^N_{0\phi}}(s_0 = \pm 1, s_\phi = \pm 1)$, $\mathbf{P}_{S^N_{\phi\theta}}(s_\phi = \pm 1, s_\theta = \pm 1)$, $\mathbf{P}_{S^N_{0\theta}}(s_0 = \pm 1, s_\theta = \pm 1)$.[22]

Remark 7.2. Formally we could introduce an infinite ensemble S of particles and define ensemble probabilities with respect to S:

$$\mathbf{P}_S(s_{\alpha_1} = \epsilon_1, s_{\alpha_2} = \epsilon_2, \ldots, s_{\alpha_n} = \epsilon_n) = \frac{|\{s \in S : s_{\alpha_1} = \epsilon_1, s_{\alpha_2} = \epsilon_2, \ldots, s_{\alpha_n} = \epsilon_n\}|}{|S|}, \tag{7.1}$$

where $\epsilon_i = \pm 1$, $n \in \mathbb{N}$. Of course, the calculations of Section 6 can be repeated for such probabilities. However, this can be done only formally. As we have already mentioned, the proportional definition of probability is meaningless for infinite ensembles in the framework of real analysis. Thus we could not perform these formal calculations on the mathematical level of rigorousness.

Remark 7.3. The proportional ensemble definition (7.1) can be used on the mathematical level of rigorousness on the basis of non-Archimedean analysis. For example, in Chapter 4 we study p-adic ensemble probabilities. All arithmetical calculations (6.5)–(6.9) can be performed in the field of p-adic numbers. But (6.9) does not imply (6.10)! Some of probabilities $\mathbf{P}_S(s_{\alpha_1} = \epsilon_1, s_{\alpha_2} = \epsilon_2, s_{\alpha_3} = \epsilon_3)$ can be negative! In fact, there is some hidden (and still unclear) logic in such an appearance of negative probabilities in models in that the formal use of infinite statistical ensembles is not justified (see Chapter 3).

Therefore we have to operate with finite ensembles $S^N_{\alpha\beta}$. The three-dimensional probabilities used in (6.5)–(6.7) must be also considered as probabilities with respect to these ensembles. Thus in (6.5)–(6.7) we use probabilities $\mathbf{P}_{S^N_{0\phi}}(\ldots), \ldots, \mathbf{P}_{S^N_{0\theta}}(\ldots)$.

[22] In Kolmogorov's model ensemble indexes are omitted. In fact, this manipulation which looks quite innocent is the origin of Bell's mystification.

Hence (6.5) and (6.6) give

$$\begin{aligned}
&\mathbf{P}_{S^N_{0\phi}}(\mathbf{s}_0 = +1, \mathbf{s}_\phi = +1) + \mathbf{P}_{S^N_{\phi\theta}}(\mathbf{s}_\phi = -1, \mathbf{s}_\theta = +1)\\
&= \mathbf{P}_{S^N_{0\phi}}(\mathbf{s}_0 = +1, \mathbf{s}_\phi = +1, \mathbf{s}_\theta = +1) + \mathbf{P}_{S^N_{0\phi}}(\mathbf{s}_0 = +1, \mathbf{s}_\phi = +1, \mathbf{s}_\theta = -1)\\
&\quad + \mathbf{P}_{S^N_{\phi\theta}}(\mathbf{s}_0 = +1, \mathbf{s}_\phi = -1, \mathbf{s}_\theta = +1) + \mathbf{P}_{S^N_{\phi\theta}}(\mathbf{s}_0 = -1, \mathbf{s}_\phi = -1, \mathbf{s}_\theta = +1).
\end{aligned}$$

But in the opposite to calculations with abstract Kolmogorov probabilities in Section 6 the first and the third terms on the right hand side of this equation are not those which when added together make up the term

$$\begin{aligned}
\mathbf{P}_{S^N_{0\theta}}(\mathbf{s}_0 = +1, \mathbf{s}_\theta = +1) = {}& \mathbf{P}_{S^N_{0\theta}}(\mathbf{s}_0 = +1, \mathbf{s}_\phi = +1, \mathbf{s}_\theta = +1)\\
&+ \mathbf{P}_{S^N_{0\theta}}(\mathbf{s}_0 = +1, \mathbf{s}_\phi = -1, \mathbf{s}_\theta = +1).
\end{aligned}$$

To obtain (6.9), we have to identify $\mathbf{P}_{S^N_{0\phi}}(\mathbf{s}_0 = +1, \mathbf{s}_\phi = +1, \mathbf{s}_\theta = +1)$ and $\mathbf{P}_{S^N_{0\theta}}(\mathbf{s}_0 = +1, \mathbf{s}_\phi = +1, \mathbf{s}_\theta = +1)$, $\mathbf{P}_{S^N_{\phi\theta}}(\mathbf{s}_0 = +1, \mathbf{s}_\phi = -1, \mathbf{s}_\theta = +1)$ and $\mathbf{P}_{S^N_{0\theta}}(\mathbf{s}_0 = +1, \mathbf{s}_\phi = -1, \mathbf{s}_\theta = +1)$. But (and this is the crucial point!) there are no reasons to do this in the general case.

For example, in quantum experiments with correlated particles it is possible to measure only two-dimensional probabilities $\mathbf{P}_{S^N_{\alpha_1\alpha_2}}(s_{\alpha_1} = \pm 1, s_{\alpha_2} = \pm 1)$ (by using correlations between particles). The physical experience is that this ensemble probabilities stabilize when $N \to \infty$. However, there are no reasons that three-dimensional probabilities $\mathbf{P}_{S^N_{\alpha_1\alpha_2}}(s_{\alpha_1} = \pm 1, s_{\alpha_2} = \pm 1, s_{\alpha_3} = \pm 1)$ must also stabilize when $N \to \infty$. They could depend essentially on statistical ensembles. Therefore the identification of probabilities with respect to different ensembles is not justified at all.

Practically the same considerations can be repeated in the framework of von Mises' frequency theory. There we have to consider three different collectives, $x_{0\phi}, x_{\phi\theta}, x_{0\theta}$, instead of ensembles $S^N_{0\phi}, S^N_{\phi\theta}, S^N_{0\theta}$. These are collectives for two-dimensional labels $(s_0, s_\phi), (s_\phi, s_\theta), (s_0, s_\theta)$. The principle of statistical stabilization can be applied only to these labels. The frequencies $\nu_N(s_\alpha = \pm 1, s_\beta = \pm 1; x_{\alpha\beta})$ stabilize when $N \to \infty$. However, the frequencies $\nu_N(s_\alpha = \pm 1, s_\beta = \pm 1, s_\gamma = \pm 1; x_{\alpha\beta})$ need not stabilize when $N \to \infty$. Moreover, they may be not defined at all. Therefore *there is no Bell's inequality in von Mises probability model.*

If we keep to f-realism, we have to use von Mises' frequency probability theory. Therefore we could not obtain Bell's inequality. There are no problems with violations of this inequality.

Remark 7.4. Typically Bell's inequality is associated with the use of so called hidden variables, see Section 9. As it has been noticed in [51], [54], it can be derived without any reference to hidden variables. As the reader has seen, it was only supposed that there exists a Kolmogorov probability space $\mathcal{P} = \{\Omega, \mathcal{F}, \mathbf{P})$ such that three spin projections $\mathbf{s}_0$, $\mathbf{s}_\phi$, $\mathbf{s}_\theta$ can be represented by random variables on this space. Under this assumption it is possible to define the joint probability distribution $\mathbf{P}_{ijk} = \mathbf{P}(\mathbf{s}_0 = i, \mathbf{s}_\phi = j, \mathbf{s}_\theta = k)$, $i, j, k = \pm 1$. On the other hand, the existence of the joint probability distribution $\mathbf{P}_{ijk}$ implies the existence of the Kolmogorov space with $\mathbf{P} = \{\mathbf{P}_{ijk}\}$. This connection between Bell's inequality and existence of joint probability distribution was discussed by A. Fine [61], P. Rastall [155], W. de Muynck and H. Martens [53] (see also [54]). Typically nonexistence of the joint probability distribution is interpreted as the impossibility to use the objective realism (at least its i-version). From our viewpoint this is just the impossibility to apply the Kolmogorov model of probability theory (i.e., to use abstract symbolic probabilities without to regard to concrete ensembles or collectives). In the frequency approach such nonexistence only demonstrates the absence of the statistical stabilization for relative frequencies $\nu_N(\mathbf{s}_0 = i, \mathbf{s}_\phi = j, \mathbf{s}_\theta = k)$ for three different projections of spin. However, there are no (!) experimental reasons to suppose such a stabilization. In the ensemble approach such nonexistence only demonstrates the absence of the reproducibility of the 'property' $(\mathbf{s}_0 = i, \mathbf{s}_\phi = j, \mathbf{s}_\theta = k)$ in statistical ensembles used for quantum experiments. However, there are no(!) reasons to suppose such a reproducibility.

8 Bell's inequality for covariations

We have considered Bell's inequality for probabilities. The original Bell's inequality [26], [27] was proved for covariations.

Theorem 8.1. *Let $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$ be a Kolmogorov probability space and $A, B, C \in \mathrm{RV}(\mathcal{P})$ be discrete random variables, $A, B, C = \pm 1$. Then Bell's inequality*

$$|\langle A, B\rangle - \langle C, B\rangle| \leq 1 - \langle A, C\rangle \tag{8.1}$$

holds true.

Proof. Set $\Delta = \langle A, B\rangle - \langle C, B\rangle$. By linearity of Lebesgue integral we obtain

$$\begin{aligned}\Delta &= \int_\Omega A(\omega)B(\omega)\,d\mathbf{P}(\omega) - \int_\Omega C(\omega)B(\omega)\,d\mathbf{P}(\omega)\\ &= \int_\Omega [A(\omega) - C(\omega)]B(\omega)\,d\mathbf{P}(\omega).\end{aligned} \tag{8.2}$$

As $A(\omega)^2 = 1$,

$$|\Delta| = |\int_\Omega [1 - A(\omega)C(\omega)]A(\omega)B(\omega)\,d\mathbf{P}(\omega)| \tag{8.3}$$

$$\leq \int_\Omega [1 - A(\omega)C(\omega)]\,d\mathbf{P}(\omega) = 1 - \langle A, C\rangle. \quad \square \tag{8.4}$$

Of course, this is the rigorous mathematical proof of (8.1) for Kolmogorov probabilities. However, as we have mentioned, Kolmogorov's model does not provide the adequate description of some quantum measurements. The root of 'Bell–Kolmogorov mystification' is again the identification of probabilities corresponding to different statistical ensembles or collectives.

Let us consider von Mises' approach. The ensemble approach will be considered in Section 9 in connection with so called hidden variables. In the frequency formalism the covariations $\langle A, B\rangle$ and $\langle C, B\rangle$ are covariations with respect to two different collectives, x_{AB} and x_{CB}: $\langle A, B\rangle \equiv \langle A, B\rangle_{x_{AB}}$ and $\langle C, B\rangle \equiv \langle C, B\rangle_{x_{CB}}$. Thus

$$\langle A, B\rangle_{x_{AB}} = \frac{1}{N}\sum_{i=1}^{N} a_i b_i, \quad \langle C, B\rangle_{x_{CB}} = \frac{1}{N}\sum_{i=1}^{N} c_i b_i'$$

and

$$\langle A, B\rangle x_{AB} - \langle C, B\rangle x_{CB} = \frac{1}{N}\sum_{i=1}^{N}[a_i b_i - c_i b_i'].$$

There are no reasons to suppose that

$$\frac{1}{N}\sum_{i=1}^{N}[a_i b_i - c_i b_i'] = \frac{1}{N}\sum_{i=1}^{N}[a_i - c_i]b_i. \tag{8.5}$$

Hence Bell made the mistake on the first step, (8.2), of the proof by using the linearity of mean value with respect to two different collectives (or statistical ensembles).

As physicists (with a few exceptions) did not see probabilistic roots of Bell's misunderstanding, they try to find some explanations of experimental violations of Bell's inequality:

1. **Death of realism.** It is impossible to keep to realism and suppose that quantum systems have objective properties.

2. **Nonlocality.** As in quantum experiments, covariations are found via measurements for correlated particles which are separated in space, it can be supposed that 'nonclassical' behavior of these covariations is a consequence of the dependence of the state of one particle on the state of other particle.

Of course, these ideas could not be denied on the basis of our probability analysis. But our analysis has demonstrated that Bell's arguments have no relation to these ideas.

9 Hidden variables and Bell's inequality

9.1 Incompleteness of quantum mechanics

Theories based on so called *hidden variables* were developed starting with the hypothesis on incompleteness of quantum mechanics. Typically incompleteness of quantum mechanics is considered as a consequence of the EPR arguments. By these arguments both position and momentum (or projections of spin to different axes) of each quantum particle are elements of reality for the system of two correlated particles. However, the quantum formalism does not describe the simultaneous existence of these elements of reality. Thus quantum mechanics is not complete. However, we have demonstrated that the EPR arguments are based on the formal use of Kolmogorov abstract probabilities. Of course, such ‘arguments’ could not be considered as the proof of incompleteness of quantum mechanics. Nevertheless, incompleteness of quantum mechanics can be directly obtained as a consequence of keeping to realist philosophy.

9.2 Hidden variables

Let us suppose that quantum mechanics is not complete. There could be finer description of reality than given by quantum mechanics. In principle there could exist some additional variables λ, hidden variables, such that by specifying the value $\lambda = \lambda_0$ of λ we could determine the values of all physical observables: $\lambda_0 \to A(\lambda_0), \lambda_0 \to B(\lambda_0), \dots$. Compatibility or incompatibility of physical observables $A, B, \dots$, do not play any role.

Typically Bell’s inequality is considered in the framework of hidden variables. The Kolmogorov probability space $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$ which was used in Section 8 has the following interpretation: $\Omega = \Lambda$ is the set of hidden variables, $\omega = \lambda$, $\mathbf{P}$ is the probability distribution of hidden variables. The experimental violations of Bell’s inequality are interpreted as the evidence that such *hidden variables do not exists.* Other authors use nonlocality arguments. They think that, despite of nonexistence of local hidden variables, *nonlocal hidden variables* can exist.

However, all our probability arguments against Bell’s inequality can be repeated for hidden variables. Let us use von Mises’ frequency approach. As we have already seen in Section 8, Bell’s mistake is the assumption on the validity of equality (8.5).

9.3 Deterministic hidden variables model and generalized Bell’s inequality

To simplify our considerations, we suppose that the set of hidden variables is finite:

$$\Lambda = \{\lambda_1, \dots, \lambda_M\}.$$

For each physical observable U, the value λ of hidden variables determines the value

$$U = U(\lambda).$$

Here we keep to realism. It is possible to keep i-realism or f-realism. If we keep to i-realism in this model, we have to assume that the result of measurement does not depend on fluctuations of an internal state ω of a measurement device $\mathcal{M}_U$ (see the next section for the model with such a dependence).

Let U and V be physical observables, $U, V = \pm 1$. We start with the consideration of the frequency (experimental) covariation $\langle U, V\rangle_{x_{UV}}$ with respect to a collective x_{UV} induced by measurements of the pair (U, V). The x_{UV} is obtained by measurements for an ensemble S_{UV} of physical systems (for example, pairs of correlated quantum particles). Our aim is to represent experimental covariation $\langle U, V\rangle_{x_{UV}}$ as ensemble covariation $\langle U, V\rangle_{S_{UV}}$. Then we shall demonstrate that in the general case it is impossible to perform for ensemble covariations Bell's calculations, (8.2)–(8.4), which have been performed for Kolmogorov covariations. Let $S_{UV} = \{d_1, \ldots, d_N\}$, where ith measurement is performed for the system d_i. Define a function $i \to \lambda(i)$, the value of hidden variables for d_i. We set $n_k(S_{UV}) = |\{d_i \in S_{UV} : \lambda(i) = \lambda_k\}|$ and $\mathbf{p}_k^{UV} = \mathbf{P}_{S_{UV}}(\lambda = \lambda_k) = \frac{n_k(S_{UV})}{N}$. We have

$$\begin{aligned}\langle U, V\rangle_{x_{UV}} &= \frac{1}{N}\sum_{i=1}^{N} U(\lambda(i))V(\lambda(i)) = \frac{1}{N}\sum_{k=1}^{M} n_k(S_{UV})u_k v_k \\ &= \sum_{k=1}^{M} \mathbf{p}_k^{UV} u_k v_k = \langle U, V\rangle_{S_{UV}},\end{aligned}$$

where $u_k = U(\lambda_k), v_k = V(\lambda_k)$. Thus

$$\begin{aligned}\Delta &= \langle A, B\rangle_{x_{AB}} - \langle C, B\rangle_{x_{CB}} \\ &= \langle A, B\rangle_{S_{AB}} - \langle C, B\rangle_{S_{CB}} = \sum_{k=1}^{M} (\mathbf{p}_k^{AB} a_k - \mathbf{p}_k^{CB} c_k) b_k\end{aligned}$$

and

$$\langle A, C\rangle_{x_{AC}} = \langle A, C\rangle_{S_{AC}} = \sum_{k=1}^{M} \mathbf{p}_k^{AC} a_k c_k.$$

We suppose now that *probabilities of* λ_k *do not depend on statistical ensembles:*

$$\mathbf{p}_k = \mathbf{p}_k^{AB} = \mathbf{p}_k^{CB} = \mathbf{p}_k^{AC} \tag{9.1}$$

(later we shall modify this condition to obtain statistical coincidence of probabilities, instead of the precise coincidence). Hence

$$\Delta = \sum_{k=1}^{M} \mathbf{p}_k (a_k - c_k) b_k \quad \text{and} \quad \langle A, C\rangle_{x_{AC}} = \sum_{k=1}^{M} \mathbf{p}_k a_k c_k.$$

We can now apply Theorem 8.1 for the discrete probability distribution $\{\mathbf{p}_k\}_{k=1}^{M}$ and obtain Bell's inequality.

However, if condition (9.1) does not hold true, then equality (8.2) and, as a consequence, Bell's inequality can be violated. The violation of condition (9.1) is the exhibition of unstable (with respect to the real metric) statistical structure on the level of hidden variables of (at least some) quantum ensembles. In particular, the principle of the statistical stabilization ('the law of the large numbers') can be violated for hidden variables: $\lim_{N\to\infty} \nu_N(\lambda_l)$ do not exist. Thus we could not introduce the probability distribution on the set of hidden labels Λ.[23]

Nevertheless, we obtained the following mathematical result:

Theorem 9.1. *Let statistical ensembles satisfy condition (9.1). Then Bell's inequality (8.1) holds true.*

We introduce now a statistical analogue of the precise coincidence of ensemble probabilities for hidden variables. Let $\mathcal{E}_1$, $\mathcal{E}_2$ be two ensembles of physical systems and let π be a property of elements of these ensembles. The π has values $(\alpha_1, \dots, \alpha_m)$. We define

$$\delta_\pi(\mathcal{E}_1, \mathcal{E}_2) = \sum_{i=1}^{M} |\mathbf{P}_{\mathcal{E}_1}(\alpha_i) - \mathbf{P}_{\mathcal{E}_2}(\alpha_i)|,$$

where $\mathbf{P}_{\mathcal{E}}(\alpha_i) = \frac{|\{d\in\mathcal{E}:\pi(d)=\alpha_i\}|}{|\mathcal{E}|}$ are ensemble probabilities. We remark that the function δ_π is a pseudometric on the set of all ensembles which elements have the property π: (1) $\delta_\pi(\mathcal{E}_1, \mathcal{E}_2) \geq 0$; (2) $\delta_\pi(\mathcal{E}_1, \mathcal{E}_2) = \delta_\pi(\mathcal{E}_2, \mathcal{E}_1)$; (3) $\delta_\pi(\mathcal{E}_1, \mathcal{E}_2) \leq \delta_\pi(\mathcal{E}_1, \mathcal{E}_3) + \delta_\pi(\mathcal{E}_3, \mathcal{E}_2)$. The distance $\delta_\pi(\mathcal{E}_1, \mathcal{E}_2) = 0$ iff ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$ have the same probability distribution of property $\pi : \mathbf{P}_{\mathcal{E}_1}(\alpha_i) = \mathbf{P}_{\mathcal{E}_2}(\alpha_i), i = 1, 2, \dots, m$.

In our model we set $\pi = \lambda$, hidden variables. The precise repeatability of the probability distribution of hidden variables (9.1) can be written as

$$\delta(S_{AB}, S_{CB}) = \delta(S_{AB}, S_{AC}) = 0,$$

where $\delta = \delta_\lambda$. Of course, we need not use such a precise coincidence in probabilistic considerations.

Theorem 9.2. *Let statistical ensembles satisfy condition*

$$\delta(S_{AB}, S_{CB}), \delta(S_{AB}, S_{AC}) \leq \epsilon.$$

Then Bell's inequality

$$|\langle A, B\rangle_{S_{AB}} - \langle C, B\rangle_{S_{CB}}| \leq (1 + 2\epsilon) - \langle A, C\rangle_{S_{AC}} \tag{9.2}$$

holds true.

[23] All our considerations were based on the statistical stabilization with respect to the real metric. In Chapter 4 we shall consider the statistical stabilization with respect to a p-adic metric. It may be that some ensembles of hidden variables which are unstable with respect to the real metric are stable with respect to the p-adic metric, see [80, 82, 83].

Proof. We have

$$\begin{aligned}|\Delta| &\le |\sum_{k=1}^{M} \mathbf{p}_k^{AB}(a_k - c_k)b_k| + |\sum_{k=1}^{M}(\mathbf{p}_k^{AB} - \mathbf{p}_k^{CB})c_k b_k| \\ &\le \epsilon + \sum_{k=1}^{M} \mathbf{p}_k^{AB}|a_k b_k|(1 - a_k c_k) \\ &\le (1+\epsilon) - \langle A, C\rangle_{S_{AC}} + \sum_{k=1}^{M} |\mathbf{p}_k^{AC} - \mathbf{p}_k^{AB}||a_k c_k| \\ &\le (1+2\epsilon) - \langle A, C\rangle_{S_{AC}}. \end{aligned}$$ □

We use the index N to denote the cardinality of a statistical ensemble. If probabilities $\mathbf{P}_{S_{UV}^N}(\lambda_k)$ stabilize when $N \to \infty$,

$$\lim_{N\to\infty} \mathbf{P}_{S_{UV}^N}(\lambda_k) = \mathbf{P}(\lambda_k),$$

then $\delta(S_{AB}, S_{CB}), \delta(S_{AB}, S_{AC}) \to 0$, $N \to \infty$. This imply precise Bell's inequality (8.1). On one hand, experimental violations of the latter inequality can demonstrate that *probabilities of hidden variables with respect to the ideal infinite ensemble do not exist at all (they fluctuate when $N \to \infty$)*. On the other hand, these violations can be a consequence of the fact that we do not know the value of a constant ϵ in (9.2). It might be that, despite of the stabilization of probabilities for $N \to \infty$, this constant is quite large for statistical ensembles which are used in quantum physics. In fact, 'right Bell's inequality' (9.2) could not be experimentally verified.

9.4 Stochastic hidden variables model, generalized Bell's inequality

Here we keep to f-realism. Thus, for each physical observable U, its value $U_f = U(\lambda)$ is the final value of U after a measurement. Such a result of a measurement depends not only on the value λ of hidden variables, but also on the state of an equipment $\mathcal{M}_U$ which is used for measuring of U. A measurement device $\mathcal{M}_U$ is a complex macroscopic system which state depends on the huge number of fluctuating parameters. Denote the ensemble of all possible states of $\mathcal{M}_U$ by the symbol Σ_U: $\Sigma_U = \{\omega_1^U, \dots, \omega_{L_U}^U\}$. The final value U_f of an observable U depends on both λ and ω:

$$U_f = U(\omega, \lambda).$$

We call such a model *stochastic hidden variables model.* Our definition of a stochastic hidden variables model differs from the standard one, see Section 9.5. The standard definition is strongly connected with Kolmogorov's model.

Let U and V be physical observables, $U, V = \pm 1$. We start again with the consideration of the frequency covariation $\langle U, V\rangle_{x_{UV}}$ with respect to a collective x_{UV} induced by the measurement of the pair (U, V). The x_{UV} is obtained by measurements for an ensemble S_{UV} of physical systems. Our aim is again to represent the experimental covariation $\langle U, V\rangle_{x_{UV}}$ as ensemble covariation $\langle U, V\rangle_{S_{UV}}$. Then we shall demonstrate that in the general case it is impossible to perform for ensemble covariations Bell's calculations, (8.2)–(8.4).

Let $S_{UV} = \{d_1, \dots, d_N\}$, where ith measurement is performed for the system d_i. Define functions $i \to \lambda(i)$ (the same function as above) and $i \to \omega^U(i)$, $i \to \omega^V(i)$, states of apparatus $\mathcal{M}_U$ and $\mathcal{M}_V$, respectively, at the instances, t_i^U and t_i^V, of measurements of U and V for ith system. We have

$$\langle U, V\rangle_{x_{UV}} = \frac{1}{N}\sum_{i=1}^{N} U(\omega^U(i), \lambda(i))V(\omega^V(i), \lambda(i)).$$

Set $D_{ks}^U = \{i : \lambda(i) = \lambda_k, \omega^U(i) = \omega_s^U\}$ and $D_{ks}^V = \{i : \lambda(i) = \lambda_k, \omega^V(i) = \omega_s^V\}$, $1 \le k \le M$, $1 \le s \le L_U$, $1 \le q \le L_V$. Set $l_{ksq}^{UV} = |D_{ks}^U \cap D_{kq}^V|$. It is evident that

$$\sum_{k=1}^{M}\sum_{s=1}^{L_U}\sum_{q=1}^{L_V} l_{ksq}^{UV} = N.$$

Hence

$$\langle U, V\rangle_{x_{UV}} = \frac{1}{N}\sum_{ksq} l_{ksq}^{UV} u_{ks} v_{kq},$$

where $u_{ks} = U(\omega_s^U, \lambda_k)$, $v_{kq} = V(\omega_q^V, \lambda_k)$. We show that $\langle U, V\rangle_{x_{UV}}$ can be represented as ensemble covariation for an appropriative ensemble of physical systems and states of measurement devices.

First we note that $\langle U, V\rangle_{x_{UV}} \neq \langle U, V\rangle_{\Lambda\times\Sigma_A\times\Sigma_B}$ (compare with Section 9.5). For the latter covariation, we have

$$\langle U, V\rangle_{\Lambda\times\Sigma_A\times\Sigma_B} = \frac{1}{ML_A L_B}\sum_{k=1}^{M}\sum_{s=1}^{L_U}\sum_{q=1}^{L_V} u_{ks} v_{kq}$$

and in general $\mathbf{P}_{\Lambda\times\Sigma_A\times\Sigma_B}(\lambda = \lambda_k, \omega^U = \omega_s^U, \omega^V = \omega_q^V) = \frac{1}{ML_A L_B} \neq \frac{l_{ksq}}{N}$ even approximately for $M, N, L_A, L_B \to \infty$.

It is also evident that $\langle U, V\rangle_{x_{UV}} \neq \langle U, V\rangle_{S_{UV}}$. The latter covariation is simply not well defined, because the 'properties' $\omega^U(i) = \omega_s^U$, $\omega^V(i) = \omega_q^V$ are not objective properties of elements of the ensemble S_{UV}. These 'properties' are determined by fluctuations of parameters in the apparatus $\mathcal{M}_U$ and $\mathcal{M}_V$.

To find the right ensemble, we have to introduce two new ensembles, namely, ensembles of states of the apparatus $\mathcal{M}_U$ and $\mathcal{M}_V$ (in the process of measurements for the ensemble of physical systems S_{UV}):

$$S_{\mathcal{M}_U} = \{\alpha_1^U, \dots, \alpha_N^U\}, \quad \alpha_j^U \in \Sigma_U, \qquad S_{\mathcal{M}_V} = \{\alpha_1^V, \dots, \alpha_N^V\}, \quad \alpha_j^V \in \Sigma_V,$$

where $\alpha_i^U = \omega^U(i), \alpha_i^V = \omega^V(i)$ are states of $\mathcal{M}_U$ and $\mathcal{M}_V$ at the instances of ith measurements. We set

$$\mathbf{S}_{UV} = \text{diag}(S_{UV} \times S_{\mathcal{M}_U} \times S_{\mathcal{M}_V}) = \{D_1, \ldots, D_N\}, \quad D_j = (d_j, \alpha_j^U, \alpha_j^V).$$

Then $\pi(D_j) = (\lambda(j), \omega^U(j), \omega^V(j))$ is an objective property of elements of the ensemble $\mathbf{S}_{UV}$ and

$$\langle U, V\rangle_{x_{UV}} = \langle U, V\rangle_{\mathbf{S}_{UV}} = \frac{1}{N}\sum_{i=1}^{N} U(\omega^U(i), \lambda(i))V(\omega^V(i), \lambda(i)).$$

We set

$$\begin{aligned}\mathbf{p}_{ksq}^{UV} &= \mathbf{P}_{\mathbf{S}_{UV}}(D_j : \pi(D_j) = (\lambda_k, \omega_s^U, \omega_s^V)) \\ &= \frac{|\{D_j \in \mathbf{S}_{UV} : \pi(D_j) = (\lambda_k, \omega_s^U, \omega_s^V)\}|}{|\mathbf{S}_{UV}|}.\end{aligned}$$

Hence we obtained that

$$\langle U, V\rangle_{x_{UV}} = \langle U, V\rangle_{\mathbf{S}_{UV}} = \sum_{ksq} \mathbf{p}_{ksq}^{UV}\, u_{ks} v_{kq}.$$

Thus in the general case we have

$$\begin{aligned}\Delta &= \langle A, B\rangle_{x_{AB}} - \langle C, B\rangle_{x_{CB}} = \langle A, B\rangle_{\mathbf{S}_{AB}} - \langle C, B\rangle_{\mathbf{S}_{CB}} \\ &= \sum_{ksq} \mathbf{p}_{ksq}^{AB}\, a_{sk} b_{kq} - \sum_{ksq} \mathbf{p}_{ksq}^{CB}\, c_{ks} b_{kq}\end{aligned}$$

and

$$\langle A, C\rangle_{x_{AC}} = \langle A, C\rangle_{\mathbf{S}_{AC}} = \sum_{ksq} \mathbf{p}_{ksq}^{AC}\, a_{ks} c_{kq}.$$

We suppose now that probabilities $\mathbf{p}_{ksq}^{UV}$ do not depend on ensembles:

$$\mathbf{p}_{ksq} = \mathbf{p}_{ksq}^{AB} = \mathbf{p}_{ksq}^{CB} = \mathbf{p}_{ksq}^{AC}. \tag{9.3}$$

In particular, we suppose that all measurement devices have the same set of states (of parameters):

$$\Sigma = \Sigma_A = \Sigma_B = \Sigma_C \ \text{ (and } \ L = L_A = L_B = L_C\text{)}. \tag{9.4}$$

Then we obtain

$$\Delta = \sum_{ksq} \mathbf{p}_{ksq}(a_{ks} - c_{ks}) b_{kq}.$$

However, we could not repeat trick (8.3) of the proof of Bell's inequality. The equality $a_{ks}^2 = 1$ does not give the possibility to proceed the proof. Of course, we have

$$|\Delta| = |\sum_{ksq} \mathbf{p}_{ksq}(a_{ks} - a_{ks}^2 c_{ks})b_{kq}| \leq \sum_{ksq} \mathbf{p}_{ksq}|a_{ks}b_{kq}|(1 - a_{ks}c_{ks})$$
$$\leq 1 - \sum_{ksq} \mathbf{p}_{ksq} a_{ks} c_{ks}.$$

But in general $\sum_{ksq} \mathbf{p}_{ksq} a_{ks} c_{ks}$ is not larger than $\langle A, C\rangle_{x_{AC}} = \sum_{ksq} \mathbf{p}_{ksq} a_{ks} c_{kq}$.

Therefore, if we keep to f-realism, even stability condition (9.3) (for combined ensembles of physical systems and states of measurement apparatus) does not imply Bell's inequality. A new source of violation of Bell's inequality is the *inconsistency* of random fluctuations for two measurement devices $\mathcal{M}_U$ and $\mathcal{M}_V$. In general $\omega^U(i) \neq \omega^V(i)$.

Suppose that it could be possible to control states of $\mathcal{M}_U$ and $\mathcal{M}_V$ and choose ω for $\mathcal{M}_U$ and $\mathcal{M}_V$ in the consistence way:

$$\omega = \omega^U(i) = \omega^V(i).$$

Then the ensemble $\mathbf{S}_{UV}$ would contain only triples of the form $(\lambda_k, \omega_s, \omega_s)$ and

$$\mathbf{p}_{ksq}^{UV} = \mathbf{P}_{\mathbf{S}_{UV}}(\lambda_k, \omega_s^U, \omega_q^V) = 0, \quad s \neq q. \tag{9.5}$$

In such a case we obtain covariations:

$$\langle U, V\rangle_{\text{Ideal}} = \frac{1}{N}\sum_{i=1}^{N} U(\omega^U(i), \lambda(i))V(\omega^V(i), \lambda(i)) = \sum_{ks} \mathbf{p}_{ks}^{UV} u_{ks} v_{ks},$$

where $\mathbf{p}_{ks}^{UV} = \mathbf{p}_{kss}^{UV}$. If we also suppose the validity of (9.3), we obtain

$$|\Delta_{\text{Ideal}}| = |\sum_{ks} \mathbf{p}_{ks}(a_{ks} - c_{ks})b_{ks}|$$
$$\leq 1 - \sum_{ks} \mathbf{p}_{ks} a_{ks} c_{ks} = 1 - \langle A, C\rangle_{\text{Ideal}}.$$

However, ideal covariations have no direct connection to experimental frequency covariations.

Nevertheless, we can formulate the following mathematical theorem:

Theorem 9.3. *Let statistical ensembles (physical systems/measurement apparatus) satisfy conditions (9.3) and (9.5). Then Bell's inequality (8.1) holds true for covariations with respect to these ensembles.*

Therefore, to obtain Bell's inequality in the framework of f-realism, we have to suppose: (1) statistical repeatability of ensemble distribution of hidden variables λ in ensembles which are used for measurements; (3) statistical repeatability of fluctuations of states ω in ensembles of an equipment; (3) consistency of fluctuations of all measurement devices.

If the reader even deny the possibility of violations of (1) or (2), he must agree that condition (3) seems to be nonphysical: we could never control fluctuations of the huge number of parameters in the equipment.

Instead of precise coincidence (9.3), it is possible to consider (under the assumption (9.4)) the statistical coincidence based on the quantity:

$$\delta(\mathbf{S}_{AB}, \mathbf{S}_{CB}) = \sum_{k=1}^{M} \sum_{s=1}^{L} \sum_{q=1}^{L} |\mathbf{p}_{ksq}^{AB} - \mathbf{p}_{ksq}^{CB}|.$$

Here $\delta = \delta_\pi$ for the property $\pi(i) = (\lambda(i), \omega^U(i), \omega^V(i))$. We remark that condition (9.3) of the precise coincidence can be written as

$$\delta(\mathbf{S}_{AB}, \mathbf{S}_{CB}) = 0$$

for every two pairs of observable (A, B) and (C, B). We also introduce a new quantity which is a statistical measure of inconsistency of ensembles $S_{\mathcal{M}_U}$ and $S_{\mathcal{M}_V}$:

$$\sigma(\mathbf{S}_{UV}) = \sum_{s \neq q} \mathbf{P}_{\mathbf{S}_{UV}}(\omega^U = \omega_s, \omega^V = \omega_q) = \sum_{k} \sum_{s \neq q} \mathbf{p}_{ksq}^{UV}.$$

Condition (9.5) of the precise consistency for states of $\mathcal{M}_U$ and $\mathcal{M}_V$ can be written in the form:

$$\sigma(\mathbf{S}_{UV}) = 0.$$

Theorem 9.4. *Let statistical ensembles (physical systems/measurement apparatus) satisfy conditions:*

$$\delta(\mathbf{S}_{AB}, \mathbf{S}_{CB}), \delta(\mathbf{S}_{AB}, \mathbf{S}_{AC}) \leq \epsilon \quad \text{and} \quad \sigma(\mathbf{S}_{AB}), \sigma(\mathbf{S}_{CB}), \sigma(\mathbf{S}_{AC}) \leq \epsilon'.$$

Then inequality

$$|\langle A, B\rangle_{\mathbf{S}_{AB}} - \langle C, B\rangle_{\mathbf{S}_{CB}}| \leq (1 + 2\epsilon + 3\epsilon') - \langle A, C\rangle_{\mathbf{S}_{AC}}$$

holds true.

Proof. We have

$$\begin{aligned}
|\Delta| &\leq \epsilon + |\sum_{ksq} \mathbf{p}_{ksq}^{AB}(a_{ks} - c_{ks})b_{kq}| \\
&\leq \epsilon + 2\epsilon' + \sum_{ks} \mathbf{p}_{ks}^{AB}|(a_{ks} - c_{ks})b_{ks}| \leq \epsilon + 2\epsilon' + \sum_{ks} \mathbf{p}_{ks}^{AB}(1 - a_{ks}c_{ks}) \\
&\leq \epsilon + 4\epsilon' + \sum_{ksq} \mathbf{p}_{ksq}^{AB}(1 - a_{ks}c_{kq}) \leq (1 + 2\epsilon + 4\epsilon') - \sum_{ksq} \mathbf{p}_{ksq}^{AC} a_{ks}c_{kq}. \qquad \square
\end{aligned}$$

9.5 Right choice of probability distributions for stochastic hidden variables models

Typically stochastic hidden variables models are defined as models with probabilities ($\epsilon = \pm 1$)

$$\mathbf{P}(U = \epsilon) = \int_\Lambda \mathbf{P}(U = \epsilon/\lambda)\, d\rho(\lambda), \tag{9.6}$$

where $\rho(\lambda)$ is the probability distribution of hidden variables and $\mathbf{P}(U = \epsilon/\lambda)$ is the conditional probability to measure the value $U = \epsilon$ for the quantum system having the hidden state λ. Then the joint probability distribution can be defined (at least mathematically) as

$$\begin{aligned}\mathbf{P}(U_1 = \epsilon_1, U_2 = \epsilon_2, U_3 = \epsilon_3)\\ = \int_\Lambda \mathbf{P}(U_1 = \epsilon_1/\lambda)\mathbf{P}(U_2 = \epsilon_2/\lambda)\mathbf{P}(U_3 = \epsilon_3/\lambda)\, d\rho(\lambda).\end{aligned} \tag{9.7}$$

In fact, to derive Bell's inequality in the Kolmogorov framework, it is sufficient to use the existence (on the mathematical level) of the joint probability distribution (9.7). However, considerations in the framework of the ensemble probability theory demonstrated that 'probabilities' (9.6) has no physical meaning. These are probabilities with respect to the ensemble $\Lambda \times \Sigma_U$. However, physical probabilities are probabilities with respect to the ensemble $\mathbf{S}_U = \text{diag}(S_U \times S_{\mathcal{M}_U})$, where $S_U = \{d_1, \ldots, d_N\}$ is the ensemble of quantum system used in the measurement.

9.6 Individual and ensemble nonreproducibilities

Our hypothesis on the nonreproducibility of the probability distribution of hidden variables in statistical ensembles used in quantum experiments is related to De Baere's [46], [47] hypothesis on the *individual nonreproducibility.* He mentioned that there are reasons (see also [54]) that it would be impossible to prepare the quantum system with the same value λ for measurements of different observables. Thus probabilities $\mathbf{P}(U_j = \epsilon_j/\lambda)$, $j = 1, 2, 3$, could not be defined for the same λ. The latter implies that joint probability distribution (9.7) does not exist and Bell's inequality could not be derived. We remark that deterministic hidden variables models satisfy the condition of the individual reproducibility. However, the ensemble reproducibility can be violated.

9.7 Other probabilistic models which do not contradict to local realism

L. Accardi [2] used non-Kolmogorovean model without Bayes' formula to eliminate Bell's inequality from considerations related to spin's model. Recently he also developed a new non-Kolmogorovean model which gives an explanation of violations of Bell's inequality, see [3].

I. Pitowsky [148], [149] discussed the possibility that some nonmeasurable sets can be physical events, i.e, some physical observables may be nonmeasurable. There is no

Bell's inequality in this approach. Thus there is no problem with violations of Bell's inequality. This model is consistent with known polarization phenomena and the existence of macroscopic magnetism. He also proposed a thought experiment which indicates a deviation from the predictions of quantum mechanics. We noted that already A. N. Kolmogorov discussed 'generalized probabilities' on the algebra F_Ω of all subsets of Ω. Pitowsky discussed the relation of 'Banach–Tarski paradox' (Theorem 5.1, Chapter 1) to foundations of probability theory. It seems that Kolmogorov suspected that 'nonmeasurable events' could play some role in probability theory. The model of Pitowsky gives the interesting application of such 'generalized probabilities.' I. Pitowsky noticed:

"This so called 'Banach–Tarski paradox' is not a paradox at all. The pieces into which the ball is cut are nonmeasurable sets, that is, one cannot assign them numbers that indicate their volume since this will clearly violate the additivity or invariance of 'volume'. In spite of this explanation and in spite of independent proofs that nonmeasurable sets exist, the Banach-Tarski result was taken as an unfortunate consequence of the axiom of choice (which is nevertheless, essential in some fields of 'good' mathematics). Suppose, however, that we reverse this attitude and maintain that the subsets into which the ball is decomposed *exist in physical reality*. These hidden pieces could be detected in two 'states'. The first is a 'one-ball state' and the second a 'two-ball state'. In each state the pieces do have a 'volume' which depends, however, on their mutual configuration. Assume that we have a source that emits five balls in the first state. On the way from the source to a counter two of the balls spontaneously transform to the second state. The counter, which does not distinguish between the states, will detect seven ball. This rather simplistic example serves to indicate that one can 'perform miracles' if one is willing to accept the physical reality of some highly abstract set-theoretical objects. In particular, if such assumptions are made, it is possible to account for interference effects in a completely mechanistic way without introducing wavelike nonlocal components to the theory.

Mathematicians, in particular applied mathematicians, where reluctant to take nonmeasurable sets seriously. As a result there exists no mathematical theory that relates nonmeasurable distributions with relative frequencies."

Such an extension of probability theory was created by I. Pitowsky and then strongly mathematically improved by S. P. Gudder [66]. He introduced the concept of a probability manifold M. The global properties of M inherited from its local structure were then considered. It was shown that a deterministic spin model due to Pitowsky falls within this general framework. Finally, Gudder constructed a phase-space model for nonrelativistic quantum mechanics. These two models give the same global description as conventional quantum mechanics. However, they also give a local descriptions which is not possible in conventional quantum mechanics.

Remark 9.1. Non-Kolmogorovean probabilistic models of Accardi, Pitowsky and Gudder have higher level of abstraction than the original Kolmogorov model. This is one of explanations why these models are not so popular in quantum physics. On the other hand, we showed that Bell's inequality does not contradict to local realism on the basis of the primary (rather primitive from mathematical viewpoint) probabilistic models, namely, the ensemble and frequency models. It seems that our models have more close relation to physical reality.

We shall discuss in Chapter 3 the use of negative probabilities and in Chapter 4 the use of p-adic probabilities to eliminate Bell's inequality from considerations.

Conclusion. 'Bell's inequality' does not imply nonexistence of local hidden variables. Physical reality may be nonlocal. It may be a realistic description of quantum phenomena is impossible. However, both these features of physical reality could not be determined via Bell's inequality.

3 Negative probabilities

In this chapter we study possibilities to extend the probability theory to describe numerous physical models with negative probabilities. Of course, negative probabilities could not appear in Kolmogorov's probability theory in that the probabilities of events must be **positive real numbers.** Therefore we have to turn back to the original probability formalisms, namely, ensemble and frequency.

1 The origin of negative probabilities in the ensemble and frequency theories

1.1 Ensemble approach: fluctuations of finite approximations

In the ensemble framework negative probabilities could not appear for finite statistical ensembles $S_N = \{s_1, s_2, \ldots, s_N\}$. However, such generalized probabilities can naturally appear for infinite statistical ensembles S as the results of the limit procedure:

$$\mathbf{P}_S(A = \alpha) = \lim_{N\to\infty} \frac{|S(A = \alpha) \cap S_N|}{|S|} , \tag{1.1}$$

where a sequence of finite ensembles $\{S_N\}$ gives an approximation of the infinite ensemble S. If this limit does not exist in $\mathbb{R}$, then some regularization procedures (for example, the summation of divergent series or integrals) can induce negative values for $\mathbf{P}_S(A = \alpha)$. Of course, in such a situation it would be natural to leave the domain of real analysis and consider some non-Archimedean number systems which contain actual infinities. In this case the probability $\mathbf{P}_S(A = \alpha)$ can be defined directly as the proportion:

$$\mathbf{P}_S(A = \alpha) = \frac{|S(A = \alpha) \cap S_N|}{|S|} . \tag{1.2}$$

In Chapter 4 we shall use the system of p-adic numbers $\mathbb{Q}_p$ for such a purpose (another natural possibility is to use nonstandard numbers, [10]). In $\mathbb{Q}_p$ proportion (1.2) can be a negative rational number (as well as a rational number which is larger than 1).

1.2 Ensemble approach: split of conventional probabilities

Nonexistence of limits (1.1) is not the unique source of negative probabilities for infinite ensembles S. It may be situations (see, for example, the p-adic framework) such that limit (1.1) (for some α) exists and equal to zero (from the viewpoint of the real analysis). For example, for the uniform distribution on $S = \mathbb{N}$, we have $\mathbf{P}_S(A = n) = \lim_{N\to\infty} \frac{1}{N} = 0$ for all $n = 1, 2, \ldots$. However, some regularization of this limit procedure can produce nonzero coefficients $\mathbf{P}_S^{\mathrm{reg}}(\alpha)$. In the mentioned

p-adic framework such coefficients (defined by (1.1) with respect to the p-adic topology or directly by (1.2) with the aid of actual infinities) are always negative (rational) numbers[1]. Thus regularizations of (1.1) can induce the split of zero conventional probabilities in a set of new labels which can be negative numbers. These new labels can be interpreted as infinitely small probabilities. Such a split of conventional probabilities is not a feature of only zero probabilities. For example, probability one can be also split in a set of new labels which are interpreted as probabilities which differ from probability one by infinitely small probabilities. These are 'practically one probabilities'. In all p-adic examples such new probabilities are given by rational numbers which are larger than one[2]. Similar splits can be obtained for other rational probabilities $q \in (0, 1)$. If $0 < q < 1$, $q \in \mathbb{Q}$, then we have two sets of labels $L_{<q}$ and $L_{>q}$. They denote, respectively, probabilities $a = q - \lambda$ and $a = q + \lambda$, where λ is infinitely small probability. In p-adic examples we have $L_{<q} \subset \mathbb{Q} \cap (-\infty, 0)$ and $L_{>q} \subset \mathbb{Q} \cap (1, +\infty)$ (see Chapter 4 for the details).

On one hand, probabilities $q < 0$ (and $q > 1$) demonstrate irregular behaviour ($N \to \infty$) of approximations of probabilities $\mathbf{P}_S$ with respect to an infinite ensemble S by probabilities $\mathbf{P}_{S_N}$ with respect to finite sub-ensembles S_N. On the other hand, they can describe the fine internal structure of S (via split of conventional probabilities). We note that from the physical point of view the irregularity of approximations means that it is impossible to prepare for all measurements for a quantum state ϕ (describing S) finite ensembles S_N with identical statistical properties.

1.3 Frequency approach: irregularity of behaviour of frequencies

In the frequency framework negative probabilities could not appear in the classical theory of R. von Mises which is based on the principle of the statistical stabilization of frequencies with respect to the real metric. However, if we assume that for some 'quasi-random sequences' $x = (x_1, x_2, \ldots, x_n, \ldots)$ this principle can be violated, namely, the limit

$$\mathbf{P}_x(\alpha) = \lim_{N \to \infty} \nu_N(\alpha; x) \tag{1.3}$$

does not exist in $\mathbb{R}$, then some regularization procedures $\mathcal{R}_{\text{fr}}$ for (1.3) can produce negative values (as well as values which are larger than 1) for $\mathbf{P}_x$. One of the possibilities for such a regularization is to change the topology on the set of rational numbers $\mathbb{Q}$ in that we study the convergence of relative frequencies. In Chapter 4 we shall use the p-adic topology for such a purpose.

[1] In fact, we could not prove such a general theorem in the framework of p-adic analysis. But numerous examples demonstrate this feature of the p-adic split of zero conventional probabilities.

[2] This is natural: if $\mathbf{P}(A) = q < 0$ is infinitely small probability, then $\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A) = 1 - q > 1$ is probability which negligibly differs from 1 and vice versa.

1.4 Frequency approach: split of Mises' probabilities

Another source of frequency probabilities $q < 0$ and $q > 1$ is the split of Mises' probabilities. For example, the fact that frequency probability $\mathbf{P}_x^{\text{Mises}}(A) = \lim_{n\to\infty} \nu_n(A;x) = 0$ does not imply that the event A should never occur. Therefore it is reasonable to take such events into account by using new labels.

Let us consider two events A and B which have zero frequency probabilities:

$$\mathbf{P}_x^{\text{Mises}}(A) = \lim_{n\to\infty} \nu_n(A;x) = 0, \quad \mathbf{P}_x^{\text{Mises}}(B) = \lim_{n\to\infty} \nu_n(B;x) = 0, \tag{1.4}$$

in $\mathbb{R}$. We are interested in the problem: What event, A or B, has larger probability? Of course, this question is meaningless from the viewpoint of the Mises' probability theory. However, this problem can be solved by extending the set of labels for probabilities.

In the frequency framework we can obtain new sets of labels automatically by using new topologies for the statistical stabilization (by finding limits (1.3) with respect to new topologies)[3]. Each topology of the statistical stabilization induces its own set of labels for split Mises' probabilities. For example, it may be that, despite of (1.4) in $\mathbb{R}$, we have

$$\mathbf{P}_x^{\tau}(A) = \lim_{n\to\infty} \nu_n(A;x) \neq 0, \quad \mathbf{P}_x^{\tau}(B) = \lim_{n\to\infty} \nu_n(B;x) \neq 0 \tag{1.5}$$

for some topology τ on $\mathbb{Q}$. If we choose the p-adic topology $\tau = \tau_p$, then in examples studied by the author p-adic probabilities (1.5) are represented by negative rational numbers. Thus by using negative probabilities we can split zero (Mises') probability. The same split can be obtained for all Mises' probabilities $q \in [0, 1] \cap \mathbb{Q}$.

On one hand, probabilities $q < 0$ and $q > 1$ demonstrate the violation of the principle of the statistical stabilization (the law of large numbers) for some 'quasi-random' sequences. On the other hand, they describe (with the aid of new topologies on $\mathbb{Q}$) the fine internal structure of some Mises' collectives.

1.5 Where are negative probabilities?

However, the reader may ask: Why could we not find negative probabilities in physical experiments? One of reasons is that, in fact, we have never tried to find them. All our experimental methodology is based on the principle of the statistical stabilization (the law of large numbers). All experiments are prepared in such circumstances that relative frequencies must stabilize. This is the result of our cognitive evolution. In the process of evolution the brain extracted from the chaotic and (lawless) reality phenomena which satisfy the principle of the statistical stabilization (repeatability in the average). These and only these phenomena are considered by the brain as real physical

[3] The real topology is only one of many topologies on the set of rational numbers $\mathbb{Q}$ which contains frequencies $\nu_N = n/N$.

phenomena. Negative probabilities give the possibility to extend the range of physical phenomena by considering phenomena which violate the principle of the statistical stabilization. Another reason of the absence of negative probabilities in the experimental framework is the common use of real analysis for the study of the experimental statistical data. However, this data is always rational and in principle other topologies on $\mathbb{Q}$ (different from the real one) can be used for studying of this data. In particular, we have to pay more attention to events A with zero conventional (Kolmogorov or Mises) probabilities, $\mathbf{P}^{\mathrm{Conv}}(A) = 0$. From our viewpoint such events are not less physical than events with positive probabilities. By using negative probabilities we can consider in analytical calculations events A such that $\mathbf{P}^{\mathrm{Conv}}(A) = 0$. In this way we can clarify the hidden internal structure of some events B with positive conventional probabilities. We shall study this question carefully in the next section.

1.6 The formula of total probability as an average procedure

We consider a quantum measurement for quantum systems prepared in a state ϕ. We suppose that each quantum system s which is taken for this measurement has a hidden state λ which determines (with some probability) a result of the measurement for the s (see Chapter 2). The set of hidden states is denoted by Λ. The number of hidden states may be infinite.

Remark 1.1. Of course, in a laboratory we can produce only a finite ensemble $S_N = \{s_1, \ldots, s_N\}$ of quantum systems which have a finite number of hidden states $\lambda_1, \ldots, \lambda_n$, $n \leq N$. However, different finite ensembles $S_N, \tilde{S}_M, \ldots$ are used in different experiments. It is natural to assume that these finite ensembles are subensembles of one infinite ensemble S. The quantum state ϕ describes this infinite ensemble. The infinite cardinality of S induces the impression that S is just an ideal mathematical abstraction. However, suppose, for example, that each electron s has the extremely complex internal structure. Then, in fact, each s must be described by its own (individual) internal state λ. In this case the number of all possible states (for all electrons in the universe) is really infinite.

The (hidden) probability for λ in S is denoted by the symbol $\mathbf{p}_\lambda$. On the basis of our previous considerations (see Chapter 2) it is natural to suppose that some of $\mathbf{p}_\lambda$ may be nonconventional probabilities; in particular, they may be negative[4].

[4] In particular, they may be infinitely small probabilities. For example, if each electron in the universe has its own state λ, then $\mathbf{p}_\lambda = \lim \frac{1}{N} = 0$ (from the viewpoint of real analysis). Negativity of $\mathbf{p}_\lambda$ can also be a consequence of the violation of the law of large numbers. Such a violation for hidden states λ is quite natural if $|\Lambda| = \infty$. For the concrete λ, behaviour of frequencies $\nu_N(\lambda; x)$ can strongly depend on a sample x. There are no reasons to assume that two different samples of quantum systems $S_N = \{s_1, \ldots, s_N\}$ and $\tilde{S}_M = \{\tilde{s}_1, \ldots, \tilde{s}_M\}$ must produce samples $x = (\lambda_1, \ldots, \lambda_N)$ and $\tilde{x} = (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_M)$ having the same probability distribution (because our macro equipment could not control statistical behaviour of hidden parameters).

In the process of a measurement each state λ is transformed into a new state λ' (due to an interaction between the quantum system and the equipment). Denote probabilities of this transition by $\mathbf{p}_{\lambda\lambda'}$. Some of these probabilities can be negative (in particular, the law of large numbers can be violated for some transitions $\lambda \to \lambda'$). In the measurement we observe events A consisting of some sets of states λ' (in principle these sets can be infinite). By the formula of total probability we obtain:

$$\mathbf{P}(A) = \sum_{\lambda \in \Lambda} \mathbf{p}_\lambda \sum_{\lambda' \in A} \mathbf{p}_{\lambda\lambda'} \,. \tag{1.6}$$

In fact, this is the average procedure with respect to the ensemble Λ of hidden states λ, transitions $\lambda \to \lambda'$ and states λ' which are identified in the observed event A. The result of this procedure can be a conventional (Kolmogorov or Mises) probability, despite of the possibility that some of probabilities $\mathbf{p}_\lambda, \mathbf{p}_{\lambda\lambda'} < 0$ or $\mathbf{p}_\lambda, \mathbf{p}_{\lambda\lambda'} > 1$.

The ensemble and frequency explanations of this phenomenon have been already presented in Section 4, Chapter 2. For example, in the frequency framework fluctuations of frequencies $\nu_N(\lambda)$ and (or) $\nu_N(\lambda'/\lambda)$ can compensate each other and produce the statistical stabilization. Examples 4.1 and 4.2 showed that such a behaviour can be demonstrated even in the case of a finite set Λ. Thus one of the sources of conventional probabilities in (1.6) is that simultaneous (chaotic) fluctuations can produce in average the statistical stabilization. Another source are infinite statistical ensembles with infinitely small initial probabilities $\mathbf{p}_\lambda < 0$ and (or) transition probabilities $\mathbf{p}_{\lambda\lambda'} < 0$. Infinite sums of infinitely small (negative) probabilities might produce conventional positive probabilities.

In all previous considerations the formula of total probability must be regularized via some procedure (for example, by using a new number system to find the limits of fluctuating frequencies, see Chapter 4). In general we could not even suppose the validity of the Bayes' formula (even for one fixed state λ and transition $\lambda \to \lambda'$).

Example 1.1. Example 4.1 (Chapter 2) can be generalized by considering the infinite set of hidden states $\lambda \in [0, \pi]$. We choose the uniform probability distribution on $[0, \pi]$ as the initial probability distribution $\mathbf{p}_\lambda$ (these are infinitely small probabilities). However, in the framework of real analysis we could not represent $\mathbf{p}_\lambda$ as proportional probabilities (1.2). The only thing which we can do is to use normalized Lebesgue measure on $[0, \pi]$ to represent $\mathbf{p}_\lambda$. Let us consider an observable $B = 0, 1$ $(\lambda' = B)$ with conditional frequencies

$$\nu_k(0/\lambda) \approx \sin^2 k\lambda, \quad \nu_k(1/\lambda) \approx \cos^2 k\lambda, \quad k \to \infty,\ \lambda \in [0, \pi].$$

If $\lambda \neq \pi l$, $l = 0, 1, 2, \ldots$, then conditional frequency probabilities $\mathbf{P}^{\mathrm{fr}}(B = 0/\lambda) = \lim_{k\to\infty} \sin^2 k\lambda$ and $\mathbf{P}^{\mathrm{fr}}(B = 1/\lambda) = \lim_{k\to\infty} \cos^2 k\lambda$ do not exist. But the average procedure based on the (integral) formula of total probability gives well-defined conventional probabilities for values of B:

$$\mathbf{P}(B=0)=\lim_{k\to\infty}\int_0^\pi \sin^2 k\lambda\, d\mathbf{p}_\lambda = \frac{1}{2},$$

$$\mathbf{P}(B=1)=\lim_{k\to\infty}\int_0^\pi \cos^2 k\lambda\, d\mathbf{p}_\lambda = \frac{1}{2}.$$

In Section 3 we shall study examples in that nonexistence of conventional conditional probabilities implies negativity of generalized conditional probabilities.

Example 1.2. The previous example can be easily modified to obtain a model in that probabilities $\mathbf{P}^{\mathrm{fr}}(\lambda)=\mathbf{p}_\lambda$ do not exist. Let $d\nu_k(\lambda)\approx \frac{2}{\pi}\sin^2 k\lambda\; d\lambda, k\to\infty$, and let $\nu_k(B=\beta/\lambda)\approx\frac{1}{2}, k\to\infty$. Then the frequency probability distribution $\mathbf{p}_\lambda$ do not exist. But via the formula of total probability we obtain in the average:

$$\mathbf{P}(B=0)=\lim_{k\to\infty}\frac{1}{2}\int_0^\pi d\nu_k(\lambda)=\frac{1}{2},$$

$$\mathbf{P}(B=1)=\lim_{k\to\infty}\frac{1}{2}\int_0^\pi d\nu_k(\lambda)=\frac{1}{2}.$$

1.7 Negative probabilities and the principle of complementarity

The considerations of the previous section on the formula of total probability as an average procedure are based on ideas of P. Dirac [55] and R. Feynman [59]. In particular, R. Feynman considered a roulette which has two internal (non-observed) states λ_1 and λ_2 and three observed states 1,2,3. By simple numerical examples (that the reader can produce by himself) he demonstrated that observed events can have positive conventional probabilities $\mathbf{p}_j > 0$, $j = 1,2,3$, despite of negativity of some hidden probabilities $\mathbf{p}_{\lambda_1}, \mathbf{p}_{\lambda_2}$ or conditional probabilities $\mathbf{p}_{\lambda_1 j}, \mathbf{p}_{\lambda_2 j}$, $j = 1,2,3$. However, neither Dirac nor Feynman could propose a mathematical explanation of the origin of negative probabilities (they considered negative probabilities as just formal quantities which could be useful in some calculations). I have found the frequency and ensemble roots of negative probabilities. For example, we can build Feynman's roulette by using 'quasi-random' generators for states λ_1 and λ_2 or for transitions $\lambda_1\to j$ and $\lambda_2\to j$ which simulate the statistical models of Examples 4.1 and 4.2 (Chapter 2), respectively.

On the basis of our interpretation of negative probabilities it would be interesting to discuss the idea of R. Feynman on a connection between negative probabilities and the principle of complementarity in quantum mechanics, see [59]. As I could understood, R. Feynman is an adherent of i-realism (at least in this paper).

In the framework of Subsection 1.6 we consider two physical properties A and B. Thus (despite of possible fluctuations of frequencies and conditional frequencies for hidden variables) frequencies

$$\nu_N(A_\alpha) = \sum_{\lambda\in\Lambda} \nu_N(\lambda) \sum_{\lambda'\in A_\alpha} \nu_N(\lambda'/\lambda), \qquad \text{where } A_\alpha = \{A = \alpha\}, \tag{1.7}$$

$$\nu_N(B_\beta) = \sum_{\lambda\in\Lambda} \nu_N(\lambda) \sum_{\lambda'\in B_\beta} \nu_N(\lambda'/\lambda), \qquad \text{where } B_\beta = \{B = \beta\}, \tag{1.8}$$

stabilize (when $N \to \infty$) to conventional probabilities $\mathbf{P}^{\text{Conv}}(A_\alpha), \mathbf{P}^{\text{Conv}}(B_\beta)$.

However, in general there are no reasons to suppose that the frequency

$$\nu_N(A_\alpha \cap B_\beta) = \sum_{\lambda\in\Lambda} \nu_N(\lambda) \sum_{\lambda'\in A_\alpha\cap B_\beta} \nu_N(\lambda'/\lambda) \tag{1.9}$$

also stabilize (when $N \to \infty$)[5]. If (1.9) does not stabilize, then conventional probability $\mathbf{P}^{\text{Conv}}(A = \alpha, B = \beta)$ is not defined.

Remark 1.2. Suppose that we could find some procedure $\mathcal{R}_{\text{fr}}$ to regularize fluctuating frequencies $\nu_N(\lambda)$ and (or) $\nu_N(\lambda'/\lambda)$. By $\mathcal{R}_{\text{fr}}$ we obtain generalized probabilities $\mathbf{p}_\lambda$ and (or) $\mathbf{p}_{\lambda\lambda'}$ (which in principle can be negative numbers). Suppose that (in the case of the infinite set Λ) we could find some procedure $\mathcal{R}_{\text{conv}}$ to regularize (probably diverging) series

$$\sum_{\lambda\in\Lambda} \mathbf{p}_\lambda \sum_{\lambda'\in A_\alpha} \mathbf{p}_{\lambda\lambda'}\,, \sum_{\lambda\in\Lambda} \mathbf{p}_\lambda \sum_{\lambda'\in B_\beta} \mathbf{p}_{\lambda\lambda'} \tag{1.10}$$

in such a way that their sums coincide with conventional probabilities $\mathbf{P}^{\text{Conv}}(A_\alpha)$ and $\mathbf{P}^{\text{Conv}}(B_\beta)$, respectively. We now apply $\mathcal{R}_{\text{conv}}$ to series

$$\sum_{\lambda\in\Lambda} \mathbf{p}_\lambda \sum_{\lambda'\in A_\alpha\cap B_\beta} \mathbf{p}_{\lambda\lambda'}\,. \tag{1.11}$$

In principle there may be different variants: (1) the procedure $\mathcal{R}_{\text{conv}}$ does not work for series (1.11); here we could not assign any real number to (1.11); (2) despite of fluctuations of frequencies (1.9), the $\mathcal{R}_{\text{conv}}$ still works for series (1.11) and gives a real number; but this number is not related to the statistical limit of frequencies (1.9) (in particular, it may be a negative number).

This simple statistical consideration explains the origin of difficulties with 'simultaneous existence' of incompatible properties of quantum systems. Therefore the presence of incompatible properties does not demonstrate some essentially new 'quantum'

[5]If $|\Lambda| = \infty$, then events A_α and B_β may differ rather slightly: $\nu_N(A/\lambda) \approx \nu_N(B/\lambda)$ for each $\lambda \in \Lambda$. But the infinite average over Λ can produce behaviour of frequencies (1.9) which essentially differs from behaviour of frequencies (1.7) and (1.8).

properties of reality. It only demonstrates that the law of large numbers is violated for internal (hidden) properties of so called quantum systems (mainly because we could not control the statistical behaviour of these properties in our (macro) preparation procedures). For some events fluctuations on the microlevel can compensate each other and produce the statistical stabilization of observed frequencies (1.7) and (1.8). At the present time such events are called physical events. For other events fluctuations on the microlevel cannot compensate each other; there is no statistical stabilization of observed frequencies (1.9). At the present time such events are called nonphysical.

There are also no reasons to suppose that (in general generalized) initial probability distribution $\mathbf{p}_\lambda$ and conditional probabilities $\mathbf{p}_{\lambda\lambda'}$ can be chosen in such a way that fluctuations in both expressions (1.7) and (1.8) could be compensated so that, for some values $A = \alpha_0$ and $B = \beta_0$, both frequencies $\nu_N(A_{\alpha_0})$ and $\nu_N(B_{\beta_0})$ stabilize to probability 1. This is nothing than the statistical explanation of the principle of the complementarity. It seems that (rather unclear) considerations of R. Feynman [59] can be interpreted in such a way.

Thus we proposed the purely statistical explanation of the phenomenon of incompatibility for some quantum observables. Here the problem of disturbance effects of measurements is totally excluded from considerations. Our approach implies that even the possibility to perform measurements on quantum systems without any disturbance effect would not imply that incompatible properties can be measured simultaneously[6]. Different structure of sets $\{\lambda' \in A_\alpha\}$, $\{\lambda' \in B_\beta\}$ and $\{\lambda' \in A_\alpha \cap B_\beta\}$ might still imply fluctuations of frequencies (1.9).

Thus the careful probabilistic considerations show that there may exist physical (in the sense of the verification by the law of large numbers) properties A, B such that the simultaneous existence of these properties could not be verified on the physical level. In such a situation one of the possibilities is to exclude pairs $C = (A, B)$ of incompatible properties from considerations (this is the modern quantum viewpoint)[7]. However, there is another possibility, namely, consider some regularization procedure $\mathcal{R}$ for (1.9). If (1.9) could be regularized via $\mathcal{R}$, then C can be considered as $\mathcal{R}$-physical property. Thus we can essentially extend physical reality by considering new $\mathcal{R}$-elements of reality. As we have already remarked, in many cases one of the simplest ways to regularize (1.9) is to use the p-adic topology, instead of the real. Here frequencies $\nu_N(A_\alpha \cap B_\beta)$ may have the limit in $\mathbb{Q}_p$, despite of fluctuations in $\mathbb{R}$. However, the possibility of a p-adic (and any other) regularization of (1.9) need not imply the possibility to use the same regularization for (1.7) and (1.8). In principle A and B need not be elements of new reality (despite of the fact that $C = (A, B)$ is an element

[6]The idea that the presence of incompatible observables in the quantum formalism (and, in particular, the Heisenberg uncertainty relation) is not a consequence of disturbance effects of the process of a measurement, but a consequence of the internal statistical structure of a quantum state (or a preparation procedure), has been intensively discussed in quantum physics (Prugovecki [154], Ballentine [21]).

[7]E. Prugovecki pointed out [154] that, far from restricting simultaneous measurements of noncommuting observables, quantum theory does not deal with them at all; its formalism being capable only of statistically predicting the results of measurements of one observable (or a commuting set of observables).

of this reality). Nevertheless, there may be coincidences such that all series

$$\mathbf{P}^{\mathcal{R}}(A_\alpha) = \sum_{\lambda\in\Lambda} \mathbf{p}_\lambda \sum_{\lambda'\in A_\alpha} \mathbf{p}_{\lambda\lambda'}, \quad \mathbf{P}^{\mathcal{R}}(B_\beta) = \sum_{\lambda\in\Lambda} \mathbf{p}_\lambda \sum_{\lambda'\in B_\beta} \mathbf{p}_{\lambda\lambda'}, \tag{1.12}$$

$$\mathbf{P}^{\mathcal{R}}(A_\alpha \cap B_\beta) = \sum_{\lambda\in\Lambda} \mathbf{p}_\lambda \sum_{\lambda'\in A_\alpha\cap B_\beta} \mathbf{p}_{\lambda\lambda'} \tag{1.13}$$

converge with respect to $\mathcal{R}$. In such a situation all events, A_α, B_β, $A_\alpha \cap B_\beta$ are $\mathcal{R}$-physical events. It could be that $\mathcal{R}$-probabilities $\mathbf{P}^{\mathcal{R}}(A_\alpha)$ and $\mathbf{P}^{\mathcal{R}}(B_\beta)$ coincide with conventional probabilities $\mathbf{P}^{\text{Conv}}(A_\alpha)$ and $\mathbf{P}^{\text{Conv}}(B_\beta)$. However, in general $\mathbf{P}^{\mathcal{R}}(A_\alpha) \neq \mathbf{P}^{\text{Conv}}(A_\alpha)$, and (or) $\mathbf{P}^{\mathcal{R}}(B_\beta) \neq \mathbf{P}^{\text{Conv}}(B_\beta)$. In the ensemble framework the previous considerations can be interpreted in the following way. The system of events $F(\pi_S)$ for the ensemble S need not be an algebra. The sets $C_{\alpha\beta} = A_\alpha \cap B_\beta$ need not belong to $F(\pi_S)$. However, we may try to extend the ensemble probability to larger class of sets by using some regularization procedures. Sometimes it is possible and sometimes it is impossible to define ensemble probabilities for $C_{\alpha\beta}$ and preserve ensemble probabilities for sets A_α and B_β.

Thus the modern physics is based the *Kolmogorov physical reality.* This model of physical reality can be extended by considering *non-Kolmogorov physical realities.* We conclude our considerations by the equality:

Model of Reality = Model of Probability.

We now consider the principle of complementarity in the framework of f-realism. The main difference between i-realism and f-realism is that in the first case we can assume that conditional probabilities $\mathbf{p}_{\lambda\lambda'}$ do not depend on a measured property and in the second case a measurement of a property D produces $\mathbf{p}_{\lambda\lambda'} = \mathbf{p}^D_{\lambda\lambda'}$. Here

$$\nu_N(A_\alpha) = \sum_{\lambda\in\Lambda} \nu_N(\lambda) \sum_{\lambda'\in A_\alpha} \nu_N(\lambda'/\lambda; A), \tag{1.14}$$

$$\nu_N(B_\beta) = \sum_{\lambda\in\Lambda} \nu_N(\lambda) \sum_{\lambda'\in B_\beta} \nu_N(\lambda'/\lambda; B), \tag{1.15}$$

$$\nu_N(A_\alpha \cap B_\beta) = \sum_{\lambda\in\Lambda} \nu_N(\lambda) \sum_{\lambda'\in A_\alpha\cap B_\beta} \nu_N(\lambda'/\lambda; C), \quad C = (A, B). \tag{1.16}$$

Here (even for finite sets Λ of hidden variables) the statistical stabilization of frequencies (1.14) and (1.15) need not imply the statistical stabilization of frequencies (1.16).

1.8 History of negative probabilities in physics

The possibility to obtain negative probabilities via a regularization of ensemble and frequency approximations (1.1) and (1.3), respectively, is so natural that the negative attitude against negative probabilities in physics can be only explained by the common use of Kolmogorov' theory of probability. From the frequency viewpoint this use imply the common viewpoint that relative frequencies must always stabilize; from the ensemble viewpoint this use imply that statistical ensembles of physical systems must always have a homogeneous structure with respect to all their nonobserved properties. A negative psychological reaction to the appearance of negative probabilities in physical models implies the desire to forget papers in that negative probabilities play the fundamental role.

Although it is well known, for instance, that P. A. M. Dirac was the first to introduce explicitly the concept of negative energy, the number of those who know his investigations [55] about negative probability – closely related to negative energy and invented simultaneously-seems to be very restricted. This concept is used with reservations but, as it seems, not without a certain kind of sympathy. Said paper (see Section 4 for the details) is not the only one on this topic meanwhile has been forgotten, at least as far as negative probability is concerned.

Another example is the famous Wigner distribution [174] $W(q, p)$ which had been introduced as a probability distribution (see Section 4). And it has no other physical interpretation than a probability distribution. However, the appearance of negative probabilities for some quantum states implies that Wigner's distribution is not more interpreted as a probability distribution (many physicists prefer to call $W(q, p)$ Wigner's function).

In the framework of the EPR experiments violations of Bell's inequality could be easily explained if we suppose that there exist negative probability distributions. However, papers on negative probability description of the EPR experiments (see, for example, the review of W. Mückenheim [146]) did not play large role in the polemics on the EPR experiments. Physicists prefer to accept the death of reality (namely, the impossibility to use realism in quantum world; thus, in fact, the absence of objective laws in reality) or nonlocality of space-time than to use negative probabilities.

The existence of quantum observables with continuous spectra is in the evident contradiction with the discreteness of results of real physical measurements. E. Prugovecki [154] developed a theory of quantum measurements with a finite precision (which takes into account reading errors of individual measurements). One of the great advantages of this theory is the possibility to describe simultaneous measurements of incompatible observables. However, there appear again negative probabilities[8]. As always, this implied the extremely strong critic of the theory.

2 Signed 'probabilistic' measures and Einstein–Podolsky–Rosen paradox

We start this section with brief mathematical introduction to the theory of signed measures (charges). Let Ω be a set and let $\mathcal{F}$ be a σ-algebra of its subsets. A

[8]This has the natural explanation on the basis of our interpretation of negative probabilities: the violation of the law of large numbers for such measurements.

σ-additive function $\mu : \mathcal{F} \to \mathbb{R}$ is said to be a *signed measure* (*charge*). Thus $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ for any sequence $A_n \in \mathcal{F}$, $A_j \cap A_i = \varnothing$, $i \neq j$.

Example 2.1 (Discrete measures). Let $\Omega = \{x_1, x_2, \ldots, x_n, \ldots\}$ be a countable set and let $\mathcal{F}$ be a σ-algebra of all subsets of Ω. Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers such that $\sum_{n=1}^{\infty} |a_n| < \infty$. We set $\mu(\{x_n\}) = a_n$ and $\mu(A) = \sum_{x_n \in A} \mu(\{x_n\})$ for $A \in \mathcal{F}$. The $\mu : \mathcal{F} \to \mathbb{R}$ is a signed measure. On the basis of this simple example we illustrate some important notions of the general theory of signed measures. Set $\Omega_- = \{x_j \in \Omega : \mu(\{x_j\}) < 0\}$, $\Omega_+ = \{x_j \in \Omega : \mu(\{x_j\}) > 0\}$ and $\Omega_0 = \{x_j \in \Omega : \mu(\{x_j\}) = 0\}$. It is evident that for any $E \in \mathcal{F}$:

$$\mu(E \cap \Omega_-) \leq 0 \quad \text{and} \quad \mu(E \cap \Omega_+) \geq 0. \tag{2.1}$$

Let $U, V \in \mathcal{F}$, $U \cap V = \varnothing$ and let $\Omega_0 = U \cup V$. Set $\Omega'_- = \Omega_- \cup U$ and $\Omega'_+ = \Omega_+ \cup V$ (thus $\Omega = \Omega'_- \cup \Omega'_+$). Then the sets Ω'_- and Ω'_+ has same property (2.1) as the sets Ω_- and Ω_+. Set

$$\mu^-(E) = -\mu(E \cap \Omega'_-) = \sum_{x_n \in E \cap \Omega'_-} |a_n|$$

and

$$\mu^+(E) = \mu(E \cap \Omega'_+) = \sum_{x_n \in E \cap \Omega'_+} a_n.$$

Then $\mu(E) = \mu^+(E) - \mu^-(E)$. This representation of μ is unique (in spite of nonuniqueness of a representation $\Omega = \Omega'_- \cup \Omega'_+$). We can associate with a signed measure μ the positive measure $|\mu| = \mu^+ + \mu^-$, $\mu(A) = \sum_{x_n \in A} |a_n|$.

In fact, this particular example demonstrated all main features of signed measures. We consider now the general case.

Definition 2.1. Let μ be a signed measure defined on a σ-algebra $\mathcal{F}$ of subsets of a space Ω. Then the set $A \subset \Omega$ is said to be *negative* with respect to μ if $E \cap A \in \mathcal{F}$ and $\mu(E \cap A) \leq 0$ for every $E \in \mathcal{F}$. Similarly, A is said to be *positive* with respect to μ if $E \cap A \in \mathcal{F}$ and $\mu(E \cap A) \geq 0$ for every $E \in \mathcal{F}$.

Theorem 2.1 (Hahn–Jordan). *Given a signed measure μ on a σ-algebra $\mathcal{F}$ of subsets of Ω, there exists a set $\Omega_- \in \mathcal{F}$ such that Ω_- is negative and $\Omega_+ = \Omega \setminus \Omega_-$ is positive with respect to μ.*

Proof. Let $a = \inf \mu(A)$ where the greatest lower bound is taken over all negative sets $A \in \mathcal{F}$. Let $A_n \in \mathcal{F}$, $n = 1, 2, \ldots$, be a sequence of negative sets such that $\lim_{n \to \infty} \mu(A_n) = a$. Then the set $\Omega_- = \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ is a negative set such that $\mu(\Omega_-) = a$ (this is a consequence of σ-additivity of μ). To show that Ω_- is the required set, we must only show that $\Omega_+ = \Omega \setminus \Omega_-$ is positive. It is possible to show that the assumption Ω_+ is not positive will imply the contradiction (see, for example, [136] for the details). □

Thus we can represent Ω as a union

$$\Omega = \Omega_+ \cup \Omega_- \tag{2.2}$$

of two disjoint measurable sets Ω_+ and Ω_-, where Ω_+ is positive and Ω_- is negative with respect to the signed measure μ. The representation (2.2) is called the *Hahn decomposition* of Ω, and may be not unique. However, if

$$\Omega = \Omega_+^1 \cup \Omega_-^1, \quad \Omega = \Omega_+^2 \cup \Omega_-^2$$

are two distinct Hahn decompositions of Ω, then

$$\mu(E \cap \Omega_-^1) = \mu(E \cap \Omega_-^2), \quad \mu(E \cap \Omega_+^1) = \mu(E \cap \Omega_+^2) \tag{2.3}$$

for every $E \in \mathcal{F}$. In fact, $E \cap (\Omega_-^1 \setminus \Omega_-^2) \subset E \cap \Omega_-^1$ and at the same time $E \cap (\Omega_-^1 \setminus \Omega_-^2) \subset E \cap \Omega_+^2$. This imply that

$$\mu(E \cap (\Omega_-^1 \setminus \Omega_-^2)) \leq 0 \quad \text{and} \quad \mu(E \cap (\Omega_-^1 \setminus \Omega_-^2)) \geq 0.$$

Thus $\mu(E \cap (\Omega_-^1 \setminus \Omega_-^2)) = 0$, and similarly $\mu(E \cap (\Omega_-^2 \setminus \Omega_-^1)) = 0$. Therefore

$$\begin{aligned}\mu(E \cap \Omega_-^1) &= \mu(E \cap (\Omega_-^1 \setminus \Omega_-^2)) + \mu(E \cap (\Omega_-^1 \cap \Omega_-^2)) \\ &= \mu(E \cap (\Omega_-^2 \setminus \Omega_-^1)) + \mu(E \cap (\Omega_-^1 \cap \Omega_-^2)) = \mu(E \cap \Omega_-^2),\end{aligned}$$

which proves the first of the formulas (2.3). The second formula is proved in exactly the same way.

Thus a signed measure μ on the space Ω uniquely determines two nonnegative set functions, namely

$$\mu^+(E) = \mu(E \cap \Omega_+), \quad \mu^-(E) = -\mu(E \cap \Omega_-)$$

called the *positive variation* and *negative variation* of μ, respectively. It is clear that

1) $\mu = \mu^+ - \mu^-$;
2) μ^+ and μ^- are nonnegative σ-additive set functions, i.e., measures;
3) The set function $|\mu| = \mu^+ + \mu^-$, called the *total variation* of μ, is also a measure.

The representation $\mu = \mu^+ - \mu^-$ is called the *Jordan decomposition* of μ.

We can present a formal generalization of Kolmogorov measure-theoretical approach. We define a *signed probability space* as the triple $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$, where Ω is an arbitrary set (points ω of Ω are said to be *elementary events*), $\mathcal{F}$ is an arbitrary σ-algebra of subsets of Ω (elements of $\mathcal{F}$ are said to be *events*), $\mathbf{P}$ is a σ-additive signed measure (a charge) on $\mathcal{F}$ normalized by the condition $\mathbf{P}(\Omega) = 1$.

It is a generalization of probability. There can be events which have negative probabilities and probabilities which are larger than 1. However, our consideration in Section 1.1 give strong motivations to use signed probability spaces in physics. Moreover, there are analogues of the law of large numbers and central limit theorem for signed probabilities (see [24], [76], [95]) which also improve the use of signed probability spaces.

There are no physical reasons to assume that even in the case of signed probabilities the system of events has the structure of a set algebra (see also Chapter 4). It is natural to consider signed probability semi-measures defined on set semi-algebras (the reader can obtain the definition of signed probability semi-measure by analogue to Definition 5.2 of Chapter 1). However, it seems that the corresponding mathematical formalism is not yet developed. In particular, I do not know anything about the possibility to obtain the Jordan decomposition for signed semi-measures.

As it has been already mentioned, some physicists (see, for example, [146]) assume that probability distributions involved in Bell's considerations are signed probability measures. This assumption implies that we could not use the standard probabilistic estimates. Therefore there is *no Bell's inequality at all.* From this viewpoint experiments for testing Bell's inequality can be considered as *experiments for testing foundations of probability theory.*

We discuss now carefully the origin of negative probabilities in the EPR framework. Let us follow the ideology of hidden variables. Consider a number N of particles prepared in a pure quantum state and possessing hidden variables λ_k, $k = 0, \dots, n$. Assume that the different values λ_k are taken with (probably generalized) probabilities $\mathbf{p}_k$. By an interaction (the nature of which need not be specified) the values of these hidden variables change from λ_i to λ'_j, $j = 0, \dots, m$, the transition probability being denoted by $\mathbf{p}_{kj}$. By this interaction the pure state may split into $l \leq m$ experimentally distinguishable states. Let A be one of such states. The set of values of j such that λ'_j form the state A is denoted by the symbol $j(A)$. The result of a measurement exhibits $N(A)$ particles in the state A and gives relative frequencies $\nu_N(A) = \frac{N(A)}{N}$. By statistical stabilization of these frequencies we obtain frequency probabilities: $\mathbf{P}^{\text{Mises}}(A) = \lim_{N\to\infty} \nu_N(A)$. The combined transition probability for the state A can be found with the aid of the formula of total probability:

$$\mathbf{P}^{\text{com}}(A) = \lim_{n\to\infty} \sum_{k=0}^{n} \mathbf{p}_k \sum_{j\in j(A)} \mathbf{p}_{kj} = \sum_{k=0}^{\infty} \mathbf{p}_k \sum_{j\in j(A)} \mathbf{p}_{kj}. \tag{2.4}$$

All probabilistic considerations on Bell's inequality are based on the assumption that the observed frequency probabilities $\mathbf{P}^{\text{Mises}}(A)$ must coincide with combined transition probabilities $\mathbf{P}^{\text{com}}(A)$ (defined by (2.4)). By this assumption we can use hidden probabilities $\mathbf{p}_k, \mathbf{p}_{kj}$, in calculations related to Bell's inequality. However, as it has been already mentioned in Section 1, the formula of total probability (2.4) can contain some pathologies. These pathologies could be in principle eliminated by some regularization procedure $\mathcal{R}$.[9] However, $\mathcal{R}$ can produce nonconventional probabilities $\mathbf{p}_j, \mathbf{p}_{ij}$.

[9]In fact, the $\mathcal{R}$ consists of two regularization procedures: 1) $\mathcal{R}_{\text{fr}}$ gives a regularization ($N \to \infty$)

The problem of fluctuating of frequencies $\nu_N(\lambda_k)$ and (or) $\nu_N(\lambda'_j/\lambda_k)$ have been already discussed in Section 1 (see also Chapter 2). We pay now attention to the average over an infinite set of hidden variables Λ. So let $|\Lambda| = \infty$. We have

$$\mathbf{P}^{\text{Mises}}(A) = \lim_{N\to\infty} \nu_N(A) = \lim_{N\to\infty} \lim_{n\to\infty} \sum_{k=1}^{n} \nu_N(\lambda_k) \sum_{j\in j(A)} \nu_N(\lambda'_j/\lambda_k) , \quad (2.5)$$

where $\lambda_1, \dots, \lambda_n$, $n = n_N$ are hidden states of particles $s_1, \dots, s_N$. On the other hand, we have

$$\mathbf{P}^{\text{com}}(A) = \lim_{n\to\infty} \sum_{k=1}^{n} \lim_{N\to\infty} \nu_N(\lambda_k) \sum_{j\in j(A)} \lim_{N\to\infty} \nu_N(\lambda'_j/\lambda_k)$$
$$= \lim_{n\to\infty} \lim_{N\to\infty} \sum_{k=1}^{n} \nu_N(\lambda_k) \sum_{j\in j(A)} \nu_N(\lambda'_j/\lambda_k).$$

To obtain the equality $\mathbf{P}^{\text{Mises}}(A) = \mathbf{P}^{\text{com}}(A)$, we have to change the order of limits

$$\lim_{N\to\infty} \lim_{n\to\infty} \to \lim_{n\to\infty} \lim_{N\to\infty} .$$

However, we could not do this in the general case. First of all, as we have already discussed, it may be that $\mathbf{P}^{\text{Mises}}(A) = \lim_{N\to\infty} \nu_N(A)$ exists but some of limits $\lim_{N\to\infty} \nu_N(\lambda_k)$ or $\lim_{N\to\infty} \nu_N(\lambda'_j/\lambda_k)$ do not exists. On the other hand, it may be that, for example, all $\mathbf{p}_k^{\text{Mises}} = \lim_{N\to\infty} \nu_N(\lambda_k) = 0$, i.e.,

$$\lim_{n\to\infty} \lim_{N\to\infty} \sum_{k=1}^{n} \nu_N(\lambda_k) \sum_{j\in j(A)} \nu_N(\lambda'_j/\lambda_k) = 0.$$

But at the same time $\mathbf{P}^{\text{Mises}}(A) \neq 0$. However, it is possible to justify (in some cases) the change of the order of limits with the aid of some regularization procedure.

We now consider the ensemble approach to find the origin of negative probabilities. Let us start with the following example.

Example 2.2 (Negative distribution of hidden variables). The hidden variable λ has the infinite number of values $\lambda = \lambda_0, \dots, \lambda_n, \dots$. A statistical ensemble S contains $n(\lambda_l) = 2^l$, $l = 0, 1, \dots,$ particles with $\lambda = \lambda_l$. Let us consider the sub-ensemble $S^{(n)}$ of S which contains all particles with $\lambda \in \{\lambda_0, \dots, \lambda_n\}$. Thus $|S^{(n)}| = 1 + \dots + 2^n = 2^{n+1} - 1$ and $\mathbf{p}_k^{(n)} = \mathbf{P}_{S^{(n)}}(\lambda_k) = \frac{2^k}{2^{n+1}-1}$, $0 \le k \le n$. The formula of total probability for the ensemble $S^{(n)}$ has the form:

$$\mathbf{P}_{S^{(n)}}(A) = \sum_{k=0}^{n} \mathbf{p}_k^{(n)} \sum_{j\in j(A)} \mathbf{p}_{kj}$$

(here it is assumed that conditional probabilities $\mathbf{p}_{kj}$ depend only on the interaction;

of (in general fluctuating) frequencies; 2) $\mathcal{R}_{\text{conv}}$ gives a regularization ($n \to \infty$) of (in general infinite) average over Λ.

they do not depend on n). If $n \to \infty$, then $S^{(n)} \to S$ and $\lim_{n\to\infty} \mathbf{P}_{S^{(n)}}(A) = \mathbf{P}_S(A)$. However, for probabilities $\mathbf{p}_k$ with respect to the ensemble S, we have $\mathbf{p}_k = \lim_{n\to\infty} \mathbf{p}_k^{(n)} = 0$. Thus, in general,

$$\mathbf{P}_S(A) = \lim_{n\to\infty} \sum_{k=0}^{n} \mathbf{p}_k^{(n)} \sum_{j\in j(A)} \mathbf{p}_{kj} \neq \sum_{k=0}^{\infty} \lim_{n\to\infty} \mathbf{p}_k^{(n)} \sum_{j\in j(A)} \mathbf{p}_{kj} = 0.$$

We make some formal computations (which, of course, has no meaning in the framework of real analysis). First, we find the 'number of particles' in S:

$$|S| = \sum_{k=0}^{\infty} 2^k = \frac{1}{1-2} = -1. \tag{2.6}$$

Then we find probabilities

$$\mathbf{p}_k = \frac{|S(\lambda = \lambda_k)|}{|S|} = -2^k . \tag{2.7}$$

Here $\sum_{k=0}^{\infty} \mathbf{p}_k = 1$. Thus we obtained negative ensemble probabilities. We can apply the ensemble formula of total probability to these probabilities:

$$\mathbf{P}_S(A) = \sum_{k=0}^{\infty} \mathbf{p}_k \sum_{j\in j(A)} \mathbf{p}_{kj} \tag{2.8}$$

(at the moment we assume that conditional probabilities are ordinary positive probabilities, $\mathbf{p}_{kj} \geq 0$). Let, for example, $\mathbf{p}_{kj} = q_j \geq 0$ do not depend on k. Then $\mathbf{P}_S(A) = (\sum_{k=0}^{\infty} \mathbf{p}_k)(\sum_{j\in j(A)} q_j) = \sum_{j\in j(A)} q_j \geq 0$ is ordinary probability (in spite of the presence of negative probabilities). We can also consider k-dependent conditional probabilities $\mathbf{p}_{kj}$. Let, for example, $A = \{\lambda_0'\}$ and $\mathbf{p}_{k0} = 0$, $k = 2l+1$, $\mathbf{p}_{k0} = 1/2^{l+s}, k = 2l$, where $s = 0, 1, \ldots$ is some (fixed) parameter of the model. Then

$$\mathbf{P}_S(A) = \sum_{l=0}^{\infty} \frac{-2^{2l}}{2^{l+s}} = -\frac{1}{2^s} \sum_{l=0}^{\infty} 2^l = \frac{1}{2^s}$$

is the ordinary probability.

Let $\mathbf{p}_{k0}$ be the same as above and let $\mathbf{p}_{k1} = 1$, $k = 2l+1$, $\mathbf{p}_{k1} = (1 - 1/2^{l+s})$, $k = 2l$. Set $B = \{\lambda_1'\}$. Then

$$\begin{aligned}\mathbf{P}_S(B) &= \sum_{l=0}^{\infty} \mathbf{p}_{2l+1}\mathbf{p}_{(2l+1)1} + \sum_{l=0}^{\infty} \mathbf{p}_{2l}\mathbf{p}_{(2l)1} \\ &= -\Big(\sum_{l=0}^{\infty} 2^{2l+1} + \sum_{l=0}^{\infty} 2^{2l}\Big(1 - \frac{1}{2^{l+s}}\Big)\Big) = -\Big(-\frac{2}{3} - \frac{1}{3} + \frac{1}{2^s}\Big) = 1 - \frac{1}{2^s}.\end{aligned}$$

Of course, these 'generalized probabilities' have some properties which are in contradiction with the common probability intuition. For example, let $s = 0$. Then $\mathbf{P}_S(A) = 1$, $\mathbf{P}(B) = 0$ (despite of the fact that $\mathbf{p}_{k1} = 1$ and $\mathbf{p}_k \neq 0$, $k = 2l+1$).

In Chapter 4 we shall see that all these formal manipulations can be realized on the mathematical level of rigorousness in the p-adic probabilistic framework. In particular, from the p-adic viewpoint probabilities $\mathbf{p}_k = -2^k$ are infinitely small probabilities. Thus in the ensemble S the proportion of systems having the fixed value λ_j of λ is infinitely small. All these infinitely small probabilities must be identified with zero probability in the conventional probability theory.

Example 2.3 (Negative conditional probabilities and negative probabilities for hidden variables). Negative conditional probabilities $\mathbf{p}_{kj}$ may also appear in quite natural statistical ensembles. We assume that the interaction which determines the transition $\lambda_k \to \lambda'_l$ can be represented as a finite chain of steps (trajectory), $(x)_n$ and at each step a particle can have one of two states, 0 or 1. Thus a trajectory of the interaction with n steps has the form $(x)_n = (u_1, \ldots, u_n)$, $u_j = 0, 1$. In our model we simply assume that the transition $\lambda_k \to \lambda'_l$ is realized via a trajectory of the length l (thus, for fixed l, conditional probabilities $\mathbf{p}_{kl}$ do not depend on k). Consider the statistical ensemble G_l of trajectories having the length l, where $l = 0, 1, 2, \ldots$ (we consider also a 'trajectory' of the length $l = 0$, which describes direct transition $\lambda_k \to \lambda'_0$). Set $G^{(n)} = \bigcup_{l \leq n} G_l$. Then $|G_l| = 2^l$ and $|G^{(n)}| = 2^{n+1} - 1$ and $|G| = \lim_{n\to\infty} |G^{(n)}| = \sum_{k=0}^{\infty} 2^k = -1$. Thus

$$\mathbf{p}_{kl} = \frac{|G_l|}{|G|} = -2^l.$$

Suppose that as in the above examples $\mathbf{p}_k = -2^k$ and that an experimentally distinguishable state A is determined by values λ'_{2k}, $k = 0, 1, \ldots$, i.e., $A = \{\lambda'_0, \ldots, \lambda'_{2k}, \ldots\}$. By the formula of total probability we have

$$\mathbf{P}_S(A) = \Big(\sum_{l=0}^{\infty} -2^l\Big)\Big(\sum_{j=0}^{\infty} -2^{2j}\Big) = \frac{1}{3},$$

$$\mathbf{P}_S(\bar{A}) = \Big(\sum_{l=0}^{\infty} -2^l\Big)\Big(\sum_{j=0}^{\infty} -2^{2j+1}\Big) = \frac{2}{3}.$$

However, for $A_j = \{\lambda'_j\}$, $\mathbf{P}_S(A_j) = -2^j < 0$.

We shall see in the p-adic framework that such probabilities can be interpreted as infinitely small (but nonzero!) quantities. Thus in this model not only probability to obtain $\lambda = \lambda_j$ for fixed j is infinitely small, but also probability of each transition $\lambda_k \to \lambda'_l$ is infinitely small.

We can easily modify the above example and introduce conditional probabilities $\mathbf{p}_{kj}$ which depend on k.

Example 2.4 (Negative conditional probabilities and positive probabilities for hidden variables). Assume that the interaction which determines the transition $\lambda_k \to \lambda'_l$ can be represented as a chain of the length l of steps (trajectory), $(x)_l$. Assume that at each step a particle can have one of the states $d \in D_l = \{d_1, \dots, d_l\}$ and each state $d \in D_l$ can appear in a trajectory $(x)_l$ only one time. Thus a trajectory for the transition $\lambda_k \to \lambda'_l$ has the form $(x)_l = (u_1, \dots, u_l)$, $u_j \in D_l$, $u_i \neq u_j$, $i \neq j$, i.e., $(x)_n = \sigma(d_1, \dots, d_l)$ is a permutation of elements of the set D_l. It is also assumed that sets of states D_l satisfy the condition of consistency: $D_{l+1} = D_l \cup \{d_{l+1}\}$. We consider now the following statistical ensembles: G_l, $l = 1, 2, \dots$ (all trajectories of the length l); $G^{(n)} = \cup_{l=0}^{n} G_l$ (all trajectories of the length $\leq n$); $G = \cup_{l=0}^{\infty} G_l$ (all trajectories of a finite length). Then $|G_l| = l!$, $|G^{(n)}| = \sum_{k=0}^{\infty} k!$. Therefore we obtain that in the framework of real analysis

$$\mathbf{P}_{G^{(n)}}(\lambda'_l/\lambda_k) = \frac{|G_l|}{|G^{(n)}|} \to 0, \quad n \to \infty,$$

i.e., $\mathbf{P}_G(\lambda'_l/\lambda_k) = 0$ in the convectional probability theory. In such a situation (even if hidden variable λ has the ordinary Kolmogorov probability distribution; for example, $\mathbf{p}_k = 1/2^{k+1}$, $k = 0, 1, \dots$) we obtain (of course, only formally) that

$$\begin{aligned}\mathbf{P}_S(A) &= \lim_{n\to\infty} \sum_{k=0}^{n} \mathbf{p}_k \sum_{l \in l(A)} \mathbf{P}_{S_{(n)}}(\lambda'_l/\lambda_k) \\ &= \sum_{k=0}^{\infty} \mathbf{p}_k \sum_{l \in l(A)} \lim_{n\to\infty} \mathbf{P}_{G^{(n)}}(\lambda'_l/\lambda_k) = 0.\end{aligned}$$

However, if we justify (via some summation procedure) the calculation $|G| = \sum_{k=0}^{\infty} k!$ (in particular, in the p-adic framework), then (nonconventional) probabilities

$$\mathbf{p}_{kl} = \frac{l!}{\sum_{k=0}^{\infty} k!} \neq 0 \tag{2.9}$$

are well defined and the formula of total probability can be applied to these probabilities.

3 Wigner phase-space distribution and negative probability

Even in non-relativistic quantum mechanics negative probabilities creep into the picture. To formulate a conventional (Maxwellian) probability distribution of the coordinates $\mathbf{x}$ and momenta $\mathbf{p}$, similarly to statistical mechanics, is plainly excluded by the corresponding uncertainty relation which prevents at least the simultaneous knowledge of these quantities. Wigner and Szilard, however, found a distribution function which for the first time was applied by Wigner in order to calculate the quantum correction to the gas pressure formula. If a wave function $\psi(x_1, \ldots, x_n)$, abbreviated by $\psi(\mathbf{x})$, is given, the corresponding Wigner function reads

$$P(\mathbf{x}, \mathbf{p}) = (\pi h)^{-n} \int_{-\infty}^{\infty} d^n \mathbf{y} \bar{\psi}(\mathbf{x}+\mathbf{y}) \psi(\mathbf{x}-\mathbf{y}) \exp\{2i(\mathbf{p}, \mathbf{y})/h\}, \tag{3.1}$$

with $\mathbf{x}, \mathbf{y}$ and $\mathbf{p}$ vectors having as many components as has the configuration space of the ψ, namely n; where $(\mathbf{p}, \mathbf{y})$ denotes the scalar product. In order to demonstrate the fundamental features of the Wigner function, relevant for the present purpose, it is sufficient to consider a single particle in linear motion. Thus $n = 1$ and the vector symbols will be dropped henceforth. The Wigner function exhibits remarkable similarities to a probability distribution in that it leads to the correct probabilities for the coordinates when integrated with respect to the momenta (the integration range is always understood to be $(-\infty, \infty)$ unless indicated otherwise),

$$\int \mathbf{P}(x, p)\, dp = |\psi(x)|^2, \tag{3.2}$$

and, vice versa, it gives the proper probabilities for the momenta when integrated over the coordinates,

$$\int \mathbf{P}(x, p)\, dx = (2\pi h)^{-1} \left| \int dx \psi(x) \exp\{-ipx/h\} \right|^2. \tag{3.3}$$

Although Wigner calls it the probability function of the simultaneous values for the coordinates and momenta (in more recent papers the notation 'quasi-probability' is adopted) he stresses in the same context, that it cannot really be interpreted in this way "as is clear from the fact, that it may take negative values. But of course this must not hinder the use of it in calculations as an auxiliary function which obeys many relations we would expect from such a probability" [174]. The existence of Wigner functions taking negative values is firmly proved by imposing two very general conditions on $\mathbf{P}$ which can be said to define this type of probability distributions, namely:

(i) $\mathbf{P}(x, p)$ should be a Hermitian form of the state vector $\psi(x)$, i.e., with $\hat{M}(x, p)$ a self-adjoint operator,

$$\mathbf{P}(x, p) = (\psi,\ \hat{M}(x, p)\psi). \tag{3.4}$$

This condition makes $\mathbf{P}(x, p)$ a real number.

(ii) $\mathbf{P}(x, p)$ should give the proper expectation values for all operators which are sums of a function of p and a function of x,

$$\iint \mathbf{P}(x, p)[f(p) + g(x)]\, dp\, dx = \left(\psi, \left[f\left(\frac{h}{i} \frac{\partial}{\partial x} \right) + g(x) \right] \psi \right). \tag{3.5}$$

This condition is a somewhat milder form of (3.2) and (3.3) which properly have to be understood as axioms of the Wigner function and, in any case, must be satisfied.

Further, it suffices to consider such ψ which are linear combinations $\psi = a\psi_1 + b\psi_2$ of any two fixed functions, vanishing in certain intervals of x. Now, by requiring

$$\mathbf{P}(x, p) \geq 0 \text{ for all } x \text{ and } p \tag{3.6}$$

for all x and p for every such ψ, Wigner obtains a contradiction which in short runs as follows:

Consider an interval I, inside of which $\psi(x) = 0$ and $g(x) \geq 0$, while $g(x) = 0$ outside and $f(p) \equiv 0$ everywhere. Then (3.5) leads to

$$\iint \mathbf{P}(x, p) g(x)\, dp\, dx = 0. \tag{3.7}$$

Thus

$$\int \mathbf{P}(x, p) g(x)\, dx = 0 \tag{3.8}$$

for all p (except a set of measure zero).

From (3.6) and the condition imposed on $g(x)$ we obtain (Wigner's lemma): *If $\psi(x)$ vanishes in an interval I, the corresponding $\mathbf{P}(x, p)$ vanishes (except for a set of measure zero) for all values of x in that interval.* Now, consider two functions $\psi_1(x)$ and $\psi_2(x)$ which vanish outside of two non-overlapping intervals I_1, and I_2, respectively. Because of (3.4) $\mathbf{P}(x, p)$ corresponding to $\psi = a\psi_1 + b\psi_2$ will have the form

$$\mathbf{P} = |a|^2\, \mathbf{P}_1 + \bar{a}b\, \mathbf{P}_{12} + a\bar{b}\, \mathbf{P}_{21} + |b|^2\, \mathbf{P}_2. \tag{3.9}$$

By setting $b = 0$ it is obvious that $\mathbf{P}_1$, is the Wigner function of ψ_1, (and $\mathbf{P}_2$ of ψ_2). The meaning of $\mathbf{P}_{12}$ and $\mathbf{P}_{21}$ is less obvious, but we need not bother, because both must be identically zero. This can be seen by considering any interval I' outside I_1. Since, according to the above lemma, $\mathbf{P}$, vanishes almost everywhere in interval I', (3.9) cannot be positive for every choice of a and b unless $\mathbf{P}_{12} = \mathbf{P}_{21} = 0$ outside I_1. The same proof applies to I_2. Thus, instead of (3.9) we have

$$\mathbf{P} = |a|^2\mathbf{P}_1 + |b|^2\mathbf{P}_2 \tag{3.10}$$

almost everywhere. In order to complete the contradiction, let us denote the Fourier transforms of ψ_1 and ψ_2 by $\phi_1(p)$ and $\phi_2(p)$, respectively. Equation (3.3) then reads

$$|a|^2 \int \mathbf{P}_1(x, p)\, dx + |b|^2 \int \mathbf{P}_2(x, p)\, dx = |a|^2|\phi_2(p)|^2 + |b|^2|\phi_2(p)|^2 + 2\Re[a\bar{b}\phi_1(p)\phi_2(p)].$$

Since this must be valid for all a and b, we must have identically in $p:\ \phi_1(p)\bar{\phi}_2(p) = 0$. This is, however, impossible since ϕ_1 and ϕ_2, being Fourier transforms of functions restricted to finite intervals, are analytic functions of their arguments and cannot vanish over any finite interval.

In order to illustrate this result, the Wigner function formalism may be applied to the paradigm of quantum theory, the linear harmonic oscillator (see W. Mückenheim [146]). From its Hamiltonian

$$H(x, p) = p^2/2m + m\omega^2 x^2/2 \tag{3.11}$$

and the equation for eigenfunctions of this Hamiltonian:

$$\hat{H}\left(x, \frac{h}{i}\frac{\partial}{\partial x}\right)\psi(x) = E\psi(x). \tag{3.12}$$

It is easy to find the wave function of the ground state

$$\psi_0(x) = (m\omega/h)^{1/4} \exp(-x^2 m\omega/2h) \tag{3.13}$$

corresponding to the energy $E_0 = h\omega/2$. Inserting (3.13) in (3.1) and integrating out in y leads to

$$\mathbf{P}_0(x, p) = (\pi h)^{-1} \exp(-x^2 m\omega/h - p^2/m\omega h), \tag{3.14}$$

which does not exhibit any anomaly in that it is non-negative and, when integrated with respect to x, supplies the proper distribution of the momentum

$$\int \mathbf{P}_0(x, p)\, dx = (m\omega\pi h)^{-1/2} \exp(-p^2/m\omega h), \tag{3.15}$$

which is a Gaussian distribution with expectation zero and standard deviation $(\Delta p)^2 = m\omega h/2$. Integrating with respect to p yields, as expected, the square of (3.13),

$$\int \mathbf{P}_0(x, p)\, dp = (m\omega/\pi h)^{-1/2} \exp(-x^2 m\omega/h), \tag{3.16}$$

also a Gaussian distribution with expectation zero and standard deviation $(\Delta x)^2 = h/2m\omega$. Gaussian distributions satisfy Heisenberg's uncertainty relation in its marginal form, i.e., as an equality. From (3.15) and (3.16) we obtain $(\Delta x)(\Delta p)^2 = h/2$. It may also be noted that the distributions of momentum and position are statistically independent, because $\int \mathbf{P}_0\, dp \int \mathbf{P}_0\, dx = \mathbf{P}_0$. A fortiori, the covariance coefficient is zero. Clearly, this example does not contradict Wigner's 'negativity proof' because the latter only says that there are state functions for which the corresponding $\mathbf{P}(x, p)$ cannot be everywhere non-negative. One of those is the first excited state of the harmonic oscillator. Using the state function of the first excited level, the same formalism as described above will lead to the corresponding Wigner function.

With H the Hamiltonian of (3.11) and L_n the nth Laguerre polynomial, the Wigner function corresponding to the nth excited state can be expressed by

$$\mathbf{P}_n(x, p) = (\pi h)^{-1}(-1)^n \exp(-2H/h\omega) L_n(4H/h\omega) \tag{3.17}$$

or, using (3.14),

$$\mathbf{P}_n(x,p) = (-1)^n \mathbf{P}_0(x,p) L_n(4H/h\omega). \tag{3.18}$$

As $\mathbf{P}_0$ was found to be non-negative everywhere, we have to examine the remaining expression

$$\mathbf{P}_n/\mathbf{P}_0 = (-1)^n L_n(4H/h\omega). \tag{3.19}$$

The first-order Laguerre polynomial is

$$L_1(u) = 1 - u. \tag{3.20}$$

Hence, $\mathbf{P}_1(x,p)$ goes negative for

$$H \equiv p^2/2m + m\omega^2 x^2/2 < h\omega/4. \tag{3.21}$$

Therefore the Wigner distribution $\mathbf{P}_1(x,p)$ becomes negative only in the extremely small domain (ellipse (3.21)). As the energy of the first excited state $E_1 = \frac{3}{2}h\omega$, probability of an energy measurement $\mathbf{P}(E < h\omega/4)$ (where E is the energy of quantum harmonic oscillator) is equal zero. Hence in this example negative values of the Wigner distribution $\mathbf{P}_1(x,p)$ correspond to events which have zero conventional probability. The use of the Wigner distribution can be interpreted as a kind of splitting of conventional zero probabilities by using negative numbers (as a class of labels to denote probabilities of events which are identified in the conventional framework with the label '0').

We now consider the Wigner function of the second excited state. The second-order Laguerre polynomial is

$$L_2(u) = 2 - 4u + u^2. \tag{3.22}$$

Using (3.19) we obtain that $\mathbf{P}_2(x,p)$ goes negative for

$$\frac{1}{2}h\omega(1 - 2^{-1/2}) < H < \frac{1}{2}h\omega(1 + 2^{-1/2}). \tag{3.23}$$

As the energy of the second excited state $E_2 = \frac{5}{2}h\omega$, probability of an energy measurement $\mathbf{P}(E < \frac{1}{2}h\omega(1 + 2^{-1/2}))$ is equal zero. Hence in this example negative values of the Wigner distribution $\mathbf{P}_1(x,p)$ can be also interpreted as additional labels for probabilities (which are identified with the label '0' in the conventional probability theory) corresponding to events which have zero conventional probability. For $H = 0$ and $H \to \infty$ however, $\mathbf{P}_2(x,p)$ is non-negative.

We will not leave this illustrative example without noting some general features of Wigner functions of the linear harmonic oscillator. From the asymptotic equivalence of $L_n(u)$ and $(-1)^n u^n$ for $u \to \infty$ and from (3.17) we find $\mathbf{P}_n$ being positive and asymptotically approaching zero for H going to infinity. In the special case of $H = 0$, $L_n(0) = n!$ together with (3.17) makes even-order $\mathbf{P}_n$ being positive and odd-order $\mathbf{P}_n$ being negative at $H = 0$.

Most interesting in the present context is, however, that all these Wigner functions of nonzero order unavoidably will take positive as well as negative values. This can easily be seen from the orthogonality relation

$$\frac{1}{n!}\frac{1}{m!}\int_0^\infty e^{-u} L_n(u) L_m(u)\, du = \delta_{nm}. \tag{3.24}$$

Cohen (see, for example, review [146] for the details) could show that a wide class of probability distribution functions is supplied by the rather general expression

$$\mathbf{P}(x,p) =$$
$$(2\pi)^{-2}\iiint f(\theta,\tau)\exp(-i\theta x - i\tau p + i\theta u)\psi^*(u-\tau h/2)\psi(u+\tau h/2)\, d\theta\, d\tau\, du.$$

Herein f is simply a smearing function. By setting $f \equiv 1$, substituting τ by $-2y/h$ and integrating over θ and u, we obtain the original Wigner function (3.1). Other distribution functions may be built with different functions f, if only f satisfies the condition $f(0,\tau) = f(\theta,0) = 1$ in order to yield the correct quantum mechanical marginal distributions.

Cohen imposed the following conditions on a general distribution function $\mathbf{P}(x,p)$: (i) those given by (3.2) and (3.3); (ii) if the quantum mechanical mean value of the Hermitian operator $\hat{M}$ is $\langle \hat{M} \rangle$, then there should exist a function $g_M(x,p)$ such that

$$\langle \hat{M} \rangle = \iint g_M(x,p)\mathbf{P}(x,p)\, dp\, dx; \tag{3.25}$$

and, for any function K,

$$\langle K(\hat{M}) \rangle = \iint K(g_M(x,p))\mathbf{P}(x,p)\, dp\, dx. \tag{3.26}$$

And he found, that, irrespective of whether $\mathbf{P}$ is positive semidefinite or not, condition (ii) can never be satisfied. The Wigner function $\mathbf{P}$, of the harmonic oscillator, e.g., yields the correct expectation value for the mean energy, but fails to supply the zero-standard deviation, which one should expect from a quantum mechanical energy eigenstate. He concludes: "Of course, it can be argued that the classical formalism does go through as long as we do not insist that the function which must be used to obtain the mean value of a function, K, of g is not identical to $K(g)$. But this would carry us even further from the conceptual basis of classical probability theory than does quantum mechanics itself!".

Finally we note that the equality (3.25) is the direct consequence of the formula of total probability. Let $\hat{M}$ be an orthogonal projector in the Hilbert space of quantum states. It represents the physical observable $M = 0, 1$. Here $\mathbf{P}(M = 1) = \langle \hat{M} \rangle$ and (3.25) is nothing than the formula of total probability for the initial probability distribution $\mathbf{P}(x, p)$ and conditional probabilities $\mathbf{P}(M = 1/(x, p)) = g_M(x, p)$. If we follow to our interpretation of negative probabilities, then we obtain that Hermitian operators represent all physical observables which permit measurements having the property of the statistical stabilization. Non-Hermitian operators represent a new class of physical observables which do not permit measurements with the property of the statistical stabilization. Here we could obtain (for some states) the negative mean value for an observable with positive values.

On the basis of our interpretation of negative probabilities we can finish this section by

Conclusion. From the frequency viewpoint negative values of Wigner's probability distribution is nothing else than the exhibition of the absence of the statistical stabilization of relative frequencies $\nu_N((x, p) \in U)$ for some domains U of the phase space; from the ensemble viewpoint negative values of Wigner probability distribution is nothing else than the exhibition of nonregular structure of infinite statistical ensembles of hidden properties which determine the point (x, p) of the phase space.

4 Dirac's world with negative probabilities

The necessity of extended probabilities becomes most distinct if a Lorentz-invariant formulation of quantum theory is attempted. The special role that time plays in non-relativistic theory can, e.g., in the most simple case of particles with no charge and spin, be removed by means of the Klein–Gordon equation which for a single free particle of rest mass m is given by

$$\left(\frac{\partial^2}{\partial x_0^2} + \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} + m^2\right)\psi = 0,$$

where $(h = c = 1)$. Born's notion, however, according to which the square of the wave function has to be interpreted as probability density, necessarily must fail in this context, because $|\psi|^2$ as a scalar violates conservation of total probability. On the other hand, the density proposed by Gordon and Klein

$$\mathbf{P}(x_0, x_1, x_2, x_3) = \frac{1}{2im}\left(\frac{\partial\psi^*}{\partial x_0}\psi - \psi^*\frac{\partial\psi}{\partial x_0}\right), \tag{4.1}$$

satisfies as time component of a four-vector the conservation law, and thus (4.1) is evidently the correct mathematical form to use, but, clearly, it can go negative.

This is not the only difficulty. If the wave function of a plane wave

$$\psi = \exp[-i(p_0x_0 - p_1x_1 - p_2x_2 - p_3x_3)], \quad p_0 \equiv E,$$

is transformed to the momentum and energy variables, the Klein–Gordon expression (4.1) goes over

$$|\psi(p_0, p_1, p_2, p_3)|^2 p_0^{-1} dp_1 dp_2 dp_3 \tag{4.2}$$

as the probability of the momentum having a value within the small domain $dp_1dp_2dp_3$ about a value p_1, p_2, p_3 with the energy having the value p_0, which must be connected with p_1, p_2, p_3 by

$$p_0^2 - p_1^2 - p_2^2 - p_3^2 - m^2 = 0.$$

The weight factor p_0^{-1} appears in (4.2) and makes it Lorentz invariant, since $\psi(p)$ is a scalar – it is defined in terms of $\psi(x)$ to make it so – and the differential element $p_0^{-1}dp_1dp_2dp_3$ is Lorentz invariant. This weight factor may be positive or negative, and makes the probability positive or negative accordingly. Thus the two undesirable things, negative energy and negative probability, always occur together. By our interpretation of negative probabilities one of possible explanations of this fact is that the probability to observe a particle with a negative energy is infinitely small.

Dirac formulates an alternative approach to quantum electrodynamics which allows for a conventional treatment of particles with half-odd integral spin, but unavoidably entails negative probabilities when applied to particles with integral spin, in special cases even demanding probabilities of plus or minus 2, distinctly outside the usual range. On the other hand, this relativistic theory has great advantages over the usual method in that it avoids the *most artificial process of renormalization.* With respect to the latter, Dirac never changed his mind, qualifying it as a 'working rule' and considering its results, in spite of their accuracy, as not reliable. Indeed, the commonly applied method of renormalization is a thing between artificial and nonphysical. We are left between Scylla and Charybdis, in that our equations contain either probabilities as large as plus or minus 2 or electron masses exceeding that of the whole universe. Obviously, also Dirac was very sceptical about those "undesirable things, negative energy and negative probability", but he asserts: "Negative energies and probabilities should not be considered as nonsense. They are well-defined concepts mathematically, like a negative sum of money, since the equations which express the important properties of energies and probabilities can still be used when they are negative. Thus negative energies and probabilities should be considered simply as things which do not appear in experimental results. The physical interpretation of relativistic quantum mechanics that one gets by a natural development of the non-relativistic theory involves these things and is thus in contradiction with experiment. We therefore have to consider ways of modifying or supplementing this interpretation" [55].

To delete the divergences Dirac proposed considering the representation including positive and negative energies. Then to resolve the problem of negative energies he proposed considering operators of emission of photons with negative energy as absorption operators of photons with positive energy. But this picture contains negative probabilities of absorption of any odd number of photons.

Let $A^1(x)$ be operators of the quantum electrodynamics of Heisenberg and Pauli referring to emission and absorption of photons into positive energy states:

$$A^1(x) = \iiint (R_k e^{(k,x)} + \bar{R}_k e^{-(k,x)}) k_0^{-1}\, dk_1\, dk_2\, dk_3, \tag{4.3}$$

where $k_0 = +\sqrt{k_1^2 + k_2^2 + k_3^2}$ and R_k being the emission operator and $\bar{R}_k$ the absorption operator. In the same way we introduce the operators $A^2(x)$ referring to the negative energy; there is the representation similar to (4.3) but with $k_0 = -\sqrt{k_1^2 + k_2^2 + k_3^2}$. Dirac considered operators $A^3 = (1/\sqrt{2})(A^1 + A^2)$ which are expended with respect to operators R_k and $\bar{R}_k$ corresponding to positive and negative energies.

The idea was to solve all divergence problems in the symmetric $A^3(x)$ representation. Then we can obtain some information about the $A^1(x)$ representation. But we cannot apply the linear transformation between $A^3(x)$ and $A^1(x)$ representations to the wave function of the $A^3(x)$ representation. There would arise the same divergences. But we can do this with the initial Gibbs ensemble of $A^3(x)$ representation.

It is convenient to consider with $A^3(x)$ additional fields

$$B^3(x) = \frac{1}{\sqrt{2}}(A^1(x) - A^2(x)),$$

which commute with $A^3(x)$, so they are redundant variables. Now let us take B equal to the initial value of A^3. Then for the initial wave function ψ, $(B^3(x) - A^3(x))\psi = 0$ or $\bar{R}_k\psi = 0$ with k_0 either positive or negative. Thus any absorption operator applied to the initial wave function gives the result zero, which means that the corresponding state is one with no photons present.

The following natural interpretation of the wave function at some later time now appears. That part corresponding to m photons of positive energy and n photons of negative energy can be interpreted as corresponding to m photons having been emitted and n photons having been absorbed.

Dirac then considered the momentum representation of $A^3(x)$ and $B^3(x)$ operators. Let k be a momentum-energy vector, $k^2 = 0$, and $\xi_{k\mu}$, $\xi^*_{k\mu}$ be operators of emission and absorption. There $k_0 = \pm\sqrt{k_1^2 + k_2^2 + k_3^2}$. Then set $\zeta_{k\mu} = \xi_{-k\mu}$ for $k_0 > 0$ and consider the wave function ψ as $\psi = \psi(\xi, \zeta)$, $k_0 > 0$. The following commutation relations take place: $[\xi^*, \xi] = c$ and $[\zeta^*, \zeta] = -c$, $c > 0$.

The variables ξ correspond to the emission of photons of positive energy $k_0 > 0$ and the ζ correspond to the absorption of photons of positive energy $k_0 > 0$. Let us denote the space of states $\psi(\xi, \zeta)$ by the symbol $\mathcal{H}$. The inner product in $\mathcal{H}$ has the form:

$$(f, g) = \sum_{m,n=0}^{\infty} f_{mn}\bar{g}_{nm} m! c^m n! (-c)^n$$

for the functions

$$f(\xi, \zeta) = \sum_{mn} f_{mn}\xi^m\zeta^n, \quad g(\xi, \zeta) = \sum_{mn} g_{mn}\xi^m\zeta^n.$$

Now for the wave function $\psi(\xi, \zeta)$, normalized by $|\psi|^2 = (\psi, \psi) = 1$, the probability of there having been m photons emitted into momentum and energy state k (corresponding to ξ) and n photons absorbed from this state is

$$\mathbf{P}(m, n) = |\psi_{mn}|^2 c^m m!(-c)^n n!.$$

It gives a negative probability for an odd number of photons having been absorbed. But this statistical interpretation has no meaning in the framework of the ordinary theory of probability. Nevertheless, we can explain the appearance of such 'generalized' probabilities. On one hand, they may appear as a consequence of the violation of the law of large numbers. On the other hand, they demonstrate that Dirac's formalism gives a fine internal structure of theory which could not be described by conventional probabilities.

5 Negative probabilities and localization

One reason for the difficulties with quantum electrodynamics is the general Lorentz condition, according to which the four-divergence of the electromagnetic potential A must vanish

$$\frac{\partial A^0}{\partial x_0} + \frac{\partial A^1}{\partial x_1} + \frac{\partial A^2}{\partial x_2} + \frac{\partial A^3}{\partial x_3} \equiv \partial_\mu A^\mu = 0.$$

A photon density obtained from this continuity equation suffers from the same problems as the Klein–Gordon conserved density (4.1) in that it is not positive semidefinite, or, according to the opinion of the respective referee, it does not exist. This problem might be related to the fact that photons cannot be sharply localized. If they could, we could define the photon density as the number of photons per unit volume in some arbitrary small volume. However, in a relativistic field, we cannot define such a density.

Therefore it seems that there are two ways for the description of reality: (1) to assume that physical systems could not be localized with arbitrary precision and use Kolmogorov's axiomatic of probability theory; (2) to assume that physical systems could be localized with arbitrary precision, but to change Kolmogorov's axiomatic and create probability theories, where negative probabilities (as well as probabilities which are larger than 1) are mathematically well defined.

If we follow (1), then we have to deny the 'continuous' model of space-time based on real numbers. The system of real numbers $\mathbb{R}$ describes reality with an infinite precision. Here a physical quantity a is represented by the real number:

$$a = \cdots + \frac{\alpha_{-k}}{m^k} + \cdots + \frac{\alpha_{-1}}{m} + \alpha_0 + \cdots + \alpha_l m^l = \alpha_l \ldots \alpha_0, \alpha_{-1} \ldots \alpha_{-k} \ldots, \quad (5.1)$$

where $\alpha_j = 0, 1, \ldots, m-1$, and a natural number $m > 1$ gives the scale of a measurement. All digits in (5.1) can be measured (at least theoretically), thus a 'exists with the infinite precision.' It would be natural to consider other number systems based on expansions which are similar to (5.1) and describe reality with a finite precision. Here we could use a system of m-adic numbers $\mathbb{Q}_m$ which is well known in number theory (mainly in the case $m = p$ is a prime number). These are quantities of the form

$$a = \frac{\alpha_{-k}}{m^k} + \cdots + \frac{\alpha_{-1}}{m} + \alpha_0 + \cdots + \alpha_l m^l + \cdots ,$$

where $\alpha_j = 0, 1, \ldots, m-1$ (thus there is only a finite number of terms corresponding to negative powers of m). A physical formalism based on an m-adic 'finite-precision world' has been developed in [87], [88] (in fact, such a viewpoint is closely connected with the theory of measurements based on nonorthogonal operator valued measures [45], [70], [139]).

If we follow (2), then we can assume that a physical system (in particular, photon) can be localized with an arbitrary precision (i.e., we can still use the real space in quantum theory). However, we could not assume that we should obtain the ordinary (Kolmogorov or Mises) probabilities if we measure statistical distributions corresponding to a 'real localization'.

Our consideration of precision of measurements of physical quantities and negative probabilities can be illustrated by the formalism of quantum theoretical description of radiation. It is given by extending (see the review [146]) the work of Weisskopf and Wigner who calculated the natural linewidth of radiative decay of an excited atom. The corresponding transition amplitude may be rewritten

$$A(E,t) \approx \frac{e^{-t/2} - e^{iEt}}{i/2 - E}$$

with E denoting the difference between actual photon energy and mean state energy E_0 in units of the natural width of the excited state, and t denoting the time interval between excitation and decay in units of the mean lifetime of the state.

It is now very interesting to consider the spectral distribution of photons emitted in finite time intervals. For the time interval $(0,t)$ we have

$$|A(E,t)|^2 = \frac{1}{2\pi} \frac{1 - 2e^{-t/2}\cos(Et) + e^{-t}}{E^2 + 1/4}$$

which undoubtedly is non-negative for every E and t. The spectral distribution emitted at time t, however, $I(E,t) = d|A(E,t)|^2/dt$, entails negative values, as easily can be seen from

$$I(E,t) = \frac{1}{2\pi} \frac{(2E\sin(Et) + \cos(Et))e^{-t/2} - e^{-t}}{E^2 + 1/4}.$$

Further, if the quantity $|A(E, t \to \infty)|^2$

$$I_{\text{norm}}(E,t) = \frac{1}{2\pi} \frac{e^{-t}}{E^2 + 1/4}$$

is used to normalize $I(E,t)$, we obtain the normalized decay probability density $\rho(E,t) = I(E,t)/I_{\text{norm}}(E,t)$, which can take on negative values as well as values exceeding unity, and, if integrated over suited domains $\Delta E \Delta t$, small compared to unity ($= h$), the normalized probability $\rho(E,t)\Delta E \Delta t$, which is an observable quantity, may violate both the lower and the upper limit of Kolmogorov's axiom. These results have been verified by experiments.

As it has been pointed out, if the quantities E and t are measured with extremely high precision, $\Delta E \Delta t < h/2$, then it quite natural that there appear negative probabilities.

4 p-adic probability theory

The development of a non-Archimedean (especially, p-adic) mathematical physics [168], [167], [64], [77–84,87], [93,94,96], [11] induced some new mathematical structures over non-Archimedean fields. In particular, probability theory with p-adic valued probabilities was developed in [79], [80], [86], [88], [130], [131]. This probability theory appeared in connection with a model of quantum mechanics with p-adic valued wave functions [78]. The main task of this probability formalism was to present the probability interpretation for p-adic valued wave functions.

The first theory with p-adic probabilities was the frequency theory in which probabilities were defined as limits of relative frequencies $\nu_N = n/N$ in the p-adic topology[1].This frequency probability theory was a natural extension of the frequency probability theory of R. von Mises [169–171]. One of the most interesting features of the p-adic frequency theory of probability is the possibility to obtain negative (rational) probabilities as limits of relative frequencies. Thus negative probabilities which has been considered in Chapter 3 can be obtained on the mathematical level of rigorousness as p-adic probabilities. Typically p-adic frequency negative probabilities (as well as probabilities which are larger than 1) appear in the cases of violation of the ordinary Mises statistical stabilization (with respect to the real metric). In fact, in this chapter we shall only consider a p-adic generalization of Mises' principle of the statistical stabilization. Thus we shall only study a p-adic generalization of the notion of the S-sequence. The next natural step is to find a p-adic generalization of Mises' principle of randomness. This problem will be studied in Chapter 6 (on the basis of a p-adic generalization of Martin-Löf's theory of statistical tests).

The next step was the creation of p-adic probability formalism on the basis of a theory of p-adic valued probability measures. It was natural to do this by following the fundamental work of A. N. Kolmogorov [133] in which he had proposed the measure-theoretical axiomatics of probability theory. Kolmogorov used properties of the frequency (Mises) probability (non-negativity, normalization by 1 and additivity) as the basis of his axiomatics. Then he added the technical condition of σ-additivity for using Lebesgue's integration theory. In works [79], [80] we tried to follow A. N. Kolmogorov. p-adic frequency probability has also the properties of additivity, it is normalized by 1 and the set of possible values of this probability is the whole field of p-adic numbers $\mathbb{Q}_p$. Thus it was natural to define p-adic probability as a $\mathbb{Q}_p$-valued measure normalized by 1.

However, it was rather complicated problem to propose a p-adic analogue of the condition of σ-additivity. It is the well-known fact that all σ-additive $\mathbb{Q}_p$-valued measures defined on σ-rings are discrete measures [157], [164]. Therefore the creators of

[1] The following trivial fact is the cornerstone of this theory: the relative frequencies belong to the field of rational numbers $\mathbb{Q}$; we can study their behaviour not only in the real topology on $\mathbb{Q}$, but also in some other topologies on $\mathbb{Q}$ and, in particular, in the p-adic topologies on $\mathbb{Q}$.

non-Archimedean integration theory (A. Monna and T. Springer [145]) did not try to develop abstract measure theory, but they proposed an integration formalism via Bourbaki based on integrals of continuous functions. This integration theory has been used for creating p-adic probability theory in the measure-theoretical framework [80]. The main disadvantage of this probability model is the strong connection with the topological structure of a sample space. This is quite similar to the old probability formalisms of Kolmogorov [132], Fréchet [63] and Cramer [44] in which the topological structure of the sample space played the important role.

An abstract theory of non-Archimedean measures has been developed by A. van Rooji [164]. The basic idea of this approach is to study measures defined on *rings* which in principle cannot be extended to measures on σ-rings. This gives the possibility for constructing non-discrete p-adic valued measures. On the other hand, the condition of continuity for measures in [164] implies the σ-additivity in all natural cases[2].

In this chapter we develop a p-adic probability formalism based on measure theory of [164]. By probabilistic reasons we use the special case of this measure theory: measures defined on *algebras* (such measures have some special properties). However, probabilistic applications stimulate also the development of the general theory of non-Archimedean measures defined on rings. We prove the formula of the change of variables for these measures and use this formula for developing the formalism of conditional expectations for p-adic valued random variables (see also [131]).

The use of p-adic valued probabilistic measures gives the possibility to work on the mathematical level of rigorousness with all signed 'probabilities' (for example, with Wigner's distribution).

As the fields of p-adic numbers are non-Archimedean there exist infinitely large p-adic numbers (in particular, infinitely large natural numbers) in $\mathbb{Q}_p$. Thus p-adic analysis gives the possibility to use actual infinities and consider statistical ensembles with an infinite number of elements. Probabilities with respect to such ensembles are defined via the standard proportion (used in Chapter 1 for finite ensembles). One of the main features of such ensemble probabilities is the appearance of negative (rational) probabilities (as well as probabilities which are larger than 1). In this approach the origin of such 'pathological' (from the real viewpoint) probabilities is very clear. In particular, we shall see that a large set of negative probabilities is naturally interpreted as a set of infinitely small probabilities (giving the split of the conventional probability 0). We shall also see that a large set of probabilities which are larger than 1 is naturally interpreted as a set of probabilities which are negligibly differ from 1. Other interesting property of p-adic ensemble probability is that the corresponding probabilistic measure is not well defined on a set algebra. The system of events is only a set semi-algebra.

[2]Thus the σ-additivity is not a problem. The problem is find the right domain of definition of p-adic probabilistic measures.

1 Non-Archimedean number systems; p-adic numbers

Here we present a brief introduction to non-Archimedean and, in particular, p-adic analysis (see, for example, [157], [164], [167], [80], [88]).

Let F be a ring[3] (a set where addition, subtraction and multiplication are well defined). Recall that a *norm* is a mapping $|\cdot|_F : F \to \mathbb{R}_+$ satisfying the following conditions:

$$|x|_F = 0 \iff x = 0, \text{ and } |1|_F = 1, \tag{1.1}$$

$$|xy|_F \le |x|_F |y|_F, \tag{1.2}$$

$$|x+y|_F \le |x|_F + |y|_F. \tag{1.3}$$

The ring F with the norm $|\cdot|_F$ is called a *normed ring*. Set $|F| = \{r \in \mathbb{R}_+ : r = |x|_F, x \in F\}$.

The inequality (1.3) is the well-known triangle axiom. A norm is said to be non-Archimedean if the *strong triangle axiom* is valid, i.e.,

$$|x+y|_F \le \max(|x|_F, |y|_F). \tag{1.4}$$

A ring F with a non-Archimedean norm is said to be a non-Archimedean ring. We shall use the following property of a non-Archimedean norm:

$$|x+y|_F = \max(|x|_F, |y|_F), \quad \text{if } |x|_F \ne |y|_F. \tag{1.5}$$

In order to prove (1.5) we may assume $|x|_F < |y|_F$. By (1.4) we find $|y|_F \le \max(|x+y|_F, |x|_F) \le \max(|x|_F, |y|_F)$. The assumption $|x|_F < |y|_F$ gives $\max(|x|_F, |y|_F) = |y|_F$. Hence $|y|_F = \max(|x+y|_F, |x|_F)$. From $|x|_F < |y|_F$, we deduce $|y|_F = |x+y|_F$. This gives (1.5).

If a norm $|\cdot|_F$ has the property: $|xy|_F = |x|_F|y|_F$, then it is called a *valuation* (sometimes a norm is called a *pseudo-valuation*). A ring F with the valuation $|\cdot|_F$ is called a *valued ring*. The absolute value $|\cdot| \equiv |\cdot|_{\mathbb{R}}$ on the field of real numbers $\mathbb{R}$ is an example of a valuation. This valuation does not satisfy the strong triangle inequality (it satisfies only (1.3)). Valuations and norms with such a property are called Archimedean. Another example of an Archimedean valuation is the absolute value $|\cdot| \equiv |\cdot|_{\mathbb{C}}$ on the field of complex numbers $\mathbb{C}$.

Denote by $\mathbb{Z}(F)$ the ring generated in F by its unity element. If F has zero characteristic (i.e., $n \cdot 1 = 1 + \cdots + 1 \ne 0$ for any $n = 1, 2, \ldots$), then $\mathbb{Z}(F)$ is isomorphic to the ring of integers $\mathbb{Z}$. Therefore in this case we can consider $\mathbb{Z}$ as a subring of F. In what follows we consider only normed rings F which have zero characteristic.

To illustrate how we can work with the strong triangle inequality we present two simple results.

[3]By a ring we always mean a commutative ring with identity 1.

Proposition 1.1. *Let* $|\cdot|_F$ *be a non-Archimedean norm. Then* $|n|_F \le 1$ *for all elements* $n \in \mathbb{Z}$.

Proof. By the strong triangle inequality (1.4) we have:

$$|n|_F = |1 + \cdots + 1|_F \le |1|_F = 1. \qquad \square$$

Proposition 1.2. *A valuation* $|\cdot|_F$ *is a non-Archimedean valuation if and only if* $|n|_F \le 1$ *for all elements* $n \in \mathbb{Z}$.

Proof. Let $|n|_F \le 1$ for all $n = 1, 2, \ldots$. Denote by $\binom{n}{k}$ the binomial coefficients, i.e.,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad k \le n.$$

As these coefficients are integers, $|\binom{n}{k}|_F \le 1$ for all n and k. Hence we have:

$$\begin{aligned} |(x+y)^n|_F &= \Big|\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}\Big|_F \\ &\le \sum_{k=0}^{n} |x|_F^k |y|_F^{n-k} \le (n+1)(\max |x|_F, |y|_F)^n, \end{aligned}$$

i.e.,

$$|x+y|_F \le \lim_{n\to\infty} (1+n)^{1/n} \max(|x|_F, |y|_F) = \max(|x|_F, |y|_F). \qquad \square$$

Let $|\cdot|_F$ be a norm on a ring F. Then the function $\rho_F(x, y) = |x - y|_F$ is a metric on F. It is a translation invariant metric, i.e. $\rho_F(x+h, y+h) = \rho_F(x, y)$. As usual in metric spaces we define 'closed' and 'open' balls in F: $U_r(a) = \{x \in F : \rho_F(x, a) \le r\}$, $U_r^-(a) = \{x \in F : \rho_F(x, a) < r\}$, $r \in \mathbb{R}_+$. We set $U_r \equiv U_r(0)$. It should be noted that any ball $U_r(a)$, $r \in \mathbb{R}_+$, coincides with some ball $U_s(a)$, $s \in |F|$, $s \le r$. In what follows we consider only balls $U_r(a)$ with $r \in |F|$. The spheres in F are defined by $S_r(a) = \{x \in F : \rho_F(x, a) = r\}$, $r \in \mathbb{R}_+$. Of course, if $r \notin |F|$ then $S_r(a) = \varnothing$. Therefore it is meaningful to consider only spheres of radius $r \in |F|$. The normed ring F is *complete* if it is a complete metric space with respect to the metric ρ_F.

Let $|\cdot|_F$ be a non-Archimedean norm. Then the corresponding metric ρ_F satisfies the strong triangle inequality:

$$\rho_F(x, y) \le \max(\rho_F(x, z), \rho_F(z, y)). \tag{1.6}$$

Such a kind of metric is called an *ultrametric*. We note that any 'open' or 'closed' ball in an ultrametric space is a simultaneously closed and open subset. Such sets are called *'clopen'* sets. Spheres in F are also clopen. It seems strange from the point of view of our Euclidean intuition. *The balls* U_r *are additive subgroups of* F: if $|x|_F, |y|_F \le r$, then $|x+y|_F \le \max(|x|_F, |y|_F) \le r$. Moreover, the ball U_1 is a ring: if $|x|_F, |y|_F \le 1$ then $|xy|_F \le |x|_F |y|_F \le 1$.

We shall continuously use the following simple result.

Lemma 1.1 ('The dream of a bad student'). *Let F be a complete non-Archimedean normed ring. The series $\sum_{n=1}^{\infty} a_n$, $a_n \in F$, converges in F if and only if $a_n \to 0$, $n \to \infty$.*

To prove this result we use the Cauchy theorem in complete metric spaces (a sequence $\{S_n\}$ converges iff it is a fundamental sequence, i.e., $|S_n - S_m|_F \to 0$, $n, m \to \infty$) and the estimate $|\sum_{k=n+1}^{m} a_k|_F \le \max_{n+1 \le k \le m} |a_k|_F$.

One of the most important non-Archimedean fields, a system of p-adic numbers $\mathbb{Q}_p$, was constructed by K. Hensel [68]. In fact, it was the first example of a commutative number field (a system where the operations of addition, subtraction, multiplication and division are well defined) which was different from the fields of real and complex numbers. Practically during 100 years p-adic numbers were considered only as objects in pure mathematics. In recent years these numbers have been intensively used in theoretical physics see, for example, the books [167], [80], [88], [95] and papers [168], [64], [10], [77, 89, 91], [78, 82, 83], in the theory of probability [79], [88], as well as in investigations of chaos and dynamical systems [85], [88] and applications to cognitive sciences and psychology [88], [90], [94], [96].

The field of real numbers $\mathbb{R}$ is constructed as the completion of the field of rational numbers $\mathbb{Q}$ with respect to the metric $\rho_R(x, y) = |x - y|$, where $|\cdot|$ is the usual valuation given by the absolute value. The fields of p-adic numbers $\mathbb{Q}_p$ are constructed in a corresponding way, by using other valuations. For any prime number the p-adic valuation $|\cdot|_p$ is defined in the following way. First we define it for natural numbers. Every natural number n can be represented as the product of prime numbers: $n = 2^{r_2} 3^{r_3} \cdots p^{r_p} \cdots$. Then we define $|n|_p = p^{-r_p}$, we set in addition $|0|_p = 0$ and $|-n|_p = |n|_p$. We extend the definition of the p-adic valuation $|\cdot|_p$ to all rational numbers by setting for $m \neq 0$: $|n/m|_p = |n|_p/|m|_p$. The completion of Q with respect to the metric $\rho_p(x, y) = |x - y|_p$ is the locally compact field of p-adic numbers $\mathbb{Q}_p$. It is well known (Ostrovsky's theorem), see [157], that $|\cdot|$ and $|\cdot|_p$ are the only possible valuations on $\mathbb{Q}$. The p-adic valuation satisfies the strong triangle inequality:

$$|x + y|_p \le \max(|x|_p, |y|_p).$$

Thus the field of p-adic numbers $\mathbb{Q}_p$ is non-Archimedean and the p-adic metric ρ_p is an ultrametric. Thus any p-adic ball $U_r(0)$ is an additive subgroup of $\mathbb{Q}_p$ and the ball $U_1(0)$ is also a ring. It is called the *ring of p-adic integers* and denoted by $\mathbb{Z}_p$.

For any $x \in \mathbb{Q}_p$ we have a unique canonical expansion (converging in the $|\cdot|_p$-norm) of the form

$$x = \alpha_{-n}/p^n + \cdots + \alpha_0 + \cdots + \alpha_k p^k + \cdots ,$$

where $\alpha_j \in \{0, 1, \ldots, p-1\}$, are the "digits" of the p-adic expansion. The elements $n \in \mathbb{Z}_p$ have the expansion:

$$n = \alpha_0 + \alpha_1 p + \cdots + a_k p^k + \cdots , \tag{1.7}$$

i.e., they can be identified with sequences of digits

$$n = (\alpha_0, \dots, \alpha_k, \dots), \quad \alpha_j \in \{0, 1, \dots, p-1\}. \tag{1.8}$$

If $n \in \mathbb{Z}_p$, $n \neq 0$, and canonical expansion (1.7) contains only a finite number of nonzero digits α_j, than n is natural number (and vice versa). It is natural to interpret a number $n \in \mathbb{Z}_p$ such that expansion (1.7) contains an infinite number of nonzero digits α_j as an *infinitely large natural number.* Thus the ring of p-adic integers contains actual infinities $n \in \mathbb{Z}_p \setminus \mathbb{N}$, $n \neq 0$. This is one of the most important features of non-Archimedean number systems (compare with nonstandard numbers [10]). In Section 3 we introduce a partial order structure on $\mathbb{Z}_p$ which extends the standard order structure on $\mathbb{N}$: for $n_1, n_2 \in \mathbb{N}$ $n_1 \leq n_2$ in $\mathbb{N}$ iff $n_1 \leq n_2$ in $\mathbb{Z}_p$. Each finite natural number is less than any infinite number: $n \leq m$ for $n \in \mathbb{N}$ and $m \in \mathbb{Z}_p \setminus \mathbb{N}$, $m \neq 0$. This order structure will be used to compare p-adic probabilities.

If, instead of a prime number p, we start from an arbitrary natural number $m > 1$, we construct the system of the so called m-adic numbers $\mathbb{Q}_m$ (by completing $\mathbb{Q}$ with respect to the m-adic metric $\rho_m(x, y) = |x - y|_m$). However, this system is not in general a field. There exist in general divisors of zero in $\mathbb{Q}_m$, thus $\mathbb{Q}_m$ is only a ring. Elements of $\mathbb{Z}_m = U_1(0)$ can be identified with sequences (1.8) with the digits $\alpha_k = 0, 1, \dots, m-1$. We can also use more complicated number systems corresponding to non-homogeneous scales: $M = (m_1, m_2, \dots, m_k, \dots)$, where $m_j > 1$ are natural numbers. In this case we obtain the number system $\mathbb{Q}_M$. The elements $x \in \mathbb{Z}_M = U_1(0)$ can be presented as sequences (1.8) with digits $a_j = 0, 1, \dots, m_j - 1$. The structure of $\mathbb{Q}_M$ is rather complicated from the mathematical point of view. In general the number system $\mathbb{Q}_M$ is not a ring. However, $\mathbb{Z}_M$ is always a ring.

Number systems $\mathbb{Q}_m$ and $\mathbb{Q}_M$ can be also used to develop new non-Kolmogorovean probabilistic models. However, the absence of the well-developed mathematical formalism does not give such a possibility.

Let K be a non-Archimedean field with the valuation $|\cdot|_K$. Here the function $n \to 1/|n!|_K$ increases (as $|n|_K \leq 1$). The following estimate holds in the field $\mathbb{Q}_p$:

$$(1/np)p^{n/(p-1)} \leq \frac{1}{|n!|_p} \leq p^{(n-1)/(p-1)}. \tag{1.9}$$

This estimate is a consequence of the following mathematical fact:

Lemma 1.2. *Let the natural number n be written in the base p*

$$n = a_0 + a_1 p + \dots + a_m p^m, \quad a_j = 0, 1, \dots, p-1.$$

Define the sum of the digits of n by $S_n = \sum_{j=0}^m a_j$. *Then*

$$|n!|_p = p^{(S_n - n)/(p-1)}. \tag{1.10}$$

Proof. There are $[n/p]$ numbers in $\{1, 2, \ldots, n\}$ that are divisible by p. Here, as usual, $[a]$ is the integer part of a. Then there are $[n/p^2]$ numbers that are divisible by p^2, etc. By definition $|n!|_p = p^{-\gamma(n)}$, where $\gamma(n) = \sum_{j=0}^{m}[n/p^j]$. For $j = 1, 2, \ldots, m$ we have

$$[n/p^j] = a_j + a_{j+1}p + \cdots + a_m p^{m-j} = p^{-j}\sum_{i=j}^{m} a_i p^i.$$

Thus,

$$\begin{aligned}\gamma(n) &= \sum_{j=1}^{m} p^{-j}\sum_{i=j}^{m} a_i p^i = \sum_{i=1}^{m} a_i p^i \sum_{j=1}^{i} p^{-j} \\ &= \sum_{i=1}^{m} a_i p^i \frac{(p^i - 1)}{p^i(p-1)} = (p-1)^{-1}\sum_{i=1}^{m} a_i(p^i - 1) \\ &= (p-1)^{-1}(n - S_n).\end{aligned}$$

□

By Ostrovsky's theorem the restriction of the valuation $|\cdot|_K$ to $\mathbb{Q}$ is equivalent to one of p-adic valuations: there exists p such that $|x|_K = |x|_p^l$, $l > 0$, for $x \in \mathbb{Q}$. Thus (1.9) implies that

$$a^n \le \frac{1}{|n!|_K} \le b^n, \tag{1.11}$$

where $a = a(p,l) > 0$ and $b = b(p,l) > 0$.

The exponent in K is defined by the standard power series $e^x = \sum_{n=0}^{\infty} x^n/n!$. This series converges if $|x|_K < b^{-1}$. In particular, in the p-adic case it converges if $|x|_p < p^{1/(1-p)}$. This is equivalent to $|x|_p \le r_p$, where $r_p = 1/p$ for $p \neq 2$ and $r_2 = 1/4$. Trigonometric functions over the field K are defined by the standard power series: $\sin x = \sum(-1)^n x^{2n+1}/(2n+1)!$ and $\cos x = \sum x^{2n}/(2n)!$. These series have the same radius of convergence as the series for the exponential function.

2 Frequency probability theory

Let us provide a generalization of the von Mises frequency theory of probability. Our main idea is very clear and it is based on the following two remarks: (1) relative frequencies $\nu_N = n/N$ always belong to the field of rational numbers $\mathbb{Q}$; (2) there exist many topologies τ on $\mathbb{Q}$ which are different from the usual real topology τ_R (corresponding to the real metric $\rho_R(x, y) = |x - y|$).

As in ordinary Mises' theory, we also consider infinite sequences

$$x = (x_1, \ldots, x_N, \ldots), \quad x_j \in L, \tag{2.1}$$

of observations (here $L = \{\alpha_1, \ldots, \alpha_k\}$ is a label set). But a new topological principle of the statistical stabilization of relative frequencies is proposed: *The statistical stabilization of relative frequencies* $\nu_N(\alpha_i; x)$ *can be considered not only in the real topology on the field of rational numbers* $\mathbb{Q}$ *but also in any other topology* τ *on* $\mathbb{Q}$.

This topology is said to be the *topology of statistical stabilization.* Limiting values $\mathbf{P}(\alpha_i) \equiv \mathbf{P}_x^\tau(\alpha_i)$ of $\nu_N(\alpha_i; x)$, $i = 1, \ldots, k$, are said to be τ-probabilities. These probabilities belong to the completion $\mathbb{Q}_\tau$ of $\mathbb{Q}$ with respect to the topology τ. The choice of the topology τ of statistical stabilization is connected with the concrete probabilistic model. Sequence (2.1), for which the principle of statistical stabilization of relative frequencies for the topology τ is valid, is said to be a (S, τ)-sequence (in particular, (S, τ_R)-sequences, where τ_R is the real topology, are ordinary (von Mises) S-sequences which were considered in Chapter 1). At the moment we do not use any τ-analogue of the principle of randomness.

We are mainly interested in the following situation. The real topology τ_R is not a topology of statistical stabilization for the sequence (2.1), but another topology τ is. In this case we cannot consider (2.1) as a von Mises S-sequence. But there is a new possibility for studying (2.1) as a (S, τ)-sequence.

Set $U_Q = \{q \in \mathbb{Q} : 0 \leq q \leq 1\}$. We denote the closure of the set U_Q in the completion $\mathbb{Q}_\tau$ by U_{Q_τ}. The following theorem is an evident consequence of the topological principle of the statistical stabilization:

Theorem 2.1. *The probabilities* $\mathbf{P}(\alpha_i)$ *belong to the set* U_{Q_τ} *for an arbitrary* (S, τ)*-sequence* x.

As usual, let us consider the algebra F_L of all subsets of L. As in the frequency theory of von Mises we define probabilities $\mathbf{P}(A) = \sum_{\alpha_i \in A} \mathbf{P}(\alpha_i)$ for $A \in F_L$. By Theorem 2.1 the probability $\mathbf{P}(A)$ belongs to the set U_{Q_τ} for every $A \in F_L$.

Theorem 2.2. *Let the completion* $\mathbb{Q}_\tau$ *of* $\mathbb{Q}$ *with respect to the topology of statistical stabilization* τ *be an additive topological group. Then for every* (S, τ)*-sequence* x *the probability is an additive function on* F_L*:* $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$, $A, B \in F_L$, $A \cap B = \varnothing$.

Here we have used only $\lim(u_N + v_N) = \lim u_N + \lim v_N$ in an additive topological group.

Theorem 2.3. *The probability* $\mathbf{P}(L) = 1$ *for every topology of the statistical stabilization* τ *on* $\mathbb{Q}$.

As in Chapter 1 we define a conditional frequency probability $\mathbf{P}(A/B)$.

Theorem 2.4. *Let* $\mathbb{Q}_\tau$ *be a multiplicative topological group. Then for arbitrary* $A, B \in F_L$, $\mathbf{P}(B) \neq 0$, *the Bayes formula* $\mathbf{P}(A/B) = \mathbf{P}(A \cap B)/\mathbf{P}(B)$ *holds.*

Here we have used $\lim u_N/v_N = \lim u_N / \lim v_N$ if $\lim v_N \neq 0$ in a multiplicative topological group.

However, we may choose the topology of statistical stabilization τ such that $\mathbb{Q}_\tau$ is not an additive group. In this case we obtain non-additive probabilities. Further, $\mathbb{Q}_\tau$ may be not a topological multiplicative group. In this case we have violations of Bayes' formula for conditional probabilities[4]. Moreover, there are possibilities of different combinations of these properties. For example, there exist additive probabilities without Bayes' formula.

Now (following to Kolmogorov) we can present an axiomatics corresponding to the properties of frequency probabilities. Of course, this axiomatics depends on the topology τ. Thus we have an infinite set of axiomatic theories $A(\tau)$. The simplest case (and the one most similar to the Kolmogorov axiomatics) is that $\mathbb{Q}_\tau$ is a topological field. There, by definition, *a τ-probability is a U_{Q_τ}-valued measure with the normalization condition* $\mathbf{P}(\Omega) = 1$. There should be technical restrictions on $\mathbf{P}$ to provide a fruitful theory of integration (compare with Kolmogorov's condition of σ-additivity).

We obtain a large class of non-Kolmogorov probabilistic models if we choose a metrizable topology τ such that the corresponding metric has the form $\rho_\tau(x, y) = |x - y|_\tau$, where $|\cdot|_\tau$ is a valuation on $\mathbb{Q}$. According to the Ostrovsky theorem, every valuation on $\mathbb{Q}$ is equivalent to the ordinary real absolute value $|\cdot|_R$ or one of the p-adic valuations $|\cdot|_p$. Therefore we may obtain only two classes of probabilistic models: 1) the ordinary theory of probability (with the topology of the statistical stabilization τ_R); 2) one of the p-adic valued probabilistic models (with topologies of the statistical stabilization τ_p).

The most interesting property of p-adic probabilities is that $U_{\mathbb{Q}_p} = \mathbb{Q}_p$, see [80]. To prove this fact we need only to show that every $x \in \mathbb{Q}_p$ can be realized as a limit of frequencies $\nu_N = n/N$, where n, N are natural numbers, $n \leq N$. Thus any p-adic number x may be a p-adic probability.

For example, every rational number may be taken as a p-adic probability. There are such 'pathological' probabilities (from the point of view of the usual theory of probability) as $\mathbf{P}(A) = 2$, $\mathbf{P}(A) = 100$, $\mathbf{P}(A) = 5/3$, $\mathbf{P}(A) = -1$. If $p = 1 \bmod 4$, then $i = \sqrt{-1}$ belongs to $\mathbb{Q}_p$. Thus 'complex quantities' can be obtained as frequency probabilities; for example, $\mathbf{P}(A) = i = \sqrt{-1}$ or $\mathbf{P}(A) = 1 \pm i$.

Thus *negative (and even complex) probabilities can be realized as **p**-adic frequency probabilities.*

We have presented [80] a large number of statistical models where frequencies oscillate with respect to the real metric ρ_R and stabilize with respect to one of p-adic metrics ρ_p. There p is a parameter of the statistical model. The corresponding statistical simulation was carried out on a computer.

Thus Mises' principle of the statistical stabilization of frequencies can be essentially extended by considering (S, τ)-sequences for topologies τ on $\mathbb{Q}$. It would be natural to extend second Mises' principle, namely, the principle of randomness and introduce an analogue of Mises' collective, namely, a τ-collective. However, I could not obtain any meaningful extension of the principle of randomness for p-adic topologies τ_p. It is still

[4] A simple realization of Accardi's idea.

not clear how we can define a class of place selections which would not disturb the p-adic statistical stabilization. On the other hand, it is well known that in ordinary (real) probability theory it is possible to develop the mathematical theory of randomness by using Martin-Löf statistical recursive tests [142–144]. In Chapter 6 we shall follow to P. Martin-Löf and develop a p-adic theory of recursive statistical tests[5].

3 Ensemble probability

Our interpretation of p-adic numbers

$$N = l_0 + l_1 p + \cdots + l_s p^s + \cdots, \tag{3.1}$$

where $l_s = 0, 1, \ldots, p-1$, with an infinite number of nonzero digits n_s as infinitely large numbers gives the possibility of considering numerous actual infinities. Therefore we can study ensemble probabilities on ensembles of an infinite volume or consider classical probabilities for an infinite number of equally possible cases.

3.1 Ensembles of infinite volumes

We shall study some ensembles $S = S_N$ which have a p-adic 'volume' N, where N is the p-adic integer (3.1). If N is finite then S is the ordinary finite ensemble, if N is infinite then S has essentially p-adic structure. Consider a sequence of ensembles M_j having volumes $l_j p^j$, $j = 0, 1, \ldots$. Set

$$S = \bigcup_{j=0}^{\infty} M_j.$$

Then $|S| = N$. This split of S will play the crucial role in our probabilistic considerations. Thus S is not just an arbitrary ensemble of the cardinality N. It is an ensemble of the cardinality N constructed via the hierarchical structure corresponding to this split. We may imagine an ensemble S as being the population of a tower $T = T_S$, which has an infinite number of floors with the following distribution of population through floors: population of jth floor is M_j. Set $T_k = \bigcup_{j=0}^{k} M_j$. This is population of the first $k+1$ floors.

Let $A \subset S$ and let there exist:

$$n(A) = \lim_{k\to\infty} n_k(A), \quad \text{where } n_k(A) = |A \cap T_k|. \tag{3.2}$$

The quantity $n(A)$ is said to be a *p-adic volume* of the set A.

[5] Of course, we understood that Martin-Löf's theory does not give the fruitful notion of randomness for an individual sequence of trials.

We define the probability of A by the standard proportional relation:

$$\mathbf{P}(A) \equiv \mathbf{P}_S(A) = \frac{n(A)}{N}. \tag{3.3}$$

Denote the family of all $A \subset S$, for which (3.3) exists, by $\mathcal{G}_S$. The sets $A \in \mathcal{G}_S$ are said to be events. Later we shall study some properties of the family of events. First we consider the set algebra F which consists of all finite subsets and their complements.

Proposition 3.1. *$F \subset \mathcal{G}_S$.*

Proof. Let A be a finite set. Then $n(A) = |A|$ and (3.3) has the form:

$$\mathbf{P}(A) = \frac{|A|}{|S|}. \tag{3.4}$$

Now let $B = \bar{A}$. Then $|B \cap T_k| = |T_k| - |A \cap T_k|$. Hence there exists $\lim_{k\to\infty} |B \cap T_k| = N - |A|$. This equality implies the standard formula:

$$\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A). \qquad \square \tag{3.5}$$

In particular, we have : $\mathbf{P}(S) = 1$.

Proposition 3.2. *Let $A_1, A_2 \in \mathcal{G}_S$ and $A_1 \cap A_2 = \varnothing$. Then $A_1 \cup A_2 \in \mathcal{G}_S$ and*

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2). \tag{3.6}$$

Proposition 3.3. *Let $A_1, A_2 \in \mathcal{G}_S$. The following conditions are equivalent:*

(1) $A_1 \cup A_2 \in \mathcal{G}_S$; (2) $A_1 \cap A_2 \in \mathcal{G}_S$;

(3) $A_1 \setminus A_2 \in \mathcal{G}_S$; (4) $A_2 \setminus A_1 \in \mathcal{G}_S$.

There are standard formulas:

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 \cap A_2); \tag{3.7}$$

$$\mathbf{P}(A_1 \setminus A_2) = \mathbf{P}(A_1) - \mathbf{P}(A_1 \cap A_2). \tag{3.8}$$

Proof. We have $n_k(A_1 \cup A_2) = n_k(A_1) + n_k(A_2) - n_k(A_1 \cap A_2)$. Therefore, if, for example, $A_1 \cap A_2 \in \mathcal{G}_S$ then there exists a limit of the right hand side. It implies $A_1 \cup A_2 \in \mathcal{G}_S$ and (3.7) holds. Other implications are proved in the same way. $\square$

Corollary 3.1. *The family $\mathcal{G}_S$ is a semi-algebra.*

In general $A_1, A_2 \in \mathcal{G}_S$ does not imply $A_1 \cup A_2 \in \mathcal{G}_S$. To show this, by Proposition 3.3 it suffices to find $A_1, A_2 \in \mathcal{G}_S$ such that $A_1 \cap A_2 \notin \mathcal{G}_S$. It is easy to do: let $A_1, A_2 \in \mathcal{G}_S$ are such that $|A_1 \cap A_2 \cap M_l| = 1$ for nonempty M_l (there is only one element $x \in A_1 \cap A_2$ on each nonempty floor). If N is infinite then $\lim_{k\to\infty} n_k(A_1 \cap A_2)$ does not exist. Thus

$$\mathcal{G}_S \textit{ is not a set algebra.}$$

It is closed only with respect to a finite unions of sets which have empty intersections. However, $\mathcal{G}_S$ is not closed with respect to countable unions of such sets: in general $(A_j \in \mathcal{G}_S,\ j = 1, 2, \ldots,\ A_i \cap A_j = \emptyset,\ i \neq j)$ does not imply $\bigcup_{j=1}^{\infty} A_j \in \mathcal{G}_S$. The natural additional assumptions (A) $\sum_{j=1}^{\infty} \mathbf{P}(A_j)$ converges in $\mathbb{Q}_p$ or (more strong assumption), (B) $\sum_{j=1}^{\infty} |\mathbf{P}(A_j)|_p < \infty$, also do not imply $A \in \mathcal{G}_S$.

Example 3.1. Let $m = 2$, $N = -1 = 1 + 2 + 2^2 + \cdots + 2^n + \cdots$. Suppose that the sets A_j have the following structure: $|A_j \cap M_{3(j-1)}| = 1$, $|A_j \cap M_{3j-1}| = 2^{3j-1} - 1$ and $A_j \cap M_i = \emptyset$, $i \neq 3(j-1), 3j-1$, i.e., the set A_j is located on two floors of the tower T. In particular, $A_i \cap A_j = \emptyset$, $i \neq j$. As $A_j \in F$, then $A_j \in \mathcal{G}_S$; the probability $\mathbf{P}(A_j) = -2^{3j-1}$, $j = 1, 2, \ldots$. The series $\sum_{j=1}^{\infty} |\mathbf{P}(A_j)|_2 < \infty$. We show that $A = \bigcup_{j=1}^{\infty} A_j \notin \mathcal{G}_S$. We have:

$$n_{3(j-1)}(A) = |A_j \cap T_{3(j-1)}| + \left| \bigcup_{s=1}^{j-1} A_s \cap T_{3(j-1)} \right| = 1 + \gamma,$$

where $|\gamma|_2 < 1$. Thus $|n_{3(j-1)}(A)|_2 = 1$. But $|n_{3j-1}(A)|_2 < 1$.

We note the following useful formula for computing probabilities:

$$\mathbf{P}(A) = \sum_{j=0}^{\infty} \mathbf{P}(A \cap M_j)$$

(probability to find in the tower T an inhabitant $\mathcal{I}$ with the property A is equal to the sum of probabilities to find an inhabitant with this property on the fixed floor).

Definition 3.1. The system $\mathcal{P} = (S, \mathcal{G}_S, \mathbf{P}_S)$ is called a *p-adic ensemble probability space for the ensemble S.*

If N is a finite natural number then we obtain the ensemble probability space which was considered in Chapter 1 (with $\mathcal{G}_S = F_S$). In fact, any ensemble probability space $\mathcal{P}$ can be approximated by ensemble probability spaces $\mathcal{P}_k$ having ensembles of finite volumes. Set

$$n_k = l_0 + l_1 p + \cdots + l_k p^k$$

for N which has the expansion (3.1). Let l_s be the first nonzero digit in (3.1). Consider finite ensembles S_{n_k}, $|S_{n_k}| = n_k$ $(k = s, s+1, \ldots)$, and ensemble probability spaces $\mathcal{P}_{n_k} = (S_{n_k}, \mathcal{G}_{S_{n_k}}, \mathbf{P}_{S_{n_k}})$. There $\mathcal{G}_{S_{n_k}}$ coincides with the algebra $F_{S_{n_k}}$ of all subsets of the finite ensemble S_{n_k} and definition (3.3) of ensemble probability coincides with the definition of Chapter 1:

$$\mathbf{P}_{S_{n_k}}(A) = \frac{|A|}{|S_{n_k}|}, \quad A \in F_{S_{n_k}}. \tag{3.9}$$

We identify S_{n_k} with the population of the first $k+1$ floors of the tower T_S.

Proposition 3.4. *Let $A \in \mathcal{G}_S$. Then*

$$\mathbf{P}_S(A) = \lim_{k\to\infty} \mathbf{P}_{S_{n_k}}(A \cap S_{n_k}). \tag{3.10}$$

To prove (3.10) we have only used that $\mathbb{Q}_p$ is a topological group. This approximation depends essentially on the rule of a measurement, which is defined by the sequence $\{n_k\}$ which gives an approximation of the infinite ensemble S by finite ensembles $\{S_{n_k}\}$. In principle the change of this rule may change the limiting result (see [80] for the details).

Proposition 3.5 (The image of ensemble probability). *The probability* $\mathbf{P}$ *maps* $\mathcal{G}_S$ *into the ball* $U_{r_S}(0)$, *where* $r_S = 1/|N|_p$.

To study conditional probabilities we have to extend the notion of the p-adic ensemble probability to consider more general ensembles.

Let S be the population of the tower T_S with an infinite number of floors M_j, $j = 0, 1, \ldots$, and the following distribution of population: there are m_j elements on the jth floor, $m_j \in \mathbb{N}$ and the series $\sum_{j=1}^{\infty} m_j$ converges in $\mathbb{Z}_p$ to a nonzero number $N = |S|$. We define the p-adic ensemble probability of a set $A \subset S$ by (3.2), (3.3); $\mathcal{G}_S$ is the corresponding family of events. It is easy to check that Propositions 3.1–3.5 hold for this more general ensemble probability.

Let $A \in \mathcal{G}_S$ and $\mathbf{P}(A) \neq 0$. We can consider A as a new ensemble with the p-adic hierarchical structure $A = \bigcup_{j=0}^{\infty} M_{Aj}$, where $M_{Aj} = A \cap M_j$, and introduce the corresponding family of events $\mathcal{G}_A$.

Proposition 3.6 (Conditional probability). *Let $A \in \mathcal{G}_S, \mathbf{P}(A) \neq 0$ and $B \in \mathcal{G}_A$. Then $B \in \mathcal{G}_S$ and Bayes' formula*

$$\mathbf{P}_A(B) = \frac{\mathbf{P}_S(B)}{\mathbf{P}_S(A)} \tag{3.11}$$

holds true.

Proof. The tower T_A of the A has the following population structure: there are M_{Aj} elements on the jth floor. In particular, $T_{Ak} = T_k \cap A$. Thus

$$n_{Ak}(B) = |B \cap T_{Ak}| = |B \cap T_k| = n_k(B) \tag{3.12}$$

for each $B \subset A$. Hence the existence of $n_A(B) = \lim_{k\to\infty} n_{Ak}(B)$ implies the existence of $n_S(B) = \lim_{k\to\infty} n_k(B)$. Moreover, $n_S(B) = n_A(B)$. Therefore,

$$\mathbf{P}_A(B) = \frac{n_A(B)}{n_S(A)} = \frac{n_A(B)/|S|}{n_S(A)/|S|}. \qquad \square$$

By (3.12) we obtain the following consequence:

Corollary 3.2. *Let $A, B \in \mathcal{G}_S$, $\mathbf{P}(A) \neq 0$, and $B \subset A$. Then $B \in \mathcal{G}_A$.*

Thus we obtain

$$\mathcal{G}_A = \{B \in \mathcal{G}_S : B \subset A\}.$$

Let $A, B, A \cap B \in \mathcal{G}_S, \mathbf{P}(A) \neq 0$. We set by definition $\mathbf{P}_A(B) = \mathbf{P}_A(A \cap B)$. Then

$$\mathbf{P}_A(B) = \frac{\mathbf{P}_S(B \cap A)}{\mathbf{P}_S(A)}. \tag{3.13}$$

If we set $\mathbf{P}_A(B) = \mathbf{P}(B/A)$ and omit the index S for the probabilities for an ensemble S, then we obtain Bayes' formula.

Remark 3.1. We have discussed many times the domain of applications of Bayes' formula. This question has the exact and simple mathematical answer in the p-adic ensemble probability theory. We can use Bayes' formula for events A and B iff $A \cap B$ is also the event, i.e., $A \cap B \in \mathcal{G}_S$.

Remark 3.2. It is important for our physical considerations that $\mathcal{G}_S$ is not a set algebra and $\mathbf{P}_S$ can in principle take any value $x \in U_{r_S}$. The manipulations which were used to prove Bell's inequality (Chapter 2) are not legal for the ensemble probability space $\mathcal{P} = (S, \mathcal{G}_S, \mathbf{P}_S)$. For instance, if there are tree sets $B_\phi, B_\theta, B_0 \in \mathcal{G}_S$, then in principle it may be that $B_\phi \cap B_\theta$, $B_\phi \cap B_0$, $B_0 \cap B_\theta \in \mathcal{G}_S$, but $B_\phi \cap B_\theta \cap B_0 \notin \mathcal{G}_S$. Moreover, probabilities can in principle be negative. In this case we cannot use the standard estimate for Kolmogorov probabilities.

3.2 The rules for working with p-adic probabilities

One of the main tools of the ordinary theory of probability is based on the order structure on the field of real numbers $\mathbb{R}$. It gives the possibility of comparing probabilities of different events; events E with probabilities $\mathbf{P}(E) \ll 1$ are considered as negligible and events E with probabilities $\mathbf{P}(E) \approx 1$ are considered as practically certain. However, the use of these relations in concrete applications is essentially based on our (real) probability intuition. What is a large probability? What is a small probability? Moreover, it is not easy to compare two arbitrary probabilities. For instance, do you prefer to win with the probability $\mathbf{P}(E_1) = \frac{11}{17}$ or $\mathbf{P}(E_2) = \frac{13}{19}$. Formally, because $\mathbf{P}(E_1) < \mathbf{P}(E_2)$ it would be better to choose E_2. But in practice this choice does not give many advantages. Thus ordinary probability intuition is based more on centuries of human experiment than on exact mathematical theory.

If we want to work with p-adic probabilities we have to develop some kind of a p-adic probability intuition. However, there arises a mathematical problem which does not give the possibility of generalizing the real scheme directly. This is the absence of an order structure on $\mathbb{Q}_p$. Of course, we can also do something without an order structure. For example, we can classify (split) different events with the aid of their p-adic probabilities. For instance, it works sufficiently successful in the frequency probability theory. If there are two sequences x and y (generated by some statistical experiment) which are not S-sequences in the ordinary von Mises' frequency theory, then we could not split properties of x and y. Both these sequences seem to be totally chaotic from the real point of view. However, if they are (S, τ_p)-sequences, then it would be possible to classify them with the aid of p-adic probability distributions, $\mathbf{P}_x(\alpha_i)$, $\mathbf{P}_y(\alpha_i)$. In the ensemble approach different p-adic probabilities, $\mathbf{P}_S(E_1) \neq \mathbf{P}_S(E_2)$, mean that the events E_1 and E_2 have different p-adic volumes.

However, we could do much more with p-adic probabilities by using the partial order structure which exists on the ring of p-adic integers.

$(\mathcal{O})$ Let $x = x_0 x_1 \dots x_n \dots$ and $y = y_0 y_1 \dots y_n \dots$ be the canonical expansions of two p-adic integers $x, y \in \mathbb{Z}_p$. We set $x < y$ if there exists n such that $x_n < y_n$ and $x_k \leq y_k$ for all $k > n$.

This partial order structure on $\mathbb{Z}_p$ is the natural extension of the standard order structure on the set of natural numbers $\mathbb{N}$. It is easy to see that $x < y$ for any $x \in \mathbb{N}$ and $y \in \mathbb{Z}_p \setminus \mathbb{N}$, i.e., any finite natural number is less that any infinite number. But we could not compare any two infinite numbers.

Example 3.2. Let $p = 2$ and let $x = -1/3 = 10101\dots1010\dots$, $z = -2/3 = 0101\dots0101\dots$ and $y = -16 = 0001\dots1111\dots$. Then $x < y$ and $z < y$, but the numbers x and z are incompatible.

It is important to remark that there exists the maximal number $N_{\max} \in \mathbb{Z}_p$. It is easy to see:

$$N_{\max} = -1 = (p-1) + (p-1)p + \cdots + (p-1)p^n + \cdots .$$

Therefore the ensemble S_{-1} is the largest ensemble which can be considered in the p-adic framework.

Remark 3.3. It seems to be natural to suppose that the volume of the ensemble increases with the increase of p, i.e., $|S_{-1}^p| < |S_{-1}^q|$, $p < q$.

Proposition 3.7. *Let $N \in \mathbb{Z}_p$, $N \neq 0$. Then $S_N \in \mathcal{G}_{S_{-1}}$ and*

$$\mathbf{P}_{S_{-1}}(S_N) = \frac{|S_N|}{|S_{-1}|} = -N. \tag{3.14}$$

Corollary 3.3. *Let $N \in \mathbb{Z}_p$, $N \neq 0$. Then $\mathcal{G}_{S_N} \subset \mathcal{G}_{S_{-1}}$ and probabilities $\mathbf{P}_{S_N}(A)$ are calculated as conditional probabilities with respect to the sub-ensemble S_N of ensemble S_{-1}:*

$$\mathbf{P}_{S_N}(A) = \mathbf{P}_{S_{-1}}(A/S_N) = \frac{\mathbf{P}_{S_{-1}}(A)}{\mathbf{P}_{S_{-1}}(S_N)}, \quad A \in \mathcal{G}_{S_N}. \tag{3.15}$$

But $A \in \mathcal{G}_{S_{-1}}$ does not imply $A \cap S_N \in \mathcal{G}_{S_N}$.

By Corollary 3.3 we can, in fact, restrict our considerations to the case of the maximal ensemble S_{-1}. Therefore we shall study this case $S \equiv S_{-1}$.

The (partial) order $\mathcal{O}$ on the set of p-adic integers $\mathbb{Z}_p$ gives the possibility to compare p-adic volumes $n(A)$ of sets $A \in \mathcal{G}_S$. It is natural to say that probability $\mathbf{P}(B)$ is larger than probability $\mathbf{P}(A)$ if the p-adic volume $n(B)$ of B is larger than the p-adic volume $n(A)$ of A. Thus we obtain the following (partial) order on the set of probabilities:

($\tilde{\mathcal{O}}$) $\mathbf{P}(B) > \mathbf{P}(A)$ iff $n(B) > n(A)$.

We use the same symbols $>$, $<$ for this new order on $\mathbb{Z}_p$. We hope that the reader would not mix these two orders on $\mathbb{Z}_p$: $\mathcal{O}$-order is used to compare p-adic volumes, $\tilde{\mathcal{O}}$-order is used to compare probabilities. For example, let $p = 2$ and let $n(B) = -2 (= 011\dots1\dots)$, $n(A) = -3 (= 1011\dots1\dots)$. Then $n(B) > n(A)$ (with respect to $\mathcal{O}$) and consequently $\mathbf{P}(B) = 2 > \mathbf{P}(A) = 3$ (with respect to $\tilde{\mathcal{O}}$).

We study some properties of probabilities.

(1) As we have only a partial order structure we cannot compare probabilities of arbitrary two events A and B.

(2) As $x \leq -1$ with respect to $\mathcal{O}$ for any $x \in \mathbb{Z}_p$, we have $\mathbf{P}(A) \leq 1 = \mathbf{P}(S)$ for any $A \in \mathcal{G}_S$.

(3) As $x \geq 0$ with respect to $\mathcal{O}$ for any $x \in \mathbb{Z}_p$, we have $\mathbf{P}(A) \geq 0$ for any $A \in \mathcal{G}_S$.

To illustrate further properties of p-adic probabilities, we shall use the third order structure, namely, the usual real order structure on the set $\mathbb{Z}_p \cap \mathbb{Q}$. In this case we shall say r-increase or r-decrease. This r-order on $\mathbb{Z}_p \cap \mathbb{Q}$ has no probabilistic meaning. We consider this order, because we want to use the 'real intuition' to imagine the location of rational probabilities $\mathbf{P}(A)$, $A \in \mathcal{G}_S$, on the real line. We shall use the symbols $[a, b], \dots, (a, b)$ for corresponding intervals of the real line. For example, let $p = 2$ and let $\mathbf{P}(B) = 2$ and $\mathbf{P}(A) = 3$. Then $\mathbf{P}(B) > \mathbf{P}(A)$, but from the viewpoint of the r-order $\mathbf{P}(B)$ is less than $\mathbf{P}(A)$.

(4) Set $F^f = \{A \in \mathcal{G}_S : n(A) \in \mathbb{N}\}$.[6]

The restriction of the order $\mathcal{O}$ on the set of natural numbers $\mathbb{N}$ coincides with the standard (real) order on $\mathbb{N}$. Thus $n(A) < n(B)$, $A, B \in F^f$, iff the natural number $n(A)$ is less than the natural number $n(B)$. This implies (by definition of the order $\tilde{\mathcal{O}}$ on the set of probabilities) that $\mathbf{P} : F^f \to (-\infty, 0) \cap \mathbb{Z}$ and $\mathbf{P}(A)$ is increasing if $\mathbf{P}(A)$ is r-decreasing. Therefore, for example, probabilities $\mathbf{P}(A) = -1$ or -3 are rather small with respect to probabilities $\mathbf{P}(B) = -100$ or -300.

(5) Set $\bar{F}^f = \{B = \bar{A} : A \in F^f\}$ (in particular, $\bar{F}^f$ contains complements of all finite subsets of Ω). Then $\mathbf{P} : \bar{F}^f \to \mathbb{N}$ and $\mathbf{P}(B)$ is decreasing if $\mathbf{P}(B)$ is r-increasing. Therefore, for example, probabilities $\mathbf{P}(E) = 100$ or 200 are rather small with respect to probabilities $\mathbf{P}(C) = 1$ or 2.

[6] In particular, F^f contains all finite subsets of S. The F^f contains also some infinite subsets $A \in \mathcal{G}_S$ which have finite p-adic volumes. For example, let $|A \cap T_k| = 1 + p^k$, $k = 1, 2, \dots$ ($1 + p^k$ inhabitants of the first $(k + 1)$ floors have the property A). Then $n(A) = 1$ and hence $A \in F^f$.

We can use these rules for conditional probabilities. For example, let $\mathbf{P}(B) = 100$, $\mathbf{P}(B') = 200$, $\mathbf{P}(A) = 2$ and $B, B' \subset A$. Then $\mathbf{P}(B/A) = 50 > \mathbf{P}(B'/A) = 100$.

By (4) and (5) we can work with probabilities belonging to $F^f \cup \bar{F}^f$.

(6) Now consider events $A \notin F^f \cup \bar{F}^f$. We can develop our intuition only by examples.

Example 3.3. Let $p = 2$. Let $|A \cap M_{2k}| = 2^{2k}$ and $A \cap M_{2k+1} = \varnothing$, $k = 0, 1, \ldots$. Then $n(A) = -1/3$ $(= 1010 \ldots 10 \ldots)$ and $\mathbf{P}(A) = 1/3$. Let $B \subset A$ and $B \cap M_{4k} = A \cap M_{4k}$, $B \cap M_j = \varnothing$, $j \neq 4k$. Then $n(B) = -1/15(= 100010001 \ldots 10001 \ldots)$ and $\mathbf{P}(B) = 1/15$. It is evident that $-1/15 < -1/3$ in $\mathbb{Z}_2$. Hence $\mathbf{P}(B) = 1/15 < \mathbf{P}(A) = 1/3$.

Thus it seems to be that the probabilistic order relation on the set $[0, 1] \cap \mathbb{Q}$ coincides with the standard real order. Moreover, it seems to be reasonable to use this relation also in the case where the numbers $n(A)$ and $n(B)$ are incompatible in $\mathbb{Z}_2$.[7]

Example 3.4. Let p and A be the same as above. Let $|C \cap M_{2k+1}| = 2^{2k+1}$, $C \cap M_{2k} = \varnothing$, $k = 0, 1, \ldots$. Then $n(C) = -2/3$ and $\mathbf{P}(C) = 2/3$. The numbers $n(A) = -1/3$ and $n(C) = -2/3$ are incompatible in $\mathbb{Z}_2$. But heuristically it seems to be evident that we can use the r-order structure on [0, 1] to compare the probabilities of the events A and C. Therefore the probability of $\omega \in C$ is two times larger than the probability $\omega \in A$. These heuristic reasons were also confirmed by some frequency statistical models, see [80] for the details.

Further we have that a probability $x \in (-\infty, 0) \cap \mathbb{Z}$ is practically negligible with respect to any probability $y \in (0, 1] \cap \mathbb{Q}$. The intuitive argument is the following. A probability $\mathbf{P}(A) \in (-\infty, 0) \cap \mathbb{Z}$ is probability of an event A with a finite p-adic volume in the infinitely large ensemble S. Probability $\mathbf{P}(A) \in (0, 1] \cap \mathbb{Q}$ is probability of an event A with an infinite p-adic volume in the infinitely large ensemble S.

Therefore, p-adics gives the possibility to split probability 0 to a set of probabilities, $0 \to D_0^+$; in particular, $(-\infty, 0) \cap \mathbb{Z} \subset D_0^+$.

Remark 3.4. A probability $\mathbf{P}$ on a Boolean algebra $\mathcal{A}$ is non-degenerated: $\mathbf{P}(A) = 0$, $A \in \mathcal{A}$ iff $A = \varnothing$. The p-adic split of probability 0 can be considered as a step in the direction to Boolean probabilities. The set of new labels D_0^+ gives the possibility to split many probabilities which must be equal to probability 0 from the viewpoint of real analysis. However, we still have not obtained a Boolean probability. There are numerous events $A \in \mathcal{G}_S$, $A \neq \varnothing$, which have probability 0. For example, let $|A \cap T_k| = p^k$, $k = 1, 2, \ldots$. Then $\mathbf{P}(A) = 0$.

We can also use these rules for conditional probabilities. For example, let $\mathbf{P}(B) = 1/15 < \mathbf{P}(B') = 2/15$, $\mathbf{P}(A) = 1/5$ and $B, B' \subset A$. Then $\mathbf{P}(B/A) = 1/3 <$

[7] However, probably it is the wrong extrapolation and we must assume existence of events with incompatible probabilities.

$\mathbf{P}(B'/A) = 2/3$. Moreover, for example, let $\mathbf{P}(B) = -1 < \mathbf{P}(B') = -5$, $\mathbf{P}(A) = -100$ and $B, B' \subset A$. Then $\mathbf{P}(B/A) = 1/100 < \mathbf{P}(B'/A) = 1/20$. Thus the r-order structure on $(0, 1] \cap \mathbb{Q}$ reproduces the rule (4).

Proposition 3.8. *If* $\mathbf{P}(B) \in \mathbb{N}$, *then* $n(\bar{B}) \in \{0\} \cup \mathbb{N}$; *if* $\mathbf{P}(B) \in (0,1) \cap \mathbb{Q}$ *then* $n(\bar{B}) \in \mathbb{Z}_p \setminus \mathbb{N}$.

Proof. If $k = \mathbf{P}(B) \in \mathbb{N}$, then $n(B) = -k$, $k = 1, 2, \dots$, and $n(\bar{B}) = -1 + k$. If $a = \mathbf{P}(B) \in (0,1) \cap \mathbb{Q}$ then $n(B) = -a$ and $n(\bar{B}) = a - 1 \notin \mathbb{N}$. □

Thus if $\mathbf{P}(B) \in \mathbb{N}$, then the set $\bar{B}$ has a finite p-adic volume, $n(\bar{B})$. On the other hand, if $\mathbf{P}(B) \in (0,1) \cap \mathbb{Q}$, then the set $\bar{B}$ has an infinite p-adic volume, $n(\bar{B})$. It is natural to assume that probability $\mathbf{P}(B) \in \mathbb{N}$ is larger than any probability $\mathbf{P}(C) \in (0,1) \cap \mathbb{Q}$.

Therefore, p-adics gives the possibility to split probability 1 to a set of probabilities, $1 \to D_1^-$. In particular, $\mathbb{N} \subset D_1^-$. However, the probability 1 is still not totally split. There are numerous events $A \neq \varnothing$ with $\mathbf{P}(A) = 1$. For example, let $|A \cap M_k| = p^{[(k+1)/2]} - 1$, $k = 1, 2, \dots$ (here $[x]$ denotes the integer part of x). Then $n(A) = -1$ and $\mathbf{P}(A) = 1$. But $\bar{A} \neq \varnothing$.

We can also split all probabilities $x = \mathbf{P}(A) \in (0,1) \cap \mathbb{Q}$.

Let $A \in \mathcal{G}_S$, $x = \mathbf{P}(A) \in (0,1) \cap \mathbb{Q}$, $C \in F^f$, $A \cap C = \varnothing$, and let $B = A \cup C$. Then $\lambda = \mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(C) = x - k$, where $\mathbf{P}(C) = -k$, $k \in \mathbb{N}$. As the p-adic volume of the set C is finite (and the ensemble S is infinite) probability $\mathbf{P}(C) = -k$ is infinitely small. Thus the probability x can be split in a set of probabilities D_x^+. Each probability $\lambda \in D_x^+$ is larger than probability x and probability $\Delta = \lambda - x = -k$ is infinitely small.

Let $B \in \mathcal{G}_S$, $C \in F^f$, $B \cap C = \varnothing$, and let $A = B \cup C$, $x = \mathbf{P}(A) \in (0,1) \cap \mathbb{Q}$. Then $\lambda = \mathbf{P}(B) = \mathbf{P}(A) - \mathbf{P}(C) = x + k$, where $\mathbf{P}(C) = -k$, $k \in \mathbb{N}$, is infinitely small probability. Thus the probability x can be split in a set of probabilities D_x^-. Each probability $\lambda \in D_x^-$ is less than probability x and probability $\Delta = x - \lambda = -k$ is infinitely small.

Thus probability x is split in a set of probabilities $D_x = D_x^- \cup D_x^+$.

We now consider probabilities with respect to an ensemble S_N for an arbitrary $N \in \mathbb{Z}_p$, $N \neq 0$. By using formula (3.15) we can translate to the general case results obtained for the ensemble $S = S_{-1}$. In the general case probability 0 is split in a set D_0^+ which contains the set $\{\lambda = \frac{k}{N} : k \in \mathbb{N}\}$; probability 1 is split in a set D_1^- which contains the set $\{\lambda = 1 - \frac{k}{N} : k \in \mathbb{N}\}$; probability $x \in (0,1) \cap \mathbb{Q}$ is split in a set $D_x = D_x^- \cup D_x^+$, where D_x^-, in particular, contains the set $\{\lambda = x - \frac{k}{N} : k \in \mathbb{N}\}$ and D_x^+, in particular, contains the set $\{\lambda = x + \frac{k}{N} : k \in \mathbb{N}\}$.

3.3 Negative probabilities and p-adic ensemble probabilities

Let us consider Example 2.2 of Chapter 3 from the p-adic viewpoint. The series $|S| = 1 + 2 + \cdots + 2^k + \cdots = -1$ converges in $\mathbb{Q}_2$. Thus the statistical ensemble

S of Example 2.2 has the 2-adic maximal volume -1. Probabilities $\mathbf{p}_k = |S(\lambda = \lambda_k)|/|S| = -2^k$ are infinitely small probabilities. p-adic approach implies that the distribution of quantum systems regarding to values $\lambda = \lambda_j$ of hidden variables has the 2-adic hierarchical structure. The ensemble S has the form of a tower in that the jth floor is 'populated' by quantum systems s with the property $\lambda = \lambda_j$. If we assume that a preparation procedure $\mathcal{E}$ produces portions of quantum systems in the accordance to this tower structure, then there will be extremely unstable behaviour of properties $\lambda = \lambda_j$ in quantum data which will be used in an experiment (compare with [93]).

The summation in the formula of total probability

$$\mathbf{P}_S(A) = \sum_{k=0}^{\infty} \mathbf{p}_k \sum_{j \in j(A)} \mathbf{p}_{kj} \tag{3.16}$$

is meaningful from 2-adic viewpoint for conditional probabilities $\mathbf{p}_{kj}$ which do not depend on k (for finite sets A).

We now consider Example 2.3 of Chapter 3. Here conditional probabilities $\mathbf{p}_{kl} = -2^l$ are well defined in $\mathbb{Q}_2$. These are infinitely small probabilities. The summation in (3.16) is meaningful. For example, for $A = \{\lambda'_0, \ldots, \lambda'_{2k}, \ldots\}$ we have

$$\mathbf{P}_S(A) = \left(\sum_{l=0}^{\infty} -2^l\right)\left(\sum_{j=0}^{\infty} -2^{2j}\right) = \frac{1}{3},$$

$$\mathbf{P}_S(\bar{A}) = \left(\sum_{l=0}^{\infty} -2^l\right)\left(\sum_{j=0}^{\infty} -2^{2j+1}\right) = \frac{2}{3}.$$

All above series converge in $\mathbb{Q}_2$.

Finally we consider Example 2.4 of Chapter 3. By equality (1.10) the factorial series $\sum_{k=0}^{\infty} k!$ converges in each field $\mathbb{Q}_p$. Thus conditional probabilities

$$\mathbf{p}_{kl} = \frac{l!}{\sum_{k=0}^{\infty} k!}$$

are well defined in each $\mathbb{Q}_p$.

4 Measures

Let X be an arbitrary set and let $\mathcal{R}$ be a ring of subsets of X. The pair $(X, \mathcal{R})$ is called a *measurable space.* The ring $\mathcal{R}$ is said to be *separating* if for every two distinct elements, x and y, of X there exists an $A \in \mathcal{R}$ such that $x \in A$, $y \notin A$. We shall consider measurable spaces only over separating rings which cover X.

Every ring $\mathcal{R}$ can be used as a base for the zero-dimensional topology[8] which we shall call the $\mathcal{R}$-topology. This topology is Hausdorff iff $\mathcal{R}$ is separating.

Throughout this section, $\mathcal{R}$ is a separating covering ring of a set X.

A subcollection $\mathcal{S}$ of $\mathcal{R}$ is said to be *shrinking* if the intersection of any two elements of $\mathcal{S}$ contains an element of $\mathcal{S}$. If $\mathcal{S}$ is shrinking, and if f is a map $\mathcal{R} \to K$ or $\mathcal{R} \to \mathbb{R}$, we say that $\lim_{A \in \mathcal{S}} f(A) = 0$ if for every $\epsilon > 0$, there exists an $A_0 \in \mathcal{S}$ such that $|f(A)| \le \epsilon$ for all $A \in \mathcal{S}$, $A \subset A_0$.

Let K be a non-Archimedean field with the valuation $|\cdot|_K$.

A *measure* on $\mathcal{R}$ is a map $\mu : \mathcal{R} \to K$ with the properties:

(i) μ is additive;

(ii) for all $A \in \mathcal{R}$, $\|A\|_\mu = \sup\{|\mu(B)|_K : B \in \mathcal{R}, B \subset A\} < \infty$;

(iii) if $\mathcal{S} \subset \mathcal{R}$ is shrinking and has empty intersection, then $\lim_{A \in \mathcal{S}} \mu(A) = 0$.

We call these conditions respectively *additivity*, *bounded*, *continuity*. The latter condition is equivalent to the following: $\lim_{A \in \mathcal{S}} \|A\|_\mu = 0$ for every shrinking collection $\mathcal{S}$ with empty intersection.

Condition (iii) is the replacement for σ-additivity. Clearly (iii) implies σ-additivity. Moreover, we shall see that for the most interesting cases (iii) is equivalent to σ-additivity. Of course, we could in principle restrict our attention to these cases and use the standard condition of σ-additivity. However, in that case we should use some topological restriction on the space X. This implies that we must consider some topological structure on a p-adic probability space. We do not like to do this. We shall develop the theory of p-adic probability measures in the same way as A. N. Kolmogorov (1933) developed the theory of real valued probability measures by starting with an arbitrary set algebra.

Further, we shall briefly discuss the main properties of measures, see [164] for the details. As in Chapter 1, for any set D, we denote its characteristic function by the symbol I_D. For $f : X \to K$ and $\phi : X \to [0, \infty)$, put

$$\|f\|_\phi = \sup_{x \in X} |f(x)|_K \phi(x).$$

We set

$$N_\mu(x) = \inf_{U \in \mathcal{R}, x \in U} \|U\|_\mu$$

for $x \in X$. Then $\|A\|_\mu = \|I_A\|_{N_\mu}$ for any $A \in \mathcal{R}$. We set $\|f\|_\mu = \|f\|_{N_\mu}$.

A *step function* (or $\mathcal{R}$-step function) is a function $f : X \to K$ of the form $f(x) = \sum_{k=1}^N c_k I_{A_k}(x)$ where $c_k \in K$ and $A_k \in \mathcal{R}$, $A_k \cap A_l = \varnothing, k \ne l$. We set for such a function

$$\int_X f(x)\, \mu(dx) = \sum_{k=1}^N c_k \mu(A_k).$$

[8] A topological space $(X; \tau)$ is zero-dimensional if each point $x \in X$ has a basis of clopen (i.e., at the same time open and closed) neighborhoods.

Denote the space of all step functions by the symbol $S(X)$. The integral $f \to \int_X f(x)\,\mu(dx)$ is the linear functional on $S(X)$ which satisfies the inequality

$$\Big|\int_X f(x)\,\mu(dx)\Big|_K \le \|f\|_\mu. \tag{4.1}$$

A function $f : X \to K$ is called *μ-integrable* if there exists a sequence of step functions $\{f_n\}$ such that $\lim_{n\to\infty} \|f - f_n\|_\mu = 0$. The μ-integrable functions form a vector space $L_1(X, \mu)$ (and $S(X) \subset L_1(X, \mu)$). The integral is extended from $S(X)$ on $L_1(X, \mu)$ by continuity. The inequality (4.1) holds for $f \in L_1(X, \mu)$.

Let $\mathcal{R}_\mu = \{A : A \subset X, I_A \in L_1(X, \mu)\}$. This is a ring. Elements of this ring are called μ-measurable sets. By setting $\mu(A) = \int_X I_A(x)\,\mu(dx)$ the measure μ is extended to a measure on $\mathcal{R}_\mu$. This is the *maximal extension* of μ, i.e., if we repeat the previous procedure starting with the ring $\mathcal{R}_\mu$, we will obtain this ring again.

Set $X_\epsilon = \{x \in X : N_\mu(x) \ge \epsilon\}$, $X_0 = \{x \in X : N_\mu(x) = 0\}$, $X_+ = X \setminus X_0$. Every $A \subset X_0$ belongs to $\mathcal{R}_\mu$. We call such sets *μ-negligible.*

Now we construct product measures. Let μ_j, $j = 1, 2, \dots, n$, be measures on (separating) rings $\mathcal{R}_j$ of subsets of sets X_j. The finite unions of the sets $A_1 \times \cdots \times A_n$, $A_j \in \mathcal{R}_j$, form a (separating) ring $\mathcal{R}_1 \times \cdots \times \mathcal{R}_n$ of $X_1 \times \cdots \times X_n$. Then there exists a unique measure $\mu_1 \times \cdots \times \mu_n$ on $\mathcal{R}_1 \times \cdots \times \mathcal{R}_n$ such that $\mu_1 \times \cdots \times \mu_n(A_1 \times \cdots \times A_n) = \mu_1(A_1) \times \cdots \times \mu_n(A_n)$. We have

$$N_{\mu_1\times\cdots\times\mu_n}(x_1, \dots, x_n) = N_{\mu_1}(x_1) \times \cdots \times N_{\mu_n}(x_n).$$

Let X be a zero-dimensional topological space[9]. We denote the ring of *clopen* (i.e., at the same time open and closed) subsets of X by the symbol $B(X)$ (in fact, this is an algebra). We denote the space of continuous bounded functions $f : X \to K$ by the symbol $C_b(X)$. We use the norm $\|f\|_\infty = \sup_{x\in X} |f(x)|_K$ on this space.

First we remark that *if X is compact and $\mathcal{R} = B(X)$ then the condition (iii) in the definition of a measure is redundant.* If X is not compact then there exist bounded additive set functions which are not continuous.

Let X be zero-dimensional $\mathbb{N}$-compact topological space, i.e., there exists a set S such that X is homeomorphic to a closed subset of $\mathbb{N}^S$. We remark that every product of $\mathbb{N}$-compact spaces is $\mathbb{N}$-compact; every closed subspace of an $\mathbb{N}$-compact space is $\mathbb{N}$-compact. Then every bounded σ-additive function $\mu : B(X) \to K$ is a measure. On the other hand, if X is a zero-dimensional space such that every bounded σ-additive function $B(X) \to K$ is a measure, then X is $\mathbb{N}$-compact.

In the theory of integration a crucial role is played by the $\mathcal{R}_\mu$-topology, i.e., the (zero-dimensional) topology that has $\mathcal{R}_\mu$ as a base. Of course, $\mathcal{R}_\mu$-topology is stronger that $\mathcal{R}$-topology. Every μ-negligible set is $\mathcal{R}_\mu$-clopen. The following two theorems [164] will be important for our considerations.

[9]We consider only Hausdorff spaces.

Theorem 4.1. *(i) If μ is a measure on $\mathcal{R}$, then N_μ is $\mathcal{R}$-upper semicontinuous, (hence, $\mathcal{R}_\mu$-upper semicontinuous) and for every $A \in \mathcal{R}_\mu$ and $\epsilon > 0$ the set $A_\epsilon = A \cap X_\epsilon$ is $\mathcal{R}_\mu$-compact.*

(ii) Conversely, let $\mu : \mathcal{R} \to K$ be additive. Assume that there exists an $\mathcal{R}$-upper semicontinuous $\phi : X \to [0, \infty)$ such that $|\mu(A)|_K \le \sup_{x \in A} \phi(x)$, $A \in \mathcal{R}$, and $\{x \in A : \phi(x) \ge \epsilon\}$ is $\mathcal{R}$-compact ($A \in \mathcal{R}, \epsilon > 0$). Then μ is a measure and $N_\mu \le \phi$.

Theorem 4.2. *Let $\mu : \mathcal{R} \to K$ be a measure. A function $f : X \to K$ is μ-integrable iff it has the following two properties: (1) f is $\mathcal{R}_\mu$-continuous; (2) for every $\epsilon > 0$, the set $\{x : |f(x)|_K N_\mu(x) \ge \epsilon\}$ is $\mathcal{R}_\mu$-compact.*

We shall also use the following fact.

Theorem 4.3. *Let $f \in L_1(X, \mu)$ and let*

$$\int_A f(x)\,\mu(dx) = 0 \quad \textit{for every} \quad A \in \mathcal{R}. \tag{4.2}$$

Then supp $f \subset X_0$.

Proof. Let us assume that f satisfies (4.2) and there exists $x_0 \in X_+$ (hence $N_\mu(x_0) = \alpha > 0$) such that $|f(x_0)|_K = c > 0$. Let $\{f_k\}$ be a sequence of $\mathcal{R}$-step functions which approximates f. For every $\epsilon > 0$ there exist N_ϵ such that $\|f - f_k\|_\mu < \alpha\epsilon$ for all $k \ge N_\epsilon$. In particular, this implies that $|f_k(x_0)|_K \ge c - \epsilon$, $k \ge N_\epsilon$. Then we have

$$\begin{aligned}\Delta_{B,k} &= \Big|\int_B f_k(x)\,\mu(dx)\Big|_K \\ &= \Big|\int_B f_k(x)\,\mu(dx) - \int_B f(x)\,\mu(dx)\Big|_K < \alpha\epsilon, \quad B \in \mathcal{R}.\end{aligned}$$

Let

$$f_k(x) = \sum_j c_{kj} I_{B_{kj}}(x), \quad c_{kj} \in K, \quad B_{kj} \in \mathcal{R}, \quad B_{kj} \cap B_{ki} = \varnothing, \quad i \ne j,$$

and let $x_0 \in B_{kj_0}$. If $B \subset B_{kj_0}$, $B \in \mathcal{R}$, then we have $\Delta_{B,k} = |c_{kj}|_K |\mu(B)|_K = |f_k(x_0)|_K |\mu(B)|_K < \alpha\epsilon$. On the other hand, as $\|B_{kj_0}\|_\mu \ge \alpha$, then for every $\delta > 0$, there exists $B \subset B_{kj_0}$, $B \in \mathcal{R}$, such that $|\mu(B)|_K \ge (\alpha - \delta)$. Thus we obtain for this B: $\Delta_{B,k} \ge (\alpha - \delta)(c - \epsilon)$. By choosing $\epsilon > 0$, $\delta > 0$, such that $(\alpha - \delta)(c - \epsilon) > \alpha\epsilon$ arrive to the contradiction. □

We shall use the following simple fact.

Lemma 4.1. *Let $(X_j, \mathcal{R}_j)$, $j = 1, 2$, be measurable spaces and let $f : X_1 \to X_2$ be measurable. If $\mathcal{S}$ is shrinking in $\mathcal{R}_2$ then $f^{-1}(\mathcal{S})$ is shrinking in $\mathcal{R}_1$. If $\mathcal{S}$ has empty intersection, then $f^{-1}(\mathcal{S})$ has also empty intersection.*

Lemma 4.2. *Let* $(X_j, \mathcal{R}_j)$, $j = 1, 2$, *be measurable spaces and let* $\eta : X_1 \to X_2$ *be a measurable function. Then, for every measure* $\mu : \mathcal{R}_1 \to K$, *the function* $\mu_\eta : \mathcal{R}_2 \to K$ *defined by the equality* $\mu_\eta(A) = \mu(\eta^{-1}(A))$ *is a measure on* $\mathcal{R}_2$ *and, for every* $\mathcal{R}_2$*-continuous function,* $h : X_2 \to K$ *the following inequality holds:*

$$\|h\|_{\mu_\eta} \le \|h \circ \eta\|_\mu. \tag{4.3}$$

Proof. We have for every $A \in \mathcal{R}_2$,

$$\|A\|_{\mu_\eta} = \sup\{|\mu(\eta^{-1}(B)) : B \in \mathcal{R}_2, B \subset A\} \le \|\eta^{-1}(A)\|_\mu < \infty. \tag{4.4}$$

Thus μ_η is bounded. We now prove that μ_η is continuous on $\mathcal{R}_2$. Let $\mathcal{S}$ be shrinking in $\mathcal{R}_2$ which has the empty intersection. By Lemma 4.1 $\eta^{-1}(\mathcal{S})$ is shrinking in $\mathcal{R}_1$ which has also the empty intersection. By (4.4) we obtain that $\lim_{A \in \mathcal{S}} \|A\|_{\mu_\eta} = 0$.

We prove inequality (4.3). Let $h : X_2 \to K$ be $\mathcal{R}_2$-continuous. We wish to prove that $|h(b)|_K N_{\mu_\eta}(b) \le \|h \circ \eta\|_\mu$ for all $b \in X_2$. So we choose $b \in X_2$ with $h(b) \neq 0$. Then the set $C_b = \{y \in X_2 : |h(y)|_K = |h(b)|_K\}$ is $\mathcal{R}_2$-open. Hence there is a $B \in \mathcal{R}_2$ with $b \in B \subset C_b$. Then

$$\begin{aligned} |h(b)|_K N_{\mu_\eta}(b) &\le |h(b)|_K \|B\|_{\mu_\eta} \le |h(b)|_K \|\eta^{-1}(B)\|_\mu \\ &= \sup_{x \in \eta^{-1}(B)} |h(b)|_K N_\mu(x) \le \sup_{x \in \eta^{-1}(B)} |(h \circ \eta)(x)|_K N_\mu(x) \\ &\le \|h \circ \eta\|_\mu. \end{aligned}$$

□

The following theorem on the change of variables will be important in our probabilistic considerations.

Theorem 4.4 (Khrennikov–van Rooij). *Let* $(X_j, \mathcal{R}_j)$, $j = 1, 2$, *be measurable spaces and let* $\eta : X_1 \to X_2$ *be a measurable function, and let* $\mu : \mathcal{R}_1 \to K$ *be a measure. If* $f : X_2 \to K$ *is an* $\mathcal{R}_2$*-continuous function such that the function* $f \circ \eta$ *belongs to* $L_1(X_1, \mu)$, *then* $f \in L_1(X_2, \mu_\eta)$ *and*

$$\int_{X_1} f(\eta(x))\, \mu(dx) = \int_{X_2} f(y)\, \mu_\eta(dy).$$

Proof. It suffices to prove that for every $\epsilon > 0$ there exists a $\mathcal{R}_2$-step function g such that $\|f - g\|_{\mu_\eta} \le \epsilon$ and $\|f \circ \eta - g \circ \eta\|_\mu \le \epsilon$. By (4.3) the first follows from the second. So we fix $\epsilon > 0$.

By Theorem 4.2 the set

$$A = \{x \in X_1 : |(f \circ \eta)(x)|_K N_\mu(x) \ge \epsilon\}$$

is $\mathcal{R}_1$-compact and therefore contained in an element of $\mathcal{R}_1$. But N_μ is bounded on every element of $\mathcal{R}_1$, so N_μ is bounded on A. We choose $\delta > 0$ so that

$$\delta N_\mu(x) \le \epsilon \quad \text{for all } x \in A.$$

As A is compact, $f(\eta(A))$ is also compact. We can cover $f(\eta(A))$ by disjoint closed balls of radius δ: $f(\eta(A)) \subset U_\delta(\alpha_0) \cup \cdots \cup U_\delta(\alpha_N)$, where α_0 is chosen to be 0 in order to obtain:

$$|\alpha_n|_K \le |t|_K \quad \text{for } t \in U_\delta(\alpha_n),\ n = 0, 1, \ldots, N. \tag{4.5}$$

For each n, $\mathcal{C}_n = \{C \in \mathcal{R}_2 : C \subset f^{-1}(U_\delta(\alpha_n))\}$ is a collection of open sets covering the compact set $\eta(A) \cap f^{-1}(U_\delta(\alpha_n))$. Thus, for each n there is a $C_n \in \mathcal{C}_n$ such that $\eta(A) \cap f^{-1}(U_\delta(\alpha_n)) \subset C_n$. We now have

$$C_0, \ldots, C_N \in \mathcal{R}_2,$$
$$C_n \subset f^{-1}(U_\delta(\alpha_n)), \quad n = 0, 1, \ldots, N,$$
$$\eta(A) \subset C_0 \cup \cdots \cup C_N.$$

Put $g(x) = \sum_{n=0}^{N} \alpha_n I_{C_n}(x)$. Then g is a $\mathcal{R}_2$-step function. We wish to show that, for all $a \in X$,

$$\Delta(a) = |(f \circ \eta)(a) - (g \circ \eta)(a)|_K N_\mu(a) \le \epsilon.$$

Thus, take $a \in X$:

(1) If $a \in A$, then there is a unique n with $\eta(a) \in C_n$. Then $\Delta(a) = |(f \circ \eta)(a) - \alpha_n|_K N_\mu(a) \le \delta N_\mu(a) \le \epsilon$.

(2) If $a \notin A$, but $\eta(a) \in C_n$ for some n, then by (4.5) we obtain that $\Delta(a) = |(f \circ \eta)(a) - \alpha_n|_K N_\mu(a) \le |(f \circ \eta)(a)|_K N_\mu(a) \le \epsilon$.

(3) If $a \notin C_0 \cup \cdots \cup C_N$, then $g(\eta(a)) = 0$. Thus $\Delta(a) = |(f \circ \eta)(a)|_K N_\mu(a) \le \epsilon$ (as $a \notin A$). □

Open Problem. Find a condition for functions f which is weaker than continuity, but implies the formula of the change of variables.

Further we shall obtain some properties of measures which are specific for measures defined on algebras.

Throughout this section, $\mathcal{A}$ is a separating algebra of a set X. First we remark that if we start with a measure μ defined on the algebra $\mathcal{A}$ then the system $\mathcal{A}_\mu$ of μ-integrable sets is again an algebra.

Proposition 4.1. *Let $\mu : \mathcal{A} \to K$ be a measure. Then for each $\epsilon > 0$, the set X_ϵ is $\mathcal{A}_\mu$-compact.*

This fact is a consequence of Theorem 4.1.

Proposition 4.2. *Let $\mu : \mathcal{A} \to K$ be a measure. Then the algebra $B(X)$ of $\mathcal{A}_\mu$-clopen sets coincides with the algebra $\mathcal{A}_\mu$.*

Proof. We use Theorem 4.2 and the previous proposition. Let $B \in B(X)$. Then I_B is $\mathcal{A}_\mu$-continuous and $\{x : |I_B(x)|_K N_\mu(x) \geq \epsilon\} = B \cap X_\epsilon$. As B is closed and X_ϵ is compact, $B \cap X_\epsilon$ is compact. Thus $B(X) \subset \mathcal{A}_\mu$. □

As a consequence of Proposition 4.2, we obtain that $C_b(X) \subset L_1(X, \mu)$ (for the space X endowed with $\mathcal{A}_\mu$-topology) and the following inequality holds:

$$\Big| \int_X f(x)\, \mu(dx) \Big|_K \leq \|f\|_\infty \|X\|_\mu, \quad f \in C_b(X).$$

Let X be a zero-dimensional topological space. A measure μ defined on the algebra $B(X)$ of the clopen sets is called a *tight* measure. Thus by Proposition 4.2 every measure $\mu : \mathcal{A} \to K$ is extended to a tight measure on the space X endowed with the $\mathcal{A}_\mu$-topology.

Proposition 4.3. *Let $\mu : \mathcal{A} \to K$ be a measure and let $f \in L_1(X, \mu)$. Then f is $(\mathcal{A}_\mu, B(K))$-measurable.*

Proof. By Theorem 4.2 f is $\mathcal{A}_\mu$-continuous. Thus $f^{-1}(B(K)) \subset B(X)$. But by Proposition 4.2 we have that $\mathcal{A}_\mu = B(X)$. □

5 p-adic probability space

Let $\mu : \mathcal{A} \to \mathbb{Q}_p$ be a measure defined on a separating algebra $\mathcal{A}$ of subsets of the set Ω which satisfies the normalization condition $\mu(\Omega) = 1$. We set $\mathcal{F} = \mathcal{A}_\mu$ and denote the extension of μ on $\mathcal{F}$ by the symbol $\mathbf{P}$. A triple $(\Omega, \mathcal{F}, \mathbf{P})$ is said to be a p-adic *probability space* (Ω is a *sample space*, $\mathcal{F}$ is an algebra of *events*, $\mathbf{P}$ is a *probability*).

As in general measure theory we set

$$\Omega_\alpha = \{\omega \in \Omega : N_{\mathbf{P}}(\omega) \geq \alpha\}, \quad \alpha > 0, \quad \Omega_+ = \bigcup_{\alpha > 0} \Omega_\alpha, \quad \Omega_0 = \Omega \setminus \Omega_+.$$

If a property Ξ is valid on the subset Ω_+ we say that Ξ is valid a.e. (mod $\mathbf{P}$).

Everywhere below (G, Γ) denotes a measurable space over the algebra Γ. Functions $\xi : \Omega \to G$ which are $(\mathcal{F}, \Gamma)$-measurable are said to be random variables.

Everywhere below Y is a zero-dimensional topological space. We consider Y as the measurable space over the algebra $B(Y)$. Every random variable $\xi : \Omega \to Y$ is continuous in the $\mathcal{F}$-topology. In particular, $\mathbb{Q}_p$-valued random variables are $(\mathcal{F}, B(\mathbb{Q}_p))$-measurable functions. If $\xi \in L_1(\Omega, \mathbf{P})$, we introduce an *expectation* of this random variable by setting $\mathbf{E}\xi = \int_\Omega \xi(\omega)\, \mathbf{P}(d\omega)$. We note that every bounded random variable $\xi : \Omega \to \mathbb{Q}_p$ belongs to $L_1(\Omega, \mathbf{P})$.

Let $\eta : \Omega \to G$ be a random variable. The measure $\mathbf{P}_\eta$ is said to be a *distribution* of the random variable. By Theorem 4.4 we have that

$$\mathbf{E} f(\eta) = \int_{\mathbb{Q}_p} f(y)\, \mathbf{P}_\eta(dy) \tag{5.1}$$

for every Γ-continuous function $f : G \to \mathbb{Q}_p$ such that $f \circ \eta \in L_1(\Omega, \mathbf{P})$. In particular, we have the following result.

Proposition 5.1. *Let $\eta : \Omega \to Y$ be a random variable and let $f \in C_b(Y)$. Then the formula (5.1) holds.*

We shall also use the following technical result.

Proposition 5.2. *Let $\eta : \Omega \to Y$ be a random variable and let $\zeta \in L_1(\Omega, \mathbf{P})$, and let $f \in C_b(Y)$. Then $\xi(\omega) = \zeta(\omega) f(\eta(\omega))$ belongs $L_1(\Omega, \mathbf{P})$ and*

$$\mathbf{E}\xi = \int_{\mathbb{Q}_p \times Y} x f(y)\, \mathbf{P}_z(dxdy), \quad z(\omega) = (\zeta(\omega), \eta(\omega)).$$

Proof. We have only to show that $\xi \in L_1(\Omega, \mathbf{P})$. This fact is a consequence of Theorem 4.2. □

The random variables $\xi, \eta : \Omega \to G$ are called independent if

$$\mathbf{P}(\xi \in A, \eta \in B) = \mathbf{P}(\xi \in A)\mathbf{P}(\eta \in B) \quad \text{for all } A, B \in \Gamma. \tag{5.2}$$

Proposition 5.3. *Let $\xi, \eta : \Omega \to Y$ be independent random variables and functions $f, g \in C_b(Y)$. Then we have:*

$$\mathbf{E} f(\xi) g(\eta) = \mathbf{E} f(\xi) \mathbf{E} g(\eta). \tag{5.3}$$

Proof. If f and g are locally constant functions then (5.3) is a consequence of (5.2). Arbitrary functions $f, g \in C_b(Y)$ can be approximated by locally constant functions (with the convergence of corresponding integrals) by using the technique developed in the proof of Theorem 4.4. □

Remark 5.1. In fact, the formula (5.3) is valid for the continuous f, g such that the random variables $f(\xi)$, $g(\eta)$ and $f(\xi)g(\eta)$ belong $L_1(\Omega, \mathbf{P})$.

Proposition 5.4. *Let ξ and η be independent random variables. Then the random vector $z = (\xi, \eta)$ has the probability distribution $\mathbf{P}_z = \mathbf{P}_\eta \times \mathbf{P}_\xi$.*

This fact is the direct consequence of (5.2).

Let ξ and η be respectively $\mathbb{Q}_p$ and G valued random variables and $\xi \in L_1(\Omega, \mathbf{P})$. A *conditional expectation* $\mathbf{E}[\xi | \eta = y]$ is defined as a function $m \in L_1(G, \mathbf{P}_\eta)$ such that

$$\int_{\{\omega \in \Omega : \eta(\omega) \in B\}} \xi(\omega)\, \mathbf{P}(d\omega) = \int_B m(y)\, \mathbf{P}_\eta(dy) \quad \text{for every } B \in \Gamma.$$

Proposition 5.5. *The conditional expectation is defined uniquely a.e.* mod $\mathbf{P}_\eta$.

Proof. We assume that there exist two conditional expectations $m_j \in L_1(G, \mathbf{P}_\eta)$ and $m_1(x_0) \neq m_2(x_0)$ at some point x_0 and $N_{\mathbf{P}_\eta}(x_0) > 0$. Set $m(x) = m_1(x) - m_2(x)$. We have: $\int_B m(x)\, \mathbf{P}_\eta(dx) = 0$ for every $B \in \Gamma$. To obtain the contradiction, it is sufficient to use Theorem 4.3. □

As there is no analogue of the Radon–Nikodym theorem in the non-Archimedean case [164], it may happens that a conditional expectation does not exist. Everywhere below we assume that $m(y) = \mathbf{E}[\xi|\eta = y]$ is well defined and moreover, that it belongs to the class $C_b(Y)$.

Proposition 5.6. *Let* $\xi : \Omega \to \mathbb{Q}_p$, $\eta : \Omega \to Y$ *be random variables, and* $\xi \in L_1(\Omega, \mathbf{P})$. *The equality*

$$\mathbf{E} f(\eta)\xi = \mathbf{E} f(\eta(\omega))\mathbf{E}[\xi(\omega)|\eta = \eta(\omega)]$$

holds for every function $f \in C_b(Y)$.

Proof. By Proposition 5.2 we obtain $\mathbf{E}\xi f(\eta) = \int_{\mathbb{Q}_p \times Y} x f(y)\, \mathbf{P}_z(dxdy)$, where $z(\omega) = (\xi(\omega), \eta(\omega))$. Set for $A \in B(Y)$,

$$\lambda(A) = \int_{\mathbb{Q}_p \times Y} x I_A(y)\, \mathbf{P}_z(dxdy).$$

As $\lambda(A) = \int_{\eta^{-1}(A)} \xi(\omega)\, \mathbf{P}(d\omega) = \int_Y m(y)\, \mathbf{P}_\eta(dy)$, it is a tight measure on Y. Then

$$\int_{\mathbb{Q}_p \times Y} x f(y)\, \mathbf{P}_z(dxdy) = \int_Y f(y)\, \lambda(dy) = \int_Y f(y) m(y)\, \mathbf{P}_\eta(dy)$$
$$= E f(\eta) m(\eta).$$

□

5 Tests for randomness for p-adic probability theory

The first p-adic probability models [79], [80] were attempts to extend R. von Mises frequency probability theory to the p-adic case (see Chapter 4, Section 2). As relative frequencies $\nu_N = \frac{n}{N} \in \mathbb{Q}$, we can study their behavior not only in $\mathbb{R}$, but also in $\mathbb{Q}_p$. It is well know that von Mises' theory is based on two principles: (1) the principle of the statistical stabilization of relative frequencies and (2) the principle of randomness.

As we have seen, the first principle can be naturally generalized to the p-adic case and p-adic probabilities are defined as limits of relative frequencies with respect to p-adic topology. However, as in the ordinary real probability theory, there is the large problem with the principle of randomness. In the p-adic case the situation with stability of limits of relative frequencies with respect to place selections is even worse than in the real case, because the p-adic metric is very unstable: if $|n|_p = \varepsilon < 1$, then $|n+1|_p = 1$. In the p-adic case we have not even the possibility to restrict our considerations to a countable number of place selections (as we can do in the real case by Tornier theorem). To obtain the reasonable definition of p-adic randomness, we tried also to apply the theory of algorithmic complexity (see, for example, [38], [135], [134], [160], [137]). However, there was no large progress, see [80], [91], [89]. We present now a p-adic generalization of Martin-Löf's theory [142–144] based on tests for randomness[1]. Such a generalization looks as the most natural approach to p-adic randomness. Here we find natural tests for randomness for p-adic valued uniform probability distribution. Each test for randomness induces a series of limit theorems. On the other hand, individual limit theorems are not good candidates for test for randomness, because each theorem describes behavior of a subsequence $S_{n_k}(\omega)$ of the sequence $S_n(\omega) = \xi_1(\omega) + \cdots + \xi_n(\omega)$ of independent equally distributed random variables.

We proved that it is possible to enumerate effectively all p-adic test for randomness. However, in the opposite to Martin-Löf's theorem for real probabilities we proved that a universal p-adic test for randomness does not exist.

We shall use the standard terminology of the book of M. Li and P. Vitànyi [137]. The abbreviation r.e. is used for "recursive enumeration."

1 p-adic probability measures on the space of binary sequences

We set $X = \{0, 1\}$, $X^n = \{x = (x_1, \dots, x_n) : x_j \in X\}$, $X^* = \bigcup_n X^n$, $X^\infty = \{\omega = (\omega_1, \dots, \omega_n, \dots) : \omega_j \in X\}$. For $x \in X^n$, we set $l(x) = n$. For $x \in X^*$, $l(x) = n$, we

[1] This theory was developed by A. Khrennikov and S. Yamada [130].

define a cylinder U_x with basis x by $U_x = \{\omega \in X^\infty : \omega_1 = x_1, \ldots, \omega_n = x_n\}$. We denote by the symbol $\mathcal{F}_{\text{cyl}}$ an algebra of subsets of X^∞ generated by all cylinders.

The map $j : X^\infty \to \mathbb{Z}_2$, $j(\omega) = \sum_{j=0}^{\infty} \omega_j 2^j$, gives one to one correspondence between X^∞ and $\mathbb{Z}_2$. Thus we can identify these sets. The algebra of cylindric sets $\mathcal{F}_{\text{cyl}}$ coincides with the algebra $B(\mathbb{Z}_2)$ of all clopen subsets of $\mathbb{Z}_2$ (see Chapter 4).

A function $\mu : \mathcal{F}_{\text{cyl}} \to \mathbb{Q}_p$ is a p-adic (valued) measure if the following properties holds true: (i) additivity: $\mu(A \cup B) = \mu(A) + \mu(B)$, $A \cap B = \emptyset$, $A, B \in \mathcal{F}_{\text{cyl}}$; (ii) boundedness: $\|\mu\|_p = \sup\{|\mu(A)|_p : A \in \mathcal{F}_{\text{cyl}}\} < \infty$. As it has been noticed in Chapter 4, the condition of continuity (iii) is redundant as $\mathcal{F}_{\text{cyl}} = B(\mathbb{Z}_2)$.

A function $f : X^* \to \mathbb{Q}_p$ is said to be *recursive* iff there is a recursive function $g : X^* \times \mathbb{N} \to \mathbb{Q}$ such that $|f(x) - g(x,k)|_p < \frac{1}{k}$. A p-adic measure $\mu : \mathcal{F}_{\text{cyl}} \to \mathbb{Q}_p$ is said to be recursive iff the function $f_p : X^* \to \mathbb{Q}_p$, $f_p(x) = \mu(U_x)$, is recursive.

The uniform p-adic measure μ_p $(p \neq 2)$ on X^∞ is defined by

$$\mu_p(U_x) = \frac{1}{2^{l(x)}}, \quad x \in X^*. \tag{1.1}$$

If X^* is realized as $\mathbb{Z}_2$ and $\mathcal{F}_{\text{cyl}}$ as $B(\mathbb{Z}_2)$, then μ_p is the p-adic valued Haar measure (translation invariant measure) on $\mathbb{Z}_2$.

As $\left|\frac{1}{2^{l(x)}}\right|_2 = 2^{l(x)}$, the additive set function μ_2 defined by (1.1) is not bounded. Therefore **we shall consider only the case $p \neq 2$.**

The simple considerations show that the function

$$N_{\mu_p}(x) = \inf\{\|U\|_\mu : x \in U \in B(\mathbb{Z}_2)\} = 1$$

for all $x \in \mathbb{Z}_2$. This implies that $L_1(\mathbb{Z}_2, \mu_p) = C(\mathbb{Z}_2)$ (because all $B(\mathbb{Z}_2)$-step functions are continuous and each continuous function can be uniformly approximated by a sequence of $B(\mathbb{Z}_2)$-step functions). This implies that the algebra $(B(\mathbb{Z}_2))_{\mu_p} = \{A \subset \mathbb{Z}_2 : I_A \in L_1(\mathbb{Z}_2, \mu_p)\} = B(\mathbb{Z}_2)$. Thus the Haar measure μ_p cannot be extended from the algebra $B(\mathbb{Z}_2)$ to any larger algebra. In particular, the μ_p cannot be extended on the Borel σ-algebra generated by the algebra of clopen subsets $B(\mathbb{Z}_2)$.

Let a measure $\mu : \mathcal{F}_{\text{cyl}} \to \mathbb{Q}_p$ be normalized: $\mu(X^\infty) = 1$. Then we can consider the p-adic probability space $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$, where $\Omega = X^\infty$, $\mathcal{F} = (\mathcal{F}_{\text{cyl}})_\mu$(the set algebra which is obtained via the μ-extension of $\mathcal{F}_{\text{cyl}}$), $\mathbf{P} = \bar{\mu}$ is a p-adic probability measure. As for the p-adic uniform measure μ_p (the $\mathbb{Q}_p$-valued Haar measure on $\mathbb{Z}_2$) the extension $(\mathcal{F}_{\text{cyl}})_{\mu_p}$ coincides with $\mathcal{F}_{\text{cyl}}$ and the extension $\bar{\mu}_p$ coincides with μ_p, the corresponding probability space is $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P}_p)$, where $\Omega = X^\infty$, $\mathcal{F} = \mathcal{F}_{\text{cyl}}$ and $\mathbf{P}_p = \mu_p$. The $\mathbf{P}_p$ is called a *uniform p-adic probability distribution.* We remark that values of $\mathbf{P}_p$ on cylinders coincide with values of the standard (real-valued) uniform probability distribution $\mathbf{P}_\infty$ on X^∞. As $\mathbb{Q} \subset \mathbb{R}$ and $\mathbb{Q} \subset \mathbb{Q}_p$, we can interpret rational numbers $\frac{1}{2^{l(x)}}$ both as real and as p-adic numbers.

In fact, we shall not use general recursive p-adic probabilities (see only definitions). We shall consider only the uniform p-adic probability distribution $\mathbf{P}_p$, $p \neq 2$ (which is, of course, recursive).

2 Some technical p-adic results

The results which are obtained in this section will be used to construct p-adic tests and prove limit theorems for p-adic probabilities.

For any $n, k \in \mathbb{N}$, (n, k) denotes the greatest common divisor of n and k; for any $n \in \mathbb{N}$, $M_p(n)$ denotes the mod p residue of n: $n = M_p(n) \mod p$. We set

$$\Theta_p(n) = \begin{cases} |n - M_p(n)|_p, & n \geq p, \\ 1, & 1 \leq n \leq p-1. \end{cases}$$

Lemma 2.1. *Let $n, k \in \mathbb{N}$, $k \leq n$ and let $M_p(n) \geq M_p(k)$. Then*

$$\left|\binom{n}{k}\right|_p = \frac{\Theta_p(n)}{\Theta_p(k)}.$$

Proof. Let $n = \alpha + ip^N$, $k = \beta + jp^l$, where $0 \leq \alpha, \beta \leq p-1$, $i, j, N, l \in \mathbb{N}$ and $(i, p) = (j, p) = 1$. We have:

$$\begin{aligned} \left|\binom{n}{k}\right|_p &= \left|(ip^N) \cdot \frac{(ip^N - p)}{p} \cdot \frac{(ip^N - 2p)}{2p} \cdots \frac{(ip^N - jp^l + p)}{(jp^l - p)} \cdot \frac{1}{(jp^l)}\right|_p \\ &= \left|\frac{p^N}{p^l}\right|_p = p^{l-N}. \end{aligned} \tag{2.1}$$

To obtain (2.1), we have used that $n - k + 1 = ip^N - jp^l + (\alpha + 1 - \beta)$ and $0 < \alpha + 1 - \beta \leq p$; hence the last term in the nominator of $\binom{n}{k} = \frac{n \cdots (n-k+1)}{1 \cdots k}$, which is divisible by p is $(ip^N - jp^l + p)$. The cases in that $n = \alpha$ or $k = \beta$, $0 \leq \alpha$, $\beta \leq p - 1$ are considered in the same way. □

Lemma 2.2. *Let $n, k \in \mathbb{N}$, $k \leq n$, and let $M_p(n) + 1 \leq M_p(k)$. Then*

$$\left|\binom{n}{k}\right|_p = \Theta_p(n).$$

Proof. Let $n = \alpha + ip^N$, $k = \beta + jp^l$, where $(i, p) = (j, p) = 1$, $0 \leq \alpha, \beta \leq p-1$. We have

$$\begin{aligned} \left|\binom{n}{k}\right|_p &= \left|(ip^N) \cdot \frac{(ip^N - p)}{p} \cdot \frac{(ip^N - 2p)}{2p} \cdots \frac{(ip^N - jp^l)}{(jp^l)}\right|_p \\ &= \left|p^N\right|_p = p^{-N}. \end{aligned} \tag{2.2}$$

To obtain (2.2), we have used that $n - k + 1 = (ip^N - jp^l) - (\beta - \alpha - 1)$ and $0 \leq \beta - \alpha - 1 < p$; hence the last term in the nominator of $\binom{n}{k} = \frac{n \cdots (n-k+1)}{1 \cdots k}$ which is divisible by p is $(ip^N - jp^l)$. The cases in that $n = \alpha$ or $k = \beta$, $0 \leq \alpha, \beta \leq p-1$, are considered in the same way. □

3 p-adic tests for randomness

We use the following notations. For each set $M \subset X^*$, we set $M^{(n)} = \{x \in M : l(x) = n\}, n = 1, 2, \ldots$. For each set $W \subset X^* \times \mathbb{N}$, we set $W_m = \{x \in X^* : (x, m) \in W\}$. Thus $W_m^{(n)} = \{x \in X^* : l(x) = n, (x, m) \in W\}$.

Everywhere in this chapter the *cardinality of a (finite) set A is denoted by the symbol $\sigma(A)$*. We do not use the standard symbol $|A|$, because we do not want to use expressions of the form $|\,|A|\,|_p$.

The following definition of a p-adic test for randomness is a natural generalization of Martin-Löf's definition of a test for randomness for ordinary real probabilities (in fact, in our particular case for the uniform distribution).

Definition 3.1. Let $\mathbf{P}$ be a p-adic recursive probability. A recursively enumerable (r.e.) set $V \subset X^* \times \mathbb{N}$ is called a *p-adic* $\mathbf{P}$*-test* (p-adic test for randomness for the probability distribution $\mathbf{P}$) if it possesses the following two properties: for all $n, m \in \mathbb{N}$, we have:

$$V_{m+1} \subset V_m,$$
$$\Big| \sum_{x \in V_m^{(n)}} \mathbf{P}(U_x) \Big|_p \leq \frac{1}{p^m}. \tag{3.1}$$

The use of p-adic tests for randomness gives the possibility to formalize (in fact, to create) p-adic statistics. We are given the sample space X^* with an associated p-adic probability distribution $\mathbf{P}$. Given an element x of the sample space, we want to test hypothesis "x is a typical outcome". Practically speaking, the property of being typical is the property of belonging to reasonable majority. To ascertain whether a given element of the sample space belongs to a particular reasonable majority we use the notation of a test. As in the ordinary probability theory, a test is given by a prescription that, for every level of significance $\varepsilon = \frac{1}{p^m}$, tells us for what elements $x \in X^*$ the hypothesis "x belongs to majority M in X^*" should rejected where $\varepsilon = 1 - \mathbf{P}(M)$. The set V_m is a *critical region* on the *significance level* $\varepsilon = \frac{1}{p^m}$. If $x \in V_m$ then the hypothesis "x belongs to majority M" is rejected with the significance level ε. We say that x fails the test at the level of critical region V_m. Of course, there is a large difference between 'p-adic majority' and the ordinary 'real majority'. Populations which are very large from the point of view of ordinary real probability may be very small from the point of view of p-adic probability and vice versa.

We shall study *only the uniform p-adic probability distribution.* Everywhere below $\mathbf{P} = \mathbf{P}_p$, $p \neq 2$. Tests for randomness for this probability distribution we shall simply call p-adic test. Here condition (3.1) can be reformulated in the following way:

$$\Big|\sigma(V_m^{(n)})\Big|_p \leq \frac{1}{p^m} \tag{3.2}$$

(as $\mathbf{P}(U_x) = \frac{1}{2^n}$ for $x \in V_m^{(n)}$ and $|2^n|_p = 1$ for $p \neq 2$, (3.1) has the form $\big|\sum_{x \in V_m^{(n)}} 1\big|_p \leq \frac{1}{p^m}$).

Proposition 3.1. *Let V be a p-adic test. Then, for each $(x,m) \in V$, we have*

$$l(x)(\log_p 2) > m \geq 1. \tag{3.3}$$

Proof. Set $n = l(x)$. As $x \in V_m$, we have $V_m^{(n)} \neq \varnothing$ and by (3.2) $\sigma(V_m^{(n)})$ is divisible by p^m. Thus $2^n = \sigma(X^n) \geq \sigma(V_m^{(n)}) \geq p^m$. This implies inequality (3.3). □

Proposition 3.2. *Let V be a p-adic test. Then, for each $k \geq m$, $n \in \mathbb{N}$,*

$$\left|\sigma(V_m^{(n)} \setminus V_k^{(n)})\right|_p \leq \frac{1}{p^m}.$$

Proof. As $V_k^{(n)} \subset V_m^{(n)}$, we have:

$$\sigma(V_m^{(n)}) = \sigma(V_k^{(n)}) + \sigma(V_m^{(n)} \setminus V_k^{(n)}).$$

By the strong triangle inequality we get:

$$|\sigma(V_m^{(n)} \setminus V_k^{(n)})|_p \leq \max(|\sigma(V_m^{(n)})|_p |\sigma(V_k^{(n)})|_p) = \frac{1}{p^m}.$$ □

As usual, we denote the integer part of a real number x by $[x]$. Condition (3.3) can be rewritten in the form

$$[l(x)\log_p 2] \geq m.$$

The function $\lambda(n) = [n \log_p 2]$, $n \in \mathbb{N}$, will play the important role in our further considerations. For any p-adic test V and $n \in \mathbb{N}$, only sets $V_m^{(n)}$, $m = 1, \dots, \lambda(n)$, can be nonempty.

We give now a few examples of p-adic tests for randomness. All these tests are related to behavior of sums:

$$S(x) = x_1 + \cdots + x_n, \quad x \in X^*, \quad n = l(x).$$

Example 3.1. We set

$$\begin{aligned} V_m = \{x \in X^* : \Theta_p(S(x)) \geq p^m \Theta_p(l(x)), S(x) \neq 0 \text{ and} \\ M_p(S(x)) \leq M_p(l(x))\}. \end{aligned} \tag{3.4}$$

To show that the set $V = \{(x,m) : x \in V_m\}$ is a p-adic test, we need only to show that (3.2) holds true. We have:

$$\sigma(V_m^{(n)}) = \sum_k \binom{n}{k},$$

where $0 \leq k \leq n$ and $M_p(k) \leq M_p(n)$, $\frac{\Theta_p(n)}{\Theta_p(k)} \leq \frac{1}{p^m}$. To obtain (3.2), it is sufficient to use the strong triangle inequality and Lemma 2.1.

Example 3.2. We set

$$\overline{V}_m = \left\{ x \in X^* : \Theta_p(l(x)) \leq \frac{1}{p^m} \text{ and } M_p(S(x)) \geq M_p(l(x)) + 1 \right\}. \tag{3.5}$$

By using Lemma 2.2 we obtain that (3.2) holds true for $\overline{V}_m$. Thus the set $\overline{V} = \{(x,m) : x \in \overline{V}_m\}$ is a p-adic test.

Example 3.3 (Finite tests). Let $n \in \mathbb{N}$ be a fixed number. Let T be some subset of X^n, $\sigma(T) = p^{-\lambda(n)}$. We set $W_m^{(n)} = T$ for $m = 1, \ldots, \lambda(n)$ and $V_j^{(n)} = \varnothing$, $j > \lambda(n)$, and $V_j^{(n)} = \varnothing$, $k \neq n$, for all $j = 1, 2, \ldots$. Then $V = \{(x,m) : x \in V_m\}$, $V_m = \bigcup_{k=1}^{\infty} V_m^{(k)}$ is a finite p-adic test.

To illustrate the statistical meaning of tests (3.4) and (3.5), it is useful to consider some subsets of them corresponding to fixed values of $M_p(n)$ and $M_p(S(x))$.

We start with test (3.4). We set

$$\begin{aligned} V_m(1,0) &= \{x \in V_m : M_p(l(x)) = 1 \text{ and } M_p(S(x)) = 0\} \quad \text{and} \\ V(1,0) &= \{(x,m) : x \in V_m(1,0)\}. \end{aligned} \tag{3.6}$$

This test is connected with samples of the form

$$x = (x_1, \ldots, x_{1+jp^N}), \quad j, N \in \mathbb{N}, \quad (j,p) = 1. \tag{3.7}$$

Such a sample must be rejected with the level of significance $\varepsilon = \frac{1}{p^m}$ if $1 > |S(x)|_p \geq p^m |l(x) - 1|_p = p^{m-N}$. Thus the test $V(1,0)$ rejects all samples of the form $x = (x_1, \ldots, x_{1+jp^N})$, $(j,p) = 1$, in that the sum $S(x) = x_1 + \cdots + x_{1+jp^N}$ is not divisible by a sufficiently high degree of p (but divisible by p^1).

A sample x of form (3.7) with $S(x) = ip^k$, $(i,p) = 1, k \geq 1$, is rejected with the level of significance $\epsilon = 1/p^m$ if $k < N - m$.

For test (3.4) and $M_p(l(x)) = 1$, we can also fix $M_p(S(x)) = 1$ and obtain a new test:

$$\begin{aligned} V_m(1,1) &= \{x \in V_m : M_p(l(x)) = 1 \text{ and } M_p(S(x)) = 1\} \quad \text{and} \\ V(1,1) &= \{(x,m) : x \in V_m(1,1)\}. \end{aligned}$$

A sample x of the form (3.7) must be rejected with the level of significance $\varepsilon = \frac{1}{p^m}$ if

$$1 > |S(x) - 1|_p \geq p^m |l(x) - 1|_p = p^{m-N}.$$

Thus the test $V(1,1)$ rejects all samples x of the form (3.7) for that $S(x) - 1$ is not divisible by a sufficiently high degree of p (but divisible by p^1).

In the same way by fixing $M_p(n) = s \in \{0, \ldots, p-1\}$ we obtain tests $V_m(s,q)$, $q = 0, \ldots, s$. The $V(s,q)$ rejects some samples of the form

$$x = (x_1, \ldots, x_{s+jp^N}), \quad j, N \in \mathbb{N}, \quad (j,p) = 1, \tag{3.8}$$

namely, samples for which $S(x) - q$ is not divisible by a sufficiently high degree of p (but divisible by p^1). A sample x of form (3.8) with $S(x) = q + ip^k$, $(i, p) = 1$, $k \geq 1$, is rejected with the level of significance $\epsilon = 1/p^m$ if $k < N - m$.

We study now test (3.5). The condition $M_p(S(x)) \geq M_p(l(x)) + 1 \geq 0$ implies that this test is used to reject (with some level of significance) some samples for that the sum $S(x)$ is not divisible by p (compare with (3.6)). We set

$$\overline{V}_m(0,1) = \{x \in \overline{V}_m : M_p(l(x)) = 0 \text{ and } M_p(S(x)) = 1\} \quad \text{and}$$
$$\overline{V}(0,1) = \{(x,m) : x \in \overline{V}_m(0,1)\}.$$

By this test we reject with the level of significance $\varepsilon = \frac{1}{p^m}$ all samples of the form $x = (x_1, \dots, x_{jp^N})$, $(j, p) = 1$, for that $N < m$ and $M_p(S(x)) = 1$. We can compare the test $\overline{V}(0,1)$ with the test $V(0,0)$. The latter test is used to reject samples of the same form, but with $S(x)$ divisible by p: $S(x) = ip^k$, $(i, p) = 1$, $k \geq 1$. A sample is rejected with the level of significance $\varepsilon = \frac{1}{p^m}$ if $k < N - m$.

It is possible to introduce a p-adic test O which covers all cases of divisibility by p of $S(x)$. We start with the following simple fact:

Proposition 3.3. *Let Φ and Ψ be two p-adic tests such that $\Phi \cap \Psi = \varnothing$. Then the set $\Gamma = \Phi \cup \Psi$ is a p-adic test with critical regions $\Gamma_m = \Phi_m \cup \Psi_m$ on the significance level $\varepsilon = \frac{1}{p^m}$.*

Proof. We need only to prove that (3.2) holds true: We have $|\sigma(\Gamma_m^{(n)})|_p = |\sigma(\Phi_m^{(n)}) + \sigma(\Psi_m^{(n)})|_p \leq \max(|\sigma(\Phi_m^{(n)})|_p, |\sigma(\Psi_m^{(n)})|_p) \leq \frac{1}{p^m}$. □

We now turn back to tests V and $\overline{V}$ defined in Examples 3.1, 3.2. It is evident that $V_m \cap \overline{V}_m = \varnothing$ for all m. Thus sets $\Sigma_m = V_m \cup \overline{V}_m$ give critical regions (with $\varepsilon = \frac{1}{p^m}$) of a p-adic test $\Sigma = \{(x,m) : x \in \Sigma_m\}$.

4 Some limit theorems

As in ordinary real probability theory tests V and $\overline{V}$ of Examples 3.1, 3.2 are related to some limit theorems for p-adic probability. Let $\mathcal{P} = (\Omega, \mathcal{F}_{\text{cyl}}, \mathbf{P})$ be the probability space based on the uniform p-adic distribution $\mathbf{P}$ on algebra $\mathcal{F}_{\text{cyl}}$ of cylindric subsets of $\Omega = X^\infty$, $p \neq 2$. For $\omega \in \Omega$, we set $S_n(\omega) = \omega_1 + \dots + \omega_n$.

Theorem 4.1. *For each $l \in \mathbb{N}$ the probability*

$$\mathbf{P}\left(\left\{\omega \in \Omega : |S_n(\omega) - M_p(S_n(\omega))|_p = \frac{1}{p^l}, M_p(S_n(\omega)) \leq M_p(n)\right\}\right) \to 0$$

in $\mathbb{Q}_p$, when $|n - M_p(n)|_p \to 0$, $n \neq M_p(n)$.

Proof. By using considerations of Example 3.1 we obtain that

$$\mathbf{P}\left(\left\{\omega \in \Omega : |S_n(\omega) - M_p(S_n(\omega))|_p = \frac{1}{p^l}, M_p(S_n(\omega)) \le M_p(n)\right\}\right)$$
$$\le p^l |n - M_p(n)|_p. \qquad \square$$

In particular, we obtain the following limit theorems:

Corollary 4.1. *For each $l \in \mathbb{N}$, the probability*

$$\mathbf{P}\left(\left\{\omega \in \Omega : S_n(\omega) \in \mathcal{S}_{\frac{1}{p^l}}(0)\right\}\right) \to 0$$

in $\mathbb{Q}_p$,when $|n|_p \to 0$.

Corollary 4.2 (see [88]). *For each $l \in \mathbb{N}$, the probabilities*

$$\mathbf{P}\left(\left\{\omega \in \Omega : S_n(\omega) \in \mathcal{S}_{\frac{1}{p^l}}(0)\right\}\right) \quad \textit{and} \quad \mathbf{P}\left(\left\{\omega \in \Omega : S_n(\omega) \in \mathcal{S}_{\frac{1}{p^l}}(1)\right\}\right)$$

tend to zero in $\mathbb{Q}_p$, when $|n-1|_p$ tends to zero.

Formally we can interpret Corollary 4.2 in the following way. The sum $S_n(\omega)$ can be considered as the sum $S_n(\omega) = \xi_1(\omega) + \cdots + \xi_n(\omega)$ of independent equally distributed random variables $\xi_j(\omega) = 0, 1$ with probabilities $1/2$. By Corollary 4.2 the probability distribution of random variable $S_{\lim}(\omega) = \lim_{n \to 1} S_n(\omega)$ is concentrated at the points $a_0 = 0$ and a_1 of $\mathbb{Q}_p$. By symmetry reasons $\mathbf{P}_{S_{\lim}}(\{a_0\}) = \mathbf{P}_{S_{\lim}}(\{a_1\}) = 1/2$. Of course, this is just a formal statement, because Corollary 4.2 gives convergence only for spheres of $\mathbb{Q}_p$.

Theorem 4.2. *The probability*

$$\mathbf{P}(\{\omega \in \Omega : M_p(S_n(\omega)) \ge M_p(n) + 1\}) \to 0$$

when $|n - M_p(n)|_p \to 0$.

As in the case of Theorem 4.1, we can, for example, put $M_p(n) = 0$ or $M_p(n) = 1$ and obtain the following consequences of Theorem 4.2:

Corollary 4.3. *The probability*

$$\mathbf{P}(\{\omega \in \Omega : M_p(S_n(\omega)) \ge 1\}) \to 0$$

in $\mathbb{Q}_p$, when $|n|_p \to 0$.

Corollary 4.4. *The probability*

$$\mathbf{P}(\{\omega \in \Omega : M_p(S_n(\omega)) \ge 2\}) \to 0$$

when $|n - 1|_p \to 0$.

We note that

$$\mathbf{P}(\{\omega \in \Omega : M_p(S_n(\omega)) \geq 1\}) = \mathbf{P}(\{\omega \in \Omega : S_n(\omega) \in \mathcal{S}_1(0)\}).$$

Thus by Corollaries 5.1 and 5.3 we obtain that

$$\mathbf{P}(\{\omega \in \Omega : S_n(\omega) \in \mathcal{U}_{\frac{1}{p^m}}(0)\}) \to 1,$$

$|n|_p \to 0$, for any $m \in \mathbb{N}$. Hence formally we obtain that the probability distribution $\mathbf{P}_{S_{\lim}}$ of $S_{\lim}(\omega) = \lim_{n\to 0} S_n(\omega)$ is concentrated at the point $a_0 = 0 \in \mathbb{Q}_p$, $\mathbf{P}_{S_{\lim}}(\{0\}) = 1$.

It seems that in the p-adic case it is more natural to use tests for randomness than limit theorems. In the opposite to ordinary real probability theory in the p-adic case we have no general limit theorems for $n \to \infty$ (in the sense of the order on $\mathbb{N}$). All limit theorems give the convergence of probabilities for some sequences $n_k \to \infty$, $k \to \infty$. For example, $|n_k|_p \to 0$, $n_k \neq 0$, implies that $n_k = jp^N$, $(j, p) = 1$, $N \to \infty$, and $|n_k - 1|_p \to 0$, $n_k \neq 1$, implies that $n_k = 1 + jp^N$, $(j, p) = 1$, $N \to \infty$, and so on.

5 Recursive enumeration of the set of p-adic tests

Here we shall prove that the set of all p-adic tests is recursively enumerable. The general scheme of the proof is the same as in the case of real probabilities. However, the main part of the proof (an algorithm for constructing a p-adic test on the basis of a partial recursive function) strongly differs from the standard one (see [137]).

We start with the following well-known lemma (see, for example, [137]).

Lemma 5.1. *There exists a partial recursive function $f : \mathbb{N} \times \mathbb{N} \to X^* \times \mathbb{N}$ with the following properties:*

(a1) for all $i, j \in \mathbb{N}$ such that $f(i, j) \neq \infty$, we have $f(i, k) \neq \infty$, for all $k \leq j$;

(a2) a set $A \subset X^* \times \mathbb{N}$ is r.e. iff $A = \{f(i, j) : j = 1, 2, \ldots\} \setminus \{\infty\}$, for some $i \geq 1$.

Theorem 5.1. *The set of all p-adic tests is r.e.*

Proof. Through the proof we shall use the fixed partial recursive function $\varphi = \varphi_i = f(i, \cdot)$ given by Lemma 5.1. We set $A_\varphi = \varphi(\mathbb{N})$. As in the standard case, we shall construct for each φ some total recursive function $g : \mathbb{N} \to X^* \times \mathbb{N}$ such that $T = A_g = g(\mathbb{N})$ is a p-adic test and if φ is a p-adic test by itself, then $T = A_\varphi$. We construct T step by step using an algorithm which produces a p-adic test at each step. In the following algorithm we shall use sets $\mathcal{D}_m^{(n)}$ which give approximations for sets $T_m^{(n)}$ in the process of building of T (as usual $T_m = \{x \in X^* : (x, m) \in T\}$ and $T_m^{(n)} = \{x \in T_m : l(x) = n\}$). We shall also use sets $R_m^{(n)}$ which are registers for collecting elements of $(A_\varphi)_m^{(n)}$. The main difference with the standard algorithm is due to the fact that we cannot increase sets $\mathcal{D}_m^{(n)}$ at each step when φ produces a value $\varphi(j) \in (A_\varphi)_m^{(n)} = \{x : \phi(j) = (x, m)$ for some j and $l(x) = n\}$ (because the p-adic metric is changed discontinuously: $|x|_p \leq \frac{1}{p^m} \Rightarrow |x+1|_p = 1, m \geq 1$). We collect (in $R_m^{(n)}$) elements of $(A_\varphi)_m^{(n)}$ until $\sigma(R_m^{(n)})$ becomes divisible by p^m. After this we set $\mathcal{D}_m^{(n)} = R_m^{(n)}$.

To be sure that the result of our construction will be a r.e. set, we construct parallel a function $g: \mathbb{N} \to X^* \times \mathbb{N}$ such that $T = g(\mathbb{N})$ and g is a total recursive function if T is an infinite set.

Algorithm

1 **Put** $T = \varnothing$, $\mathcal{D}_m^{(n)} = R_m^{(n)} = \varnothing$; put $j = 0$, $i = 0$, $t_m^{(n)} = 0$.

% j is the argument of φ, i is the argument of g; $t_m^{(n)} = \sigma(R_m^{(n)})$.

2 **Put** $j = j + 1$

3 If $\varphi(j) = \infty$ continual indefinitely.

4 Find $\varphi(j) = (x, m)$ and $n = l(x)$.

5 If $m > [n \log_p 2]$, then $T = \varnothing$ and stop.

6 **Put** $R_m^{(n)} = R_m^{(n)} \cup \{x\}$ and $t_m^{(n)} = t_m^{(n)} + 1$.

7 If $|t_m^{(n)}|_p > \frac{1}{p^m}$, then go to step 2.

8 If $m \geq 2$ and $\mathcal{D}_{m-1}^{(n)} \not\supset R_m^{(n)}$, then go to step 2.

9 **Put** $\mathcal{D}_m^{(n)} = R_m^{(n)}$.

% We must make step 8 before step 9 to get $T_{m-1} \supset T_m$.

10 **(a)** Enumerate elements if $\mathcal{D}_m^{(n)} = \{z_1, \ldots, z_{t_m^{(n)}}\}$;

(b) for $l = 1, \ldots, t_m^{(n)}$, put $g(i + l) = (z_l, m)$;

(c) put $i = i + t_m^{(n)}$.

% The previous step is not related to the construction if T; here we construct the function g which gives recursive enumeration for T.

11 **Put** $s = m$.

12 **Put** $s = s + 1$.

13 If $s > [n \log_p 2]$, go to 18.

14 If $|t_s^{(n)}|_p > \frac{1}{p^s}$, go to 18.

15 If $\mathcal{D}_{s-1}^{(n)} \not\supset R_s^{(n)}$, go to 18.

16 **Put** $\mathcal{D}_s^{(n)} = R_s^{(n)}$.

% We explain the meaning of steps 11–16. By step 9 the set $\mathcal{D}_m^{(n)}$ has been increased. Thus condition 8 must be reconsidered for sets $\mathcal{D}_s^{(n)}$ with $s > m$. It can be that occasionally some of sets $R_s^{(n)}$ has the number of elements which is divisible by p^s. If they pass step 15, then we increase sets $\mathcal{D}_j^{(n)}$ by 16.

17 Repeat step 10 for $m = s$.

18 **Put** $T = T \bigcup_{m \le s \le [n \log_p 2]} \mathcal{D}_s^{(n)} \times \{s\}$ and go to step 2.

We prove now that the set T which is constructed by the algorithm is a p-adic test.

(A1) We use the parameter j to denote the step (determined by 2) of the algorithm. We have $T_m^{(n)} = \bigcup_{j=1}^{\infty} \mathcal{D}_m^{(n)}(j)$. As $\mathcal{D}_m^{(n)}(j+1) \supset \mathcal{D}_m^{(n)}(j)$ and $\sigma(\mathcal{D}_m^{(n)}(j))$ is divisible by p^m, we get that $\sigma(T_m^{(n)})$ is divisible by p^m. Thus $|\sigma(T_m^{(n)})|_p \le \frac{1}{p^m}$.

(A2) By step 8 and 15 we get that $\mathcal{D}_m^{(n)} \supset \mathcal{D}_{m+1}^{(n)}$, $n, m \in \mathbb{N}$. Thus $T_m^{(n)} \supset T_{m+1}^{(n)}$, $n, m \in \mathbb{N}$.

(A3) If steps 10 and 16 are passed an infinite number of times, then g is the total recursive function and, hence, $T = A_\varphi$ is r.e. If the steps 10 and 16 are passed only a finite number of times, then the set T is finite and, hence, r.e.

We prove now that if $V = A_\varphi$ is a p-adic test, then $T = A_\varphi$.

It is evident that $T \subset A_\varphi$. We have only to prove that $V \subset T$. It is sufficient to prove that, for each n, $V_m^{(n)} \times \{m\} \subset T_m^{(n)} \times \{m\}$ for all $m \le [n \log_p 2]$.

For each n, the set $V^{(n)} = \{(x, m) \in V : l(x) = n\}$ is finite (since $m \le [n \log_p 2]$). Thus φ produces all elements of $V^{(n)}$ after a finite numbers of steps $J = J(n, \varphi)$.[2] Let $\varphi(J) = (x_J, m_J)$ (here $l(x_J) = n$ and $m_J \le [n \log_p 2]$). We have: $\mathcal{D}_1^{(n)} \supset \mathcal{D}_2^{(n)} \supset \cdots \supset \mathcal{D}_M^{(n)}$ and $|\sigma(\mathcal{D}_s^{(n)})|_p \le \frac{1}{p^s}$ for $s = 1, \ldots, M = [n \log_p 2]$. We also have: $R_s^{(n)} = V_s^{(n)}$ (because $V^{(n)} \subset \varphi(\{1, 2, \ldots, J\})$ and $\varphi(\{1, 2, \ldots, J\})_s^{(n)} = R_s^{(n)}$). Thus, for all s, $|\sigma(R_s^{(n)})|_p \le \frac{1}{p^s}$. In particular, this holds for $s = m_J$. Hence, for $m = m_J$, step 7 is passed.

We prove that $\mathcal{D}_s^{(n)} = R_s^{(n)} = V_s^{(n)}$ for all $s = 1, \ldots, m_J - 1$. As $|\sigma(R_1^{(n)})|_p \le \frac{1}{p}$, $R_1^{(n)}$ has passed step 7. But step 8 is trivial for $m = 1$. Thus by step 9 we get $\mathcal{D}_1^{(n)} = R_1^{(n)} = V_1^{(n)}$. For $s = 2$, we have $|\sigma(R_2^{(n)})|_p \le \frac{1}{p^2}$ and step 7 is passed. As

[2]Of course, some points $(x, m) \in V^{(n)}$ can appear again on some steps $J' > J$.

$\mathcal{D}_1^{(n)} = V_1^{(n)}$ and $R_2^{(n)} = V_2^{(n)}$, we have $\mathcal{D}_1^{(n)} \supset R_2^{(n)}$ and step 8 is passed. By step 9 we get $\mathcal{D}_2^{(n)} = R_2^{(n)} = V_2^{(n)}$. We can repeat such considerations until s takes value $m_J - 1$.

As $\mathcal{D}_{m_J-1}^{(n)} = V_{m_J-1}^{(n)} \supset V_{m_J}^{(n)} = R_{m_J}^{(n)}$, step 8 is passed for $m = m_J$ and we get $\mathcal{D}_{m_J}^{(n)} = R_{m_J}^{(n)} = V_{m_J}^{(n)}$. Thus we arrive to step 11 with $m = m_J$. For all $m_J < s \le M = [n \log_p 2]$, step 14 is passed automatically. For $s = m_J + 1$ we have $\mathcal{D}_{m_J}^{(n)} = V_{m_J}^{(n)} \supset V_{m_J+1}^{(n)} = R_{m_J+1}^{(n)}$. Hence step 15 is passed and we put $\mathcal{D}_{m_J+1}^{(n)} = R_{m_J+1}^{(n)} = V_{m_J+1}^{(n)}$. Repeating these considerations, we prove that $\mathcal{D}_s^{(n)} = R_s^{(n)} = V_s^{(n)}$ for all $s = m_J, \dots, M$. Hence $V^{(n)} = T^{(n)}$. □

6 No p-adic universal test

A natural generalization of the definition of a universal test for randomness is the following one:

Definition 6.1. A p-adic test U is said to be *universal* if for every p-adic test V we can effectively find $c \in \mathbb{N}$ (depending upon U and V) such that $V_{m+c} \subset U_m$ for all m.

It is well known that in the ordinary real probability theory there exists a universal test for randomness (which is, of course, not unique). We shall show that in p-adic probability theory there is no universal recursive tests. We start with some technical considerations. We have to study more carefully properties of the function $\lambda(n) = [n \log_p 2]$. As $p > 2$, we have $\log_p 2 < 1$. We set $L_k = \left[\frac{k}{\log_p 2}\right]$. If $0 < n \le L_1$, then $n \log_p 2 < 1$ and $\lambda(n) = 0$; in the same way we have: if $L_{k-1} < n \le L_k$, then $\lambda(n) = k - 1$, $k \ge 2$. We set $n_k = L_k + 1$.

Lemma 6.1. *The inequality*

$$p^{\lambda(n_k)} > 2^{n_k - 1} \tag{6.1}$$

holds true for all $k = 1, 2, \dots$.

Proof. We have $\lambda(n_k) = k$ and $\lambda(n_k - 1) = k - 1$. By definition $\lambda(n) = \max\{l : p^l < 2^n\}$. Hence, for all n, $p^{\lambda(n)+1} > 2^n$. In particular, $p^{\lambda(n_k-1)+1} = p^k > 2^{n_k-1}$. Hence $p^k = p^{\lambda(n_k)} > 2^{n_k-1}$. □

We construct now two p-adic tests W and $\widetilde{W}$ by using the following procedure.

For $k = 1, 2, \dots$ and $j = 1, \dots, \lambda(n_k)$, we set

$$W_j^{(n_k)} = W_{\lambda(n_k)}^{(n_k)} = \{x_1, \dots, x_{p^{\lambda(n_k)}}\}$$

and

$$\widetilde{W}_j^{(n_k)} = \widetilde{W}_{\lambda(n_k)}^{n_k} = \{x_{2^{n_k} - p^{\lambda(n_k)} + 1}, \dots, x_{2^{n_k}}\}$$

and $W_l^{(n_k)} = \widetilde{W}_l^{(n_k)} = \varnothing$ for $n \neq n_k$ and $l = 1, 2, \ldots$. Here we have used the lexicographic enumeration of elements of X^{n_k}, $k = 1, 2, \ldots$: $x_1, x_2, \ldots, x_{2^{n_k}}$. Since $\sigma(W_j^{(n_k)}) = \sigma(\widetilde{W}_j^{(n_k)}) = p^{\lambda(n_k}$, $j = 1, 2, \ldots, \lambda(n)$, by (6.1) we obtain $W_j^{(n_k)} \cap \widetilde{W}_j^{(n_k)} \neq \varnothing$ and hence

$$X^{n_k} = W_j^{(n_k)} \cup \widetilde{W}_j^{(n_k)}, \quad j = 1, \ldots, \lambda(n_k).$$

Theorem 6.1. *A universal p-adic test does not exist.*

Proof. Let us suppose that there exists a universal p-adic test U. Thus we can effectively find $c_1, c_2 \in \mathbb{N}$ such that $W_{m+c_1} \subset U_m$ and $\widetilde{W}_{m+c_2} \subset U_m$, where W and $\widetilde{W}$ are p-adic tests constructed before this theorem. Let k be so large that $\lambda(n_k) - c_1 \geq 1$ and $\lambda(n_k) - c_2 \geq 1$. Thus $W_{1+c_1}^{(n_k)} = W_{\lambda(n_k)}^{(n_k)}$, $\widetilde{W}_{1+c_2}^{(n_k)} = \widetilde{W}_{\lambda(n_k)}^{(n_k)}$. Hence $U_1^{(n_k)} \supset W_{1+c_1}^{(n_k)} \cup \widetilde{W}_{1+c_2}^{(n_k)} = X^{n_k}$. This implies that $|\sigma(U_1^{(n_k)})|_p = |\sigma(X^{n_k})|_p = 1$. This contradicts to (3.2). □

7 Randomness of infinite sequences

Let $V \subset X^* \times \mathbb{N}$ be a p-adic **P**-test, where **P** is an arbitrary p-adic recursive probability. We set

$$O_m = \cup\{\mathcal{U}_y : y \in V_m\} \subset X^\infty \text{ and } O = \{(\omega, m) : \omega \in O_m\} \subset X^\infty \times \mathbb{N}. \quad (7.1)$$

If the set V_m is infinite, then in general O_m does not belong to the set algebra $\mathcal{F}_{\text{cyl}}$. Therefore probability $\mathbf{P}(O_m)$ may be not defined.

Thus we could not generalize the standard condition for real probabilities (namely $\mathbf{P}(O_m) \leq \frac{1}{p^m}$) to define a p-adic sequential test. It seems that the only possibility to define a p-adic sequential test is to use all tests O obtained via (7.1) from p-adic **P**-tests $V \subset X^* \times \mathbb{N}$.

Definition 7.1. Let $\mathbf{P}$: $\mathcal{F}_{\text{cyl}} \to \mathbb{Q}_p$ be a recursive probability and let $V \subset X^* \times \mathbb{N}$ be a p-adic **P**-test. The set O defined on the basis of V via (7.1) is said to be a *p-adic sequential* **P**-*test*.

Definition 7.2. Let O be a p-adic sequential test. A sequence ω is said to be **P**-*random* with respect to the test O if

$$\omega \notin O_\infty = \bigcap_{m=1}^{\infty} O_m.$$

In general, the set $O_\infty \notin \mathcal{F}_{\text{cyl}}$ and **P** cannot be extended on the σ-algebra $\mathcal{B}(X^\infty)$ containing O_∞. Therefore in general $\mathbf{P}(O_\infty)$ is not defined.

A sequence $\omega \in O_\infty$ is considered as non-random with respect to **P**.

As usual, we restrict our considerations to the case of the uniform p-adic distribution $\mathbf{P} = \mathbf{P}_p$, $p \neq 2$.

Example 7.1. Let O be the p-adic sequential test based on the p-adic test V of Example 3.1. We consider a few examples of sequences $\omega \in X^\infty$ which are random or non-random with respect to O:

(1) Let only a finite number $k \geq 1$ of coordinates of $\omega = (\omega_j)$ be equal to 1: $\omega_{j_1} = \cdots = \omega_{j_k} = 1$. We show that ω is not non-random with respect to O. We have to show that for each m there exist n such that $\omega_{1:n} \in V_m^{(n)}$ where $\omega_{1:n} = (\omega_1, \ldots, \omega_n)$. Let $k = \beta = 1, \ldots, p-1$. We set $n = \beta + p^N$, where $N \geq m$, $N > \log_p(j_k - \beta)$. Then $\Theta_p(n)/\Theta_p(k) = p^{-N} \leq p^{-m}$ and $\omega_{1:n} \in V_m^{(n)}$. Let $k = \beta + jp^l$, $j, l \in \mathbb{N}$, $(j, p) = 1$. We set $n = \beta + p^N$, where $N \geq m + l$, $N > \log_p(j_k - \beta)$. Then $\Theta_p(n)/\Theta_p(k) = p^{-N+l} \leq p^{-m}$ and $\omega_{1:n} \in V_m^{(n)}$.

(2) The sequence $\omega = (0, \ldots, 0, \ldots)$ is random with respect to O, because $\omega \notin O_1$ ($\omega_{1:n} \notin V_1^{(n)}$ for all n); the sequence $\omega = (1, \ldots, 1, \ldots)$ is random with respect to O, because $k = S(\omega_{1:n}) = n$ and $\Theta_p(n)/\Theta_p(k) = 1$ and hence $\omega_{1:n} \notin V_1^{(n)}$ for all n.

(3) Here we present an example of a random sequence $\omega \in X^\infty$ with respect to O which contains the infinite number both of zeros and ones. Any $\omega \in X^\infty$ can be represented as a sequence of blocks $\omega = b_1 b_2 \ldots b_m \ldots$, where $l(b_j) = p^{2j}$. Let $S(b_1) = p$ and $S(b_j) = p^j - p^{j-1}$, $j > 1$. Set $x = b_1 \ldots b_m, m \geq 1$. Then $x \in V_m^{(p^{2m})}$. Here $l(x) = p^{2m}$ and $S(x) = p + (p^2 - p) + \cdots + (p^m - p^{m-1}) = p^m$, thus: $\Theta_p(l(x))/\Theta_p(S(x)) = p^{-m}$.

Example 7.2. Let $\overline{O}$ be a p-adic sequential test based on the p-adic test $\bar{V}$ of Example 3.2.

(1) We consider the same sequence ω as in (1) of Example 7.1.

(a) Let $k = \beta$ or $k = \beta + jp^l$, $(j, p) \neq 1$ and $\beta = 1, \ldots, p-1$. We show that such an $\omega \in X^\infty$ is non-random with respect to $\overline{O}$. Let $n = (\beta - 1) + p^N$, where $N \geq m$ and $N > \log_p(j_k - \beta + 1)$. Then $S(\omega_{1:n}) = k$ and hence $M_p(S(\omega_{1:n})) = \beta \geq M_p(n) + 1$ and $\Theta_p(n) = p^{-N} \leq p^{-m}$. Thus $\omega_{1:n} \in \overline{V}_m^{(n)}$ and $\omega \in \overline{O}_m$.

(b) Let $k = jp^l$, $(j, p) \neq 1$. We show that if $p^m \geq j_k$, then $\omega_{1:n} \notin V_m^{(n)}$, $n \geq 1$ (the condition $\Theta_p(l(x)) \leq p^{-m}$ implies that $l(x) \geq p^m$; but, for $\omega_{1:n}$ with $n \geq p^m$, we have $S(\omega_{1:n}) = k$ and, as $M_p(S(\omega_{1:n})) = 0$, there is no n such that $M_p(S(\omega_{1:n})) \geq M_p(n) + 1$). Thus in the opposite to the test O any sequence $\omega \in X^\infty$ in that only a finite number $k = p^t$, $t = 1, 2, \ldots$, of coordinates are equal to 1 is considered as random with respect to $\overline{O}$.

(2) The sequence $\omega = (0, \ldots, 0, \ldots)$ is random with respect to $\overline{O}$ (because, for all $x = (0, \ldots, 0)$, $0 = M_p(S(x)) < M_p(l(x)) + 1$; the sequence $\omega = (1, \ldots, 1, \ldots)$ is also random with respect to $\overline{O}$ (because, for all $x = (1, \ldots, 1)$, $M_p(S(x)) = M_p(l(x))$).

(3) We consider the same sequence as in (3) of Example 7.1. We show that some of such sequences are random with respect to $\overline{O}$ and some are non-random. Let $\omega = b_1 b_2 \ldots b_m \ldots$ and in each block b_j the first $p^j - p^{j-1}$ elements are equal 1 and $\omega_1 = \cdots = \omega_p = 1$ in b_1 (other elements in each block are equal to 0). If, for $\omega_{1:n}$,

$\Theta_p(n) \leq \frac{1}{p^m}$, then $M_p(n) = 0$ and, hence, $M_p(S(\omega_{1:n})) = 0$. Thus $\omega_{1:n} \notin \overline{V}_m^{(n)}$. Let $\omega = b_1 b_2 \ldots b_m \ldots$ and the distribution of ones in blocks have the following structure. For $b_1 = (x_1, \ldots, x_{p^2})$, $x_1 = \cdots = x_{p-1} = 1$, $x_p = 0$, $x_{p+1} = 1$; for $b_j = (x_1, \ldots, x_{p^{2j}})$, $x_1 = \cdots = x_{p^j - p^{j-1} - 1} = 1$, $x_{p^j - p^{j-1}} = 0$, $x_{p^j + p^{j-1} + 1} = 1$. Then $\omega_{1:p} \in V_1^{(p)}$ (since $M_p(S(\omega_{1:p})) = p - 1 > 1 + M_p(p) = 1$); $\omega_{1:n} \in V_{p-j+1}^{(n)}$ for $n = p^{2j} + p^j - p^{j-1}$ (since $S(\omega_{1:n}) = p^j - p^{j-1} - 1$ implies $M_p(S(\omega_{1:n})) = 1$).

As consequence of Theorem 5.1 we obtain the following theorem:

Theorem 7.1. *The set of all p-adic sequential tests is r.e.*

A p-adic sequential test $\mathcal{D}$ is said to be universal if, for every p-adic sequential test O, we can effectively find $c \in \mathbb{N}$ (depending upon $\mathcal{D}$ and O) such that $O_{m+c} \subset \mathcal{D}_m$ for all m.

At first sight, it seems to be natural to consider the set

$$O_\infty^{\max} = \bigcup_{i=1}^{\infty} O_{(i),\infty}, \tag{7.2}$$

where $O_{(i)}$, $i = 1, 2, \ldots$, is a recursive enumeration of p-adic sequential tests, as the maximal set of p-adic non-random sequences (with respect to the p-adic uniform distribution) and call a sequence $\omega \in X^\infty \setminus O_\infty^{\max}$ a p-adic random sequence. However, Theorem 6.1 (nonexistence of universal p-adic test) is the sign that such a procedure could not be successful.

Proposition 7.1. *The set $O_\infty^{\max}$ defined as (7.2) is equal to X^∞.*

Proof. Let W and $\widetilde{W}$ be p-adic tests defined in Section 6 and let O and $\widetilde{O}$ be the corresponding p-adic statistical tests. We have that $W_j^{(n)} \cup \widetilde{W}_j^{(n)} = X^n$, $j = 1, \ldots, \lambda(n)$, for all n. As $\lambda(n) \to \infty$, $n \to \infty$, then $\forall l, m \in \mathbb{N}\ \exists N = N(l, m)$: $\lambda(N) \geq l, m$. As we also have that $W_j^{(n)} = W_{\lambda(n)}^{(n)}$ and $\widetilde{W}_j^{(n)} = \widetilde{W}_{\lambda(n)}^{(n)}$, $j = 1, \ldots, \lambda(n)$, then, for $N = N(l, m)$, we obtain $W_l^{(N)} \cup \widetilde{W}_m^{(N)} = X^N$. We also have:

$$O_\infty \cup \widetilde{O}_\infty = \Big(\bigcap_{l=1}^{\infty} O_l\Big) \cup \Big(\bigcap_{m=1}^{\infty} \widetilde{O}_m\Big) = \bigcap_{l=1}^{\infty} \bigcap_{m=1}^{\infty} (O_l \cup \widetilde{O}_m).$$

Finally we show that $O_l \cup \widetilde{O}_m = X^\infty$ for every l and m:

$$O_l \cup \widetilde{O}_m \supset \bigcup \{\mathcal{U}_x : x \in W_m^{(N)} \cup \widetilde{W}_l^{(N)}\} = X^\infty,$$

where $N = N(l, m)$. □

Thus in the opposite the real case the existence of the recursive enumeration of the set of all p-adic sequential tests does not imply the possibility of the fruitful development of the theory of randomness based on the maximal constructive set of non-random sequences. In some sense the situation here is similar to the ordinary (real) nonconstructive probability theory where any $B \subset X^\infty$, $\mathbf{P}(B) = 1$, may be considered as a 'law of randomness' (thus the maximal set of non-random sequences coincides with X^∞).

Definition 7.3. A p-adic sequential test $\mathcal{O}$ is said to be a *universal* if for every p-adic sequential test O we can effectively find $c \in \mathbb{N}$ such that $O_{m+c} \subset \mathcal{O}_m$ for all m.

Lemma 7.1. *Let n_j, $j = 1, 2, \ldots$, be numbers associated with the function λ. Then, for $m_j = \lambda(n_j)$,*

$$N_j = (2^{n_j} - p^{m_j})2^{n_{j+1}-n_j} < p^{m_{j+1}}, \quad j = 1, 2, \ldots .$$

Proof. By (6.1) we obtain: $p^{m_j}2^{-n_j} > 1/2$. Thus $(1 - p^{m_j}2^{-n_j}) < 1/2$ and hence

$$N_j = (1 - p^{m_j}2^{-n_j})2^{n_{j+1}} < \frac{2^{n_{j+1}}}{2}.$$

But by (6.1) we also have $p^{m_{j+1}} > \frac{2^{n_{j+1}}}{2}$. □

Proposition 7.2. *The trivial p-adic sequential test $\mathcal{O}$ with $\mathcal{O}_m = X^\infty$, $m \geq 1$, is the (unique) universal p-adic sequential test.*

Proof. We prove that the $\mathcal{O}$ with $\mathcal{O}_m = X^\infty$ for all $m \geq 1$ is a p-adic sequential test: there exists a p-adic test $V \subset X^* \times \mathbb{N}$ such that V induces $\mathcal{O}$. Let n_j be natural numbers associated with λ and let $m_j = \lambda(n_j)$.

We represent $X^{n_j} = A_{m_j}^{(n_j)} \cup B_{m_j}^{(n_j)}$, $A_{m_j}^{(n_j)} \cap B_{m_j}^{(n_j)} = \varnothing$ and $\sigma(A_{m_j}^{(n_j)}) = p^{m_j}$ (and, consequently, $\sigma(B_{m_j}^{(n_j)}) = 2^{n_j} - p^{m_j}$), where the sets $A_{m_j}^{(n_j)}$ are constructed by the following procedure. We set $A_{m_1}^{(n_1)} = \{x_1, \ldots, x_{p^{m_1}}\}$, where $X^{n_1} = \{x_1, \ldots, x_{p^{m_1}}, \ldots, x_{2^{n_1}}\}$. Suppose that the set $A_{m_j}^{(n_j)}$ has been constructed. We set

$$\begin{aligned} C_{m_{j+1}}^{n_{j+1}} &= \{x \in X^{n_j} : x \text{ has a prefix } y \text{ belonging to the set } B_{m_j}^{(n_j)} = X^{n_j} \setminus A_{m_j}^{(n_j)}\} \\ &= B_{m_j}^{(n_j)} \times X^{n_{j+1}-n_j}. \end{aligned}$$

The set $A_{m_{j+1}}^{(n_{j+1})}$ is the union of the set $C_{m_{j+1}}^{n_{j+1}}$ and $p^{m_{j+1}} - \sigma(C_{m_{j+1}}^{n_{j+1}}) = p^{m_{j+1}} - N_j$ (where $N_j = (2^{n_j} - p^{m_j})2^{n_{j+1}-n_j}$) first elements of $X^{n_{j+1}} = (x_1, \ldots, x_{2^{n_{j+1}}})$ which do not belong to $C_{m_{j+1}}^{n_{j+1}}$. By (7.2) $p^{m_{j+1}} - N_j > 0$ for all $j \geq 1$. Thus this procedure define sets $A_{m_j}^{(n_j)}$ for all $j \geq 1$.

We set $V_{m_k} = \bigcup_{j=k}^\infty A_{m_j}^{(n_j)}$ and $V_m = V_{m_k}$ for $m_{k-1} < m \leq m_k$. We prove that $V = \{(x, m) : x \in V_m\}$ is a p-adic test. The set V is r.e. and $V_m \supset V_{m+1}$ by the procedure of construction. We also have: $V_{m_k}^{(n_j)} = A_{m_j}^{(n_j)}$, $j \geq k$, and $V_{m_k}^{(n)} = \emptyset$, $n \neq n_j$, $j \geq k$. Thus

$$|\sigma(V_{m_k}^{(n_j)})|_p = |\sigma(A_{m_j}^{(n_j)})|_p = p^{-m_j} \leq p^{-m_k}.$$

On the other hand, for each m_k, we have:

$$\begin{aligned}
&\Big(\bigcup\{\mathcal{U}_x : x \in A_{m_k}^{(n_k)}\}\Big) \cup \Big(\bigcup\{\mathcal{U}_x : x \in A_{m_{k+1}}^{(n_{k+1})}\}\Big) \\
&\qquad \supset \Big(\bigcup\{\mathcal{U}_x : x \in A_{m_k}^{(n_k)}\}\Big) \cup \Big(\bigcup\{\mathcal{U}_{ya} : y \in B_{m_k}^{n_k}, a \in X^{n_{k+1}-n_k}\}\Big) \\
&\qquad = \Big(\bigcup\{\mathcal{U}_x : x \in A_{m_k}^{(n_k)}\}\Big) \cup \Big(\bigcup\{\mathcal{U}_y : y \in B_{m_k}^{n_k}\}\Big) = X^\infty. \qquad \square
\end{aligned}$$

The previous result implies that in the p-adic case (similar to Schnorr's theory of randomness [158]) the only reasonable approach to randomness of infinite sequences is to use randomness with respect to the concrete p-adic sequential test O. Of course, the use of O-randomness has extremely different origins in our theory and Schnorr's theory. It seems that in the p-adic case this situation is a consequence of the impossibility to define σ-additive (non-discrete) probability on the σ-algebra generated by $\mathcal{F}_{\text{cyl}}$. Thus we have no other possibility than to identify a sequential tests O with tests $V \subset X^* \times \mathbb{N}$. In Schnorr's theory this situation is a consequence of the use of total recursive null-sets.

6 Contextual probability and interference

In the classical Kolmogorov or von Mises theory the *formula of total probability* holds, see Chapter 1. We recall that in the case of two dichotomous random variables $a = \alpha_1, \alpha_2$ and $b = \beta_1, \beta_2$ it has the form:

$$\mathbf{P}(b = \beta_i) = \mathbf{P}(a = \alpha_1)\ \mathbf{P}(b = \beta_i/a = \alpha_1) + \mathbf{P}(a = \alpha_2)\ \mathbf{P}(b = \beta_i/a = \alpha_2). \quad (0.1)$$

On the other hand, we have a quantum analogue of this formula, see Chapter 2, (1.5). We call it the *formula of total probability with the interference term*. It can be written in the form of a perturbation of the classical formula of total probability:

$$\begin{aligned}\mathbf{P}(b = \beta_i) \\ &= \mathbf{P}(a = \alpha_1)\ \mathbf{P}(b = \beta_i/a = \alpha_1) + \mathbf{P}(a = \alpha_2)\ \mathbf{P}(b = \beta_i/a = \alpha_2) \quad (0.2)\\ &\quad + 2\cos\theta\ \sqrt{\mathbf{P}(a = \alpha_1)\mathbf{P}(b = \beta_i/a = \alpha_1)\mathbf{P}(a = \alpha_2)\mathbf{P}(b = \beta_i/a = \alpha_2)},\end{aligned}$$

where θ is the phase angle. This formula was derived, see Chapter 2, in the formalism of complex Hilbert space on the basis of the Born's postulate.

The Hilbert space derivation of the formula (0.2) might induce the impression that we deal with something rather strange and impossible from the point of view of classical probability theory. The appearance of the *interference term* has led to the use of the term "quantum probability" in contradiction to what could be called "regular" or "classical" probability.

However, there is only one type of physical probability and it is one that is subject to measurement via counting and the generation of *relative frequencies.* It is the relative frequency probability (of von Mises) that is directly connected with data from experiment.

In this chapter we provide contextual probabilistic analysis making a contribution to the understanding of probability, the formula of total probability and its violation, cf. Chapter 2.

Our analysis begins with the contextual definition of the relevant probabilities. The probability for the value of one observable is then expressed in terms of the conditional (contextual) probabilities involving the values of a second ("supplementary") observable. In this way the interference term in the generalized formula of total probability gives a measure of *supplementarity* of information which can be obtained through measurements of observables a and b.[1]

[1] Of course, it should be better to use the terminology "complementarity of information". However, N. Bohr had already reserved the notion of complementarity in quantum physics. The crucial in the Bohr's complementarity (Copenhagen complementarity) [72] is *mutual exclusivity* (*incompatibility*), see A. Plotnitsky [32, 150–152] for an extended discussion. And in our approach *supplement information* which need not be based on incompatibility plays the crucial role, see Section 4 of this chapter. The

The perturbing term in the generalized formula of total probability is then expressed in terms of a coefficient λ (probabilistic measure of supplementarity) whose absolute value can be less or equal to one, or it can be greater than one for each of the values of the observable. This range of values for the coefficient λ then introduces three distinct types of perturbations[2]:

a) *trigonometric,*

b) *hyperbolic,*

c) *hyper-trigonometric.*

Each case is then examined separately. Classical ($\lambda = 0$) and quantum cases ($|\lambda| \leq 1$) are then special cases of more general results. Later it will be shown that in the quantum case it is possible to reproduce a Hilbert space in which the probabilities are found in the usual way, but there is a case in which this is not possible, though the space is linear it is not a Hilbert space. In general it is not a complex linear space. In the case of hyperbolic probabilistic behavior we have to use linear representation of probabilities over so called hyperbolic numbers.

1 Växjö model: contextual probabilistic description of observables

A general statistical realistic model for observables based on the contextual viewpoint on probability will be presented. It will be shown that classical as well as quantum probabilistic models can be obtained as particular cases of our general contextual model, the *Växjö model*. Realism is one of the main distinguishing features of the Växjö model. Despite the presence of such essentially quantum effects as, e.g., the interference of probabilities and violation of Bell's inequality, there is still a possibility to go beyond quantum mechanics, see Chapter 2.

From the mathematical point of view our probabilistic model is quite close to the well-known Mackey's model. George Mackey [153] presented a program of huge complexity and importance:

To deduce the probabilistic formalism of quantum mechanics starting with a system of natural probabilistic axioms.

(Here "natural" has the meaning of a natural formulation in classical probabilistic terms.) Mackey tried to realize this program starting with a system of 8 axioms –

Bohr's principle of complementarity is the basis for the Copenhagen interpretation of quantum mechanics which is not a realistic interpretation. Our *principle of supplementarity* (or to say the Växjö principle of complementarity) is compatible with any realistic interpretation of quantum mechanics and in particular with the so called *the Växjö interpretation* – the contextual statistical realistic interpretation [100, 103, 105], see Section 2.

[2]In this book "perturbation" has the meaning of perturbation of a probability distribution and not perturbation of an individual system.

Mackey axioms, see [153]. This was an important step in clarification of the probabilistic structure of quantum mechanics. However, he did not totally succeed, see [153] for details. The crucial axiom (about the complex Hilbert space) was not formulated in natural (classical) probabilistic terms.

In [98, 99, 101, 109–111, 113, 140] I presented a new attempt of realization of the Mackey's program, see also [112, 127–129] for further developments. In my approach the probabilistic structure of quantum mechanics (including the complex Hilbert space) can be derived on the basis of two axioms formulated in classical contextual probabilistic terms. My variant of realization of Mackey's program gives the possibility to combine realism and quantum or better to say quantum-like (QL) probabilistic behavior.

As Mackey [153] pointed out, probabilities cannot be considered as abstract quantities defined outside any reference to a concrete complex of physical conditions C. All probabilities are conditional or better to say contextual.[3] Mackey did a lot to unify classical and quantum probabilistic description and, in particular, demystify quantum probability. One crucial step is however missing in Mackey's work. In his book Mackey [153] introduced the quantum probabilistic model (based on the complex Hilbert space) by means of a special axiom (Axiom 7, p. 71) that looked rather artificial in his general conditional probabilistic framework.

Mackey's model is based on a system of eight axioms, when our own model requires only two axioms. Let us briefly mention the content of Mackey's first axioms. The first four axioms concern conditional structure of probabilities, that is, they can be considered as axioms of a classical probabilistic model. The fifth and sixth axioms are of a logical nature (about questions). We reproduce below Mackey's "quantum axiom", and Mackey's own comments on this axiom (see [153], pp. 71–72):

Axiom 7 (Mackey). *The partially ordered set of all questions in quantum mechanics is isomorphic to the partially ordered set of all closed subsets of a separable, infinite-dimensional Hilbert space.*[4]

Our activity can be considered as an attempt to find a list of physically plausible assumptions from which the Hilbert space structure can be deduced. We show that this list can consist in two axioms (see our Axioms 1 and 2) and that these axioms can be formulated in the same natural probabilistic manner as Mackey's Axioms 1–4.

[3] We remark that the same point of view can be found in the works of A. N. Kolmogorov and R. von Mises. However, it seems that Mackey's book [153] was the first thorough presentation of a program of conditional probabilistic description of measurements, both in classical and quantum physics.

[4] "This axiom has rather a different character from Axioms 1 through 4. These all had some degree of physical naturalness and plausibility. Axiom 7 seems entirely *ad hoc*. Why do we make it? Can we justify making it? What else might we assume? We shall discuss these questions in turn. The first is the easiest to answer. We make it because it "works", that is, it leads to a theory which explains physical phenomena and successfully predicts the results of experiments. It is conceivable that a quite different assumption would do likewise but this is a possibility that no one seems to have explored. Ideally one would like to have a list of physically plausible assumptions from which one could deduce Axiom 7."

1.1 Contexts

We start with the basic definition:

Definition 1.1. Context C is a *complex of conditions*.[5]

In particular, a physical context C is a complex of physical conditions. We consider contexts as *basic elements of reality.* To construct a concrete model M of reality, we should fix some set of contexts $\mathcal{C}$, see Definition 1.2.

Remark 1.1 (Contextuality). Before having a closer look at our model, it is perhaps necessary to discuss the meaning of the term *contextuality*, as it can obviously be interpreted in many different ways. The most common meaning (in QM and quantum logic and especially in consideration of Bell's inequality) is that the outcome for a measurement of an observable u under a contextual model is calculated using a different (albeit hidden) measure space, depending on whether or not compatible observables $v, w, \ldots$ were also made in the same experiment. We remark that the well-known "no-go" theorems (of e.g. Bell) cannot be applied to such contextual models. In our approach the term contextuality has an essentially more general meaning. Physical context is any complex of physical conditions *preceding measurement.*[6]

1.2 Observables

A set of contexts $\mathcal{C}$ which will be used in our model was already fixed. Now also a set of observables $\mathcal{O}$ is given. We shall denote observables by Latin letters, $a, b, \ldots$, and their values by Greek letters, $\alpha, \beta, \ldots$. We suppose that any observable $a \in \mathcal{O}$ can be measured under a complex of physical conditions C for any $C \in \mathcal{C}$.

We remark that our general Växjö-representation of reality does not contain physical systems. At the moment we do not (and need not) consider observables as observables on physical systems. It is only supposed that if a context C is fixed then for any instant of time t we can perform a measurement of any observable $a \in \mathcal{O}$.

For an observable $a \in \mathcal{O}$, we denote the set of its possible values ("spectrum") by the symbol X_a.

We do not assume that all observables or even pairs of them can be measured simultaneously; so they need not be compatible. To simplify considerations, we shall consider only discrete observables and, moreover, all concrete investigations will be performed for *dichotomous observables.*

---

[5] In principle, the notion of context can be considered as a generalization of a widely used notion of *preparation procedure*, see Chapter 2. However, identification of context with preparation procedure would restrict essentially our theory. In applications outside physics (e.g., in psychology and cognitive science) we will consider mental contexts. Such contexts are not simply preparation procedures. The same can be said about economical, political and social contexts. In this book we shall not provide a deeper formalization of the notion of context. In our model the notion of context is basic and irreducible.

[6] In particular, one can create a context by fixing the values of observables $v, w, \ldots$ which are compatible with u we determine some context. However, in this way we can obtain only a very special class of contexts.

1.3 Probabilistic representation of contexts

Axiom 1. *For any context $C \in \mathcal{C}$ and any observable $a \in O$ contextual probabilities $\mathbf{P}(a = \alpha/C)$, $\alpha \in X_a$, are well defined.*

For any observable $a \in \mathcal{O}$ and its value $\alpha \in X_a$, we define the $[a = \alpha]$-selection context C_α as such that

$$\mathbf{P}(a = \alpha/C_\alpha) = 1.$$

Observable a under the context C_α takes the value α with the unit probability, "practically always". Such a context can be obtained via completing measurement of the observable a by the procedure of selection of results coinciding with the value $a = \alpha$. Families of observables and contexts are coupled in the following way:

Axiom 2. *For any observable $a \in O$ and any its value $\alpha \in X_a$, the system of contexts $\mathcal{C}$ contains one (and only one) selection context C_α.*

Postulated uniqueness of the selection context C_α (inside a system of contexts $\mathcal{C}$ which was chosen for a model) is an analogue of nondegeneration of spectrum of an observable $\widehat{a}$ in QM, cf. Example 1.2.

We prefer to call probabilities $\mathbf{P}(b = \beta/C)$ with respect to a context $C \in \mathcal{C}$ *contextual probabilities.* Of course, it would be also possible to call them conditional probabilities, but the latter term was already used in other approaches (e.g., Bayes–Kolmogorov, von Mises).

In contrast to the Bayes–Kolmogorov model, the contextual probability is not probability that an event, say B, occurs under the condition that another event, say C, occurred. The contextual probability $\mathbf{P}(b = \beta/C)$ is probability to get the result $b = \beta$ under the complex of physical conditions C. We can say that this is the probability that the event $B_\beta = \{b = \beta\}$ occurs under the complex of physical conditions C. Thus in our approach not event, but context should be considered as a condition.

Let $a, b \in \mathcal{O}$ and let $\alpha \in X_a$, $\beta \in X_b$. We consider the $[a = \alpha]$-selection context C_α. The contextual probability

$$p^{b/a}(\beta/\alpha) \equiv \mathbf{P}(b = \beta/C_\alpha)$$

is not probability that the event $B_\beta = \{b = \beta\}$ occurs under the condition that the event $A_\alpha = \{a = \alpha\}$ occurred. To find probability $p^{b/a}(\beta/\alpha)$, it is not sufficient to observe the event B_β following the event A_α. It should be verified that the complex of physical conditions C_α was really created. Then there should be performed measurements of the observable b under this context.

At the moment we do not fix a definition of probability. Depending on the choice of a probability theory we can obtain different models. By Axiom 1 the set of probabilities

$$W(\mathcal{O}, C) = \{\mathbf{P}(a = \alpha/C) : a \in \mathcal{O}, \alpha \in X_a\}$$

is well defined for any context $C \in \mathcal{C}$.

We remark that in general the probabilistic data $W(\mathcal{O}, C)$ does not contain the joint probability distributions for pairs of observables belonging to $\mathcal{O}$ (because for $a, b \in \mathcal{O}$ the vector-observable $c = (a, b)$ need not belong to the $\mathcal{O}$). The data $W(\mathcal{O}, C)$ provides a probabilistic image of the context C through the system of observables $\mathcal{O}$.

The matrices of probabilities, "transition probabilities", for pairs of observables $a, b \in \mathcal{O}$

$$\mathbf{P}^{b/a} = (p^{b/a}(\beta/\alpha))$$

will play the fundamental role in our further studies.

For any context $C \in \mathcal{C}$, we complete the probabilistic data $W(\mathcal{O}, C)$ by the data contained in the matrices $\mathbf{P}^{b/a}$ for all pairs $a, b \in \mathcal{O}$. We obtain a collection of contextual probabilities which will be denoted by the symbol $D(\mathcal{O}, C) \equiv D(\mathcal{O}, C)$.

We remark that in further considerations we shall mainly consider the set of observables $\mathcal{O}$ containing only two observables: $\mathcal{O} = \{a, b\}$. By exploring analogy with classical and quantum mechanics one may call them "position" and "momentum." Here

$$D(\mathcal{O}, C) = \{\mathbf{P}(a = \alpha/C), \mathbf{P}(b = \beta/C), \mathbf{P}(a = \alpha/C_\beta), \mathbf{P}(b = \beta/C_\alpha)\},$$

where $\alpha = \alpha_1, \alpha_2$ and $\beta = \beta_1, \beta_2$. Thus the statistical data about the context C collected with the aid of the observables a and b is completed by the "transition probabilities" for these observables.

In the general case we denote by the symbol $\mathcal{D}(\mathcal{O}, \mathcal{C})$ the collection of probabilistic data $D(\mathcal{O}, C)$ for all possible contexts $C \in \mathcal{C}$. In general, the map

$$\pi : \mathcal{C} \to \mathcal{D}(\mathcal{O}, \mathcal{C}), \quad \pi(C) = D(\mathcal{O}, C). \tag{1.1}$$

is not one-to-one. Thus the π-image of contextual reality is very rough:

In general, contexts can not be distinguished with the aid of probabilistic data produced by the class of observables $\mathcal{O}$.

Mathematically such probabilistic data can be represented in various ways. For some ("classical-like" – CL) contexts data $\mathcal{D}(\mathcal{O}, \mathcal{C})$ can be represented by a probability distribution on the phase-space. But there exist contexts (non-CL or we can say QL) for which data $\mathcal{D}(\mathcal{O}, \mathcal{C})$ cannot be represented in such a way. For some QL contexts this data can be represented by complex amplitudes, Chapter 7. In this way we obtain, in particular, the probabilistic formalism of quantum mechanics. However, one can find examples of QL contexts for which data $\mathcal{D}(\mathcal{O}, \mathcal{C})$ can not be represented by complex amplitudes. However, it is possible to represent this data by so called hyperbolic amplitudes (taking values in the algebra of hyperbolic numbers – two-dimensional Clifford algebra). In this case we obtain the probabilistic formalism of "hyperbolic quantum mechanics".

1.4 Växjö model

Definition 1.2. A *contextual statistical model of reality* is a triple

$$M = (\mathcal{C}, \mathcal{O}, \mathcal{D}(\mathcal{O}, \mathcal{C})) \tag{1.2}$$

where $\mathcal{C}$ is a set of contexts and $\mathcal{O}$ is a set of observables which satisfy to Axioms 1, 2, and $\mathcal{D}(\mathcal{O}, \mathcal{C})$ is probabilistic data about contexts belonging to $\mathcal{C}$ which is obtained with the aid of observables belonging to $\mathcal{O}$.

We call observables belonging to the set $\mathcal{O} \equiv \mathcal{O}(M)$ *reference observables.* Inside a model M observables belonging to $\mathcal{O}$ give the only possible references about a context $C \in \mathcal{C}$.

Definition 1.3. Two contexts $C_1, C_2 \in \mathcal{C}$ are *probabilistically equivalent* (with respect to the family of observables $\mathcal{O}$) if for any observable $b \in \mathcal{O}$:

$$\mathbf{P}(b = \beta / C_1) = \mathbf{P}(b = \beta / C_2), \ \beta \in X_b. \tag{1.3}$$

Hence, $W(\mathcal{O}, C_1) = W(\mathcal{O}, C_2)$ for two equivalent contexts. Two equivalent contexts are indistinguishable inside the chosen Växjö model M. However, they could be distinguished inside another model M' which is endowed with another set of reference observables $\mathcal{O}$.

Example 1.1 (Växjö models induced by a Kolmogorov probability space). Let $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$ be a Kolmogorov probability space. It induces various Växjö models via various choices of collections of contexts $\mathcal{C}$ and observables $\mathcal{O}$. Here the collection of contexts $\mathcal{C}$ can be chosen as some sub-family (which need not be a sub σ-algebra) of $\mathcal{F}$ consisting of sets of positive probability: $\mathbf{P}(C) > 0$, $C \in \mathcal{C}$. The crucial point is that "physically realizable" contexts need not form a σ-algebra or algebra (even in this very special case when they have the set-representation). The collection of reference variables $\mathcal{O}$ can be chosen as a subset of the space of random variables $\mathrm{RV}(\mathcal{P})$. At the moment we consider only discrete random variables. The crucial point is that not all random variables can be observed. Therefore in general $\mathcal{O}$ is only a proper subset of $\mathrm{RV}(\mathcal{P})$. Contextual probabilities are given by the Bayes' formula. For an observable (random variable) a and its value α the $[a = \alpha]$-selection context C_α can be chosen as $C_\alpha = \{\omega \in \Omega : a(\omega) = \alpha\}$ – "maximal selection context." By Axiom 2 to get the Växjö model we should assume that all selection contexts $C_\alpha \in \mathcal{C}$. We remark that two contexts C_1 and C_2 such that $\mathbf{P}(C_1 \Delta C_2) = 0$, where the symmetric difference of two sets is defined by $C_1 \Delta C_2 = (C_1 \setminus C_2) \cup (C_2 \setminus C_1)$, are equivalent with respect to any system of observables $\mathcal{O}$. However, this is not a necessary condition of equivalence of contexts. For some systems of observables two contexts can be equivalent even if the symmetric difference of the sets representing these contexts has nonzero probability.

Example 1.2 (Växjö models induced by QM). Here the set of contexts $\mathcal{C}$ can be chosen as a subset of the unit sphere S of complex Hilbert space $\mathcal{H}$. Thus each context $C \in \mathcal{C}$ is encoded by a vector $\psi \in S$: $C \equiv C_\psi$. The set of observables $\mathcal{O}$ can be chosen as a subset of the set of self-adjoint operators having purely discrete[7] nondegenerate[8] spectra. Contextual probabilities are defined by Born's rule. Let an operator $\widehat{a} \in \mathcal{O}$ have the spectrum $X_a = \{\alpha_1, \dots, \alpha_N, \dots\}$, $\alpha_i \neq \alpha_j$ and let $e^a_\alpha, \alpha \in X_a$, be corresponding eigenvectors. The $[a = \alpha]$-selection contexts C_α are represented by these eigenvectors: $C_\alpha \equiv C_{e^a_\alpha}$. Suppose now that $\mathcal{H}$ is the two-dimensional Hilbert space, $\mathcal{O} = \{\widehat{a}, \widehat{b}\}$, where self-adjoint operators $\widehat{a}$ and $\widehat{b}$ do not commute. Take a context which is represented by a normalized vector ψ. If one knows the probabilistic data

$$D(\mathcal{O}, \psi): \quad \mathbf{P}(a = \alpha_i / C_\psi) = |\langle \psi, e^a_{\alpha_i} \rangle|^2, \quad \mathbf{P}(b = \beta_i / C_\psi) = |\langle \psi, e^b_{\beta_i} \rangle|^2,$$
$$\mathbf{P}(b = \beta_i / C_{\alpha_j}) = \mathbf{P}(a = \alpha_j / C_{\beta_i}) = |\langle e^a_{\alpha_j}, e^b_{\beta_i} \rangle|^2, \quad i, j = 1, 2,$$

he can easily reconstruct the vector ψ (up to the complex factor $c : |c| = 1$). Our aim is to show that the same reconstruction, $C \to \psi_C$, can be done for any contextual statistical model, i.e., in the case when a priori we do not have any Hilbert space or operators, but only probabilities. One of the problems in realization of this program is to find a purely probabilistic contextual analogue of the condition of noncommutativity. This condition is crucial in reconstruction of a state on the basis of probabilistic data in ordinary QM. Suppose now that we do not have any Hilbert space, nor operators. We have only physical observables and probabilistic data. Which restriction on this data will induce the complex Hilbert space representation and noncommutativity of corresponding operators?[9]

1.5 Role of reference observables

The reader has already understood that the reference observables play the special role in our model. We interpret the set $\mathcal{O}$ as a family of observables which represent some fixed class of properties of contexts belonging to the class $\mathcal{C}$.

Consider a class of observers (e.g., a class of cognitive systems or measurement devices). Suppose that they collect only the $\mathcal{O}$-properties of contexts belonging to the class $\mathcal{C}$. It can happen, because those observers are able to measure only observables belonging to the family $\mathcal{O}$. However, it may be that those observers simply

[7] We recall that at the moment we defined the Växjö model only for discrete observables.

[8] We consider nondegenerate spectra to escape the problem of non-unique choice of selection contexts.

[9] We emphasize that in opposite to e.g. Heisenberg we do not want to operate with such a notion as incompatibility of observables, i.e., impossibility of simultaneous measurement. Incompatibility is a metaphysical concept. How could one know that two observables could be never measured simultaneously? We shall proceed in the really physical framework. A physical analog of incompatibility will be formulated as the condition of statistical supplementarity. The latter could be checked on the basis of experimental statistical data.

ignore information about the $\mathcal{C}$-contexts which cannot be obtained with help of the $\mathcal{O}$-observables. There can exist other properties of the $\mathcal{C}$-contexts which are not represented by the $\mathcal{O}$-observables.

The same set of contexts $\mathcal{C}$ can be basic for a few models of contextual reality corresponding to different choices of families of observables $\mathcal{O}_i R, Ri = 1, 2, \ldots$: $M_i = (\mathcal{C}, \mathcal{O}_i, \mathcal{D}(\mathcal{O}_i, \mathcal{C}))$. For example, human beings operates with one family of observables, but bats with another.

Remark 1.2. In both most important physical models – in classical and quantum – the set $\mathcal{O}$ of reference observables consists of **two observables:** *position and momentum.*

1.6 Växjö model in biology, sociology, and economy

Our contextual statistical realistic models can be used not only in physics, but in any domain of natural and social sciences. Instead of complexes of physical conditions, we can consider complexes of biological, social, economic,... conditions – contexts – as elements of reality. Such elements of reality are represented (roughly) by probabilistic data obtained with the aid of some reference observables (biological, social, economic,...).

In some special cases it is possible to encode such data by complex amplitudes (as it was done in QM). In this way we obtain representations of some biological, social, economic,... models in complex Hilbert spaces. We call them complex QL-models. These models describe the usual cos-interference of probabilities, see [106].

In other cases it is possible to encode probabilistic data $\mathcal{D}(\mathcal{O}, \mathcal{C})$ by hyperbolic amplitudes. These are amplitudes taking values in so called hyperbolic algebras – two-dimensional Clifford algebras. In this way we obtain representations of some biological, social, economic,... models in hyperbolic Hilbert spaces. We call them hyperbolic QL-models. These models describe the cosh-interference of probabilities.

Remark 1.3 (On macroscopic quantum systems). The study of macroscopic quantum systems is the subject of the greatest interest for foundations of quantum mechanics as well as its applications. However, it is not clear: *"What kind of a system can be called a macroscopic quantum system?"* Of course, this question is closely related to the old question: *"What kind of a system can be called a quantum system?"* We point out that there is no common point of view on such notions as quantization, quantum theory. For me, opposite to N. Bohr, the presence of quanta (of, e.g., action) is not the main distinguishing feature of quantum theory. The presence of observables with discrete spectra is an important feature of quantum theory. However, the basic quantum observables, the position and the momentum, still have continuous ranges of values. In this sense there is no difference between classical and quantum mechanics.

I think that the crucial point is that quantum theory is a *statistical theory.* Therefore it should be characterized in statistical terms. We should find the basic feature of quantum theory which distinguishes this theory from classical statistical mechanics.

The *interference of probabilities* is such a basic statistical feature of quantum theory. Therefore any system (material or not) which exhibits for some observables the interference of probabilities should be considered as a QL system.

1.7 Choice of probability model

As was mentioned, any Växjö model M should be based on some concrete mathematical probabilistic formalism describing probabilistic data $\mathcal{D}(\mathcal{O}, \mathcal{C})$. Of course, the Kolmogorov measure-theoretical approach dominates in modern physics. However, it is not the only possible mathematical theory of probability, see Chapter 1.

I like the contextual frequency approach to probability – the contextual extension of the classical von Mises approach. In this book we shall show that this frequency formalism can be used to describe probabilistic data obtained via measurements of "supplementary of observables."

In the next section we shall present the frequency probabilistic description of data $\mathcal{D}(\mathcal{O}, C)$.

2 S-sequences (collectives) and corresponding probabilities

We shall consider everywhere in this book a set of reference observables $\mathcal{O} = \{a, b\}$ consisting of *two observables a and b.* Another important assumption is that *reference observables are dichotomous:*

$$a = \alpha_1, \alpha_2, \quad \text{and} \quad b = \beta_1, \beta_2.$$

In fact, the representation of probabilistic data by complex or hyperbolic amplitudes in the case of observables taking a finite number of values can be reduced to the case of dichotomous observables.

Let C be some context. In a series of observations of b under this context we obtain a sequence of values of b:

$$x \equiv x(b/C) = (x_1, x_2, \ldots, x_N, \ldots), \quad x_j = \beta_1, \beta_2. \tag{2.1}$$

In a series of observations of a under this context we obtain a sequence of values of a:

$$y \equiv y(a/C) = (y_1, y_2, \ldots, y_N, \ldots), \quad y_j = \alpha_1, \alpha_2. \tag{2.2}$$

It is not assumed that the observables b and a can be measured simultaneously under the context C. Moreover, each measurement of a or b can disturb the C. To produce the sequence (2.1) or the sequence (2.2), one should be able to reproduce C many (in theory infinitely many) times.

We suppose – Axiom 1 completed by the frequency probabilistic model – that these are S-sequences (or even von Mises collectives). Thus the principle of statistical stabilization holds and the frequency probabilities are well defined:

$$p^b(\beta) \equiv \mathbf{P}_x(b = \beta) = \lim_{N\to\infty} \nu_N(\beta; x), \qquad \beta = \beta_1, \beta_2; \tag{2.3}$$

$$p^a(\alpha) \equiv \mathbf{P}_y(a = \alpha) = \lim_{N\to\infty} \nu_N(\alpha; y), \qquad \alpha = \alpha_1, \alpha_2, \tag{2.4}$$

where $\nu_N(\beta; x) = \frac{n(\beta;x)}{N}$, $\nu_N(\alpha; x) = \frac{n(\alpha;x)}{N}$ are relative frequencies of realizations of labels $b = \beta$ and $a = \alpha$ in sequences of observations x and y, respectively (relative frequencies of observations of the results $b = \beta$ and $a = \alpha$ under the context C).

Let C_{α_1} and C_{α_2} be two contexts corresponding to $[a = \alpha_1]$ and $[a = \alpha_2]$-selections, see Axiom 2. By observation of b under the context C_α we obtain a sequence:

$$x^\alpha \equiv x(b/C_\alpha) = (x_1, x_2, \ldots, x_N, \ldots), \quad x_j = \beta_1, \beta_2. \tag{2.5}$$

It is also assumed (see Axiom 1) that the x^α, $\alpha = \alpha_1, \alpha_2$, are von Mises collectives or at least S-sequences. Thus the frequency probabilities with respect to the x^α are well defined:

$$p^{b/a}(\beta/\alpha) \equiv \mathbf{P}_{x^\alpha}(b = \beta) = \lim_{N\to\infty} \nu_N(\beta; x^\alpha), \tag{2.6}$$

where $\nu_N(\beta; x^\alpha) = \frac{n(\beta;x^\alpha)}{N}$ are relative frequencies of realizations of the label β in x^α (relative frequencies of observations of the result $b = \beta$ under the context C_α). One has four different S-sequences (collectives), $x, y, x^{\alpha_1}, x^{\alpha_2}$ which produce four different probability distributions:

$$\mathbf{P}_x(\beta), \quad \mathbf{P}_y(\alpha); \quad \mathbf{P}_{x^{\alpha_1}}(\beta), \quad \mathbf{P}_{x^{\alpha_2}}(\beta). \tag{2.7}$$

We can repeat all previous considerations by changing b/a-conditioning to a/b-conditioning. We consider contexts C_β corresponding to the selections with respect to the values of the observable b (see Axiom 2) and the corresponding S-sequences (collectives) $y^\beta \equiv y(a/C_\beta)$ induced by observations of a in contexts C_β. There are well-defined probabilities $p^{a/b}(\alpha/\beta) \equiv \mathbf{P}_{y^\beta}(\alpha)$, see Axiom 1.

3 Formula of total probability and measures of supplementarity

Let $M = (\mathcal{C}, \mathcal{O}, \mathcal{D}(\mathcal{O}, \mathcal{C}))$ be a Växjö model such that the set of observables $\mathcal{O} = \{a, b\}$ and a, b are dichotomous observables. Let $C \in \mathcal{C}$. There are no reasons to assume that all probability distributions in $D(a, b, C)$ should be described by a single Kolmogorov probability space $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$. Thus the classical (Kolmogorovean) formula of total probability, Chapter 1:

$$\mathbf{P}(b = \beta) = \sum_\alpha \mathbf{P}(a = \alpha)\mathbf{P}(b = \beta/a = \alpha) \tag{3.1}$$

can be violated. We explain this point in more detail. In fact, all probabilities in (3.1) are contextual. In (3.1) we omitted the indexes of contexts. However, in reality three different contexts were involved. There was chosen a (reproducible) context C for observations of a or b. Here

$$\mathbf{P}(a=\alpha) \equiv \mathbf{P}_C(a=\alpha) = \mathbf{P}(a=\alpha/C),$$
$$\mathbf{P}(b=\beta) \equiv \mathbf{P}_C(b=\beta) = \mathbf{P}(b=\beta/C).$$

There are also involved the selection contexts C_{α_1} and C_{α_2}. The correct contextual definition of conditional probabilities in (3.1) is given by

$$\mathbf{P}(b=\beta/a=\alpha) = \mathbf{P}(b=\beta/C_\alpha).$$

In general there are no reasons to assume that probabilities with respect to the different contexts C, C_{α_1} and C_{α_1} can be mathematically described by the same Kolmogorov probability space. We recall that in the conventional (noncontextual) Kolmogorov framework it is assumed that all probabilities are taken with respect the same space of elementary events Ω. This Ω plays the role of a common context:

$$\begin{aligned}\mathbf{P}(a=\alpha) &= \mathbf{P}_\Omega(A_\alpha),\\ \mathbf{P}(b=\beta) &= \mathbf{P}_\Omega(B_\beta),\\ \mathbf{P}(b=\beta/a=\alpha) &= \mathbf{P}_\Omega(B_\beta/A_\alpha)\\ &= \mathbf{P}_\Omega(\omega\in\Omega:\omega\in A_\alpha\cap B_\beta)/\mathbf{P}_\Omega(\omega\in\Omega:\omega\in A_\alpha),\end{aligned}$$

where $B_\beta=\{\omega\in\Omega: b(\omega)=\beta\}$, $A_\alpha=\{\omega\in\Omega: a(\omega)=\alpha\}$.

In the same way, in the von Mises framework the formula of total probability holds for a partition $\{A_k\}$ of the label set L of a fixed S-sequence (collective) u:

$$\mathbf{P}_u(b=\beta) = \sum_\alpha \mathbf{P}_u(a=\alpha)\mathbf{P}_u(b=\beta/a=\alpha). \tag{3.2}$$

This formula can not be derived in the contextual frequency approach, where the conditional probabilities $\mathbf{P}(b=\beta/a=\alpha)$ are defined as contextual probabilities $\mathbf{P}(b=\beta/C_\alpha)$. In this approach, see Section 2, four different S-sequences (collectives) are involved:

$$x=x(b/C),\quad y=y(a/C),\quad x^\alpha=x(b/C_\alpha),\quad \alpha=\alpha_1,\alpha_2.$$

In the contextual Kolmogorov and frequency approaches one could not exclude the possibility that the following statistical coefficient:

$$\delta(\beta/a,C) = \mathbf{P}_x(\beta) - \sum_\alpha \mathbf{P}_y(\alpha)\mathbf{P}_{x^\alpha}(\beta) \neq 0 \tag{3.3}$$

As was mentioned, in the noncontextual Kolmogorov or von Mises models, see (3.1), (3.2), we have:

$$\delta(\beta/a, C) = 0. \tag{3.4}$$

Hence, in the noncontextual Kolmogorov or von Mises models by using the Bayesian sum of the probabilities $\mathbf{P}(a = \alpha)$ and the conditional probabilities $\mathbf{P}(b = \beta/a = \alpha)$ we find nothing new, but the original probabilities $\mathbf{P}(b = \beta)$. However, in the contextual approach we obtain new information via conditional observations: the Bayesian sum of the probabilities $\mathbf{P}(a = \alpha)$ and the contextual probabilities $\mathbf{P}(b = \beta/a = \alpha)$ does not coincide with $\mathbf{P}(b = \beta)$. Here conditional observations give us *supplementary information.*[10]

Definition 3.1. The quantity $\delta(\beta/a, C)$ is said to be a *probabilistic measure of b/a-supplementarity in the context C.*

We can write the equality (3.3) in the form which is similar to the classical formula of total probability:

$$\mathbf{P}_x(\beta) = \sum_\alpha \mathbf{P}_y(\alpha)\mathbf{P}_{x^\alpha}(\beta) + \delta(\beta/a, C), \tag{3.5}$$

or by using shorter notations:

$$p^b(\beta) = \sum_\alpha p^a(\alpha) p^{b/a}(\beta/\alpha) + \delta(\beta/a, C). \tag{3.6}$$

This formula has the same structure as the quantum formula of total probability:

[classical part] + additional term,

cf. (0.2). To write the additional term in the same form as in the quantum representation of statistical data, we perform the normalization of the probabilistic measure of supplementarity by the square root of the product of all probabilities:

$$\lambda(\beta/a, C) = \frac{\delta(\beta/a, C)}{2\sqrt{\prod_\alpha p^a(\alpha) p^{b/a}(\beta/\alpha)}}. \tag{3.7}$$

The coefficient $\lambda(\beta/a, C)$ also will be called the probabilistic measure of supplementarity.

By using this coefficient we rewrite (3.6) in the QL form:

$$p^b(\beta) = \sum_\alpha p^a(\alpha) p^{b/a}(\beta/\alpha) + 2\lambda(\beta/a, C)\sqrt{\prod_\alpha p^a(\alpha) p^{b/a}(\beta/\alpha)}. \tag{3.8}$$

[10] We remark again that information need not be complementary in the sense of N. Bohr's definition of complementarity (i.e., mutually exclusive).

The coefficient $\lambda(\beta/a, C)$ is well defined only in the case when all probabilities $p^a(\alpha)$, $p^{b/a}(\beta/\alpha)$ are strictly positive. We consider the matrix

$$\mathbf{P}^{b/a} = (p^{b/a}(\beta/\alpha)).$$

Traditionally this matrix is called the matrix of transition probabilities. In our approach $p^{b/a}(\beta/\alpha) \equiv \mathbf{P}_{x^\alpha}(b = \beta)$ is the probability to obtain the value $b = \beta$ for the collective x^α. Thus in general we need not speak about states of physical systems and interpret $p^{b/a}(\beta/\alpha)$ as the probability of the transition from the state α to the state β. We remark that the matrix $\mathbf{P}^{b/a}$ is always *stochastic*:

$$\sum_\beta p^{b/a}(\beta/\alpha) = 1 \tag{3.9}$$

for any $\alpha \in X_a$, because for any S-sequence (collective) x^α:

$$\sum_\beta \mathbf{P}_{x^\alpha}(b = \beta) = 1,$$

see Chapter 1. We defined a nondegenerate S-sequence (collective) y as such as

$$p^a(\alpha) \equiv \mathbf{P}_y(\alpha) \neq 0 \quad \text{for all } \alpha.$$

Definition 3.2. A context C is said to be *a-nondegenerate* (*b-nondegenerate*) if the corresponding collective $y \equiv y(a/C)$ ($x \equiv x(b/C)$) is nondegenerate.

We remark that the contexts C_α (collectives x^α) are b-nondegenerate iff

$$p^{b/a}(\beta/\alpha) \neq 0. \tag{3.10}$$

The representation (3.8) is the basis of transition to a (complex or hyperbolic) Hilbert space representation of probabilistic data $D(a, b, C)$. The representation (3.8) can be used only for nondegenerate contexts C and C_α.

We can repeat all previous considerations by changing b/a-conditioning to a/b-conditioning. We consider contexts C_β corresponding to selections with respect to values of the observable b and the corresponding collectives $y^\beta \equiv y(a/C_\beta)$. There can be defined probabilistic measures of supplementarity $\delta(\alpha/b, C)$ and $\lambda(\alpha/b, C)$, $\alpha \in X_a$. We remark that the contexts C_β (collectives y^β) are a-nondegenerate iff

$$p^{a/b}(\alpha/\beta) \neq 0. \tag{3.11}$$

For nondegenerate contexts C and C_β, we have:

$$p^a(\alpha) = \sum_\beta p^b(\beta) p^{a/b}(\alpha/\beta) + 2\lambda(\alpha/b, C)\sqrt{\prod_\beta p^b(\beta) p^{a/b}(\alpha/\beta)}. \tag{3.12}$$

Definition 3.3. Reference observables a and b are called (statistically) *conjugate* if (3.10) and (3.11) hold.

Theorem 3.1. *Let reference observables be conjugate and let a context $C \in \mathcal{C}$ be both a- and b-nondegenerate. Then quantum-like formulas of total probability (3.8) and (3.12) hold.*

4 Supplementary physical observables

Definition 4.1. Reference observables a and b are called *b/a-supplementary in a context C* if

$$\delta(\beta/a, C) \neq 0 \quad \text{for some } \beta \in X_b. \tag{4.1}$$

Lemma 4.1. *For any context $C \in \mathcal{C}$, we have:*

$$\sum_{\beta \in X_b} \delta(\beta/a, C) = 0. \tag{4.2}$$

Proof. We have

$$1 = \sum_{\beta \in X_b} p^b(\beta) = \sum_{\beta \in X_b} \sum_{\alpha \in X_a} p^a(\alpha) p^{b/a}(\beta/\alpha) + \sum_{\beta \in X_b} \delta(\beta/a, C).$$

Since $\mathbf{P}^{b/a}$ is always a stochastic matrix, we have for any $\alpha \in X_a$:

$$\sum_{\beta \in X_b} p^{b/a}(\beta/\alpha) = 1.$$

By using that $\sum_{\alpha \in X_a} p^a(\alpha) = 1$ we obtain (4.2). □

We point out that by Lemma 4.1 the coefficient $\delta(\beta_1/a, C) = 0$ iff $\delta(\beta_2/a, C) = 0$. Thus b/a-supplementarity is equivalent to the condition $\delta(\beta/a, C) \neq 0$ both for β_1 and β_2.

Definition 4.2. Reference observables a and b are called *supplementary in a context C* if they are b/a or a/b supplementary:

$$\delta(\beta/a, C) \neq 0 \quad \text{or} \quad \delta(\alpha/b, C) \neq 0 \quad \text{for some } \beta \in X_b,\ \alpha \in X_a. \tag{4.3}$$

By Lemma 4.1 observables are supplementary iff the coefficient $\delta(\beta/a, C) \neq 0$ for all $\beta \in X_b$ or the coefficient $\delta(\alpha/b, C) \neq 0$ for all $\alpha \in X_a$.

Let us consider a contextual model M with the set of contexts $\mathcal{C}$. Observables a and b are said to be supplementary in the model M if there exists $C \in \mathcal{C}$ such that they are supplementary in the context C.

Reference observables a and b are called *nonsupplementary* in the context C if they are neither b/a nor a/b-supplementary:

$$\delta(\beta/a, C) = 0 \quad \text{and} \quad \delta(\alpha/b, C) = 0 \quad \text{for all } \beta \in X_b,\ \alpha \in X_a. \tag{4.4}$$

Thus in the case of b/a-supplementarity we have (for $\beta \in X_b$):

$$p^b(\beta) \neq \sum_\alpha p^a(\alpha) p^{b/a}(\beta/\alpha); \tag{4.5}$$

in the case of a/b-supplementarity we have (for $\alpha \in X_a$):

$$p^a(\alpha) \neq \sum_\beta p^b(\beta) p^{a/b}(\alpha/\beta); \tag{4.6}$$

in the case of supplementarity we have (4.5) or (4.6). In the case of nonsupplementarity we have both representations:

$$p^b(\beta) = \sum_\alpha p^a(\alpha) p^{b/a}(\beta/\alpha), \qquad \beta \in X_b, \tag{4.7}$$

$$p^a(\alpha) = \sum_\beta p^b(\beta) p^{a/b}(\alpha/\beta), \qquad \alpha \in X_a. \tag{4.8}$$

5 Principle of supplementarity

At the first stage of development of the Växjö model I used (following Heisenberg) the terminology *incompatible observables*. However, after a careful analysis of discussions between N. Bohr and W. Heisenberg on the difference between principles of complementarity and uncertainty, I decided to choose the side of N. Bohr and to use *complementarity* instead of *incompatibility.* I think that N. Bohr was totally right by underlying complementarity of information given by some quantum observables. However, after more careful study of Bohr's views to complementarity and especially discussions with A. Plotnitsky (see also his works on Bohr's complementarity [32, 150–152], I found that by complementarity N. Bohr understood complementarity based on *mutual exclusivity.* In the Växjö approach mutual exclusivity of experimental conditions is not important. The crucial role is played by supplementarity of information in the sense of "additional information."

Finally, we remark that observables are incompatible if they cannot be measured simultaneously (e.g., because of the mutual disturbance). However, this need not imply that such observables should be supplementary. The impossibility to be measured simultaneously does not imply the presence of supplementary information about a physical context C. We formulate the following principle:

THE PRINCIPLE OF SUPPLEMENTARITY:

There exist physical observables, say a and b, such that for some context C they produce supplementary statistical information; in the sense that the contextual probability distribution of, e.g., the observable, b could not be reconstructed on the basis of the probability distribution of a. The classical formula of total probability is violated. Supplementarity of the observables a and b under the context C induces interference of probabilities $\mathbf{P}(b = \beta/C)$ *and* $\mathbf{P}(a = \alpha/C)$.

We remark that, opposite to Bohr's principle of complementarity which is a purely philosophic statement, our principle of supplementarity is experimentally testable for pairs of observables.

6 Supplementarity and Kolmogorovness

Let us consider a Växjö model with the set of observables $\mathcal{O} = \{a, b\}$, where a and b are dichotomous.

Definition 6.1. Probabilistic data $D(a, b, C)$ is said to be *Kolmogorovean* if there exists a Kolmogorov probability space $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$ and random variables ξ_a and ξ_b on $\mathcal{P}$ such that:

$$p^a(\alpha) = \mathbf{P}(\xi_a = \alpha), \qquad p^b(\beta) = \mathbf{P}(\xi_b = \beta); \tag{6.1}$$

$$p^{b/a}(\beta/\alpha) = \mathbf{P}(\xi_b = \beta/\xi_a = \alpha), \quad p^{a/b}(\alpha/\beta) = \mathbf{P}(\xi_a = \alpha/\xi_b = \beta). \tag{6.2}$$

Here the conditional probabilities are defined by the Bayes' formula.

If data $D(a, b, C)$ is Kolmogorovean then the observables a and b can represented by Kolmogorovean random variables ξ_a and ξ_b. We remark that Kolmogorovness of statistical data in the sense of Definition 6.1 is not so natural from the physical viewpoint. In fact, probabilities p^a, p^b, $p^{b/a}$, $p^{a/b}$ correspond to different complexes of physical conditions (contexts) C, C_α, C_β. It would be more natural to assume that each context determines its own Kolmogorov probability measure. Nevertheless, Kolmogorovean data appear in many models, e.g., in classical statistical physics.

Definition 6.2. Data $D(a, b, C)$ is *Kolmogorovean* if and only if

$$p^a(\alpha) p^{b/a}(\beta/\alpha) = p^b(\beta) p^{a/b}(\alpha/\beta). \tag{6.3}$$

Proof. a) If data $D(a, b, C)$ is Kolmogorovean then (6.3) is reduced to the equality $\mathbf{P}(O_1 \cap O_2) = \mathbf{P}(O_2 \cap O_1)$ for $O_1, O_2 \in \mathcal{F}$.

b) Let (6.3) holds true. We set $\Omega = X_a \times X_b$, where $X_a = \{\alpha_1, \alpha_2\}$, $X_b = \{\beta_1, \beta_2\}$. We define the probability distribution on Ω by

$$\mathbf{P}(\alpha, \beta) = p^b(\beta) p^{a/b}(\alpha/\beta) = p^a(\alpha) p^{b/a}(\beta/\alpha);$$

and define the random variables $\xi_a(\omega) = \alpha, \xi_b(\omega) = \beta$ for $\omega = (\alpha, \beta)$. We have

$$\mathbf{P}(a = \alpha) = \sum_\beta \mathbf{P}(\alpha, \beta) = \sum_\beta p^a(\alpha) p^{b/a}(\beta/\alpha)$$
$$= p^a(\alpha) \sum_\beta p^{b/a}(\beta/\alpha) = p^a(\alpha).$$

Analogously, it follows $\mathbf{P}(b = \beta) = p^b(\beta)$. Thus

$$\mathbf{P}(a = \alpha/b = \beta) = \frac{\mathbf{P}(a = \alpha, b = \beta)}{\mathbf{P}(b = \beta)} = \frac{p^b(\beta) p^{a/b}(\alpha/\beta)}{p^b(\beta)} = p^{a/b}(\alpha/\beta).$$

And in the same way we prove that $p^{b/a}(\beta/\alpha) = \mathbf{P}(b = \beta/a = \alpha)$. □

We now investigate the relation between Kolmogorovness and nonsupplementarity. If $D(a, b, C)$ is Kolmogorovean then the formula of total probability holds true and we have (4.4). Thus observables a and b are nonsupplementary (in the context C). Thus:

Kolmogorovness implies nonsupplementarity

or as we also can say:

Supplementarity implies non-Kolmogorovness.

However, in the general case nonsupplementarity does not imply that probabilistic data $D(a, b, C)$ is Kolmogorovean. Let us investigate in more detail the case when both matrices $\mathbf{P}^{a/b}$ and $\mathbf{P}^{b/a}$ are *doubly stochastic*. We recall that a matrix $\mathbf{P}^{b/a} = (p^{b/a}(\beta/\alpha))$ is doubly stochastic if it is stochastic (so (3.9) holds) and, moreover,

$$\sum_\alpha p^{b/a}(\beta/\alpha) = 1, \beta = \beta_1, \beta_2. \tag{6.4}$$

Remark 6.1 (Doubly stochasticity as the law of statistical balance). As was mentioned, the equality (3.9) holds automatically. This is a consequence of additivity and normalization by 1 of the probability distribution of any collective x^α. But the equality (6.4) is an additional condition on the observables a and b. Thus by *considering doubly stochastic matrices we choose a very special pair of reference observables.* I propose the following physical interpretation of the equality (3.9). Since

$$p^{b/a}(\beta/\alpha_2) = 1 - p^{b/a}(\beta/\alpha_1),$$

the C_{α_1} and C_{α_2} contexts compensate each other in "preparation of the property" $b = \beta$. Thus the equation (6.4) could be interpreted as the *law of statistical balance* for the property $b = \beta$. If both matrices $\mathbf{P}^{b/a}$ and $\mathbf{P}^{a/b}$ are doubly stochastic then we have laws of statistical balance for both properties: $a = \alpha$ and $b = \beta$.

Definition 6.3. Observables a and b are said to be *statistically balanced* if both matrices $\mathbf{P}^{b/a}$ and $\mathbf{P}^{a/b}$ are doubly stochastic.

It is useful to recall the following well-known result about double stochasticity for Kolmogorovean random variables:

Lemma 6.1. *Let ξ_a and ξ_b be random variables on a Kolmogorov space $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$. Then the following conditions are equivalent:*

(1) The matrices $\mathbf{P}^{a/b} = (\mathbf{P}(\xi_a = \alpha/\xi_b = \beta))$, $\mathbf{P}^{b/a} = (\mathbf{P}(\xi_b = \beta/\xi_a = \alpha))$ are doubly stochastic.

(2) Random variables are uniformly distributed:

$$\mathbf{P}(\xi_a = \alpha) = \mathbf{P}(\xi_b = \beta) = \frac{1}{2}.$$

(3) Random variables are "symmetrically conditioned" in the following sense:

$$\mathbf{P}(\xi_a = \alpha/\xi_b = \beta) = \mathbf{P}(\xi_b = \beta/\xi_a = \alpha), \tag{6.5}$$

so $\mathbf{P}^{a/b} = \mathbf{P}^{b/a}$.

Proof. We set $A_i = \{\omega \in \Omega : \xi_a(\omega) = \alpha_i\}$ and $B_j = \{\omega \in \Omega : \xi_b(\omega) = \beta_j\}$, $j = 1, 2$ (we recall that we consider dichotomous random variables). First we prove that (3) is equivalent to (2):

(a1). Let $\mathbf{P}(A_i/B_j) = \mathbf{P}(B_j/A_i)$. Then

$$\frac{\mathbf{P}(A_1B_1)}{\mathbf{P}(B_1)} = \frac{\mathbf{P}(B_1A_1)}{\mathbf{P}(A_1)}, \quad \frac{\mathbf{P}(A_2B_2)}{\mathbf{P}(B_2)} = \frac{\mathbf{P}(B_2A_2)}{\mathbf{P}(A_2)}, \quad \frac{\mathbf{P}(A_1B_2)}{\mathbf{P}(B_2)} = \frac{\mathbf{P}(B_2A_1)}{\mathbf{P}(A_1)}. \tag{6.6}$$

Thus we obtain: $\mathbf{P}(B_1) = \mathbf{P}(A_1)$, $\mathbf{P}(A_2) = \mathbf{P}(B_2)$, $\mathbf{P}(B_2) = \mathbf{P}(A_1)$. Thus

$$\mathbf{P}(B_1) = \mathbf{P}(B_2) = 1/2 \text{ and } \mathbf{P}(A_1) = \mathbf{P}(A_2) = 1/2. \tag{6.7}$$

(b1). Starting with (6.7) we obtain (6.6) and consequently a, b-symmetry of transition probabilities.

We now prove that (1) is equivalent to (2):

(a2). Let $\mathbf{P}(B_1/A_1) = \mathbf{P}(B_2/A_2)$, $\mathbf{P}(B_1/A_2) = \mathbf{P}(B_2/A_1)$. Then

$$\frac{\mathbf{P}(B_1A_1)}{\mathbf{P}(A_1)} = \frac{\mathbf{P}(B_2A_2)}{\mathbf{P}(A_2)}, \qquad \frac{\mathbf{P}(B_1A_2)}{\mathbf{P}(A_2)} = \frac{\mathbf{P}(B_2A_1)}{\mathbf{P}(A_1)}; \tag{6.8}$$

$$\frac{\mathbf{P}(A_1B_1)}{\mathbf{P}(B_1)} = \frac{\mathbf{P}(A_2B_2)}{\mathbf{P}(B_2)}, \qquad \frac{\mathbf{P}(A_1B_2)}{\mathbf{P}(B_2)} = \frac{\mathbf{P}(A_2B_1)}{\mathbf{P}(B_1)}. \tag{6.9}$$

By these equations we obtain:

$$\frac{\mathbf{P}(B_1)}{\mathbf{P}(A_1)} = \frac{\mathbf{P}(B_2)}{\mathbf{P}(A_2)}, \quad \frac{\mathbf{P}(B_2)}{\mathbf{P}(A_1)} = \frac{\mathbf{P}(B_1)}{\mathbf{P}(A_2)}.$$

So $\frac{\mathbf{P}(A_2)}{\mathbf{P}(A_1)} = \frac{\mathbf{P}(A_1)}{\mathbf{P}(A_2)}$. Thus we obtain (6.7).

(b2). Let (6.7) hold true. Then we have already proved that transition probabilities are symmetric. Thus

$$\mathbf{P}(B_i/A_1) + \mathbf{P}(B_i/A_2) = \mathbf{P}(A_1/B_i) + \mathbf{P}(A_2/B_i) = 1$$

(since every matrix of transition probabilities is always stochastic). □

In general the Kolmogorovean characterization of statistically balanced random variables is not valid for observables of a Växjö model – contextual statistical model.

Proposition 6.1. *A Kolmogorov model for data $D(a, b, C)$ need not exist even in the case of nonsupplementary statistically balanced observables having the uniform probability distribution (for the context C).*

Proof. Let us consider probabilistic data $D(a, b, C)$ such that $p^a(\alpha) = p^b(\beta) = 1/2$ (here $p^a(\alpha) \equiv \mathbf{P}(a = \alpha/C)$, $p^b(\beta) \equiv \mathbf{P}(b = \beta/C)$) and both matrices $\mathbf{P}^{a/b}$ and $\mathbf{P}^{b/a}$ are doubly stochastic. Let us assume that $p^{a/b}(\alpha/\beta) \neq p^{b/a}(\beta/\alpha)$. Then by Lemma 6.1 data $D(a, b, C)$ is non-Kolmogorovean, but

$$2\delta(\alpha/\beta, C) = 1 - \sum_\beta p^{a/b}(\alpha/\beta) = 0, \quad 2\delta(\beta/\alpha, C) = 1 - \sum_\alpha p^{b/a}(\beta/\alpha) = 0. \quad \square$$

It seems to be that symmetrical conditioning plays the crucial role in these considerations. Let M be a Växjö model.

Definition 6.4. Observables $a, b \in \mathcal{O}$ are called *symmetrically conditioned* if

$$p^{a/b}(\alpha/\beta) = p^{b/a}(\beta/\alpha). \tag{6.10}$$

Lemma 6.2. *If observables a and b are symmetrically conditioned, then they are statistically balanced (so the matrices $\mathbf{P}^{a/b}$ and $\mathbf{P}^{b/a}$ are doubly stochastic).*

Proof. We have

$$\sum_\beta p^{a/b}(\alpha/\beta) = \sum_\beta p^{b/a}(\beta/\alpha) = \sum_\beta \mathbf{P}_{x^\alpha}(b = \beta) = 1;$$

$$\sum_\alpha p^{b/a}(\beta/\alpha) = \sum_\alpha \mathbf{P}_{y^\beta}(a = \alpha) = 1. \quad \square$$

As we have seen in Proposition 6.1, statistically balanced observables need not be symmetrically conditioned, cf. Lemma 6.2.

Proposition 6.2. *Let observables a and b be symmetrically conditioned. Probabilistic data $D(a, b, C)$ is Kolmogorovean iff the observables a and b are nonsupplementary in the context C.*

Proof. Suppose that a and b are nonsupplementary. We set

$$p^{b/a}(1/1) = p^{b/a}(2/2) = p \quad \text{and} \quad p^{b/a}(1/2) = p^{b/a}(2/1) = 1 - p$$

(we recall that by Lemma 6.3 the matrix $\mathbf{P}^{b/a}$ is doubly stochastic).

By (4.7), (4.8) we have

$$p^a(\alpha_i) = \sum_\beta p^b(\beta)p^{a/b}(\alpha_i/\beta) = \sum_\beta \sum_\alpha p^a(\alpha)p^{b/a}(\beta/\alpha)p^{a/b}(\alpha_i/\beta)$$
$$= \sum_\alpha p^a(\alpha) \sum_\beta p^{b/a}(\beta/\alpha)p^{b/a}(\beta/\alpha_i).$$

Let us consider the case $i = 1$:

$$p^a(\alpha_1) = p^a(\alpha_1)(p^2 + (1-p)^2) + 2p^a(\alpha_2)p(1-p)$$
$$= p^a(\alpha_1)(1 - 4p + 4p^2) + 2p(1-p).$$

Thus $p^a(\alpha_1) = 1/2$. Hence $p^a(\alpha_2) = 1/2$. In the same way we get that $p^b(\beta_1) = p^b(\beta_2) = 1/2$. Thus the condition (6.3) holds true and there exist a Kolmogorov model $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$ for probabilistic data $D(a, b, C)$. □

Conclusion. In the case of symmetrical conditioning Kolmogorovness is equivalent to nonsupplementarity.

Corollary 6.1. *For symmetrically conditioned observables probabilistic data $D(a, b, C)$ is Kolmogorovean iff the observables a and b are uniformly distributed:*

$$p^a(\alpha_1) = p^a(\alpha_2) = 1/2; \quad p^b(\beta_1) = p^b(\beta_2) = 1/2.$$

7 Quantum-like representation

As was pointed out in Chapter 6, starting with the formula of total probability with the interference term we can construct the representation of a special class of contexts of the Växjö model, so called *trigonometric contexts*, in complex Hilbert space. Then we obtain the Born's rule and the representation of the reference observables by (noncommutative) self-adjoint operators $\widehat{a}$ and $\widehat{b}$. (Noncommutativity of operators is equivalent to consideration of statistically conjugate reference observables.) Thus the quantum probabilistic formalism can be derived from the Växjö model on the basis of the formula of total probability with the interference term. In this chapter we shall realize this program of derivation of the quantum probabilistic formalism.

We shall present a simple algorithm for transferring the probabilistic data $D(a, b, C)$ about a context C into a complex probabilistic amplitude: *quantum-like representation algorithm,* QLRA. The main distinguishing feature of QLRA is that classical probabilistic data is coupled with its QL-image by the Born's rule.

1 Trigonometric, hyperbolic, and hyper-trigonometric contexts

Let $M = (\mathcal{C}, \mathcal{O}, \mathcal{D}(\mathcal{O}, \mathcal{C}))$ be a contextual statistical model such that $\mathcal{O} = \{a, b\}$. Here a and b are dichotomous reference observables. The formula of total probability with the interference term, Chapter 6, plays the fundamental role in further considerations. It was shown that for reference observables a, b and a context C such that

(Na): all elements of the matrix of $\mathbf{P}^{b/a}$ are strictly positive: $p^{b/a}(\beta/\alpha) > 0$,

(Nb): the context C is a-nondegenerate, i.e, $p^a(\alpha) \equiv \mathbf{P}(a = \alpha/C) > 0$ for all $\alpha \in X_a$,

we have the following interfering-representation of probabilities

$$p^b(\beta) = \sum_\alpha p^a(\alpha)p^{b/a}(\beta/\alpha) + 2\lambda(\beta/\alpha, C)\sqrt{\prod_\alpha p^a(\alpha)p^{b/a}(\beta/\alpha)}, \qquad (1.1)$$

where $\lambda(\beta/\alpha, C)$ is the coefficient of b/a-supplementarity with respect to the context C. Depending on the magnitude of this coefficient the generalized formula of total probability can be rewritten either in the form of the well-known trigonometric cos-interference or in the form of so called hyperbolic cosh-interference.

(1) Suppose that the coefficients of b/a-supplementarity $\lambda(\beta/\alpha, C)$ with respect to the context C are relatively small:

$$|\lambda(\beta/a, C)| \leq 1, \quad \beta \in X_b. \qquad (1.2)$$

In this case we can introduce new statistical parameters ("probabilistic angles") $\theta(\beta/\alpha, C) \in [0, 2\pi]$ and represent the coefficients in the trigonometric form:

$$\lambda(\beta/a, C) = \cos\theta(\beta/a, C).$$

Parameters $\theta(\beta/\alpha, C)$ are said to be *b/a-relative phases* with respect to the context C. In this case we obtain the following interference formula of total probability:

$$p^b(\beta) = \sum_\alpha p^a(\alpha)p^{b/a}(\beta/\alpha) + 2\cos\theta(\beta/\alpha, C)\sqrt{\prod_\alpha p^a(\alpha)p^{b/a}(\beta/\alpha)}. \quad (1.3)$$

This is nothing else than the famous *formula of interference of probabilities.* Typically this formula is derived by using the Hilbert space (unitary) transformation corresponding to the transition from one orthonormal basis to another and Born's probability postulate, see Chapter 2. The orthonormal basis under quantum consideration consist of eigenvectors of operators $\hat{a}$ and $\hat{b}$ (noncommutative) corresponding to quantum physical observables a and b. We demonstrated that, opposite to the common (especially in quantum physics) opinion, nontrivial interference of probabilities is not related to some special (and even mystical) "quantum features" of a model (observables a and b and a context C). In the Växjö approach all probabilistic considerations are purely classical. We shall not consider waves or appeal to wave-particle duality. Interference of probabilities for observables a and b in a context C is a consequence of the presence in these observables some supplementary information about the context C. The coefficient λ gives the measure of this supplementarity. Thus by the Växjö interpretation the interference of probabilities is exhibition of the presence in b-observations some additional information which could not be obtained on the basis of a-observations.

Definition 1.1. Let C be an a-nondegenerate context and let (1.2) hold. Such a context is called *b/a- trigonometric*.

For a trigonometric context C, starting from (1.3) and applying the QLRA we shall construct a complex probability amplitude ψ_C. We shall introduce a Hilbert space structure on the space of complex amplitudes, and represent the reference observables a, b by noncommutative operators $\hat{a}$, $\hat{b}$ in this Hilbert space. The QLRA is consistent with the Born's rule.

(2) Suppose that the coefficients of b/a-supplementarity $\lambda(\beta/\alpha, C)$ with respect to the context C are relatively large:

$$|\lambda(\beta/a, C)| \geq 1, \quad \beta \in X_b. \quad (1.4)$$

In this case we can introduce new statistical parameters ("hyperbolic probabilistic angles") $\theta(\beta/a, C) \in (-\infty, +\infty)$ and represent the coefficients in the hyperbolic form:

$$\lambda(\beta/a, C) = \pm\cosh\theta(\beta/a, C).$$

Parameters $\theta(\beta/a, C)$ are said to be (hyperbolic) b/a-*relative phases* with respect to the context C. In this case we obtain the following hyperbolic interference formula of total probability:

$$p^b(\beta) = \sum_\alpha p^a(\alpha) p^{b/a}(\beta/\alpha) \pm 2\cosh\theta(\beta/\alpha, C)\sqrt{\prod_\alpha p^a(\alpha) p^{b/a}(\beta/\alpha)}. \quad (1.5)$$

Definition 1.2. Let C be an a-nondegenerate context and let (1.4) hold. Such a context is called *hyperbolic*.

For a hyperbolic context C, starting from (1.5) and applying the hyperbolic version of QLRA we shall construct a hyperbolic probability amplitude ψ_C taking values in the hyperbolic algebra **G** – the two-dimensional Clifford algebra with the basis $e_1 = 1$ and $e_2 = j$, where $j^2 = +1$, see [140], [110].

(3) Suppose that the coefficients of b/a-supplementarity $\lambda(\beta/\alpha, C)$ with respect to the context C are relatively small for some β and relatively large for another. In this case we obtain the hyper-trigonometric interference of probabilities. This case has not yet been studied in detail.

2 Quantum-like representation algorithm – QLRA

2.1 Probabilistic data about context

We denote the set of trigonometric contexts by the symbol $\mathcal{C}^{\mathrm{tr}}$. We emphasize that $\mathcal{C}^{\mathrm{tr}}$ depends on the reference observables a, b:

$$\mathcal{C}^{\mathrm{tr}} \equiv \mathcal{C}^{\mathrm{tr}}_{b/a,a}.$$

Here the index b/a points to the definition of the coefficients of supplementarity through b/a- conditioning, and the index a points to consideration of contexts which are a-nondegenerate. Everywhere below we assume that the reference observables a and b are *statistically conjugate.* Hence $p^{b/a}(\beta/\alpha) > 0$ for all $\alpha \in X_a$ and $\beta \in X_b$.

Let a context $C \in \mathcal{C}^{\mathrm{tr}}$. We would like to notice the dependence of probabilities on the context C:

$$p^a(\alpha) \equiv p^a_C(\alpha), \quad p^b(\beta) \equiv p^b_C(\beta), \quad \alpha \in X_a,\ \beta \in X_b.$$

We rewrite the generalized formula of total probability (1.3) in the context dependent form:

$$p^b_C(\beta) = \sum_\alpha p^a_C(\alpha) p^{b/a}(\beta/\alpha) + 2\cos\theta(\beta/\alpha, C)\sqrt{\prod_\alpha p^a_C(\alpha) p^{b/a}(\beta/\alpha)}. \quad (2.1)$$

2.2 Construction of complex probabilistic amplitudes

By using the elementary formula:

$$D = A + B + 2\sqrt{AB}\cos\theta = |\sqrt{A} + e^{i\theta}\sqrt{B}|^2,$$

for real numbers $A, B > 0$, $\theta \in [0, 2\pi]$, we can represent the probability $p_C^b(\beta)$ as the square of the complex amplitude (Born's rule):

$$p_C^b(\beta) = |\psi_C(\beta)|^2. \tag{2.2}$$

Here

$$\begin{aligned}\psi(\beta) &\equiv \psi_C(\beta)\\ &= \sqrt{p_C^a(\alpha_1)p^{b/a}(\beta/\alpha_1)} + e^{i\theta_C(\beta)}\sqrt{p_C^a(\alpha_2)p^{b/a}(\beta/\alpha_2)}, \quad \beta \in X_b,\end{aligned} \tag{2.3}$$

where $\theta_C(\beta) \equiv \theta(\beta/\alpha, C)$.

The formula (2.3) gives the QL representation algorithm – QLRA. For any trigonometric context C by starting with the probabilistic data – $p_C^b(\beta)$, $p_C^a(\alpha)$, $p^{b/a}(\beta/\alpha)$ – QLRA produces the complex amplitude ψ_C. This algorithm can be used in any domain of science to create the QL-representation of probabilistic data (for a special class of contexts).[1]

We point out that QLRA contains the reference observables as parameters. Hence the complex amplitude give by (2.3) depends on a, b: $\psi_C \equiv \psi_C^{b/a}$.

Remark 2.1 (Choice of probabilistic phases). For each $\beta \in X_b$ the phase $\theta_C(\beta)$ can be chosen in two ways – by taking signs $+$ or $-$. Hence, the representation of contexts by complex amplitudes is not uniquely defined by the formula (2.3). In general each trigonometric context can be represented by four complex probability amplitudes based on the same formula (2.1). Suppose that we have chosen two fixed probabilistic angles:

Choice 1: $\theta_C(\beta_1) = \theta_1, \theta_C(\beta_2) = \theta_2$.

Then we can also make following choices:

Choice 2: $\theta_C(\beta_1) = -\theta_1, \theta_C(\beta_2) = \theta_2$,

Choice 3: $\theta_C(\beta_1) = \theta_1, \theta_C(\beta_2) = -\theta_2$,

Choice 4: $\theta_C(\beta_1) = -\theta_1, \theta_C(\beta_2) = -\theta_2$.

For each of the complex probability amplitudes corresponding to these choices of the phase, we obtain the Born's rule. Denote the corresponding amplitudes by the symbols $\psi_{C,j}$, $j = 1, \ldots, 4$. Then $\overline{\psi_{C,1}} = \psi_{C,4}$ as well as $\overline{\psi_{C,2}} = \psi_{C,3}$. Thus

[1] I was even thinking to get a patent for QLRA, but then I decided that QLRA should be free for all users.

the choices one and four are conjugately-equivalent as well as the choices two and three. In future we shall see that only the choices one and four induce a natural QL-representation. But, at the moment we just fix, for any trigonometric context C, one of four possible probability amplitudes.

We denote the space of functions $\psi : X_b \to \mathbb{C}$ by the symbol

$$E_b \equiv \Phi(X_b, \mathbb{C}),$$

where $\mathbb{C}$ is the field of complex numbers. Since the observable b takes only two values – $X_b = \{\beta_1, \beta_2\}$, the E_b is the two-dimensional complex linear space. The Dirac's δ-functions $\{\delta(\beta - \beta_1), \delta(\beta - \beta_2)\}$ form the canonical basis in this space. Each $\psi \in E_b$ can be expanded with respect to this basis:

$$\psi(\beta) = \psi(\beta_1)\delta(\beta - \beta_1) + \psi(\beta_2)\delta(\beta - \beta_2).$$

By using the representation (2.3) we construct the map

$$J^{b/a} : \mathcal{C}^{\mathrm{tr}} \to E_b. \tag{2.4}$$

The $J^{b/a}$ maps contexts (complexes of, e.g., physical conditions) into complex amplitudes. The representation (2.2) of probability as the square of the absolute value of the complex b/a-amplitude is nothing else than the famous *Born rule.*

Remark 2.2 (Role of reference observables). We repeat that the complex linear space representation (2.3) of the set of contexts $\mathcal{C}^{\mathrm{tr}}$ is based on a pair (a, b) of *statistically conjugate* observables. By choosing another pair we shall get a different representation.

Remark 2.3 (Origin of complex numbers). The appearance of complex numbers is one of mysteries of QM. In our contextual approach complex numbers appeared as just a special representation of the formula of total probability with trigonometric interference term. Instead of the formula (2.1) and a collection of contextual probabilities, we preferred to work with Born's rule (2.2) and complex probability amplitudes.

Definition 2.1. The complex amplitude $\psi_C(\beta)$ produced by QLRA is called a QL *wave function* (of the complex of physical conditions, context C) or a QL *state*.

Thus by the Växjö interpretation the wave function does not provide the (complete) description of the state of an individual system. The wave function is a special (incomplete) representation of context.

In fact, the multi-value structure of QLRA is even more complicated than it was pointed out in Remark 2.1. We might represent each context $C \in \mathcal{C}^{\mathrm{tr}}$ by a family of complex amplitudes:

$$\psi(x) \equiv \psi_C(\beta) = \sum_{\alpha \in X_a} \sqrt{p_C^a(\alpha) p^{b/a}(\beta/\alpha)} e^{i\xi_C(\beta/\alpha)} \tag{2.5}$$

such that

$$\xi_C(\beta/\alpha_1) - \xi_C(\beta/\alpha_2) = \theta_C(\beta).$$

For such complex amplitudes we also have Born's rule (2.2). Thus QLRA could be realized as a map from the set of trigonometric contexts into $\tilde{S}$ (the set of equivalent classes with respect to the equivalent relation: $\phi = e^{iu}\psi$). However, to simplify considerations we shall consider only the representation (2.3) and the map (2.4) induced by this representation.

3 Hilbert space representation of *b*-observable

3.1 Born's rule

We consider the following basis in the space E_b ($\beta \in X_b$):

$$e^b_\beta(x) = \delta(x - \beta).$$

The Born's rule for complex amplitudes (2.2) can be rewritten in the following form:

$$p^b_C(\beta) = |\langle \psi_C, e^b_\beta \rangle|^2, \quad \beta \in X_b. \tag{3.1}$$

Here the scalar product in the complex linear space E_b is defined by the standard formula:

$$\langle \psi_1, \psi_2 \rangle = \sum_{\beta \in X_b} \psi_1(\beta)\overline{\psi_2(\beta)}. \tag{3.2}$$

The system of functions $\{e^b_\beta\}_{x \in X_b}$ is an orthonormal basis in the Hilbert space

$$\mathcal{H} \equiv \mathcal{H}^{b/a}_b = (E_b, \langle \cdot, \cdot \rangle). \tag{3.3}$$

In the symbolic notation $\mathcal{H}^{b/a}_b$ the upper index b/a notices that the representation of contexts was created through b/a-supplementarity, the low index b notices that the observable b plays the fundamental role. Contexts are represented by functions defined on the range of values X_b of the observable b. We emphasize that the reference observables a and b are involved into the $\mathcal{H}^{b/a}_b$-representation in the asymmetric way.

3.2 Fundamental physical observable: views of De Broglie and Bohm

L. De Broglie permanently emphasized [49], [48] the exceptional role of the *position observable* q in QM. Similar views can be found in works of D. Bohm. In our approach the b-observable plays the role of the position observable q. The a-observable (which is chosen as statistically conjugate to b) plays the role of the momentum observable p. L. De Broglie wrote in [49], p. 55: "A consideration of foregoing examples, and others which we could imagine, necessary leads us to the conclusion that it is the representation in space and time which is objective, and not the Fourier analysis which only exists in the mind of the theoretician. The various Fourier components can only be observed by means of devices which change completely the initial state of affairs and modify the phase relationships. In the language of the Theory of Transformations, this can be expressed by saying that the q-representation is the only objective representation, whilst the p-representation – the abstract representation in momentum space – exists only in the mind of the theoretician. This shows, contrary to what the Theory of Transformations usually asserts, that the two representations – q and p – are by no means equivalent. It is the wave function that describes the physical reality and not the coefficients c_i considered separately. Moreover, this conclusion is the consequence of the obvious fact that the three-dimensional space is a physical reality and the essential framework of our experiment, whilst momentum space is only an abstract mathematical representation." Here "Theory of Transformations" is theory of unitary transformations performing transition from one orthonormal basis in complex Hilbert space to another.

3.3 *b*-observable as multiplication operator

Let $X_b \subset \mathbb{R}$ (so we assume that the results of measurements of the observable b are given by real numbers). By using the Hilbert space representation (3.1) of the Born's rule we obtain the Hilbert space representation of the expectation of the observable b given by the von Mises frequency model for the collective (or S-sequence) $x(b/C)$. Here $x(b/C)$ is induced by measurements of the b-observable under the context C. We also remark that this expectation can be considered as the Kolmogorov measure-theoretic expectation with respect to the probability measure $\mathbf{P}_{x(b/C)}$ corresponding to the collective (or the S-sequence) $x(b/C)$. We have:

$$\begin{aligned} E[b/C] &= \sum_{\beta \in X_b} \beta p_C^b(\beta) = \sum_{\beta \in X_b} \beta |\psi_C(\beta)|^2 \\ &= \sum_{\beta \in X_b} \beta \langle \psi_C, e_\beta^b \rangle \overline{\langle \psi_C, e_\beta^b \rangle} = \langle \hat{b}\psi_C, \psi_C \rangle. \end{aligned} \tag{3.4}$$

Here the (self-adjoint) operator $\hat{b} : \mathcal{H} \to \mathcal{H}$ is determined by its eigenvectors:

$$\hat{b} e_\beta^b = \beta e_\beta^b, \quad \beta \in X_b.$$

This is the multiplication operator in the linear space of complex-valued functions $E_b = \Phi(X_b, \mathbf{C})$:

$$\hat{b}\psi(\beta) = \beta\psi(\beta).$$

It is natural to represent the b-observable by the operator $\hat{b}$ in the Hilbert state space.

3.4 Interference

We set

$$u_j^a = \sqrt{p_C^a(\alpha_j)}, \qquad u_j^b = \sqrt{p_C^b(\beta_j)}, \qquad p_{ij}^{b/a} = p^{b/a}(\beta_j/\alpha_i),$$
$$u_{ij} = \sqrt{p_{ij}^{b/a}}, \qquad \theta_j = \theta_C(\beta_j). \tag{3.5}$$

We remark that the coefficients u_j^a, u_j^b depend on a context C; so

$$u_j^a = u_j^a(C), \quad u_j^b = u_j^b(C).$$

We consider the matrix $\mathbf{P}^{b/a} = (p_{ij}^{b/a})$. We have, see (2.5), that

$$\psi_C = v_1^b e_{\beta_1}^b + v_2^b e_{\beta_2}^b, \quad \text{where} \quad v_j^b = u_1^a u_{1j} + u_2^a u_{2j} e^{i\theta_j}. \tag{3.6}$$

Hence

$$p_C^b(\beta_j) = |v_j^b|^2 = |u_1^a u_{1j} + u_2^a u_{2j} e^{i\theta_j}|^2. \tag{3.7}$$

This is the *interference representation of probabilities* that is used, e.g., in the quantum formalism.[2]

4 Hilbert space representation of a-observable

We would like to have the Born's rule and the Hilbert space representation not only for the b-observable, but also for the a-observable. Therefore we should introduce in a natural way a basis $e^a = \{e_\alpha^a\}_{\alpha \in X_a}$ corresponding to the a-observable in the Hilbert space $\mathcal{H}_b^{b/a}$. The formula of interference of probabilities written in the form of superposition (3.6) will play the crucial role in introduction the a-basis in $\mathcal{H}_b^{b/a}$.

4.1 Conventional quantum and quantum-like representations

As we shall see, we cannot be lucky in the general case. In fact, by starting from two arbitrary (statistically conjugate) observables a and b we constructed the complex Hilbert space representation for the b-observable which was more general than the standard quantum representation. In our (more general) representation the "dual observable" a need not be mapped into a symmetric operator in the same Hilbert space $\mathcal{H}_b^{b/a}$ which was generated by the b/a-conditioning of the b-observable, see Section 6.[3] We recall that in quantum mechanics both reference observables (the position

[2]By starting with the general representation (2.5) we obtain $v_j^b = u_1^a u_{1j} e^{i\xi_{1j}} + u_2^a u_{2j} e^{i\xi_{2j}}$ and the interference representation $p_C^b(\beta_j) = |v_j^b|^2 = |u_1^a u_{1j} e^{i\xi_{1j}} + u_2^a u_{2j} e^{i\xi_{2j}}|^2$.

[3]We constructed the Hilbert space $\mathcal{H}_b^{b/a}$ by introducing the natural scalar product on the space $E_b = \Phi(X_b, \mathbb{C})$ of functions $\psi : X_b \to \mathbb{C}$. Here X_b is the range of values of b.

observable and the momentum observable) are represented in the same Hilbert space. Thus we should find conditions which would guarantee the possibility to represent both reference observables in the same Hilbert space $\mathcal{H}_b^{b/a}$ and to have the Born's rule for both of them.

We recall that QM implies that for any pair of observables a and b which are represented by self-adjoint operators $\hat{a}$ and $\hat{b}$ the corresponding transition probabilities are symmetrically conditioned, see Chapter 6: $p^{b/a}(\beta/\alpha) = p^{a/b}(\alpha/\beta)$. In the Växjö model we do not assume that in general probabilities are symmetrically conditioned. It can be that for some probabilistic data which was collected e.g. in sociology or psychology we shall have:

$$p^{b/a}(\beta/\alpha) \equiv \mathbf{P}(b = \beta/C_\alpha) \neq p^{a/b}(\alpha/\beta) \equiv \mathbf{P}(a = \alpha/C_\beta).$$

Here C_α and C_β are contexts corresponding to selections with respect to fixed values $a = \alpha$ and $b = \beta$. Really there are no reasons to expect symmetrical conditioning for any pair of reference observables. In QM the reference observables were chosen in very special way to guarantee symmetrical conditioning. Nevertheless, starting with the Växjö model we can apply QLRA for any pair of statistically conjugate observables and represent the set of contexts $\mathcal{C}^{\mathrm{tr}}_{b/a,a}$ in the complex Hilbert space $\mathcal{H}_b^{b/a}$ where the b-observable is represented by a self-adjoint operator and Born's rule holds.

We now would like to find conditions for the matrix $\mathbf{P}^{b/a}$ which would provide a possibility to represent the a-observable by a self-adjoint operator in such a way that the Born's rule would hold. The easiest way is to borrow the QM-condition, namely, symmetric conditioning. We shall see that this condition is really sufficient to construct representation for which we are looking for, see Section 8. However, we do not borrow directly the QM-condition. We would like to derive a natural necessary and sufficient condition internally from the Växjö model. We shall see that this condition is essentially weaker than symmetric conditioning.

4.2 *a*-basis from interference

For any (trigonometric) context C_0, we can represent (by using the expansion (3.6)) the corresponding wave function $\psi = \psi_{C_0}$ in the form:

$$\psi = u_1^a e_{\alpha_1}^a + u_2^a e_{\alpha_2}^a, \tag{4.1}$$

where

$$e_{\alpha_1}^a = \begin{pmatrix} u_{11} \\ u_{12} \end{pmatrix}, \quad e_{\alpha_2}^a = \begin{pmatrix} e^{i\theta_1} u_{21} \\ e^{i\theta_2} u_{22} \end{pmatrix}. \tag{4.2}$$

The system of vectors $\{e_{\alpha_i}^a\}$ will be used to represent the a-observable in the Hilbert space $\mathcal{H} \equiv \mathcal{H}_b^{b/a}$.[4]

[4] We recall that our construction is not symmetric with respect to the reference observables a and b. The observable b is the "fundamental observable" and a is its dual, cf. views of De Broglie and Bohm.

We suppose that vectors $\{e^a_{\alpha_i}\}$ are linearly independent, so $\{e^a_{\alpha_i}\}$ is a basis in $\mathcal{H}$. We have:

$$e^a_{\alpha_1} = v_{11} e^b_{\beta_1} + v_{12} e^b_{\beta_2}, \quad e^a_{\alpha_2} = v_{21} e^b_{\beta_1} + v_{22} e^b_{\beta_2}$$

Here

$$V \equiv V^{b/a}_{b\to a} = (v_{ij}) = \begin{pmatrix} u_{11} & u_{12} \\ e^{i\theta_1} u_{21} & e^{i\theta_2} u_{22} \end{pmatrix},$$

is the matrix corresponding to the transformation from the b-basis to the a-basis:

$$\begin{pmatrix} e^a_{\alpha_1} \\ e^a_{\alpha_2} \end{pmatrix} = V \begin{pmatrix} e^b_{\beta_1} \\ e^b_{\beta_2} \end{pmatrix}.$$

It is interesting to notice the expressions of the scalar products for basis vectors:

$$\begin{aligned} V = (v_{ij}) &= \begin{pmatrix} \langle e^a_{\alpha_1}, e^b_{\beta_1}\rangle & \langle e^a_{\alpha_1}, e^b_{\beta_2}\rangle \\ \langle e^a_{\alpha_2}, e^b_{\beta_1}\rangle & \langle e^a_{\alpha_2}, e^b_{\beta_2}\rangle \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{p^{b/a}(\beta_1/\alpha_1)} & \sqrt{p^{b/a}(\beta_2/\alpha_1)} \\ e^{i\theta_1}\sqrt{p^{b/a}(\beta_1/\alpha_2)} & e^{i\theta_2}\sqrt{p^{b/a}(\beta_2/\alpha_2)} \end{pmatrix}. \end{aligned} \tag{4.3}$$

The crucial point is that the matrix $V^{b/a}_{b\to a}$ can be constructed by using only probabilistic data, even phases are purely probabilistic phases.

4.3 Necessary and sufficient conditions for Born's rule

We would like to find a class of matrices $V^{b/a}_{b\to a}$ such that the Born's rule also holds for the a-basis:

$$p^a_C(\alpha) = |\langle \psi_C, e^a_\alpha\rangle|^2. \tag{4.4}$$

We have the Born's rule (4.4) iff $\{e^a_\alpha\}$ was an *orthonormal basis,* i.e., the $V^{b/a}_{b\to a}$ was a *unitary* matrix.

Remark 4.1 (On the origin of unitarity). In our model *unitarity* appeared as a consequence of the Born's rule. To construct a representation in which the Born's rule holds for both reference observables a and b, we should choose these observables in such a way that the matrix $V^{b/a}_{b\to a}$ will be unitary.

We recall that a matrix $P = (p_{ij})$ is called *doubly stochastic* if it is stochastic, i.e., $p^{b/a}_{j1} + p^{b/a}_{j2} = 1$, and, moreover,

$$p^{b/a}_{1j} + p^{b/a}_{2j} = 1, \quad j = 1, 2. \tag{4.5}$$

We remark that any matrix of transition probabilities $\mathbf{P}^{b/a}$ is stochastic:

$$p^{b/a}(b = \beta_1/\alpha_1) + p^{b/a}(b = \beta_2/\alpha_1) = 1,$$

$$p^{b/a}(b = \beta_1/\alpha_2) + p^{b/a}(b = \beta_2/\alpha_2) = 1$$

(these are simply consequences of additivity of contextual probabilities $\mathbf{P}(\cdot/C_\alpha)$). But in general it is not doubly stochastic. Thus the condition of double stochasticity

$$p^{b/a}(b=\beta_1/\alpha_1)+p^{b/a}(b=\beta_1/\alpha_2)=1, \tag{4.6}$$

$$p^{b/a}(b=\beta_2/\alpha_1)+p^{b/a}(b=\beta_2/\alpha_2)=1 \tag{4.7}$$

could be violated. We recall, see Chapter 6, that any symmetrically conditioned matrix $\mathbf{P}^{b/a}$ is doubly stochastic. But in general double stochasticity does not imply symmetrical conditioning. The following proposition can be easily proved by direct calculations.

Proposition 4.1. *In the two-dimensional case (i.e., for dichotomous observables), the matrix $V_{b\to a}^{b/a}$ is unitary iff the matrix $\mathbf{P}^{b/a}$ is doubly stochastic and additionally:*

$$e^{i\theta_1}=-e^{i\theta_2}$$

or

$$\theta_{C_0}(\beta_1)-\theta_{C_0}(\beta_2)=\pi \mod 2\pi. \tag{4.8}$$

We remark that the constraints (4.8) on phases and the double stochasticity constraint (4.6) are not independent:

Lemma 4.1. *Let the matrix of transition probabilities $\mathbf{P}^{b/a}$ be doubly stochastic. Then:*

$$\cos\theta_C(\beta_2)=-\cos\theta_C(\beta_1) \tag{4.9}$$

for any context $C\in\mathcal{C}^{\mathrm{tr}}$.

Proof. By Lemma 4.1, Chapter 6, we have:

$$\sum_{\beta\in X_b}\cos\theta_C(\beta)\sqrt{\Pi_{\alpha\in X_a}p_C^a(\alpha)p^{b/a}(\beta/\alpha)}=0.$$

But, for a doubly stochastic matrix $\mathbf{P}^{b/a}=(p^{b/a}(\alpha/\beta))$, we have:

$$\Pi_{\alpha\in X_a}p_C^a(\alpha_1)p^{b/a}(\beta_1/\alpha)=\Pi_{\alpha\in X_a}p_C^a(\alpha_2)p^{b/a}(\beta_2/\alpha).$$

Since we work with statistically conjugate reference observables a,b and the context C is a-nondegenerate, all probabilities are strictly positive. Therefore we obtain (4.9). □

4.4 Choice of probabilistic phases

By Lemma 4.1 we have two different possibilities to choose phases:

$$\theta_{C_0}(\beta_1) + \theta_{C_0}(\beta_2) = \pi \quad \text{or} \quad \theta_{C_0}(\beta_1) - \theta_{C_0}(\beta_2) = \pi \bmod 2\pi.$$

By (4.8) to obtain the Born's rule for the a-observable we should choose phases $\theta_{C_0}(\beta_i)$, $i = 1, 2$, in such a way that

$$\theta_{C_0}(\beta_2) = \theta_{C_0}(\beta_1) + \pi. \tag{4.10}$$

If $\theta_{C_0}(\beta_1) \in [0, \pi]$ then $\theta_{C_0}(\beta_2) \in [\pi, 2\pi]$ and vice versa.

Lemma 4.1 is very important: if the matrix $\mathbf{P}^{b/a}$ is doubly stochastic we can always choose $\theta_{C_0}(\beta_j)$, $j = 1, 2$, satisfying (4.10). Hence we can always assume that QLRA produces complex amplitudes of the form:

$$\psi(\beta_1) = \sqrt{p_C^a(\alpha_1) p^{b/a}(\beta_1/\alpha_1)} + e^{i\theta_C(\beta_1)} \sqrt{p_C^a(\alpha_2) p^{b/a}(\beta_1/\alpha_2)}, \tag{4.11}$$

$$\psi(\beta) = \sqrt{p_C^a(\alpha_1) p^{b/a}(\beta_2/\alpha_1)} - e^{i\theta_C(\beta_1)} \sqrt{p_C^a(\alpha_2) p^{b/a}(\beta_2/\alpha_2)}. \tag{4.12}$$

We now come back to Remark 2.1. Suppose that, for a context C_0, we have chosen the phases $\theta_{C_0}(\beta_1)$ and $\theta_{C_0}(\beta_2)$ satisfying (4.10). We denote this representation of C_0 by a complex probability amplitude by Choice 1. In Remark 2.1 we presented three other choices created by varying the signs of phases. We see that only phases given by the Choice 4 also satisfy the condition (4.10). Thus if we want to have a natural representation (with the Born's rule for both reference observables) we should take away Choices 2, 3. So the arbitrariness in choosing a complex amplitude for a fixed context is essentially reduced.

4.5 Contextual dependence of a-basis

The delicate feature of the presented construction of the a-representation is that the basis $\{e_\alpha^a\}$ depends on the context C_0:

$$e_\alpha^a = e_\alpha^a(C_0).$$

And the Born's rule, in fact, has the form:

$$p_{C_0}^a(\alpha) = |\langle \psi_{C_0}, e_\alpha^a(C_0) \rangle|^2, \quad \alpha \in X_a.$$

We would like to use (as in the conventional quantum formalism) one fixed a-basis for all contexts $C \in \mathcal{C}^{\text{tr}}$. We may try to use for all contexts $C \in \mathcal{C}^{\text{tr}}$ the basis $e_\alpha^a \equiv e_\alpha^a(C_0)$ corresponding to one fixed context C_0. We shall see that this is really the fruitful strategy.

Lemma 4.2. *Let the matrix of transition probabilities* $\mathbf{P}^{b/a}$ *be doubly stochastic and let for any context* $C \in \mathcal{C}^{\rm tr}$ *phases* $\theta_C(\beta)$, $\beta \in X_b$, *be chosen as*

$$\theta_C(\beta_2) = \theta_C(\beta_1) + \pi \bmod 2\pi. \tag{4.13}$$

Then for any context $C \in \mathcal{C}^{\rm tr}$ *we have the Born's rule for the basis* $e^a_\alpha \equiv e^a_\alpha(C_0)$ *constructed for a fixed context* C_0*:*

$$p^a_C(\alpha) = |\langle \psi_C, e^a_\alpha \rangle|^2, \ \alpha \in X_a. \tag{4.14}$$

Proof. Let C_0 be some fixed context. We take the basic $\{e^a_{\alpha_j}(C_0)\}$ (and the matrix $V(C_0)$) corresponding to this context. For any $C \in \mathcal{C}^{\rm tr}$, we would like to represent the wave function ψ_C as

$$\psi_C = v^a_1(C) e^a_{\alpha_1}(C_0) + v^a_2(C) e^a_{\alpha_2}(C_0), \quad \text{where } |v^a_j(C)|^2 = p^a_C(\alpha_j). \tag{4.15}$$

It is clear that, for any $C \in \mathcal{C}^{\rm tr}$, we can represent the wave function as

$$\psi_C(\beta_1) = u^a_1(C) v_{11}(C_0) + e^{i[\theta_C(\beta_1) - \theta_{C_0}(\beta_1)]} u^a_2(C) v_{12}(C_0),$$
$$\psi_C(\beta_2) = u^a_1(C) v_{21}(C_0) + e^{i[\theta_C(\beta_2) - \theta_{C_0}(\beta_2)]} u^a_2(C) v_{22}(C_0).$$

Thus to obtain (4.15) we should have:

$$\theta_C(\beta_1) - \theta_{C_0}(\beta_1) = \theta_C(\beta_2) - \theta_{C_0}(\beta_2) \bmod 2\pi \tag{4.16}$$

for any pair of contexts C_0 and C_1. By using the relations (4.13) between phases $\theta_C(\beta_1)$, $\theta_C(\beta_2)$ and $\theta_{C_0}(\beta_1)$, $\theta_{C_0}(\beta_2)$ we obtain:

$$\theta_C(\beta_2) - \theta_{C_0}(\beta_2) = (\theta_C(\beta_1) + \pi - \theta_{C_0}(\beta_1) - \pi) = \theta_C(\beta_1) - \theta_{C_0}(\beta_1) \bmod 2\pi. \quad \square$$

The constraint (4.13) essentially restricted the class of complex amplitudes which can be used to represent a context $C \in \mathcal{C}^{\rm tr}$. Any C can be represented only by two amplitudes $\psi(x)$ and $\bar{\psi}(x)$ corresponding to the two possible choices of $\theta_C(\beta_1)$: in $[0, \pi]$ or $(\pi, 2\pi)$.

4.6 Existence of quantum-like representation with Born's rule for both reference observables

By Lemma 4.2 we obtain the following result playing the fundamental role in our approach:

Theorem 4.1. *We can construct the QL (complex Hilbert space) representation of the set of trigonometric contexts* $\mathcal{C}^{\rm tr}$ *such that the Born's rule holds true for both reference observables* a, b *(which are assumed to be statistically conjugate) iff the matrix* $\mathbf{P}^{b/a}$ *is doubly stochastic.*

If the matrix $\mathbf{P}^{b/a}$ is doubly stochastic, then we have the QL representation not only for the conditional expectation of the observable b, see (3.4), but also for the observable a:

$$\begin{aligned} E[a/C] &= \sum_{\alpha \in X_a} \alpha p_C^a(\alpha) \\ &= \sum_{\alpha \in X_a} \alpha |\langle \psi_C, e_\alpha^a \rangle|^2 = \langle \hat{a}\psi_C, \psi_C \rangle, \end{aligned} \tag{4.17}$$

where the self-adjoint operator (symmetric matrix) $\hat{a} : E_b \to E_b$ is determined by its eigenvectors:

$$\hat{a} e_\alpha^a = \alpha e_\alpha^a.$$

By (4.17) it is natural to represent the observable a by the operator $\hat{a}$.

We also remark that in the case of doubly stochastic $\mathbf{P}^{b/a}$ the scalar products for basis vectors have the form:

$$\begin{aligned} v_{11} &= \langle e_{\alpha_1}^a, e_{\beta_1}^b \rangle = \sqrt{p^{b/a}(\beta_1/\alpha_1)}, \\ v_{12} &= \langle e_{\alpha_1}^a, e_{\beta_2}^b \rangle = \sqrt{p^{b/a}(\beta_2/\alpha_1)}, \\ v_{21} &= \langle e_{\alpha_2}^a, e_{\beta_1}^b \rangle = e^{i\theta} \sqrt{p^{b/a}(\beta_1/\alpha_2)}, \\ v_{22} &= \langle e_{\alpha_2}^a, e_{\beta_2}^b \rangle = -e^{i\theta} \sqrt{p^{b/a}(\beta_2/\alpha_2)}. \end{aligned}$$

As always, we denote the unit sphere in the Hilbert space $\mathcal{H}$ by the symbol S. In general, i.e., for an arbitrary contextual statistical model of reality, there are no reasons to expect that the representation map $J^{b/a} : \mathcal{C}^{\mathrm{tr}} \to S$ should to be one-to-one, i.e., surjection and injection. We shall study the question about injectivity of the map $J^{b/a}$ in Section 5.2.

Regarding surjectivity we can say that in principle in some physical (or mental, or economic) model the set of context $\mathcal{C}$ may be not large enough to cover the whole unit sphere S of the complex Hilbert space. However, in the conventional quantum model it is claimed that each quantum state can be prepared on the basis of some complex of physical conditions.

4.7 "Pathologies"

We remark that, although Theorem 4.1 guarantees existence of the QL representation with Born's rule for both reference observables, this representation may have features which differ essentially from features of the conventional quantum representation.

In particular, contexts C_β need not belong to the family of trigonometric contexts. In such a case, although the scalar product $\langle e_\beta^b, e_\alpha^a \rangle$ is well defined and, moreover, $|\langle e_\beta^b, e_\alpha^a \rangle|^2 = p^{b/a}(\beta/\alpha)$, we cannot write the Born's rule in our contextual form:

$|\langle\psi_{C_\beta}, e^a_\alpha\rangle|^2 = p^{b/a}(\beta/\alpha)$ (because QLRA cannot be applied to the context C_β). In principle, one might just formally extend the domain of application of QLRA by setting

$$J^{b/a}(C_\beta) = e^b_\beta. \tag{4.18}$$

Situation is even worse for contexts C_α, $\alpha \in X_a$. They are not a-nondegenerate: $\mathbf{P}(a = \alpha_i/C_{\alpha_j}) = 0$, $i \neq j$. Therefore QLRA cannot be applied to C_α. Thus the image $J^{b/a}(C_\alpha)$ cannot be defined by (2.3). In principle, one might just formally extend the domain of application of QLRA by setting

$$J^{b/a}(C_\alpha) = e^a_\alpha. \tag{4.19}$$

However, if the model is not symmetrically conditioned, then such a definition would imply the following pathology:

$$|\langle J^{b/a}(C_\alpha), e^b_\beta\rangle|^2 = |\langle e^a_\alpha, e^b_\beta\rangle|^2 = |\langle e^b_\beta, e^a_\alpha\rangle|^2 = p^{b/a}(\beta/\alpha).$$

Hence:

$$|\langle J^{b/a}(C_\alpha), e^b_\beta\rangle|^2 \neq p^{a/b}(\alpha/\beta). \tag{4.20}$$

It is clear that the latter problem would disappear if one considers only Växjö models with symmetrically conditioned reference observables. It is surprising that such a restriction would also eliminate the problem with non-trigonometrical behavior of contexts C_β, see Section 8.

5 Properties of mapping of trigonometric contexts into complex amplitudes

5.1 Classical-like contexts

Suppose that, for some context $C \in \mathcal{C}^{\text{tr}}$, the reference observables are not b/a-supplementary with respect to C. Thus:

$$\delta(\beta/a, C) = 0, \quad \beta \in X_b.$$

Thus even

$$\lambda(\beta/a, C) = 0, \quad \beta \in X_b.$$

Hence: $\theta_C(\beta_1) = \frac{\pi}{2}$ or $\theta_C(\beta_1) = \frac{3}{2}\pi$. In the first case we have

$$\psi_C(\beta_1) = \sqrt{p^a_C(\alpha_1)p(\beta_1/\alpha_1)} + i\sqrt{p^a_C(\alpha_2)p(\beta_1/\alpha_2)},$$
$$\psi_C(\beta_2) = \sqrt{p^a_C(\alpha_1)p(\beta_2/\alpha_1)} - i\sqrt{p^a_C(\alpha_2)p(\beta_2/\alpha_2)}. \tag{5.1}$$

The second choice of phases gives the representation of C by the complex amplitude ϕ_C which is conjugate to (5.1): $\phi_C = \overline{\psi_C}$. We set

$$\mathcal{C}^{\mathrm{tr}}_{\mathrm{CL}} = \{C \in \mathcal{C}^{\mathrm{tr}} : \delta(\beta/a, C) = 0\}.$$

These are trigonometric contexts for which the reference observables are not b/a-supplementary. We call them CL-contexts.

5.2 Non-injectivity of representation map

Let $C_1, C_2 \in \mathcal{C}^{\mathrm{tr}}$ be contexts such that the probability distributions of the reference observables a and b under C_1 and C_2 coincide:

$$p^a_{C_1}(\alpha) = p^a_{C_2}(\alpha), \alpha \in X_a, \quad p^b_{C_1}(\beta) = p^b_{C_2}(\beta), \beta \in X_b.$$

In such a case $\lambda(\beta/a, C_1) = \lambda(\beta/a, C_2)$ and $\theta(\beta/a, C_1) = \pm\theta(\beta/a, C_2)$. If the probability distributions coincide only for a pair of contexts (C_1, C_2), then we can represent C_1 and C_2 by two different complex amplitudes, ψ_{C_1} and $\psi_{C_2} = \bar{\psi}_{C_1}$. But if the probability distributions coincide for a triple of contexts (C_1, C_2, C_3), then it is impossible to represent them by different complex amplitudes. We should choose $\psi_{C_3} = \psi_{C_1}$ or $\psi_{C_3} = \psi_{C_2}$; so $J^{b/a}(C_3) = J^{b/a}(C_1)$ or $J^{b/a}(C_3) = J^{b/a}(C_2)$. Thus in general the map $J^{b/a}$ is not injective.

6 Non-doubly stochastic matrix: quantum-like representations

Of course, for arbitrary (statistically conjugate) observables a and b the matrix $\mathbf{P}^{b/a}$ need not be doubly stochastic. Therefore the matrix $V^{b/a}_{b\to a}$ for transition from the b-observable (which could be interpreted, cf. De Broglie and Bohm in Section 3.2, as the fundamental observable) to the “dual observable” a can be nonunitary. In this case we could not obtain Born’s rule in the Hilbert space $\mathcal{H}^{b/a}_b$ both for the b and a observables.

We now assume that the $\mathbf{P}^{b/a}$ is not doubly stochastic. For each reference observable we should introduce its own scalar product and corresponding Hilbert space in that the Born’s rule holds true:

$$\mathcal{H}^{b/a}_b = (E_b, \langle\cdot, \quad \cdot\rangle_b), \mathcal{H}^{b/a}_a = (E_b, \langle\cdot,\cdot\rangle_a), \tag{6.1}$$

where scalar products on the complex linear spaces E_b are given by

$$\langle\psi, \psi\rangle_b = \sum_j v^b_j \bar{w}^b_j \quad \text{for} \quad \psi = \sum_j v^b_j e^b_{\beta_j}, \psi = \sum_j w_j e^b_{\beta_j},$$

and

$$\langle \psi, \psi \rangle_a = \sum_j v_j^a \bar{w}_j^a \quad \text{for} \quad \psi = \sum_j v_j^a e_{\alpha_j}^a, \psi = \sum_j w_j^a e_{\alpha_j}^a.$$

We have Born's rules with respect to these scalar products:

$$p_C^b(\beta) = |\langle \psi_C, e_\beta^b \rangle_b|^2, \quad p_C^a(\alpha) = |\langle \psi_C, e_\alpha^a \rangle_a|^2.$$

The reference observables b and a are represented by symmetric matrices in the Hilbert spaces $\mathcal{H}_b^{b/a}$ and $\mathcal{H}_a^{b/a}$, respectively. Thus we do not have even mathematical equivalence (in the sense of unitary equivalence) of representations of a and b, cf. with the discussion on physical nonequivalence for the position and momentum representations in QM, cf. Section 3.2. But the appearance of different Hilbert spaces (6.1) is not the end of mathematical difficulties in the case in that the $\mathbf{P}^{b/a}$ is not doubly stochastic.

As we have already discussed, the crucial difficulty is that $e_\alpha^a = e_\alpha^a(C_0)$. In fact, for any context $C_0 \in \mathcal{C}^{\text{tr}}$ we constructed its own Hilbert space representation for the a-observable: $\mathcal{H}_a^{b/a} = \mathcal{H}_a^{b/a}(C_0)$. In the same way as in Section 3 we obtain that we would be able to use the same representation for contexts C and C_0 if the condition (4.16) holds true. Thus we should have:

$$\theta_C(\beta_2) = \theta_C(\beta_1) + \alpha \quad \text{and} \quad \theta_{C_0}(\beta_2) = \theta_{C_0}(\beta_1) + \gamma \bmod 2\pi,$$

where γ is some phase (if $\mathbf{P}^{b/a}$ is doubly stochastic then $\gamma = \pi$).

Theorem 6.1. *Suppose that $\mathbf{P}^{b/a}$ is not doubly stochastic and $\mathcal{C}^{\text{tr}} \neq \mathcal{C}_0^{\text{tr}}$. Then there is no such an γ that*

$$\theta_C(\beta_2) = \theta_C(\beta_1) + \gamma \tag{6.2}$$

for all contexts $C \in \mathcal{C}^{\text{tr}}$.

To prove this theorem we need the following generalization of Lemma 4.1 for the case in that the $\mathbf{P}^{b/a}$ is not doubly stochastic:

Lemma 6.1. *For any context $C \in \mathcal{C}^{\text{tr}}$, the following equality holds true:*

$$\cos\theta_C(\beta_2) = -k \cos\theta_C(\beta_1) \tag{6.3}$$

where

$$k \equiv k^{b/a} = \sqrt{\frac{p_{11}^{b/a} p_{21}^{b/a}}{p_{12}^{b/a} p_{22}^{b/a}}}.$$

It is also easy to obtain:

Proposition 6.1. *The coefficient $k^{b/a} = 1$ iff $\mathbf{P}^{b/a}$ is doubly stochastic.*

Proof of Theorem 6.1. By Lemma 6.1 we have: $-k\cos\theta_C(\beta_1) = \cos(\theta_C(\beta_1)+\gamma)$. We take $C = \Omega$ and obtain: $\cos(\theta_\Omega(\beta_1)+\gamma) = 0$. But $\theta_\Omega(\beta_1) = \pm\frac{\pi}{2}$. Thus $\theta_\Omega(\beta_1)+\gamma = \pm\frac{\pi}{2}$ and $\gamma = 0, \pi \mod 2\pi$.

Since $\mathcal{C}^{\text{tr}} \neq \mathcal{C}_0^{\text{tr}}$ there exists a context C such that $\cos\theta_C(\beta_1) \neq 0$. If $\gamma = 0$ then $\cos\theta_C(\beta_1)(k+1) = 0$. This contradicts to the positivity of k. Let $\gamma = \pi$. Then $\cos\theta_C(\beta_1)(k-1) = 0$. Thus $k = 1$. But this implies that $\mathbf{P}^{b/a}$ is doubly stochastic. □

Despite Theorem 6.1, we can still hope that there can be found some extended family of contexts such that (6.2) would hold true for contexts from that family. But it is impossible:

Proposition 6.2. *Let condition (6.2) hold true for two contexts C_1, C_2 such that*

$$|\lambda(\beta_1/a, C_1)| \neq |\lambda(\beta_1/a, C_2)|. \tag{6.4}$$

Then $\mathbf{P}^{b/a}$ is doubly stochastic.

Proof. We set $\theta = \theta_{C_1}(\beta_1)$ and $\theta' = \theta_{C_2}(\beta_1)$. We have: $-k\cos\theta = \cos(\theta+\gamma)$, $-k\cos\theta' = \cos(\theta'+\gamma)$. Thus

$$-k\cos\frac{\theta+\theta'}{2}\cos\frac{\theta-\theta'}{2} = \cos\left(\frac{\theta+\theta'}{2}+\gamma\right)\cos\frac{\theta-\theta'}{2}.$$

By (6.4) we have that $\cos\frac{\theta-\theta'}{2} \neq 0$ and hence $-k\cos\frac{\theta+\theta'}{2} = \cos(\frac{\theta+\theta'}{2}+\gamma)$. We also have

$$k\sin\frac{\theta+\theta'}{2}\sin\frac{\theta-\theta'}{2} = -\sin\left(\frac{\theta+\theta'}{2}+\alpha\right)\sin\frac{\theta-\theta'}{2}.$$

By (6.4) we have that $\sin\frac{\theta-\theta'}{2} \neq 0$ and hence $-k\sin\frac{\theta+\theta'}{2} = \sin(\frac{\theta+\theta'}{2}+\gamma)$. Thus $k^2 = 1$ and hence $k = 1$. Hence the matrix $\mathbf{P}^{b/a}$ is doubly stochastic. □

Thus if $\mathbf{P}^{b/a}$ is not doubly stochastic then each surface $M_t = \{C \in \mathcal{C}^{\text{tr}} : |\lambda(\beta_1/\alpha, C)| = t\}$, $0 \leq t \leq 1$, in the space of contexts is represented in its own Hilbert space $\mathcal{H}_a(t)$.

We remark that such a complicated picture arises only if we represent the dichotomous reference observables in the two-dimensional Hilbert. By using the quantum terminology we can say that we proceed under the assumption that these observables have nondegenerate spectra. By considering Hilbert spaces of higher dimensions we could proceed with matrices of transition probabilities which are not doubly stochastic. However, it would be a more general QL story.

7 Non-commutativity of operators representing observables

Let the matrix of probabilities $\mathbf{P}^{b/a}$ be doubly stochastic. We consider in this section the case of real valued observables. Here the ranges of observables b and a are subsets of $\mathbb{R}$. We set

$$q_1 = \sqrt{p_{11}^{b/a}} = \sqrt{p_{22}^{b/a}}$$

and

$$q_2 = \sqrt{p_{12}^{b/a}} = \sqrt{p_{21}^{b/a}}.$$

Thus the vectors of the a-basis, see (4.2), have the following form:

$$e_{\alpha_1}^a = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}, \quad e_{\alpha_2}^a = \begin{pmatrix} e^{i\theta_1} q_2 \\ e^{i\theta_2} q_1 \end{pmatrix}.$$

Since $\theta_2 = \theta_1 + \pi$, we get

$$e_{\alpha_2}^a = e^{i\theta_2} \begin{pmatrix} -q_2 \\ q_1 \end{pmatrix}.$$

We now find matrices of operators $\hat{a}$ and $\hat{b}$ in the b-representation. The latter one is diagonal. For $\hat{a}$ we have:

$$\hat{a} = V^\star \mathrm{diag}(\alpha_1, \alpha_2) V,$$

where V is the matrix of transition from the b-basis to the a-basis. Thus

$$\hat{a} = \begin{pmatrix} \alpha_1 q_1^2 + \alpha_2 q_2^2 & (\alpha_1 - \alpha_2) q_1 q_2 \\ (\alpha_1 - \alpha_2) q_1 q_2 & \alpha_1 q_2^2 + \alpha_2 q_1^2 \end{pmatrix}.$$

We remark that by varying the matrix $\mathbf{P}^{b/a}$ we can obtain any symmetric matrix with real coefficients. We do not obtain matrices with complex coefficients (as a consequence of the special choice of a- and b-bases).

Hence

$$[\hat{b}, \hat{a}] = \hat{m},$$

where

$$\hat{m} = \begin{pmatrix} 0 & (\alpha_1 - \alpha_2)(\beta_2 - \beta_1) q_1 q_2 \\ (\alpha_1 - \alpha_2)(\beta_2 - \beta_1) q_1 q_2 & 0 \end{pmatrix}.$$

Since $\alpha_1 \neq \alpha_2$, $\beta_1 \neq \beta_2$ and $q_j \neq 0$, we have $\hat{m} \neq 0$.

8 Symmetrically conditioned observables

We recall that in quantum mechanics matrices $\mathbf{P}^{b/a}$ and $\mathbf{P}^{a/b}$ always satisfy the following condition of the interchange symmetry:

$$p^{b/a}(\beta/\alpha) = p^{a/b}(\alpha/\beta). \tag{8.1}$$

This is a consequence of symmetry of the scalar product. We recall that in Chapter 6 we called arbitrary reference observables a and b (i.e., having no direct relation to QM) satisfying this condition *symmetrically conditioned.* We recall that by Lemma 8.2, Chapter 6, symmetrically conditioned reference observables are always *statistically balanced,* i.e., both matrices $\mathbf{P}^{b/a}$ and $\mathbf{P}^{a/b}$ are doubly stochastic. In this section we would like to study some special features of our representation of trigonometric contexts in this case.

8.1 b-selections are trigonometric contexts

Theorem 8.1. *Let the matrix $\mathbf{P}^{b/a}$ be doubly stochastic. The contexts C_β, $\beta \in X_b$, belong to $\mathcal{C}^{\rm tr}$ iff the reference observables a and b are symmetrically conditioned.*

Proof. A). We have

$$\lambda(\beta_2/a, C_{\beta_1}) = -\frac{\mu_1^2 + \mu_2^2}{2\mu_1\mu_2},$$

where $\mu_j = \sqrt{p^a_{C_{\beta_1}}(\alpha_j) p^{b/a}(\beta_2/\alpha_j)}$. So $\lambda(\beta_2/a, C_{\beta_1}) \geq 1$ and we have the trigonometric behavior only in the case $\mu_1 = \mu_2$. Thus:

$$p^a_{C_{\beta_1}}(\alpha_1) p^{b/a}(\beta_2/\alpha_1) = p^a_{C_{\beta_1}}(\alpha_2) p^{b/a}(\beta_2/\alpha_2).$$

In this case $\lambda(\beta_2/a, C_{\beta_1}) = -1$ and hence $\theta(\beta_2/a, C_{\beta_1}) = \pi$, and consequently $\theta(\beta_1/a, C_{\beta_1}) = 0$. We point out that $p^a_{C_\beta}(\alpha) = p^{a/b}(\alpha/\beta)$. Thus we have:

$$p^{a/b}(\alpha_1/\beta_1) p^{b/a}(\beta_2/\alpha_1) = p^{a/b}(\alpha_2/\beta_1) p^{b/a}(\beta_2/\alpha_2). \tag{8.2}$$

In the same way by using conditioning with respect to C_{β_2} we obtain:

$$p^{a/b}(\alpha_1/\beta_2) p^{b/a}(\beta_1/\alpha_1) = p^{a/b}(\alpha_2/\beta_2) p^{b/a}(\beta_1/\alpha_2).$$

By using double stochasticity of $\mathbf{P}^{b/a}$ we can rewrite the last equality as

$$p^{a/b}(\alpha_1/\beta_2) p^{b/a}(\beta_2/\alpha_2) = p^{a/b}(\alpha_2/\beta_2) p^{b/a}(\beta_2/\alpha_1). \tag{8.3}$$

Thus by (8.2) and (8.3) we have:

$$\frac{p^{a/b}(\alpha_1/\beta_2)}{p^{a/b}(\alpha_2/\beta_1)} = \frac{p^{a/b}(\alpha_2/\beta_2)}{p^{a/b}(\alpha_1/\beta_1)}.$$

Hence $p^{a/b}(\alpha_1/\beta_2) = tp^{a/b}(\alpha_2/\beta_1)$ and $p^{a/b}(\alpha_2/\beta_2) = tp^{a/b}(\alpha_1/\beta_1)$, $t > 0$. But $1 = p^{a/b}(\alpha_1/\beta_2) + p^{a/b}(\alpha_2/\beta_2) = t[p^{a/b}(\alpha_2/\beta_1) + p^{a/b}(\alpha_1/\beta_1)] = t$. We proved that the matrix $P^{a/b}$ is also doubly stochastic. Thus the reference observables a and b are statistically balanced. We now prove even more: they are symmetrically conditioned. By the equality (8.2) there exists $k > 0$ such that

$$\frac{p^{a/b}(\alpha_1/\beta_1)}{p^{a/b}(\alpha_2/\beta_1)} = \frac{p^{b/a}(\beta_2/\alpha_2)}{p^{b/a}(\beta_2/\alpha_1)} = k. \tag{8.4}$$

Thus

$$p^{a/b}(\alpha_1/\beta_1) = kp(\alpha_2/\beta_1), \quad p^{b/a}(\beta_2/\alpha_2) = kp^{b/a}(\beta_2/\alpha_1).$$

But $1 = p^{a/b}(\alpha_1/\beta_1) + p^{a/b}(\alpha_2/\beta_1) = (k+1)p^{a/b}(\alpha_2/\beta_1)$ (because $P^{a/b}$ is a stochastic matrix) and $1 = p^{b/a}(\beta_2/\alpha_2) + p^{b/a}(\beta_2/\alpha_1) = (k+1)p^{b/a}(\beta_2/\alpha_1)$ (because $P^{b/a}$ is a doubly stochastic matrix). Thus:

$$p^{a/b}(\alpha_2/\beta_1) = p^{b/a}(\beta_2/\alpha_1) = p^{b/a}(\beta_1/\alpha_2), \tag{8.5}$$

$$p^{a/b}(\alpha_1/\beta_1) = p^{b/a}(\beta_2/\alpha_2) = p^{b/a}(\beta_1/\alpha_1) \tag{8.6}$$

(we have used again that $P^{b/a}$ is doubly stochastic). Finally, by using double stochasticity of $P^{a/b}$ we obtain

$$p^{a/b}(\alpha_1/\beta_2) = p^{a/b}(\alpha_2/\beta_1) = p^{b/a}(\beta_2/\alpha_1), \tag{8.7}$$

$$p^{a/b}(\alpha_2/\beta_2) = p^{a/b}(\alpha_1/\beta_1) = p^{b/a}(\beta_2/\alpha_2). \tag{8.8}$$

B). Now let the reference observables a and b be symmetrically conditioned. Lemma 8.2, Chapter 6, implies that they are statistically balanced. Therefore:

$$\begin{aligned} p^{a/b}(\alpha_1/\beta_1)p^{b/a}(\beta_2/\alpha_1) &= p^{a/b}(\alpha_2/\beta_2)p^{b/a}(\beta_1/\alpha_2) \\ &= p^{a/b}(\alpha_2/\beta_1)p^{b/a}(\beta_2/\alpha_2). \end{aligned} \tag{8.9}$$

Thus we obtained the equality (8.2). It implies that C_{β_1} belongs to $\mathcal{C}^{\mathrm{tr}}$. In the same way we prove that C_{β_2} belongs to $\mathcal{C}^{\mathrm{tr}}$. □

Lemma 8.1. *Let the reference observables a and b be symmetrically conditioned. Then:*

$$\lambda(\beta/a, C_\beta) = 1, \quad \beta \in X_b. \tag{8.10}$$

Proof. Here $\delta(\beta/a, C_\beta) = 1 - p^{b/a}(\beta/\alpha_1)p^{a/b}(\alpha_1/\beta) - p^{b/a}(\beta/\alpha_2)p^{a/b}(\alpha_2/\beta) = 1 - p^{a/b}(\alpha_1/\beta)^2 - p^{a/b}(\alpha_2/\beta)^2 = 2p^{a/b}(\alpha_1/\beta)p^{a/b}(\alpha_2/\beta)$. Hence:

$$\lambda(\beta/a, C_\beta) = \frac{\delta(\beta/a, C_\beta)}{2\sqrt{p_{C_\beta}(\alpha_1)p_{C_\beta}(\alpha_2)p^{b/a}(\beta/\alpha_1)p^{b/a}(\beta/\alpha_2)}}.$$

We now remark that a and b symmetrically conditioned. Thus $\lambda(\beta/a, C_\beta) = 1$. □

By (8.10) we have

$$\lambda(\beta_i/a, C_{\beta_j}) = -1, \quad i \neq j.$$

Thus

$$\theta(\beta_i/a, C_{\beta_i}) = 0 \quad \text{and} \quad \theta(\beta_i/a, C_{\beta_j}) = \pi, \quad i \neq j.$$

Proposition 8.1. *Let the reference observables a and b be symmetrically conditioned. Then*

$$J^{b/a}(C_{\beta_j})(\beta) = \delta(\beta_j - \beta), \beta \in X_b, \quad \textit{and} \quad J^{a/b}(C_{\alpha_j})(\alpha) = \delta(\alpha_j - \alpha), \quad \alpha \in X_a.$$

Proof. Since $\theta(\beta_1/a, C_{\beta_1}) = 0$ we have:

$$\begin{aligned} J^{b/a}(C_{\beta_1})(\beta_1) &= \sqrt{p^{a/b}(\alpha_1/\beta_1)p^{b/a}(\beta_1/\alpha_1)} + e^{i0}\sqrt{p^{a/b}(\alpha_2/\beta_1)p^{b/a}(\beta_1/\alpha_2)} \\ &= p^{a/b}(\alpha_1/\beta_1) + p^{a/b}(\alpha_2/\beta_1) = 1. \end{aligned}$$

Since $\theta(\beta_2/a, C_{\beta_1}) = \pi$ we have

$$\begin{aligned} J^{b/a}(C_{\beta_1})(\beta_2) &= \sqrt{p^{a/b}(\alpha_1/\beta_1)p^{b/a}(\beta_2/\alpha_1)} + e^{i\pi}\sqrt{p^{a/b}(\alpha_2/\beta_1)p^{b/a}(\beta_2/\alpha_2)} \\ &= \sqrt{p^{a/b}(\alpha_1/\beta_1)}(\sqrt{p^{b/a}(\beta_2/\alpha_1} - \sqrt{p^{a/b}(\alpha_2/\beta_1)}) = 0. \end{aligned}$$ □

Thus in the case of symmetrically conditioned reference observables a and b we have:

$$J^{b/a}(C_\beta) = e^b_\beta, \quad \beta \in X_b,$$

and the Born's rule has the form:

$$p^b_C(\beta) = |\langle \psi_C, \psi_{C_\beta} \rangle|^2. \tag{8.11}$$

8.2 Extension of representation map

We can formally extend the map $J^{b/a}$ to contexts C_α by (4.19). We set

$$\overline{\mathcal{C}^{\text{tr}}} = \mathcal{C}^{\text{tr}} \bigcup_{\alpha \in X_a} C_\alpha.$$

Thus we have constructed the Hilbert space representation: $J^{b/a} : \overline{\mathcal{C}}^{\text{tr}} \to S$.

The domain of definition of QLRA contains now the selection contexts for both reference observables.

9 Formalization of the notion of quantum-like representation

We have constructed the QL-representation for a special class of contextual statistical models (Växjö models), namely, for the case $\mathcal{O} = \{a, b\}$. It would be interesting to construct such QL representations for more complicated models, e.g., having larger sets of observables $\mathcal{O}$. We start with formalization of the notion of the QL-representation[5]. As usual, P_γ^u denotes the spectral projector onto the eigenspace of the operator $\widehat{u}$ corresponding to its eigenvalue γ.

Definition 9.1. Let $M = (\mathcal{C}, \mathcal{O}, \mathcal{D}(\mathcal{O}, \mathcal{C}))$ be a contextual statistical model with two fixed observables $a, b \in \mathcal{O}$ – the reference observables. A QL *representation* of this model (corresponding to these reference observables) is defined by a pair of maps with the domains of definition $\mathcal{C}_{J_1} \subset \mathcal{C}$ and $\mathcal{O}_{J_2} \subset \mathcal{O}$, respectively: $J_1 : \mathcal{C}_{J_1} \to \tilde{S}$, and $J_2 : \mathcal{O}_{J_2} \to L_s(\mathcal{H})$. These maps have the following properties:

AV). For any observable $d \in \mathcal{O}_{J_2}$ and any context $C \in \mathcal{C}_{J_1}$ the contextual and quantum averages coincide:

$$E[d/C] = \langle J_2(d) J_1(C), J_1(C) \rangle \tag{9.1}$$

(if $J_1(C)$ belongs to the domain of definition of the operator $J_2(d)$).

RO). Both reference observables $u = a, b$ belong to $\mathcal{O}_{J_2}$, the corresponding selection contexts C_γ^u, $\gamma \in X_u$, belong to $\mathcal{C}_{J_1}$. Moreover,

a) the range of values X_u of the observable u coincides with the spectrum of the corresponding operator $\hat{u}$;
b) the contextual probability distribution coincides with the corresponding quantum probability distribution given by the Born rule:

$$\mathbf{P}(u = \gamma/C) = |P_\gamma^u J_1(C)|^2, \quad C \in \mathcal{C}_{J_1}. \tag{9.2}$$

If the operator $\hat{u} = J_2(u)$ has nondegenerate (purely discrete) spectrum, then

$$P_\gamma^u = J_1(C_\gamma^u) \otimes J_1(C_\gamma^u), \quad \gamma \in X_u. \tag{9.3}$$

If a) and b) hold for any observable $d \in \mathcal{O}_{J_2}$ (and not only for the reference observables $u = a, b$), then the QL representation is called *strong*. From consideration in Section 8, we obtain:

[5]For the unit sphere S of complex Hilbert space $\mathcal{H}$, the set of equivalent classes $C_\psi = \{\phi = e^{i\sigma}\psi : \sigma \in [0, 2\pi)\}$, $\psi \in S$, is denoted by the symbol $\tilde{S}$; the set of self-adjoint operators, $\widehat{d} : \mathcal{H} \to \mathcal{H}$, is denoted by the symbol $L_s(\mathcal{H})$.

Theorem 9.1. *Let the reference observables a and b be symmetrically conditioned (as well as statistically conjugate and dichotomous). Then the pair of maps* $J_1 \equiv J^{a/b} : \overline{\mathcal{C}^{\mathrm{tr}}} \to \tilde{S}$ *and* $J_2 : \{a, b\} \to L_s(\mathcal{H}_b^{b/a})$, $J_2(a) = \hat{a}$, $J_2(b) = \hat{b}$, *where the operators* $\hat{a}$, $\hat{b}$ *were defined in Section 7, give the strong QL representation.*

Of course, we would be more happy to construct strong representations. However, simple examples with $\mathcal{O}_{J_2}$ larger than just the set of reference observables $\{a, b\}$ show that, for an arbitrary observable $d \in \mathcal{O}_{J_2}$, one cannot expect more than coincidence of the contextual ("classical") and QL averages. Only for the reference observables a and b the probability distributions are also preserved.

We now proceed with *statistically conjugate and symmetrically conditioned dichotomous* reference observables. We also assume that all observables belonging to $\mathcal{O}$ take values in $\mathbb{R}$. We would like to extend the QL-representation given by Theorem 9.1.

Proposition 9.1. *For any map* $f : \mathbb{R} \to \mathbb{R}$, *we have:*

$$E[f(a)/C] = \langle f(\hat{a})\rangle_{\psi_C} \equiv \langle f(\hat{a})J^{a/b}(C), J^{a/b}(C)\rangle, \tag{9.4}$$

$$E[f(b)/C] = \langle f(\hat{b})\rangle_{\psi_C} \equiv \langle f(\hat{b})J^{a/b}(C), J^{a/b}(C)\rangle \tag{9.5}$$

for any context $C \in \overline{\mathcal{C}^{\mathrm{tr}}}$.

Proof. Since in the b-representation the Born rule holds, we obtain:

$$E[f(b)/C] = \sum_{\beta \in X_b} f(\beta) p_c^b(\beta) = \sum_{\beta \in X_b} f(\beta) |\langle \psi_C, e_\beta^b\rangle|^2 = \langle f(\hat{b})\rangle_{\psi_C},$$

where $\psi_C = J^{a/b}(C)$. The same result we have for the $f(\hat{a})$, since we have the Born probability rule both for b and a (because the matrix $\mathbf{P}^{b/a}$ is doubly stochastic). □

Proposition 9.2. *Let* $f, g : \mathbb{R} \to \mathbb{R}$ *be two arbitrary functions. Then*

$$E[f(a) + g(b)/C] = \langle f(\hat{a}) + g(\hat{b})\rangle_{\psi_C} \tag{9.6}$$

for any context $C \in \overline{\mathcal{C}^{\mathrm{tr}}}$.

Proof. By using linearity of the mathematical expectation and linearity of the Hilbert space scalar product we obtain:

$$\begin{aligned} E[f(a) + g(b)/C] &= E[f(a/C] + E[g(b)/C] \\ &= \langle f(\hat{a})\rangle_{\psi_C} + \langle g(\hat{b})\rangle_{\psi_C} = \langle f(\hat{a}) + g(\hat{b})\rangle_{\psi_C}. \end{aligned}$$ □

We recall that the frequency definition of probabilities is used. Therefore contextual averages are defined on the basis of corresponding collectives or S-sequences. Suppose that we have a collective (or just an S-sequence) $x_C(d)$ corresponding to measurements of the observable d under the context C:

$$x_C(d) = (x_1, \dots, x_N, \dots), \quad x_j \in X_d.$$

Then for any function f we can easily construct the collective $x_C(f(d))$ corresponding to measurements of the observable $f(d)$ under the context C:

$$x_C(f(d)) = (f(x_1), \dots, f(x_N), \dots), \quad x_j \in X_d.$$

The average $E[f(d)/C]$ is by definition the average with respect to the latter collective.

Therefore it would be natural to assume that for any observable d belonging to $\mathcal{O}$ any functions $f(d)$ of this observable also belongs $\mathcal{O}$ (we recall that we consider discrete observables).

However, to define the average $E[f(a) + g(b)/C]$ in the frequency framework we should assume *that collectives $x_C(a)$ and $x_C(b)$ are combinable,* Chapter 1. Thus there should exist the simultaneous probability distribution for these collectives:

$$x_C(a) = (a_1, \dots, a_N, \dots), \qquad a_j \in X_a, \tag{9.7}$$

$$x_C(b) = (b_1, \dots, b_N, \dots), \qquad b_j \in X_b, \tag{9.8}$$

and

$$\mathbf{P}_C(a = \alpha, b = \beta) = \lim_{N \to \infty} \frac{n_N(\alpha, \beta)}{N} \tag{9.9}$$

exists. Here, for any pair $(\alpha, \beta) \in X_a \times X_b$, $n_N(\alpha, \beta)$ is the number of realizations of this pair in the first N measurements.

This is a very strong restriction. We could not make such an assumption in the general case. Nevertheless, in some contextual statistical models it could happen for collectives corresponding the reference observables. The reader should not be astonished that we speak about existence of the joint probability distribution for observables which are represented in the QL-model by noncommutative operators!

We proceed under the assumption that collectives (9.7), (9.8) are combinable. Thus limit (9.9) always exists.

Denote the linear space of all observables of the form $d = f(a) + g(b)$ by the symbol $\mathcal{O}_+(a, b)$. We assume that

$$\mathcal{O}_+(a, b) \subset \mathcal{O}. \tag{9.10}$$

Proposition 9.3. *The map*

$$J_2^{a/b} : \mathcal{O}_+(a, b) \to L_s(\mathcal{H}_b^{b/a}), \quad d = f(a) + g(b) \to \hat{d} = f(\hat{a}) + g(\hat{b}), \tag{9.11}$$

preserves the conditional expectation:

$$E[d/C] = \langle J_2^{a/b}(d)\rangle_{\psi_C} \equiv \langle J^{a/b}(d) J^{a/b}(C), J^{a/b}(C)\rangle \tag{9.12}$$

for any context $C \in \overline{\mathcal{C}^{\mathrm{tr}}}$.

As a consequence of this proposition, we have:

Theorem 9.2. *Let the reference observables a and b be symmetrically conditioned (as well as statistically conjugate and dichotomous) and let (9.10) take place. Then maps* $J_1 \equiv J^{a/b} : \overline{\mathcal{C}^{\mathrm{tr}}} \to \tilde{S}$ *and* $J_2 \equiv J_2^{a/b} : \mathcal{O}_+(a,b) \to L_s(\mathcal{H}_b^{b/a})$ *give the QL representation.*

The transformation $J_2^{a/b}$ preserves the conditional expectation for observables $d \in \mathcal{O}_+(a,b)$. In general we cannot expect anything more, since in general $J_2^{a/b}$ does not preserve probability distributions. It preserves them only for the reference observables.

The important problem is to extend the map $J_2^{a/b}$ to even a larger class of observables with preserving (at least) the averages. It might be natural to define (as we always do in the conventional quantum formalism):

$$J_2^{a/b}(f)(\hat{a}, \hat{b}) = f(\hat{a}, \hat{b})$$

where $f(\hat{a}, \hat{b})$ is the pseudo-differential operator with the Weyl symbol $f(a,b)$.

It is possible to show that already for the function

$$f(a,b) = ab \to f(\hat{a}, \hat{b}) = (\hat{a}\hat{b} + \hat{b}\hat{a})/2$$

even the equality (9.12) is violated.

Finally, we remark that, of course, the Definition 9.1 of a QL representation is little bit complicated. However, this complexity is reality of interrelation between the contextual statistical and quantum models. In any event our framework is essentially simpler than Mackey's one.

10 Domain of application of quantum-like representation algorithm

In this section we collect conditions providing the possibility to apply QLRA and obtain a natural QL representation of probabilities.

R1). The reference observables a and b are symmetrically conditioned[6]:

$$p^{b/a}(\beta/\alpha) = p^{a/b}(\alpha/\beta).$$

[6]This condition induces symmetry of the scalar product and the equivalence of the b/a and a/b representations.

R2). The reference observables a and b are statistically conjugate (mutually nondegenerate)[7]:

$$p^{a/b}(\alpha/\beta) > 0, \quad p^{b/a}(\beta/\alpha) > 0.$$

R2a). Context C is nondegenerate with respect to both reference observables a and b:

$$p_C^b(\beta) > 0, \quad p_C^a(\alpha) > 0.$$

Suppose that also the following conditions hold:

R3). Coefficients of supplementarity are bounded by one[8]:

$$\left| \frac{p_C^b(\beta) - \sum_\alpha p_C^a(\alpha) p^{b|a}(\beta/\alpha)}{2\sqrt{\prod_\alpha p_C^a(\alpha) p^{b/a}(\beta/\alpha)}} \right| \leq 1,$$

$$\left| \frac{p_C^a(\alpha) - \sum_\beta p_C^b(\beta) p^{a/b}(\alpha/\beta)}{2\sqrt{\prod_\alpha p_C^b(\beta) p^{a/b}(\alpha/\beta)}} \right| \leq 1.$$

Under these conditions we can apply QLRA (to probabilistic data). The QL representation $\mathcal{H}_b^{b/a}$, see (3.3), is unitary equivalent to the representation $\mathcal{H}_a^{b/a}$, see (6.1). Thus we can identify these two representations. In the same way we can identify the representations $\mathcal{H}_a^{a/b}$ and $\mathcal{H}_b^{a/b}$.

Moreover, the condition R1) implies that even the order of conditioning can be changed peacefully. The representation $\mathcal{H}_b^{b/a}$ is equivalent the representation $\mathcal{H}_a^{a/b}$. All these representations are identified in the conventional quantum mechanics. We denote the result of such a unitary identification by $\mathcal{H}_{ab}$. In this symbol the order of a and b does not play any role, in the same way we could use the symbol $\mathcal{H}_{ba}$.

[7]This condition induces noncommutativity of operators $\hat{a}$ and $\hat{b}$ representing these observables.

[8]This condition induces the QL-representation of the context C in the complex Hilbert space. Thus complex numbers appear due to this condition.

Bibliography

[1] Aaronson S., Is quantum mechanics an island in theoryspace? in: *Quantum Theory: Reconsideration of Foundations-2*, edited by A. Yu. Khrennikov, pp. 15–28, Ser. Math. Model. 10, Växjö University Press, Växjö, 2003; electronic volume: `http://www.vxu.se/msi/forskn/publications.html`.

[2] Accardi L., The probabilistic roots of the quantum mechanical paradoxes, in: *The wave-particle dualism. A tribute to Louis de Broglie on his 90th Birthday*, edited by S. Diner, D. Fargue, G. Lochak and F. Selleri, pp. 47–55, D. Reidel Publ. Company, Dordrecht, 1970.

[3] Accardi L., *Urne e Camaleoni: Dialogo sulla realta, le leggi del caso e la teoria quantistica*, Il Saggiatore, Rome, 1997.

[4] Accardi L., Some loopholes to save quantum nonlocality, in: *Foundations of Probability and Physics-3*, edited by A. Yu. Khrennikov, pp. 1–20, American Institute of Physics, Ser. Conference Proceedings 750, Melville, NY, 2005.

[5] Accardi L., Could one now convince Einstein?, in: *Quantum Theory: Reconsideration of Foundations-3*, edited by G. Adenier, A. Yu. Khrennikov and Th. M. Nieuwenhuizen, pp. 3–18, American Institute of Physics, Ser. Conference Proceedings 810, Melville, NY, 2006.

[6] Adenier G., Fuchs C. and Khrennikov A. Yu. (eds.), *Foundations of Probability and Physics-4*, AIP Conference Proceedings 889, Melville, NY, 2007.

[7] Adenier G., Khrennikov A. Yu. and Nieuwenhuizen T. (eds.), *Quantum Theory: Reconsideration of Foundations-3*, AIP Conference Proceedings 810, Melville, NY, 2006.

[8] Aerts D., Czachor M. and Pawlovski M., Security in quantum cryptography vs. nonlocal hidden variables, in: *Foundations of Probability and Physics-4*, edited by G. Adenier, C. Fuchs and A. Yu. Khrennikov, pp. 71–78, American Institute of Physics, Ser. Conference Proceedings 89, Melville, NY, 2007.

[9] Aichele T., Herzog U., Scholz M. and Benson O., Single photon generation and simultaneous observation of wave and particle properties, in: *Foundations of Probability and Physics-3*, edited by G. Adenier and A. Yu. Khrennikov, pp. 35–41, American Institute of Physics, Ser. Conference Proceedings 750, Melville, NY, 2005.

[10] Albeverio S., Fenstad I., Hoegh-Krohn P. and Lindström T., *Nonstandard Method in the Stochastic Analysis and Mathematical Physics*, Academic Press, London, 1990.

[11] Albeverio S. and Khrennikov A. Yu., Representation of the Weyl group in spaces of square integrable functions with respect to p-adic valued Gaussian distributions, *J. of Phys. A* **29** (1996), 5515–5527.

[12] Allahverdyan A. E., Balian R. and Nieuwenhuizen Th. M., The quantum measurement process in an exactly solvable model, in: Foun*dations of Probability and Physics-3*, edited by A. Yu. Khrennikov, pp. 16–24, American Institute of Physics, Ser. Conference Proceedings 750, Melville, NY, 2005.

[13] Allahverdyan A. E., Khrennikov A. Yu. and Nieuwenhuizen Th. M., Brownian entanglement, *Phys. Rev. A* **71**(2005), 032102-1-14.

[14] Andreev V. A. and Man'ko V. I., Quantum tomography and verification of generalized Bell-CHSH inequalities, in: *Foundations of Probability and Physics-3*, edited by A. Yu. Khrennikov, pp. 42–48, American Institute of Physics, Ser. Conference Proceedings 750, Melville, NY, 2005.

[15] Appleby D. M., Concerning dice and divinity, in: *Foundations of Probability and Physics-3*, edited by G. Adenier, C. Fuchs and A. Yu. Khrennikov, pp. 30–39, American Institute of Physics, Ser. Conference Proceedings 889, Melville, NY, 2007.

[16] Aspect A., Dalibard J. and Roger G., Experimental test of Bell's inequalities using time-varying analyzers, *Phys. Rev. Lett.* **49** (1982), 1804–1807.

[17] Atmanspacher H. and Primas H., *Epistemic and ontic quantum realities*, in: *Foundations of Probability and Physics*-3, edited by A. Yu. Khrennikov, pp. 10–20, AIP Conference Proceedings, 2005.

[18] Atmanspacher H., Räth C. and Wiedenmann G., Statistics and metastatistics in the concept of complexity, *Physica A* **234** (1997), 819–829.

[19] Bacciagaluppi G., Classical extensions, classical representations and Bayesian updating in quantum mechanics, in: *Quantum Theory: Reconsideration of Foundations-2*, edited by A. Yu. Khrennikov, pp. 55–70, Ser. Math. Model. 10, Växjö University Press, Växjö, 2003.

[20] Ballentine L. E., The statistical interpretation of quantum mechanics, *Rev. Mod. Phys.* **42** (1970), 358–381.

[21] Ballentine L. E., *Quantum mechanics*, Englewood Cliffs, New Jersey, 1989.

[22] Ballentine L. E., Interpretations of probability and quantum theory, in: *Foundations of Probability and Physics*, edited by A. Yu. Khrennikov, pp. 71–84, Quantum Probability and White Noise Analysis 13, WSP, Singapore, 2001.

[23] Ballentine L. E., The classical limit of quantum mechanics and its implications for the foundations of quantum mechanics, in: *Quantum Theory: Reconsideration of Foundations-2*, edited by A. Yu. Khrennikov, pp. 71–82, Ser. Math. Model. 10, Växjö University Press, Växjö, 2003; electronic volume: `http://www.vxu.se/msi/forskn/publications.html`.

[24] Barnett M. S., Negative probability, *Proc. Cambridge Phil. Soc.* **41** (1944), 71–73.

[25] Belavkin V. P., Visible information dynamics of hidden quantum mechanics, in: *Foundations of Probability and Physics-2*, edited by A. Yu. Khrennikov, pp. 69–98, Ser. Math. Model. 5, Växjö University Press, Växjö, 2003; electronic volume: `http://www.vxu.se/msi/forskn/publications.html`.

[26] Bell J. S., On the Einstein–Podolsky–Rosen paradox, *Physics* **1** (1964), 195–200.

[27] Bell J. S., On the problem of hidden variables in quantum mechanics, *Rev. Mod. Phys.* **38** (1966), 447–452.

[28] Bell J. S., *Speakable and unspeakable in quantum mechanics*, Cambridge Univ. Press, 1987.

[29] Bialynicki-Birula I., Rényi entropy and the uncertainty relations, in: *Foundations of Probability and Physics-3*, edited by G. Adenier, C. Fuchs and A. Yu. Khrennikov, pp. 52–61, American Institute of Physics, Ser. Conference Proceedings 889, Melville, NY, 2007.

[30] Bohm D., Quantum theory as an indications of a new order in physics. Part B. Implicate and explicate order in physical law, *Found. Phys.* **1** (1971), 139–168.

[31] Bohm D. and Hiley B., *The undivided universe: an ontological interpretation of quantum mechanics*, Routledge and Kegan Paul, London, 1993.

[32] Bohr N., Can quantum-mechanical description of physical reality be considered complete?, *Phys. Rev.* **48** (1935), 696.

[33] Borel E., *Éléments de la théorie des probabilités*, Paris, 1908.

[34] Borel E., Les probabilités dénombrables et leurs applications arithmétiques, *Rend. del Circolo matem. di Palermo* (1909).

[35] Buonomano V., The double slit experiment and the non-ergodic interpretation of quantum mechanics, in: *Proceedings of the International Conference on Quantum Uncertainies*, edited by W. Honig, E. Panarella and D. W. Kraft, pp. 427–434, Plenum Press, 1987.

[36] Busch P., Less (precision) is more (information): quantum information in fuzzy probability theory, in: *Quantum Theory: Reconsideration of Foundations-2*, pp. 113–148, Ser. Math. Model. 10, Växjö University Press, Växjö, 2003.

[37] Busch P., Grabowski M. and Lahti P., *Operational Quantum Physics.* Springer-Verlag, Berlin, 1995.

[38] Chaitin G. J., On the length of programs for computing finite binary sequences, *J. Assoc. Comput. Mach.* **13** (1966), 547–569.

[39] Church A., On the concept of a random sequence, *Bull. Amer. Math. Soc.* **46** (1940), 130–135.

[40] Clauser J. F. and Horne M. A., Experimental consequences of objective local theories, *Phys. Rev. D.* **10** (1974), 526–547.

[41] Clauser J. F., Horne M. A., Schimony A. and Holt R. A., *Phys. Rev. Lett.* **23** (1969), 888–893.

[42] Clauser J. F. and Shimony A., Bell's theorem: experimental tests and implications, *Rep. Progr.Phys.* **41** (1978), 1881–1901.

[43] Conte E., Todarello O., Federici A., Vitiello F., Lopane M., Khrennikov A. Yu. and Zbilut J. P., Some remarks on an experiment suggesting quantum-like behavior of cognitive entities and formulation of an abstract quantum mechanical formalism to describe cognitive entity and its dynamics, *Chaos, Solitons and Fractals* **31** (2006), 1076–1088.

[44] Cramer H., *Mathematical theory of statistics*, Univ. Press, Princeton, 1949.

[45] Davies E. B., *Quantum theory of open systems*, Academic Press, London, 1976.

[46] De Baere W., On the significance of Bell's inequality for hidden-variable theories, *Lett. Nuovo Cimento* **39** (1984), 234–238.

[47] De Baere W., Einstein–Podolsky–Rosen paradox, *Adv. Electronics and Electron physics* **68** (1986), 245–336.

[48] De Broglie L., *La Physique Quantique Restera–t–elle Indeterministe?*, Gauthier-Villars, Paris, 1953.

[49] De Broglie L., *The current interpretation of wave mechanics. A critical study*, Elsevier Publ., Amsterdam, 1964.

[50] d'Espagnat B., *Reality and Physicist*, Cambridge Univ. Press, 1989.

[51] d'Espagnat B., *Veiled Reality. An anlysis of present-day quantum mechanical concepts*, Addison-Wesley publishing company, 1995.

[52] De Muynck W. M., Can we escape from Bell's conclusion that quantum mechanics describes a non-local reality, *Stud. Hist. Phil. Mod. Phys.* **27** (1996), no. 3, 315–330.

[53] De Muynck W. M., *Foundations of Quantum Mechanics, an Empiricists Approach*, Kluwer, Dordrecht, 2002.

[54] De Muynck W. M., De Baere W., Martens H., Interpretations of quantum mechanics, joint measurement of incompatible observables, and counterfactual definitness, *Found. of Physics* **24** (1994), 1589–1663.

[55] Dirac P. A. M., The physical interpretation of quantum mechanics. *Proc. Roy. Soc. London* **A180** (1942), 1–39.

[56] Dubischar D., Gundlach V. M., Steinkamp O. and Khrennikov A. Yu., A p-adic model for the process of thinking disturbed by physiological and information noise, *J. Theor. Biology* **197** (1999), 451–467.

[57] Einstein, A. , *The collected papers of Albert Einstein*, Princeton Univ. Press, Princeton, 1993.

[58] Einstein A., Podolsky B. and Rosen N., Can quantum mechanical description of physical reality be considered complete?, *Phys. Rev.* **47** (1935), 777–780.

[59] Feynman R. P., Negative probability, in: *Quantum Implications, Essays in Honour of David Bohm*, edited by B. J. Hiley and F. D. Peat, pp. 235–246, Routledge and Kegan Paul, London, 1987.

[60] Feynman R. P. and Hibbs A., *Quantum mechanics and path integrals*, McGraw-Hill, New York, 1965.

[61] Fine A., Hidden variables, joint probability, and Bell inequalities, *Phys. Rev. Letters* **48** (1982), 291–295.

[62] Fine T. L., *Theories of probabilities, an examination of foundations*, Academic Press, New York, 1973.

[63] Fréchet M., *Recherches théoriques modernes sur la théorie des probabilités*, Univ. Press., Paris, 1937–1938.

[64] Freund P. G. O. and Witten E., Adelic string amplitudes, *Phys. Lett. B* **199** (1987), 191–195.

[65] Fuchs C., The anti-Växjö interpretation of quantum mechanics, in: *Quantum Theory: Reconsideration of Foundations*, edited by A. Yu. Khrennikov, pp. 99–116, Ser. Math. Modelling 2, Växjö Univ. Press, Växjö, 2002, arXiv:quant-ph/0204146.

[66] Gudder S. P., Probability manifolds, *J. Math Phys.* **25** (1984), 2397–2401.

[67] Heisenberg W., *Physics and Philosophy*, Allen and Unwin, London, 1958.

[68] Hensel K., Untersuchung der Fundamentalgleichung einer Gattung für eine reelle Primzahl als Modul und Bestimmung der Theiler ihrer Discriminante, *J. Reine Angew Math.* **113** (1894), 61–83.

[69] Hida T., *Brownian Motion*, Nauka, Moscow, 1987.

[70] Holevo A. S., *Probabilistic and statistical aspects of quantum theory*, North-Holland, Amsterdam, 1982.

[71] Home D. and Selleri F., Bell's theorem and the EPR paradox, *Revista del Nuovo Cimento* **14** (1991), no. 9, 2–176.

[72] Jaynes E. T., Clearing up mysteries: the original goal, in: *Maximum Entropy and Bayesian Methods*, edited by J. Skilling, pp. 1–27, Kluwer Academic Publishers, Dordrecht, 1989.

[73] Jech T. J., *The axiom of choice*, North-Holland, Amsterdam, 1973.

[74] Kamke E., Über neuere Begründungen der Wahrscheinlichkeitsrechnung, *Jahresbericht der Deutschen Mathematiker-Vereinigung* **42** (1932), 14–27.

[75] Keynes J. M., *A treatise on probability*, Macmillan, London, 1952.

[76] Khrennikov A. Yu., Integration with respect to generalized measures on linear topological spaces, *Trudy Mosk. Mat. Obshch.* **49** (1986), 113–129.

[77] Khrennikov A. Yu., Mathematical methods of the non-Archimedean physics, *Uspekhi Mat. Nauk* **45** (1990), no. 4, 79–110.

[78] Khrennikov A. Yu., p-adic quantum mechanics with p-adic valued functions, *J. Math. Phys.* **32** (1991), no. 4, 932–937.

[79] Khrennikov A. Yu., The p-adic probability and statistics, *Dokl. Akad. Nauk SSSR, ser. Matem.* **322** (1992), no. 6, 1075–1079.

[80] Khrennikov A. Yu., *p-adic valued distributions in mathematical physics*, Kluwer Academic Publishers, Dordrecht, 1994.

[81] Khrennikov A. Yu., p-adic probability description of Dirac's hypothetical world, *Int. J. Theor. Phys.* **34** (1995), no. 12, 2423–2434.

[82] Khrennikov A. Yu., p-adic probability distribution of hidden variables, *Physica A* **215** (1995), 577–587.

[83] Khrennikov A. Yu., p-adic probability interpretation of Bell's inequality paradoxes, *Physics Letters A* **200** (1995), 119–223.

[84] Khrennikov A. Yu., Statistical interpretation of p-adic quantum theories with p-adic valued wave functions, *J. Math. Phys.* **36** (1995), no. 12, 6625–6632.

[85] Khrennikov A. Yu., p-adic description of chaos, in: *Proc. of Workshop "Nonlinear Physics:theory and experiment"*, Gallipoli, Italy, 1995, edited by E. Alfinito, M. Boti et al., pp. 177–184, World Sci. Publ., Singapure, 1996.

[86] Khrennikov A. Yu., *Time to change Kolmogorov axiomatics*, Preprint Ruhr University Bochum, SFB-237, no. 297, 1996.

[87] Khrennikov A. Yu., Ultrametric Hilbert space representation of quantum mechanics with a finite exactness, *Found. of Physics.* **26** (1996), no. 8, 1033–1054.

[88] Khrennikov A. Yu., *Non-Archimedean analysis: quantum paradoxes, dynamical systems and biological models*, Kluwer Acad. Publ., Dordreht, 1997.

[89] Khrennikov A. Yu., Description of experiments detecting p-adic statistics in quantum diffraction experiments, *Doklady Mathematics (Doklady of Russian Academy of Sc.)* **58** (1998), no. 3, 478–480.

[90] Khrennikov A. Yu., Human subconscious as the p-adic dynamical system, *J. of Theor. Biology* **193** (1998), 179–196.

[91] Khrennikov A. Yu., p-adic probability predictions of correlations between particles in the two slit and neuron interferometry experiments, *Il Nuovo Cimento B* **113** (1998), no. 6, 751–760.

[92] Khrennikov A. Yu., p-adic probability predictions of correlations between particles in the two slit and neuron interferometry experiments, *Il Nuovo Cimento B* **113** (1998), no. 6, 751–760.

[93] Khrennikov A. Yu., p-adic stochastic hidden variable model, *J. of Math. Physics* **39** (1998), no. 3, 1388–1402.

[94] Khrennikov A. Yu., Classical and quantum mechanics on information spaces with applications to cognitive, psychological, social and anomalous phenomena, *Found. of Physics* **29** (1999), no. 7, 1065–1098.

[95] Khrennikov A. Yu., *Supernalysis*, Nauka, Moscow, 1997 (in Russian). English translation: Kluwer Academic Publishers, 1999.

[96] Khrennikov A. Yu., Classical and quantum mechanics on p-adic trees of ideas, *BioSystems* **56** (2000), 95–120.

[97] Khrennikov A. Yu. (ed.), *Foundations of Probability and Physics*, Ser. Quantum Probability and White Noise Analysis 13, WSP, Singapore, 2001.

[98] Khrennikov A. Yu., Linear representations of probabilistic transformations induced by context transitions, *J. Phys. A: Math. Gen.* **34** (2001), 9965–9981.

[99] Khrennikov A. Yu., Ensemble fluctuations and the origin of quantum probabilistic rule, *J. Math. Phys.* **43** (2002), 789–802.

[100] Khrennikov A. Yu., On foundations of quantum theory, in: *Quantum Theory: Reconsideration of Foundations*, pp. 163–196, Ser. Math. Modelling 2, Växjö Univ. Press, 2002.

[101] Khrennikov A. Yu., Quantum statistics via perturbation effects of preparation procedures, *Il Nuovo Cimento B* **117** (2002), 267–281.

[102] Khrennikov A. Yu. (ed.), *Quantum Theory: Reconsideration of Foundations*, Ser. Math. Modelling 2, Växjö Univ. Press, Växjö, 2002.

[103] Khrennikov A. Yu., Växjö interpretation of quantum mechanics, 2002, arXiv:quant-ph/0202107.

[104] Khrennikov A. Yu. (ed.), *Foundations of Probability and Physics-2*, Ser. Math. Modelling 5, Växjö Univ. Press, Växjö, 2003.

[105] Khrennikov A. Yu., Växjö interpretation-2003: realism of contexts, 2003, arXiv:quant-ph/0401072.

[106] Khrennikov A. Yu., Contextual approach to quantum mechanics and the theory of the fundamental prespace, *J. Math. Phys.* **45** (2004), 902–921.

[107] Khrennikov A. Yu. (ed.), *Quantum Theory: Reconsideration of Foundations-2*, Ser. Math. Modelling, **10**, Växjö Univ. Press, Växjö, 2004.

[108] Khrennikov A. Yu. (ed.), *Foundations of Probability and Physics-3*, AIP Conference Proceedings 750, Melville, NY, 2005.

[109] Khrennikov A. Yu., Hyperbolic quantum mechanics, *Dokl. Acad. Nauk* **402** (2005), 170–172. English Translation: *Doklady Mathematics* **71** (2005), 363–365.

[110] Khrennikov A. Yu., Interference in the classical probabilistic framework, *Fuzzy Sets and Systems* **155** (2005), 4–17.

[111] Khrennikov A. Yu., Interference in the classical probabilistic model and its representation in complex Hilbert space, *Physica E* **29** (2005), 226–236.

[112] Khrennikov A. Yu., On the representation of contextual probabilistic dynamics in the complex Hilbert space: linear and nonlinear evolutions, Schrödinger dynamics, *Il Nuovo Cimento* **120** (2005), 353–366.

[113] Khrennikov A. Yu., The principle of supplementarity: a contextual probabilistic viewpoint to complementarity, the interference of probabilities, and the incompatibility of variables in quantum mechanics, *Found. Phys.* **35** (2005), 1655–1693.

[114] Khrennikov A. Yu., Classical representation of quantum means and quantum dynamics, *Doklady Mathematics* **73** (2006), no. 3, 445–448.

[115] Khrennikov A. Yu., Generalizations of quantum mechanics induced by classical statistical field theory, *Found. Phys. Letters* **18** (2006), 637–650.

[116] Khrennikov A. Yu., Nonlinear Schrödinger equations from prequantum classical statistical field theory, *Physics Letters A* **357** (2006), no. 3, 171–176.

[117] Khrennikov A. Yu., On the problem of hidden variables for quantum field theory, *Nuovo Cimento B* **121** (2006), no. 5, 505–521.

[118] Khrennikov A. Yu., Prequantum classical statistical field theory: complex representation, Hamilton–Schrödinger equation, and interpretation of stationary states, *Found. Phys. Lett.* **19** (2006), no. 4, 299–319.

[119] Khrennikov A. Yu., Quantum mechanics from time scaling and random fluctuations at the “quick time scale”, *Nuovo Cimento B* **121** (2006), no. 9, 1005–1021.

[120] Khrennikov A. Yu., Representation of quantum field theory as classical statistical mechanics for field functionals, *Doklady Mathematics* **74** (2006), no. 2, 758–761.

[121] Khrennikov A. Yu., Analysis of explicit and implicit assumptions in the theorems of J. von Neumann and J. Bell, *Journal of Russian Laser Research* **28** (2007), 244–254.

[122] Khrennikov A. Yu., Prequantum classical statistical field theory: Time scale of fluctuations, *Doklady Mathematics* **75** (2007), no. 3, 456–459.

[123] Khrennikov A. Yu., Quantum mechanics as an approximation of statistical mechanics for classical fields, *Rep. Math. Phys.*, **60** (2007), no. 3, 453–484.

[124] Khrennikov A. Yu., Quantum mechanics as the quadratic Taylor approximation of classical mechanics: The finite-dimensional case, *J. Theor. Math. Phys.* **152** (2007), no. 2, 1111–1121.

[125] Khrennikov A. Yu., To quantum averages through asymptotic expansion of classical averages on infinite-dimensional space, *J. Math. Phys.* **48** (2007), no. 1, Art. No. 013512.

[126] Khrennikov A. Yu., Quantum randomness as a result of random fluctuations at the Planck time scale?, *Int. J. Theor. Phys.* **47** (2008), no. 1, 121–137.

[127] Khrennikov A. Yu. and Loubenets E., On relation between probabilities in quantum and classical experiments, *Found. Phys.* **34** (2004), 689–704.

[128] Khrennikov A. Yu. and Smolyanov O. G., Nonclassical Kolmogorovian type models describing quantum experiments, *Dokl. Acad. Nauk* **388** (2003), 27–32. English translation: *Doklady Mathematics* **67** (2003), 93–97.

[129] Khrennikov A. Yu., Smolyanov O. G. and Truman A., Kolmogorov probability spaces describing quantum correlations, *Dokl. Acad. Nauk* **393**, 28–32 (2003). English translation: *Doklady Mathematics* **68** (2003), 452–456.

[130] Khrennikov A. Yu. and Yamada Sh., On the concept of a random sequence with respect to p-adic probabilities, *Theory of Probability and Applications* **49** (2004), no. 1.

[131] Khrennikov A. Yu., Yamada Sh. and van Rooij A., Measure-theoretical approach to p-adic probability theory, *Annals Math. Blaise Pascal* **6** (1999), no. 1, 21–32.

[132] Kolmolgorov A. N., Über die analytischen Methoden in der Wahrscheinlichtkeitsrechnung, *Math. Ann.* **104** (1931), 415–458.

[133] Kolmolgorov A. N., *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer-Verlag, Berlin, 1933; reprinted: *Foundations of the Probability Theory*, Chelsea Publ. Comp., New York, 1956.

[134] Kolmolgorov A. N., Three approaches to the quantitative definition of information, *Problems Inform. Transmition.* **1** (1965), 1–7.

[135] Kolmolgorov A. N., Logical basis for information theory and probability theory, *IEEE Trans.* **IT–14** (1968), 662–664.

[136] Kolmorogov A. N. and Fomin S. V., *Elements of Theory of Functions and Functional Analysis*, Nauka, Moscow, 1980.

[137] Li M. and Vitànyi P., *An introduction to Kolmogorov complexity and its applications,* Springer-Verlag, Berlin-Heidelberg-New York, 1997.

[138] Lochak G., De Broglie's initial conception of de Broglie waves, in: *The wave-particle dualism. A tribute to Louis de Broglie on his 90th Birthday*, edited by S. Diner, D. Fargue, G. Lochak and F. Selleri, pp. 1–25, D. Reidel Publ. Company, Dordrecht, 1970.

[139] Ludwig G., *Foundations of quantum mechanics*, Springer-Verlag, Berlin, 1983.

[140] Mackey G. W., *Mathematical Foundations of Quantum Mechanics*, W. A. Benjamin Inc., New York, 1963.

[141] Maddox J., Can chance be less than zero?, *Nature* **320** (1986), 481.

[142] Martin-Löf P., On the concept of random sequence, *Theory of Probability Appl.* **11** (1966), 177–179.

[143] Martin-Löf P., The definition of random sequence, *Inform. Contr.* **9** (1966), 602–619.

[144] Martin-Löf P., Complexity oscillations in infinite binary sequences, *Z. Wahrscheinlichkeitstheorie verw.* **19** (1971), 225–230.

[145] Monna A. and Springer T., Integration non-Archimédienne, 1, 2, *Indag. Math.* **25** (1963), 634–653.

[146] Mückenheim W., A review on extended probabilities, *Phys. Reports* **133** (1986), 338–401.

[147] Obata N., Density of natural numbers and the Levy group, *J. of Number Theory* **30** (1988), 288–297.

[148] Pitowsky I., Resolution of the Einstein–Podolsky–Rosen and Bell paradoxes, *Phys. Rev. Lett* **48** (1982), no. 10, 1299–1302.

[149] Pitowsky I., Deterministic model of spin and statistics, *Phys. Rev. D* **27** (1983), no. 10, 2316–2326.

[150] Plotnitsky A., Reading Bohr: complementarity, epistemology, entanglement, and decoherence, in: *NATO Workshop Decoherence and its Implications for Quantum Computations*, edited by A. Gonis and P. Turchi, pp. 3–37, IOS Press, Amsterdam, 2001.

[151] Plotnitsky A., *The Knowable and Unknowable: Modern Science, Nonclassical Thought, and the "Two Cultures"*, Univ. Michigan Press, 2002.

[152] Plotnitsky A., "This is an extremely funny thing, something must be hidden behind that": Quantum waves and quantum probability with Erwin Schrödinger, in: *Foundations of probability and physics-3*, edited by G. Adenier and A. Yu. Khrennikov, pp. 388–408, American Institute of Physics, Ser. Conference Proceedings 750, Melville, NY, 2005.

[153] Plotnitsky A., *Reading Bohr: Physics and Philosophy*, Springer-Verlag, Dordrecht, 2006.

[154] Prugovecki E., Simultaneous measurement of several observables, *Found. of Physics* **3** (1973), no. 1, 3–18.

[155] Rastall P., The Bell inequalities, *Found. of Physics* **13** (1983), 555–570.

[156] Rauch H., Summhammer J., Zawisky M. and Jericha E., Law-contrast and low-counting-rate measurements in neutron interferometry, *Phys. Rev. A* **42** (1990), 3726–3732.

[157] Schikhov W., *Ultrametric calculus*, Cambridge Univ. Press, Cambridge, 1984.

[158] Schnorr C. P., *Zufälligkeit und Wahrscheinlichkeit*, Lect. Notes in Math. 218, Springer-Verlag, Berlin, 1971.

[159] Shiryayev A. N. (ed), *Selected works of A. N. Kolmogorov, Volume 2: Probability theory and mathematical statistics*, Kluwer Academic Publ., Dordrecht, 1986.

[160] Solomonoff R. J., A formal theory of inductive inference, *Inform. Contr.* **7** (1964), 1–22 (Part I) and 224–254 (Part II).

[161] Summhammer J., Neutron interferometric test of the nonergodic interpretation of quantum mechanics, *Il Nuovo Cimento* **103 B** (1989), 265–280.

[162] Tornier E., *Wahrscheinlichkeitsrechnunug und allgemeine Integrationstheorie*, Univ. Press, Leipzig, 1936.

[163] Van Lambalgen M., Von Mises' definition of random sequences reconsidered, *J. of Symbolic Logic* **52** (1987), no. 3, 725–755.

[164] Van Rooij A., *Non-Archimedean Functional Analysis*, Marcel Dekker, Inc., New York, 1978.

[165] Venn J., *The logic of chance*, London, 1866; reprint: Chelsea, New York, 1962.

[166] Ville J., *Etude critique de la notion de collective*, Gauthier-Villars, Paris, 1939.

[167] Vladimirov V. S., Volovich I. V. and Zelenov E. I., *p-adic analysis and Mathematical Physics*, World Scientific Publ., Singapure, 1993.

[168] Volovich I. V., p-adic string, *Class. Quant. Grav.* **4** (1987), 83–87.

[169] von Mises R., Grundlagen der Wahrscheinlichkeitsrechnung, *Math. Z.* **5** (1919), 52–99.

[170] von Mises R., *Probability, Statistics and Truth*, Macmillan, London, 1957.

[171] von Mises R., *The mathematical theory of probability and statistics*, Academic, London, 1964.

[172] von Neuman J., *Mathematical foundations of quantum mechanics*, Princeton Univ. Press, Princenton, 1955.

[173] Wald A., Die Widerspruchsfreiheit des Kollektivbegriffs in der Wahrscheinlichkeitsrechnung, *Ergebnisse eines Math. Kolloquiums* **8**, 38–72 (1938).

[174] Wigner E., *Quantum-mechanical distribution functions revisted,* in: *Perspectives in quantum theory*, edited by W. Yourgrau and A. van der Merwe, MIT Press, Cambridge MA, 1971.

[175] Zvonkin A. K. and Levin L. A., A complexity of finite objects and justification of the notions of information and randomness on the basis of algorithms, *Uspehi Matem. Nauk* **25** (1970), No. 6, 85–127.

Index