Alexander Holevo

Probabilistic and Statistical Aspects of Quantum Theory

EDIZIONI
DELLA
NORMALE

1

QUADERNI

MONOGRAPHS

Alexander Holevo
Steklov Mathematical Institute
Moscow

Probabilistic and Statistical Aspects of Quantum Theory

Alexander Holevo

Probabilistic and Statistical Aspects of Quantum Theory

EDIZIONI
DELLA
NORMALE

ISBN: 978-88-7642-375-8

Contents

Foreword to the second English edition

In the present edition the *comments to chapters* in the main text and the References were extended and updated. During the past years the quantum (noncommutative) generalizations of probability, mathematical statistics and information theory were substantially developed. Theory of quantum stochastic processes was elaborated, unifying repeated and continuous measurement with the dynamics of open quantum systems. This material is reflected in the book [178] where the reader can find an extended bibliography. With the emergence of ideas of quantum computation the powerful impetus got the quantum information theory which was born more than half-century ago and shaped as an independent scientific discipline in the 1990-s. This progress enhanced, in particular, recent development of asymptotic methods of quantum estimation theory. An introduction to this circle of problems can be found in the books [152, 170, 179].

The present edition would not be possible without the enthusiasm and perseverance of Professor Vittorio Giovannetti and Professor Rosario Fazio to whom the author expresses his warm gratitude. The author is grateful to Dottoressa Luisa Ferrini, *Edizioni della Normale*, for her professional and most efficient assistance in preparation of the manuscript.

This book is dedicated to the memory of Academician Kamil A. Valiev who passed away in the summer 2010.

Alexander Holevo
Moscow, September 2010.

Foreword to the second Russian edition

When this book was first published in 1980 (the English edition appeared in 1982 in North Holland), the author addressed it to a broad audience of readers, both mathematicians and physicists having intention to make them acquainted with the new prospects and possibilities which emerge from the interaction of ideas of the mathematical statistics and the quantum theory. During the past period this approach became even more demanded. On one side, its advantages in the questions of foundations of quantum theory related to quantum measurements became more apparent and widely acknowledged. On the other hand, one should stress that these theoretical findings were not an end in itself: in modern high-precision physical experiments researchers become able to operate with single ions, atoms and photons which leads to potentially important applications, such as quantum communications, computation and cryptography. Of great importance is the question of extraction of maximal possible information from the state of a given quantum system. For example, in currently discussed proposals for quantum computing information is written in the states of elementary quantum memory cells, qubits, and then read off by means of quantum measurement. From the statistical point of view, measurement gives an estimate for the quantum state – as a whole, or for some of its parameters. In this way a new interest emerges to quantum estimation theory, the fundamentals of which are presented in this book.

One of essential consequences of penetration of the ideas of mathematical statistics into the theory of quantum measurement is the wide use of the mathematical notion of (non-orthogonal) resolution of the identity in the system Hilbert space (in the Western literature – POVM, positive operator-valued measures), describing the statistics of decision procedures. During the time passed resolutions of the identity became a standard tool both in mathematical and in physical literature on quantum measurements. All this, in my opinion, justifies publication of the second Russian edition of the book, moreover as the first one became a rare book.

The present edition includes Supplement discussing in some detail the problem of hidden variables in quantum mechanics which continues to provoke a keen interest. Moreover, the edition is complemented with Comments reflecting new results and achievements.

Academician K. A. Valiev
Moscow, April 2003.

Preface

The mathematical language of modern quantum mechanics is operator theory. Operators play there a role similar to functions in classical mechanics, probability theory and statistics. However, while the use of functions in classical theories is founded on premises which seem intuitively quite clear, in quantum theory the situation with operators is different.

Historically the "matrix mechanics" of Heisenberg and the "wave mechanics" of Schrödinger which gave rise to the contemporary form of quantum theory, originated from ingenious attempts to fit mathematical objects able to reflect some unusual (from the macroscopic point of view) features of microparticle behavior – in particular, a peculiar combination of continuous and discrete properties. The "probabilistic interpretation" developed later by Born and others elucidated the meaning of operator formalism by postulating rules connecting mathematical objects with observable quantities. However a good deal of arbitrariness remained in these postulates and the most convincing argument for quantum-theoretical explanations was still the "striking" coincidence of theoretical predictions with experimental data. This state of affairs gave rise to numerous attempts, on one hand, to find classical alternatives to quantum theory which would give an equally satisfactory description of the experimental data, and on the other hand, to find out physical and philosophical arguments for justifying the inevitability of the new mechanics.

Notwithstanding the impressive philosophical achievements in this field there was and still is a need for the structural investigation of quantum theory from a more mathematical point of view aimed at elucidating the connections between the entities of the physical world and the elements of operator formalism. The present book is essentially in this line of research opened by the classical von Neumann's treatise "Mathematical Foundations of Quantum Mechanics". However it differs from most subsequent investigations by the strong emphasis on the statistical rather than "logical" essence of quantum theory; it gives an account of recent

progress in the statistical theory of quantum measurement, stimulated by the new applications of quantum mechanics, particularly in quantum optics.

The first three chapters give an introduction to the foundations of quantum mechanics, addressed to the reader interested in the structure of quantum theory and its relations with classical probability. In spite of the mathematical character of the presentation it is not "axiomatic". Its purpose is to display the origin of the basic elements of operator formalism resting, as far as possible, upon the classical probabilistic concepts.

The present revision is not an end in itself – it emerged from the solution of concrete problems concerning the quantum limitations to measurement accuracy, arising in applications. So far there has been no general approach to such kind of problems. The methods of mathematical statistics adapted for classical measurements required radial quantum modification. The last chapters of the book are devoted to the recently developed quantum estimation theory, which is an analog of the corresponding branch of mathematical statistics.

We now give a more detailed account of the contents of the book. In Chapter 1 the general concepts of state and measurement are introduced on the basis of statistical analysis of an experimental situation. From the very beginning this approach leads to a substantial generalization of the Dirac-von Neumann concept of an observable. Mathematically it is reflected by the occurrence of arbitrary resolutions of identity in place of orthogonal ones (spectral measures) and the repudiation of self-adjointness as an indispensable attribute of an observable. In this way nonorthogonal resolutions of identity like the "overcomplete" system of coherent states known in physics for rather a long time find their proper place in quantum phenomenology. The new concept of quantum measurement is central for the whole book.

The notion of statistical model exploited in Chapter 1 is quite general and may find applications different from quantum theory. It gives us a new insight into the still controversial "hidden variables" problem.

In Chapter 2 the elements of operator theory in Hilbert space are introduced to provide mathematical background for the subsequent material. As compared to standard presentations relatively much attention is paid to nonorthogonal resolutions of identity and related questions. A novel feature is also the introduction of the $\mathscr{L}^2$ spaces of observables associated with a quantum state and playing a role similar to the Hilbert space of random variables with finite second moment in probability theory. These $\mathscr{L}^2$ spaces give the framework for a calculus of unbounded operators.

Of fundamental importance to quantum theory are groups of symmetries. In Chapter 3 elementary quantum mechanics is considered from this

point of view. An important result of this discussion is the isolation of the notion of covariant measurement which ties physical quantities with certain classes of resolutions of identity in the underlying Hilbert space. In this way we construct quantum measurements canonically corresponding to such quantities as time, phase of harmonic oscillator, angle of rotation and joint measurement of coordinate and velocity. Allowing the broader concept of quantum measurement enables us to resolve old troubles of quantum theory connected with the non-existence of self-adjoint operators having the required covariance properties.

Chapter 4 is devoted to a more advanced study of covariant measurements and extreme quantum limits for the accuracy of estimation of physical parameters. The latter problem becomes important in view of the progress in experimental physics. We present a unified statistical approach to "non-standard" uncertainty relations of the "angle-angular momentum" type. They appear to be related to the quantum analog of the Hunt-Stein theorem in mathematical statistics. A general conclusion which can be drawn from Chapter 4 is that the requirements of covariance and optimality, *i.e.*, extremal quantum accuracy, determine the canonical measurement of a "shift" parameter, such as angle, coordinate, time, uniquely up to a "gauge" transformation.

An example of a situation where quantum limitations are important is provided by optical communication. As it is known, "quantum noise" distorting the signal in the optical range can be much more significant than the thermal background radiation. As in ordinary communication theory the problem of signal estimation arises, but now it requires a specifically quantum-theoretic formulation and solution.

Chapter 5 is devoted to the so-called Gaussian states which, in particular, describe radiation fields in optical communication theory. The presentation is intended to make maximal use of the remarkable parallel with the Gaussian probability distributions. An important role is played here by quantum characteristic functions.

In Chapter 6 the general inequalities for the measurement mean square errors are derived, which are quantum analogs of the well-known Cramer-Rao inequality in mathematical statistics. The best unbiased measurements of the mean-value parameters of a Gaussian state are described.

Needless to say, the present book cannot (and is not intended to) replace the standard textbooks on quantum mechanics. Most of the important topics, such as perturbation theory, are apparently out of its scope. Nor does it pretend to give a full account of quantum measurements. We have discussed only those problems which concern measurement statistics and do not require consideration of state changes after measurements.

The references to the relevant work on "open" quantum systems and quantum stochastic processes can be found in the comments.

The author's intention was to write a book accessible to a wide circle of readers, both mathematicians and physicists. As a result, the presentation, being in general mathematical, is rather informal and certainly not "the most economic" from a mathematical point of view. On the other hand, it neglects some subtleties concerning measurability etc. As a rule a rigorous treatment can be found in the special papers refered to. The necessary background for the whole material is knowledge of fundamentals of probability theory. Mathematically the most elementary is Chapter 1 which uses mainly finite-dimensional linear analysis. The functional analytic minimum is given in Section 2.1-2.6 of Chapter 2, and a mathematically educated reader may just glance over it. On the other hand, a reader familiar with quantum mechanics can omit the detailed discussion of such topics as harmonic oscillators and spin in Chapter 3, included to make the presentation self-contained, and concentrate on less familiar things.

The Dirac notation is used intensively throughout the book but with round brackets for the inner product as accepted in mathematical literature. The angle brackets, associated with the averaging symbol in statistical mechanics, are reserved for the different inner product defining the correlation of a pair of observables. To denote a quantum state as well as its density operator we use the letter S (not the usual ρ) allied to the notation P for the classical state (probability distribution).

The author's thanks are due to Prof. D. P. Želobenko and the late Prof. Yu. M. Shirokov who read the manuscript and made useful comments.

In translating the book the author took the opportunity of improving the presentation which concerned mainly Chapters 3, 4. Few references were added. The author is grateful to Prof. Yu. A. Rozanov and Prof. P. R. Krishnaiah for providing the opportunity of translating this book for *North-Holland Series in Statistics and Probability*.

Chapter 1
Statistical models

1.1. States and measurements

Any theoretical model ultimately relies upon experience – the framework for a model is constituted by the array of experimental data relevant to the study of the object or phenomenon. Let us consider a very schematic and general description of an experimental situation and try to trace back the emergence of the principal components of a theoretical model.

The fundamental reproducibility condition requires at least in principle the unrestricted possibility of repetition of an experiment. Considering a sequence of identical and independent realizations of some experimental situation one always sees that practically the data obtained are not identically the same but subject to random fluctuations, the magnitude of which depends on the nature of the experiment and of the object under investigation.

There exist large classes of phenomena, for example, planetary motion or constant electric currents, in which these random fluctuations can be both practically and theoretically ignored. The corresponding theories – classical celestial mechanics and circuit theory – proceed from the assumption that the parameters describing the object can be measured with arbitrary accuracy, or, ultimately, with absolute precision. In such cases the object is said to admit deterministic description. Such a description, however self-contained it seems to be, is usually only an approximation to reality, valid in so far as it agrees with the experience.

The fruitfulness of the deterministic point of view in the classical physics of the 18-19$^{\text{th}}$ centuries gave rise to the illusion of its universality. However, with the penetration of experimental physics into the atomic domain the inapplicability of the classical deterministic approach and the relevance of statistical concepts in this domain became more and more evident. The behavior of atomic and subatomic objects is essentially probabilistic; an ordinary way to extract information about them is to observe a large number of identical objects to obtain statistical data. The

interested reader can find about the experimental evidence for statistical description in microphysics, which is now generally accepted, in any contemporary tract on quantum physics.

The possibility of statistical description presumes the fulfilment of the following statistical postulate, incorporating the previous requirement of reproducibility: *the individual results in a sequence of identical, independent realizations of an experiment may vary, but the occurrence of one or another result in a long enough sequence of realizations can be characterized by a definite stable frequency*. Then, abstracting from the practical impossibility of performing an infinite sequence of realizations, one can adopt that the results of the experiment are theoretically described by the *probabilities* of various possible outcomes. More precisely, we must distinguish an individual realization of the experiment which results in some concrete outcome from the experiment as a collection of all its possible individual realizations. In this latter sense, the final results of the experiment are theoretically described by probability distributions. The deterministic dependence of the experimental results on the initial conditions is replaced by the statistical one: the function of the initial data is now the output probability distribution.

As an example consider a beam of identical independent particles which are scattered by an obstacle and then registered by a photographic plate, so that an individual particle hitting the plate causes a blackening of the emulsion at the place of the collision. Exposing a beam which consists of a large enough number of particles will result in a photographic picture giving the visual image of the probability density for the point at which an individual particle hits the plate. The natural light is the chaotic flow of an immense number of specific corpuscules – the photons. The well-known optical diffraction pictures present the images of the probability density of an individual photon scattered by an aperture.

Of course, the statistical description is by no means subject to atomic or subatomic phenomena. When investigating a system which consists of a large number of components (*e.g.*, a gas or a liquid) the experimenter has at his disposal only a very restricted set of parameters to vary (say, pressure, volume or temperature). An immense number of parameters, giving a detailed description for the behavior of subsystems of the system are out of control; their uncontrolled changes may substantially influence the results of measurements. A study of these fluctuations is essential for understanding the mechanisms of phenomena occurring in large systems. The statistics of observations is most important in problems of information transmission, where the fluctuations in the physical carriers of information are the source of various “noises” distorting the signal.

The statistical approach is often appropriate in biometrical research. In studying the effect of a medicine, a physician can take into account a limited number of parameters characterizing his patients such as age, blood group etc. However the effect of the treatment in each individual case will depend not only on these "integral" parameters, but also on a number of other internal factors which were not, or could not be taken into account. In such cases the dependence of the effect on the "input parameters" is not deterministic and often can be successfully described statistically.

These two examples show that the origin of fluctuations in results of measurement may be uncertainty in the values of some "hidden variables" which are beyond the control of the experimenter. The nature of randomness in atomic and subatomic phenomena is still not so clear, though the relevance of the statistical approach is confirmed here by more than half a century experience of applications of quantum theory. We shall not touch here the issues concerning the nature of randomness in microscopic phenomena, but we shall comment on some mathematical aspects of the relevant "problem of hidden variables" in the Supplement. The main attention we shall pay here to the consequences of the statistical postulate irrespective of the nature of the object under consideration. We shall see that already on this very general level the notions of the state and the measurement arise, which play a basic role, in particular, in quantum theory.

In any experiment one can distinguish the two main stages. At the first stage of *preparation* a definite experimental set-up is settled, some initial conditions or "input data" of the experiment are established. At the following stage of the experiment the "prepared" object is coupled to a measuring device, resulting in these or the other output data (Figure 1.1).

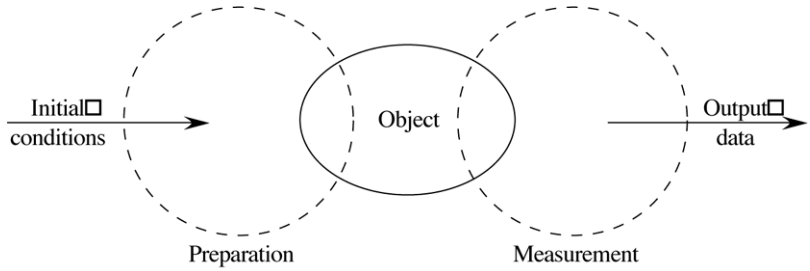

Figure 1.1.

Conventionally, one may conceive the object as a "black box" at the "input" of which one can impose some initial conditions $\widetilde{S}$. After the object has been definitely prepared, some measurement is performed, resulting in the output data u. These data may be of arbitrary nature; they may be

discrete if the measuring device registers the occurrence of some events, *e.g.*, the presence or absence of some particles; they may be represented by a scalar or vector quantity, if the measuring device has one or several scales; at last the result of a measurement may be a picture of a whole trajectory, as in a bubble chamber. To give a uniform treatment for all these possibilities we assume that the outcomes of measurement form a measurable space U with the σ-field of measurable subsets $\mathscr{A}(U)$. In the concrete cases we shall deal with, U will be usually a domain in the real n-dimensional space $\mathbb{R}^n$ with the Borel σ-field generated by open sets (or by multi-dimensional intervals). A measurable subset $B \subset U$ corresponds to the event: the result of the measurement u lies in B.

According to the statistical postulate, a result of an individual measurement can be considered as a realization of a random variable taking values in U. Let $\mu_{\widetilde{S}}(\mathrm{d}u)$ be the probability distribution of this random variable. The subscript $\widetilde{S}$ reflects the dependence of the statistics of the measurement upon the preparation procedure, *i.e.*, the initial conditions of the experiment, so that

$$\mu_{\widetilde{S}}(B) = \Pr\{u \in B|\widetilde{S}\}, \qquad B \in \mathscr{A}(U)$$

is the conditional probability of obtaining a result $u \in B$ under the initial condition $\widetilde{S}$. The map $\widetilde{S} \to \mu_{\widetilde{S}}(\mathrm{d}u)$ gives a complete statistical description for the results of the measurement. It should be stressed, however, that such a description does not contain indications either on a concrete mechanism of the measurement, or on its consequences for the object. From this point of view one should not distinguish between the measuring procedures giving the same statistics $\mu_{\widetilde{S}}$ under the same condition $\widetilde{S}$ however different their practical implementation may be. Thus the map $\widetilde{S} \to \mu_{\widetilde{S}}$ refers to this whole class of measuring procedures.

Similarly the initial conditions $\widetilde{S}_1$ and $\widetilde{S}_2$ are indiscernible from the point of view of the results of measurements if $\mu_{\widetilde{S}_1} = \mu_{\widetilde{S}_2}$ for any map $\widetilde{S} \to \mu_{\widetilde{S}}$ describing a measurement. We shall join the indiscernible preparation procedures $\widetilde{S}$ in the equivalence classes $S = [\widetilde{S}]$ which will be called *states*. Denote by $\mathfrak{S}$ the set of all possible states. Since the probability distribution $\mu_{\widetilde{S}}$ is the same for all $\widetilde{S}$ from the class S, it is a function of classes, $\mu_{\widetilde{S}} = \mu_S$. The map $S \to \mu_S$ transforming states $S \in \mathfrak{S}$ into the probability distributions on the space of outcomes U will be called a *measurement* (with values in U).

The set $\mathfrak{S}$ and the maps $S \to \mu_S$ enjoy an important structural property. Let $\{S_\alpha\}$ be a finite collection of states. Consider an infinite sequence of individual experiments in each of which the object is prepared in some of the states S_α, the occurrence of different values of α being

characterized by a probability distribution $\{p_\alpha\}$. This may reflect fluctuations in the values of some parameters in the preparation procedure. Let one and the same measurement be performed in each individual experiment. Then by the statistical postulate and the elementary properties of probabilities the occurrence of an outcome u will be described by the probability distribution $\mu(\mathrm{d}u) = \sum_\alpha p_\alpha \mu_{S_\alpha}(\mathrm{d}u)$. The situation described above can be considered as a special way of state preparation (mixing) when the value of the parameter α is not fixed but is chosen according to the prior distribution $\{p_\alpha\}$. Denoting such "mixed" state by

$$S = S(\{S_\alpha\}, \{p_\alpha\}) \tag{1.1.1}$$

we have for any measurement $S \to \mu_S$

$$\mu_S(\mathrm{d}u) = \sum_a p_\alpha \mu_{S_\alpha}(\mathrm{d}u). \tag{1.1.2}$$

Thus, we are led to adopt that for any finite set of states $\{S_\alpha\} \subset \mathfrak{S}$ and any probability distribution $\{p_\alpha\}$ there is uniquely defined "mixed" state $S(\{S_\alpha\}, \{p_\alpha\})$, which is characterized by (1.1.2). Then it turns out that the set of the states can be naturally identified with a convex set in a linear space, such that $S(\{S_\alpha\}, \{p_\alpha\}) = \sum_\alpha p_\alpha S_\alpha$. The exact formulation requires some knowledge in convexity presented in the next section.

1.2. Some facts about convex sets

Let $S_1, \ldots, S_n$ be the elements of a linear space $\mathscr{L}$, and $p_1, \ldots, p_n$ a set of real numbers satisfying

$$p_j \geq 0, \qquad j = 1, \ldots, n; \qquad \sum_{j=1}^{n} p_j = 1,$$

i.e., a finite probability distribution. Then the element

$$S = \sum_{j=1}^{n} p_j S_j$$

is called a *convex combination* of S_j with the coefficients (weights) $\{p_j\}$. The *convex hull* of a set $\mathscr{K} \subset \mathscr{L}$ is the collection of all convex combinations of all finite subsets $\{S_j\} \subset \mathscr{K}$. A set $\mathfrak{S}$ is called *convex* if it coincides with its convex hull, *i.e.*, if it contains convex combination of any finite subset of its elements. For two elements S_0, S_1, their convex combinations form the segment $[S_0, S_1]$:

$$\{S : S = p_0 S_0 + p_1 S_1;\ p_0, p_1 \geq 0,\ p_0 + p_1 = 1\}.$$

It is easy to see that the set $\mathfrak{S}$ is convex if and only if together with any two elements S_0, S_1, it contains the segment $[S_0, S_1]$.

An abstract set $\mathfrak{S}$ is called *mixture space* if there is a rule by which to any finite unordered collection $\{S_\alpha\}$ of elements of $\mathfrak{S}$ and any finite probability distribution $\{p_\alpha\}$ there corresponds a unique element $S(\{S_\alpha\}, \{p_\alpha\}) \in \mathfrak{S}$ called the mixture of the states S_α with weights p_α. It is assumed that mixing the collection consisting of copies of one and the same element S_0 gives again S_0, *i.e.*, if $S_\alpha \equiv S_0$ for all α, then $S(\{S_\alpha\}, \{p_\alpha\}) = S_0$. An example of mixture space is a convex set with the convex combination as the mixture.

Let F be a map from a mixture space $\mathfrak{S}$ into a linear space. The map is called *affine* if for any mixture $S(\{S_\alpha\}, \{p_\alpha\})$

$$F(S(\{S_\alpha\}, \{p_\alpha\})) = \sum_\alpha p_\alpha F(S_\alpha).$$

The set of affine maps is nonempty, since the map which sends any $S \in \mathfrak{S}$ into a constant vector b is affine. Clearly, the image of a convex set under an affine map is again convex. In linear space there is a close connection between affine and linear maps: namely, any affine map F of a convex set $\mathfrak{S} \subset \mathscr{L}$ has the form $F(T) = A(T) + b$, $T \in \mathfrak{S}$, where A is a linear map defined on $\mathscr{L}$. In particular, any affine functional (*i.e.*, map with values on the real line $\mathbb{R}$) is up to an additive constant a restriction to $\mathfrak{S}$ of a linear functional on $\mathscr{L}$.

A mixture space is called *separated* if for any two $S_1, S_2 \in \mathfrak{S}$ there is an affine functional φ on $\mathfrak{S}$ such that $\varphi(S_1) \neq \varphi(S_2)$.

An example of a separated mixture space is the set of states of Section 1.1. Indeed, for any measurement $S \to \mu_S$ and any subset $B \in \mathscr{A}(U)$ the functional $S \to \mu_S(B)$ is affine by (1.1.2). By construction, for any two states S_1 and S_2 there exists a measurement $S \to \mu_S$ such that $\mu_{S_1} \not\equiv \mu_{S_2}$, *i.e.*, $\mu_{s_1}(B) \neq \mu_{S_2}(B)$ for some B. The following simple statement shows that the set of states can be considered as a convex subset in a linear space, with the convex combinations as mixtures.

Proposition 1.2.1. *For any separated mixture space $\mathfrak{S}$ there is a one-to-one affine map of $\mathfrak{S}$ onto a convex subset of a linear space.*

Proof. Let $\mathfrak{A}(\mathfrak{S})$ be the linear space of all affine functional on $\mathfrak{S}$ and $\mathscr{L} = \mathfrak{A}(\mathfrak{S})'$ the dual space of all linear functionals on $\mathfrak{A}(\mathfrak{S})$. For each $S \in \mathfrak{S}$ introduce $\widehat{S} \in \mathscr{L}$, putting

$$\widehat{S}(\varphi) = \varphi(S), \qquad \varphi \in \mathscr{L}.$$

The map $S \to \widehat{S}$ is affine, since

$$\sum_j p_j \widehat{S}_j(\varphi) = \sum_j p_j \varphi(S_j) = \varphi(S(\{S_j\}, \{p_j\})) = \widehat{S}(\{S_j\}, \{p_j\})(\varphi),$$

and it is one-to-one since $\widehat{S}_1(\varphi) = \widehat{S}_2(\varphi)$ implies $\varphi(S_1) = \varphi(S_2)$ for all affine functionals φ. This proves the proposition. □

The most simple example of a convex set in an n-dimensional *simplex*, which is defined as a convex hull of $n+1$ points $S_0, \ldots, S_n$ in a space of dimension $\geq n$, such that vectors $S_0S_1, \ldots, S_0S_n$ are linearly independent. For $n = 1$ this is a segment, for $n = 2$ a triangle, for $n = 3$ a tetrahedron (Figure 1.2). The points $S_0, \ldots, S_n$ are the vertices of the simplex.

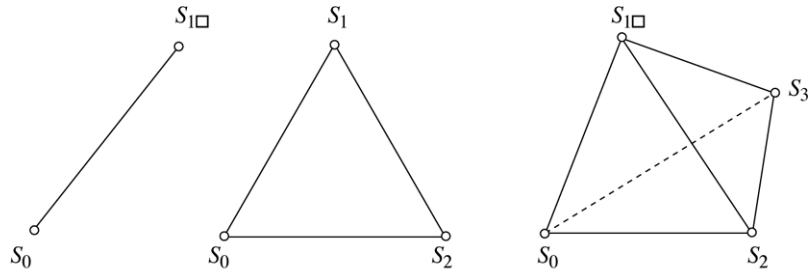

Figure 1.2.

The basic role in the theory of convex sets plays the notion of extreme point. The point S is an *extreme point* of a convex set $\mathfrak{S}$, if it is not an interior point of a segment, lying completely in $\mathfrak{S}$; that is, it cannot be represented in the form $S = p_0S_0 + p_1S_1$, where $p_0, p_1 > 0$, $p_0 + p_1 = 1$; $S_0, S_1 \in \mathfrak{S}$ and $S_0 \neq S_1$. For example, the extreme points of a simplex are its vertices. In a finite-dimensional space the following statement holds.

Theorem 1.2.2. *Any compact* (*i.e., bounded and closed*) *set* $\mathfrak{S}$ *coincides with the convex hull of the set of its extreme points.*

In general, there may be several ways to represent a point S of a convex set $\mathfrak{S}$ as a convex combination of extreme points. *The representation is unique for any point* $S \in \mathfrak{S}$ *if and only if* $\mathfrak{S}$ *is a simplex.*

Theorem 1.2.2, the definition and the characteristic property of a simplex can be generalized also to the infinite-dimensional case, but a careful treatment of these problems would require much more space; on the other hand, the finite-dimensional picture presented above will suffice for

understanding the following sections. Therefore we consider only one infinite-dimensional example, which we shall substantially need later.

Let $\mathfrak{P}(\Omega)$ be the collection of all probability distributions on a measurable space Ω. This set is convex, since any "mixture" of probability distributions $P_j(\mathrm{d}\omega)$ is again a probability distribution on Ω

$$P(A) = \sum_j p_j P_j(A), \qquad A \in \mathscr{A}(\Omega).$$

To any point $\omega \in \Omega$ there corresponds the *δ-distribution*, concentrated at ω,

$$\delta_\omega(A) = \begin{cases} 1, & A \ni \omega, \\ 0, & A \not\ni \omega. \end{cases}$$

We shall suppose, without loss of generality, that the σ-field $\mathscr{A}(\Omega)$ separates points of Ω, *i.e.*, for any two $\omega_1, \omega_2 \in \Omega$ there is $A \in \mathscr{A}(\Omega)$ such that $\omega_1 \in A$; $\omega_2 \notin A$. Then the correspondence $\omega \to \delta_\omega$ is one-to-one. The δ-distributions are precisely the extreme points of $\mathfrak{P}(\Omega)$. For any $P \in \mathfrak{P}(\Omega)$

$$P(A) = \int_\Omega \delta_\omega(A) P(\mathrm{d}\omega), \qquad A \in \mathscr{A}(\Omega). \tag{1.2.3}$$

This representation is a continual analog of a finite convex combination with respect to extreme points, the role of weights p_j is played by the probability distribution $P(\mathrm{d}\omega)$. The representation (1.2.3) is unique; so that the convex set $\mathfrak{P}(\Omega)$ has the property, characteristic of a simplex in the finite-dimensional case, and we shall keep this name for $\mathfrak{P}(\Omega)$.

If the space Ω consists of n points, $\Omega = \Omega_n$ ($\mathscr{A}(\Omega_n)$ being the Boolean field of subsets of Ω_n), the set

$$\mathfrak{P}_n = \left\{ P = [p_1, \ldots, p_n] : p_j \geq 0, \sum_{j=1}^n p_j = 1 \right\}$$

is clearly an (n–1)-dimensional simplex with the vertices [1,0,...,0], ..., [0, ..., 0, 1]. It will be convenient for us to represent P by the diagonal $n \times n$-matrix

$$P = \begin{bmatrix} p_1 & & 0 \\ & \ddots & \\ 0 & & p_n \end{bmatrix}.$$

Then the characteristic properties of P take the form

$$P \geq 0, \qquad \mathrm{Tr}\, P = 1, \tag{1.2.4}$$

where Tr denotes the *trace*, *i.e.*, the sum of diagonal elements of a matrix.

If X is a random variable on Ω_n with the values $x_1, \ldots, x_n$, then putting

$$X = \begin{bmatrix} x_1 & & 0 \\ & \ddots & \\ 0 & & x_n \end{bmatrix}$$

we get that the expectation of X with respect to the probability distribution P is equal to

$$\sum_{j=1}^{n} p_j x_j = \operatorname{Tr} PX. \tag{1.2.5}$$

Consider the set $\mathfrak{O}_n$ of random variables, satisfying $0 \leq x_j \leq 1$, *i.e.*,

$$0 \leq X \leq I \tag{1.2.6}$$

where I is the unit $n \times n$-matrix. Then $\mathfrak{O}_n$ is a convex set – a unit hypercube, the extreme points being its vertices, represented by the matrices X for which x_j is either 0 or 1. Such matrices satisfy $X^2 = X$, so that the extreme points of $\mathfrak{O}_n$ are the idempotent (projection) matrices.

This elementary consideration leads naturally to the following construction which is of principal interest in connection with quantum theory. We can consider (complex) $n \times n$-matrices as operators acting in the n-dimensional unitary space $\mathscr{H}$ of column-vectors $\varphi = [\varphi_1]$, $\psi = [\psi_1], \ldots$ The inner product of φ and ψ is defined by: $(\varphi|\psi) = \sum \overline{\varphi}_j \psi_j$. We shall use Dirac's notation $|\varphi), |\psi), \ldots$ for column-vectors $\varphi, \psi, \ldots$ and $(\varphi|, (\psi|, \ldots$ for the Hermitean conjugated row-vectors $\varphi^*, \psi^*, \ldots$ The symbol for the inner product is then simply a graphic junction of symbols for the factors $(\varphi|$ and $|\psi)$. The "outer" product $|\psi)(\varphi|$ is then the $n \times n$-matrix with the components $[\psi_j \overline{\varphi}_k]$. If ψ is the unit vector, $(\psi|\psi) = 1$, then $S_\psi = |\psi)(\psi|$ is the matrix of the (orthogonal) projection on the vector ψ.

The finite-dimensional spectral theorem says that for any Hermitean $n \times n$-matrix X there is a complete orthonormal system of vectors $\{e_j;\ j = 1, \ldots, n\}$, in which X has the diagonal form

$$X = \sum_{j=1}^{n} \lambda_j |e_j)(e_j|, \tag{1.2.7}$$

where λ_j are the eigenvalues of X, which are real. It follows that e_j is an eigenvector of X corresponding to the eigenvalue λ_j.

In (1.2.7) the λ_j's are not necessarily different. Denote by $x_1, \ldots, x_m$ ($m \leq n$) the distinct eigenvalues of X numbered in the increasing order,

and by $E_k = \sum |e_j)(e_j|$ (the sum extends over all e_j belonging to λ_k) the matrix of projection onto the invariant subspace of X, corresponding to the eigenvalue λ_k. Then we have a different form of the spectral representation

$$X = \sum_{k=1}^{m} x_k E_k. \tag{1.2.8}$$

Consider the set $\mathfrak{S}_n$ of all Hermitean $n \times n$-matrices $S = [s_{jk}]$ satisfying the conditions

$$S \geq 0, \qquad \operatorname{Tr} S = 1, \tag{1.2.9}$$

which have the same form as (1.2.4). By the finite-dimensional spectral theorem

$$S = \sum_{j=1}^{n} \lambda_j S_{\psi_j}, \tag{1.2.10}$$

where λ_j are the eigenvalues of S, and $S_{\psi_j} = |\psi_j)(\psi_j|$ are the mutually orthogonal projections on the unit eigenvectors of S. The condition (1.2.7) implies that the eigenvalues of $S \in \mathfrak{S}_n$ constitute a probability distribution

$$\lambda_j \geq 0, \qquad \sum_{j=1}^{n} \lambda_j = 1.$$

In particular, $0 \leq \lambda_j \leq 1$ and (1.2.10) implies

$$S - S^2 = \sum_{j=1}^{n} \lambda_j (1 - \lambda_j) S_{\psi_j} \geq 0,$$

with the sign of equality attained if and only if $S = S_{\psi_k}$ for some ψ_k, *i.e.*, if S is a one-dimensional projection.

Proposition 1.2.3. *The set $\mathfrak{S}_n$ in convex, its extreme points being precisely one-dimensional projections.*

Proof. The first part of the statement follows from the fact that the conditions (1.2.9), defining $\mathfrak{S}_n$ sustain forming convex combinations. To prove that any one-dimensional projection S is an extreme point of $\mathfrak{S}_n$, assume that

$$\begin{aligned} &S = p_0 S_0 + p_1 S_1; \\ &p_0, p_1 > 1, \qquad p_0 + p_1 = 1. \end{aligned}$$

Taking square of this equality, subtracting and using the inequality $S \geq S^2$ for $S \in \mathfrak{S}_n$, we obtain

$$\begin{aligned} S - S^2 &= p_0(S_0 - S_0^2) + p_1(S_1 - S_1^2) + p_0 p_1 (S_0 - S_1)^2 \\ &\geq p_0 p_1 (S_0 - S_1)^2. \end{aligned}$$

Since S is a one-dimensional projection, then $S = S^2$, which implies that $S_0 = S_1$. Therefore S is an extreme point.

To prove the converse consider the spectral representation (1.2.10). Since $S_{\psi_j} \in \mathfrak{S}_n$ and $S_{\psi_j} \neq S_{\psi_k}$ for $j \neq k$, then for an extreme point S of $\mathfrak{S}_n$ the sum (1.2.10) can have only one nonzero term. Therefore $S = S_{\psi_j}$ for some ψ_j, which proves the proposition. □

The relation (1.2.10) is one of many possible representations of S as a convex combination of the extreme points.

We consider also the convex set $\mathfrak{X}_n$ of all Hermitean $n \times n$-matrices X, satisfying (1.2.6), and show that *the extreme points of this set are the* (*orthogonal*) *projections*, *i.e.*, (*Hermitean*) *matrices satisfying* $X^2 = X$. The proof of the statement that every projection is an extreme point is the same as in Proposition 1.2.3. To prove the converse, write the spectral decomposition of the matrix X in the form (1.2.8) with $0 \leq x_k \leq 1$ where x_k are the distinct eigenvalues of X, E_k is the projection onto the invariant subspace of X, corresponding to the eigenvalue x_k. Since $x_1 < \cdots < x_m$, then using in (1.2.8) the Abel transform and taking into account the equality $E_1 + \cdots + E_m = I$, we have

$$X = (1 - x_1) \cdot 0 + \sum_{k=1}^{m-1} (x_k - x_{k+1}) \cdot E'_k + x_m \cdot I,$$

where $E'_k = E_1 + \cdots + E_k$. Since the projections 0, I and E'_k belong to $\mathfrak{X}_n$ and the coefficients are nonnegative and sum to 1, then this is a convex combination of distinct projections. If X is an extreme point, then the sum can have only one nonzero term, and the matrix X must be a projection.

The difference between the sets $\mathfrak{P}_n$ and $\mathfrak{S}_n$ (correspondingly, between $\mathfrak{O}_n$ and $\mathfrak{X}_n$) is that in the latter case we consider all Hermitean matrices, satisfying (1.2.9) (correspondingly (1.2.6)), while in the former case only the diagonal matrices. We could consider as well a commuting family of matrices which can be simultaneously diagonalized. Therefore the latter case may be called the "noncommutative" analog to the former; defining the mean value by (1.2.5) one may treat $S \in \mathfrak{S}_n$ as a "noncommutative

probability distribution", and a Hermitean matrix X as a "noncommutative random variable". We shall see this connection to be deeper than a pure formal analogy.

We took the complex matrices because of their relevance to quantum theory, but the real matrices can be treated similarly.

At the end of this section we consider in some detail the structure of the convex set $\mathfrak{S}_n$ in the simplest "noncommutative" case, $n = 2$. Any matrix $S \in \mathfrak{S}_2$ can be represented as

$$S = \frac{1}{2}\begin{bmatrix} 1+\theta_3 & \theta_1 - i\theta_2 \\ \theta_1 + i\theta_2 & 1-\theta_3 \end{bmatrix}, \tag{1.2.11}$$

where $\theta_1, \theta_2, \theta_3$ are the real numbers called the *Stokes parameters*. The condition $S \geq 0$ is equivalent to the inequality $\theta_1^2 + \theta_2^2 + \theta_3^2 \leq 1$. Thus, $\mathfrak{S}_2$ as a convex set can be represented by the unit ball in the three-dimensional real vector space; the extreme points are the matrices for which the vector $[\theta_1, \theta_2, \theta_3]$ lies on the sphere $\theta_1^2 + \theta_2^2 + \theta_3^2 = 1$.

If $n > 2$, then the set $\mathfrak{S}_n$ is a proper subset of the unit ball in the $(n^2 - 1)$-dimensional real vector space, and it cannot be represented so explicitly.

1.3. Definition of a statistical model

Motivated by consideration in Section 1.1, we define a *statistical model*[1] as a pair $(\mathfrak{S}, \mathfrak{M})$ where $\mathfrak{S}$ is a convex set and $\mathfrak{M}$ is a class of affine maps of $\mathfrak{S}$ into the collections of probability distributions on some measurable spaces U. The elements of $\mathfrak{S}$ are called *states*, and the elements of $\mathfrak{M}$ *measurements*. The problem of theoretical description of an object or a phenomenon satisfying the statistical postulate can then be described as a problem of construction of an appropriate statistical model. In more detail, the construction must first give a mathematical description of the set $\mathfrak{S}$ of theoretical states and the set $\mathfrak{M}$ of theoretical measurements and second, prescribe the rules for correspondence between the real procedures of preparation and measurement and the theoretical objects, *i.e.*, an injection of the experimental data into the statistical model.

The probability theory and statistics deal with the models in which the set of the states $\mathfrak{S}$ has the specifically simple structure. The statistical model of quantum theory is drastically different. We shall consider these models in detail in the following sections.

[1] The concept of statistical model will be considered in the Supplement in greater detail.

In this chapter we shall often simplify our consideration by using only the measurements with a finite number of outcomes. In such a case the space U is finite and the probability distribution of the results of the measurement is described by a finite collection of real affine functionals $\{\mu_S(u); u \in U\}$ on $\mathfrak{S}$, satisfying

$$\mu_S(u) \geq 0, \qquad u \in U; \qquad \sum_{u \in U} \mu_S(u) = 1. \tag{1.3.12}$$

Here $\mu_S(u)$ is the probability of the result u when the state is S. For any $B \subset U$

$$\mu_S(B) = \sum_{u \in B} \mu_S(u).$$

Technically this case is much simpler than the continual one, being still sufficient to expose the essential features of the theory. In practice such measurements correspond to the procedures resulting in some classification of the data. Furthermore, one can easily imagine a finitely-valued approximation of a measurement with a continual space of the outcomes U by making a partition of U into a finite number of "small" pieces.

A two-valued measurement is called a *test*. Denoting one of the results of the test by 0, and the other by 1 we get that any test can be described by defining only one function on $\mathfrak{S}$, say $\mu_S(1)$, the probability of getting 1, since $\mu_S(0) = 1 - \mu_S(1)$. The probability $\mu_S(1)$ is an affine functional on $\mathfrak{S}$ satisfying $0 \leq \mu_S(1) \leq 1$.

Let $S \to \mu_S(\mathrm{d}u)$ be a measurement with an arbitrary space of results U. Then to any $B \in \mathscr{A}(U)$ there corresponds the test $S \to \{\mu_S(\overline{B}), \mu_S(B)\}$, the result of which is 0 if $u \in \overline{B}$ and is 1 if $u \in B$. ($\overline{B}$ denotes the complement of the set B). Thus, any measurement can be considered as a collection of tests (satisfying apparent compatibility conditions).

1.4. The classical statistical model

We have seen in Section 1.1 that the notion of state refers to the initial conditions of the experiment. Here we shall adopt that these conditions can be formally described by the points ω of some set Ω, which will be called *phase space*.

To take into account the possibility of variations in the initial data during the repetitions of an experiment, or uncertainties in some parameters in the preparation procedure we shall consider also the probability distributions on Ω. To do this we must accept that Ω is a measurable space with a σ-field $\mathscr{A}(\Omega)$; we assume that $\mathscr{A}(\Omega)$ separates the points of Ω.

Any probability distribution P on Ω will be called *classical state*. It should be interpreted as a statistical description of the preparation stage.

To any $\omega \in \Omega$ corresponds the *pure state* described by the δ-distribution $\delta_\omega(A)$, $A \in \mathscr{A}(\Omega)$. According to Section 1.2 the collection $\mathfrak{P}(\Omega)$ of all classical states is the convex set of the most simple structure, the simplex, and the pure states are its extreme points.

A measurement with values in U is described by an affine map

$$P \to \mu_P(\mathrm{d}u) \tag{1.4.13}$$

which transforms the set of classical states $\mathfrak{P}(\Omega)$ into the set of probability distribution $\mathfrak{P}(U)$. Denote by $M_\omega(\mathrm{d}u)$ the probability distribution of the given measurement with respect to a pure state δ_ω, so that $M_\omega(\mathrm{d}u) = \mu_{\delta_\omega}(\mathrm{d}u)$, and consider the mixture of the pure states

$$P(\mathrm{d}\omega) = \sum_\alpha p_\alpha \delta_{\omega_\alpha}(\mathrm{d}\omega).$$

Since (1.4.13) is affine, the probability distribution of the results with respect to this state will be given by

$$\mu_P(B) = \int M_\omega(B) P(\mathrm{d}\omega), \qquad B \in \mathscr{A}(U). \tag{1.4.14}$$

Under some additional assumptions this relation will be valid for any classical state P. We shall not discuss this question and simply restrict our attention to measurements $P \to \mu_P$, which have the form (1.4.14) where $M_\omega(\mathrm{d}u)$ is a *conditional probability distribution* on U[2]. While P describes the uncertainty in the initial conditions of the experiment, the probability distribution $M_\omega(\mathrm{d}u)$ characterizes the disturbance due to the measuring device. The relation (1.4.14) shows how these two sources of uncertainty enter into the overall measurement statistics. We shall denote by $\boldsymbol{M}$ both the conditional probability distribution $\{M_\omega(\mathrm{d}u)\}$ and the corresponding measurement (1.4.13).

The classical statistical model which we are going to define is based on the assumption of *complete observability*, according to which the values of any parameters of the object can be established with absolute precision. To give a precise formulation we introduce the following definition. The measurement $\boldsymbol{M} = \{M_\omega(\mathrm{d}u)\}$ is called *deterministic* if for any $\omega \in \Omega$ and $B \in \mathscr{A}(U)$ either $M_\omega(B) = 0$ or $M_\omega(B) = 1$. This means that if the object is in a pure state, then for any $B \in \mathscr{A}(U)$ the result of the

[2] This means that for any $\omega \in \Omega$, $M_\omega(\mathrm{d}u)$ is a probability distribution on U, and for any $B \in \mathscr{A}(U)$, $M_\omega(B)$ is a measurable function of ω.

measurement u is either in B or not in B with probability 1. This can be written in the following form

$$M_\omega(B)^2 = M_\omega(B), \qquad B \in \mathscr{A}(U). \tag{1.4.15}$$

The nature of this condition can be made clear by discussing finitely-valued measurements. Let $\boldsymbol{M} = \{M_\omega(u);\, u \in U\}$ be such a measurement, where $M_\omega(u)$ is the probability of the outcome u if the object is in the pure state δ_ω. These probabilities satisfy

$$M_\omega(u) \geq 0, \qquad \sum_u M_\omega(u) = 1; \qquad \omega \in \Omega. \tag{1.4.16}$$

If $\boldsymbol{M}$ is deterministic measurement, then $M_\omega(u)$ is equal to either 0 or 1. Introducing the *indicator* of a set $F \subset \Omega$ as the function $\mathbf{1}_F(\omega)$, which is equal to 1 on F and 0 outside F, we have $M_\omega(U) = \mathbf{1}_{\Omega_{(u)}}(\omega)$, where $\Omega_{(u)} = \{\omega : M_\omega(u) = 1\}$. It follows from (1.4.16) that the sets $\Omega_{(u)}$ for different values of u do not intersect, and the union of all $\Omega_{(u)}$ is equal to Ω; this is expressed by saying that the sets $\{\Omega_{(u)}\}$ form a *decomposition* of the set Ω. Therefore for any ω there is a unique $u = u(\omega)$ such that $M_\omega(u(\omega)) = 1$. For any $B \subset U$

$$M_\omega(B) = \sum_{u \in B} M_\omega(u) = \mathbf{1}_B(u(\omega)). \tag{1.4.17}$$

The function $\omega \to u(\omega)$ is a random variable on Ω with values in U; the relation (1.4.17) establishes the one-to-one correspondence between the deterministic measurements and the random variables with values in U. To make this connection more transparent consider a random variable $X(\omega)$ on Ω taking values in a finite subset $\{x\}$ of the real line $\mathbb{R}$. Let $\Omega_{(x)}$ be the subset of Ω on which $X(\omega)$ is equal to x, then

$$X(\omega) = \sum_X x \cdot \mathbf{1}_{\Omega_{(x)}}(\omega) = \sum_x x M_\omega(x). \tag{1.4.18}$$

Thus, to the random variable X there corresponds the unique deterministic measurement $\boldsymbol{M} = \{M_\omega(x)\}$ such that X takes a value x if and only if x is the result of the measurement $\boldsymbol{M}$.

The case of continuous random variables is technically much more involved but the conclusion is essentially the same: under some regularity conditions the relation (1.4.17) establishes the one-to-one correspondence between the random variables and the deterministic measurements.

We can now formalize the requirement of the complete observability by adopting the following definition. The *classical statistical model* is a