I. I. Gihman A. V. Škorohod

The Theory of Stochastic Processes I

Translated from the Russian
by S. Kotz

I. I. Gihman A.V. Skorohod

The Theory of Stochastic Processes I

Translated from the Russian
by S. Kotz

Springer-Verlag
Berlin Heidelberg New York 1974

Iosif Il'ich Gihman
Academy of Sciences of the Ukrainian SSR, Institute of Applied Mathematics and Mechanics, Donetsk

Anatoliĭ Vladimirovich Skorohod
Academy of Sciences of the Ukrainian SSR, Institute of Mathematics, Kiev

Translator:

Samuel Kotz
Department of Mathematics, Temple University, Philadelphia

Title of the Russian Original Edition: Teoriya sluchainyh protsessov, Vol. I. Publisher: Nauka, Moscow 1971

AMS Subject Classification (1970): 60-02, 60 G xx

ISBN 3-540-06573-3 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-06573-3 Springer-Verlag New York Heidelberg Berlin

 Library of Congress Catalog Card Number 74-2552. Printed in Germany. Typesetting: D. Reidel, Book Manufacturers, Dordrecht · Holland. Printing and bookbinding: Konrad Triltsch, Würzburg.

Preface

We have endeavoured in this planned three volume work to present an exposition of the basic results, methods and applications of the theory of random processes. The various branches of the theory are, however, not treated in equal detail.

This volume should be of value principally to mathematicians who are interested in studying the theory of random processes. We hope that researchers who apply the methods of the theory of random processes will also find the book interesting and useful. Prerequisites to the study of this book are basic courses in probability theory, measure theory and integration, and functional analysis.

The first volume of "*The Theory of Random Processes*" is devoted to general problems of the theory of random functions and measure theory in function spaces. Some of the material presented in the authors' book "*Introduction to the Theory of Random Processes*" (Ergebnisse der Mathematik Band 72) is utilized here. Chapters III, IV, V and IX of the *Introduction* have been revised and now constitute the contents of Chapters I, III, IV and VI respectively.

In volume II, the following topics are treated: the general theory of Markov processes, the theory of processes with independent increments, jump Markov processes, semi-Markov processes and branching processes.

The third volume deals with the theory of martingales, stochastic integrals, stochastic differential equations, diffusion processes and limit theorems associated with stochastic differential equations.

I. I. Gihman and A. V. Skorohod

Table of Contents

Chapter I

Basic Notions of Probability Theory

§ 1. Axioms and Definitions

Events. The basic notions of probability theory are experiment, event and probability of events.

A formal description of these notions is usually based on the set-theoretical model of probability theory developed by A. N. Kolmogorov in 1929.

The experiments studied in probability theory (referred to as stochastic experiments) are carried out when a certain set of conditions Y is satisfied. This set of conditions does not uniquely determine the results of the experiment (also called the outcome or realization). This means that if the experiment is repeated (provided that the set of conditions Y is accurately satisfied) the results of the experiment will generally be different.

When formalizing the notions of probability theory the first fundamental assumption is that the results of a collection of experiments under investigation in a given situation can be described by means of a certain set Ω. Every meaningful event (occurring or not during the given experiment) corresponds to a certain subset A of Ω in such a manner that the probabilistic operations on events correspond to set-theoretical operations on the corresponding subsets of Ω.

Moreover, the points $\omega \in \Omega$ correspond to atoms – namely, every event is a sum of points while each point ω cannot be represented as a sum of other events. For this reason the points belonging to Ω are called elementary events.

In relation to Ω, an experiment is completely characterized by the class of those events (subsets of Ω) such that one can assert in each case whether it did or did not occur during the given experiment. These events are called observable (in the given experiment).

Henceforth we shall adhere to this model of probability theory and identify events with the corresponding subsets of Ω. The resulting dual terminology is presented below in a glossary translating set-theoretic notions into probabilistic notions.

Set theory	Probability theory
Space Ω	Sure event
ω a point of Ω	Elementary event
0 the empty set	Impossible event
A a subset of Ω, $A \subset \Omega$	Event
The set A is contained in B $(A \subset B)$	Event A implies B
C the sum (union) of sets A and B $(C = A \cup B)$	C the sum (union) of events A and B
C the intersection of sets A and B $(C = A \cap B)$	C the intersection (or product) of events A and B
$\bar{A}$ the complement of set A	$\bar{A}$ the contrary event of A
C the difference of two sets A and B $(C = A \setminus B)$	C the difference of events A and B
Sets A and B are without common points $(A \cap B = 0)$	Events A and B are disjoint

We note that any arbitrary subset of Ω is called an event. However, from both a practical as well as a purely mathematical point of view it does not make sense to regard any arbitrary subsets of Ω as events worthy of interest. Therefore one must select out of Ω a suitable class of events. This class should be sufficiently wide and contain all the events which may arise during the solution of various practical problems. On the other hand, the size of this class is limited by the feasibility of effective utilization of mathematical techniques. Obviously, the problem of selecting the corresponding class of events should be solved individually in each case, however, we shall always assume subsequently that this class forms a σ-algebra of events.

Definition 1. A class of events $\mathfrak{A}$ is called an *algebra of events* if it contains the sure event Ω, the impossible event 0 and together with each pair of events A and B belonging to the class, their sum as well as the contrary event $\bar{A}$.

Two events Ω and 0 constitute the *trivial algebra*.

The minimal algebra containing event A consists of four events: Ω, 0, A and $\bar{A}$.

Definition 2. An algebra of events which contains a sequence of events along with their sum is called a *σ-algebra*.

It is clear that in the definitions and properties above we could have referred to algebras and σ-algebras of sets of a certain abstract space Ω.

Definition 3. The space Ω along with the σ-algebra of sets $\mathfrak{A}$ defined on it is called the *measurable space* $\{\Omega, \mathfrak{A}\}$ and the subsets of Ω belonging to $\mathfrak{A}$ are called *$\mathfrak{A}$-measurable sets* ($\mathfrak{A}$-measurable events) or simply *measurable sets* (events) if no ambiguity arises concerning the σ-algebra under consideration.

The σ-algebra of all the events under consideration in a given situation is usually denoted by the letter $\mathfrak{S}$. With respect to the measurable space $(\Omega, \mathfrak{S})$ any given stochastic experiment is completely characterized by the class of events $\mathfrak{F}$ observed during this experiment. Clearly, (this class is contained in Ω and it is also evident that the class $\mathfrak{F}$ is closed with respect to the operations of addition, intersection and complementation. It is therefore natural to consider $\mathfrak{F}$ a σ-algebra of events. Therefore, formally a stochastic experiment is determined by a certain σ-algebra $\mathfrak{F}$ of $\mathfrak{S}$-measurable events. We call it the *σ-algebra corresponding to the given experiment.*

Probability. Definition 4. A triple $(\Omega, \mathfrak{S}, \mathsf{P})$ consisting of a space of elementary events Ω, a selected σ-algebra of events $\mathfrak{S}$ in Ω, and a measure P defined on $\mathfrak{S}$ such that $\mathsf{P}(\Omega)=1$ is called a probability space and the measure P is called the probability.

Probability spaces are the initial objects of probability theory. This. however, does not contradict the fact that when solving many specific problems the probability space is not given explicitly.

We present below several of the simplest well known properties of probability which easily follow from its definition (S and S_n, $n=1, 2, \ldots$ as given below all belong to $\mathfrak{S}$):

a) $\mathsf{P}(\emptyset)=0$;

b) if $S_k \cap S_r = 0$, $k \neq r$, then $\mathsf{P}\left(\bigcup_1^\infty S_k\right)=\sum_1^\infty \mathsf{P}(S_k)$;

c) if $S_1 \subset S_2$, then $\mathsf{P}(S_2 \setminus S_1)=\mathsf{P}(S_2)-\mathsf{P}(S_1)$;

d) $\mathsf{P}(\bar{S})=1-\mathsf{P}(S)$;

e) if $S_n \subset S_{n+1}$, $n=1, 2, \ldots$ then $\mathsf{P}\left(\bigcup_1^\infty S_n\right)=\lim \mathsf{P}(S_n)$;

f) if $S_n \supset S_{n+1}$, $n=1, 2, \ldots$ then $\mathsf{P}\left(\bigcap_1^\infty S_n\right)=\lim \mathsf{P}(S_n)$.

Random variables. The concept of a random variable corresponds to the description of a stochastic experiment which measures a certain numerical quantity ξ. It is assumed that for any pair of numbers a and b $(a<b)$ the event $A(a, b)$ expressing that $\xi \in (a, b)$ is an observable event.

The minimal σ-algebra $\mathfrak{F}_\xi$ containing all the events $A(a, b)$, $-\infty<a<b<\infty$ is the σ-algebra corresponding to this stochastic experiment.

Let $A_x(-\infty<x<\infty)$ denote the event $\xi=x$. This event is measurable. Indeed $A_x = \bigcap_{n=1}^{\infty} A\left(x - \frac{1}{n},\ x + \frac{1}{n}\right)$. Moreover, if $x_1 \neq x_2$, events A_{x_1} and A_{x_2} are disjoint (this follows from the single-valuedness of the measurement results) and the union of all A_x, $-\infty < x < \infty$, is the set Ω, since the measurement result is always represented by some real number. We now define a single-valued real function $f(\omega)$, $\omega \in \Omega$ by setting $f(\omega) = x$ if $\omega \in A_x$. It follows from the definition, that $\xi = f(\omega)$ in each experiment and, moreover, that the set $\{\omega : a < f(\omega) < b\} = A(a, b)$ is measurable. Recall that a real-valued function $f(\omega)$ defined on a measurable space $\{\Omega, \mathfrak{S}\}$ is called *measurable* ($\mathfrak{S}$-measurable) if for any two real numbers a and b the set $\{\omega : a < f(\omega) < b\} \in \mathfrak{S}$. Therefore, a random variable ξ can be identified with a certain measurable function on the probability space $(\Omega, \mathfrak{S}, \mathsf{P})$.

Definition 5. A $\mathfrak{S}$-measurable real-valued function of elementary events ω is called a *random variable* ξ (on a given probability space $\{\Omega, \mathfrak{S}, \mathsf{P}\}$).

Henceforth, we shall occasionally consider measurable functions on $\{\Omega, \mathfrak{S}, \mathsf{P}\}$ which may possibly take on the values $\pm\infty$ also, or functions which are defined only on a measurable subset of $\{\Omega, \mathfrak{S}, \mathsf{P}\}$. These functions are called generalized random variables.

We note the following point connected with the definition of a random variable. It is commonly assumed that from the empirical point of view one cannot distinguish between events which differ on an event of probability zero. It would therefore be natural to identify two random variables ξ and η which are equal to each other with probability 1 and hence interpret a random variable as a class of measurable functions, in which each pair of functions may differ only on a set of probability 0. Such functions are called equivalent (or P-equivalent). This point of view is also justified by the fact that the majority of notions introduced here as well as the relationships obtained refer essentially to classes of equivalent functions. However, a consistent adherence to this point of view presents certain technical as well as basic problems. For this reason it would seem more convenient to regard random variables as individual functions and use special notation for their equivalent classes.

Definition 6. Random variables ξ and η are called *equivalent* (P-equivalent) if $\mathsf{P}\{\xi \neq \eta\} = 0$. The P-equivalence of 2 random variables ξ and η is denoted by $\xi = \eta \pmod{\mathsf{P}}$.

Equivalent random variables are also referred to as satisfying $\xi = \eta$ almost surely (a.s.) or $\xi = \eta$ with probability 1.

Analogous terminology and notation is also used in more general cases. We thus say that a certain function (or certain other objects) possess property H almost surely (for almost all ω or for all ω (modP)) if the set of ω for which this property is not satisfied is of probability 0. For example, if a sequence of random variables $\xi_n = f_n(\omega)$ converges to $\xi = f(\omega)$ for each ω except for a certain set N and $\mathsf{P}(N) = 0$, we say that ξ_n converges to ξ almost surely or that

$$\xi = \lim \xi_n \quad (\mathrm{mod}\,\mathsf{P}).$$

We now present a number of basic properties of random variables which follow directly from the corresponding properties of arbitrary measurable functions. It is assumed that the random variables are defined on a fixed probability space $\{\Omega, \mathfrak{S}, \mathsf{P}\}$.

a) If $h(t_1, t_2, \ldots, t_n)$ is an arbitrary Borel function of n real variables $t_1, \ldots, t_n$, and $\xi_1, \xi_2, \ldots, \xi_n$ are random variables, then $h(\xi_1, \xi_2, \ldots, \xi_n)$ is also a random variable.

b) If $\{\xi_n; n = 1, 2, \ldots\}$ is a sequence of random variables, then, $\sup \xi_n$, $\inf \xi_n$, $\overline{\lim}\, \xi_n$, $\underline{\lim}\, \xi_n$ are also random variables.

Hence a very wide class of analytic operations commonly performed on functions transforms a random variable into a random variable independently of the specific form of the σ-algebra $\mathfrak{S}$. It is easy to see that these operations do not interfere with the equivalence relations between the random variables. More precisely:

c) If ξ_n and η_n are equivalent ($n = 1, 2, \ldots$), and $h(t_1, t_2, \ldots, t_n)$ is a Borel function of n real variables, then $h(\xi_1, \xi_2, \ldots, \xi_n)$ and $h(\eta_1, \eta_2, \ldots, \eta_n)$ are also equivalent. Moreover, the following pairs of random variables are equivalent as well: $\sup \xi_n$ and $\sup \eta_n$, $\inf \xi_n$ and $\inf \eta_n$, $\overline{\lim}\, \xi_n$ and $\overline{\lim}\, \eta_n$, $\underline{\lim}\, \xi_n$ and $\underline{\lim}\, \eta_n$.

d) Let ξ_n, $n = 1, 2, \ldots$ be a sequence of random variables. The event $S = \{\lim \xi_n \text{ exists}\}$ is $\mathfrak{S}$-measurable. It is easy to verify that this event can be represented as:

$$S = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m_1, m_2 > n} \left\{\omega : |\xi_{m_1} - \xi_{m_2}| < \frac{1}{k}\right\}.$$

Indicators of events serve as an important example of random variables. The *indicator of an event* A is a random variable $\chi_A = \chi_A(\omega)$ defined as follows:

$$\chi_A(\omega) = 1 \quad \text{if} \quad \omega \in A$$
$$\chi_A(\omega) = 0 \quad \text{if} \quad \omega \notin A.$$

If $A \in \mathfrak{S}$, then $\chi_A(\omega)$ is $\mathfrak{S}$-measurable.

Note the correspondence between set-theoretical operations on events and the analogous algebraic operations on indicators:

$$\chi_{\bigcup_{k=1}^{\infty} A_k}(\omega)=\sum_{k=1}^{\infty} \chi_{A_k}(\omega), \quad \text{if} \quad A_k \cap A_r = 0 \quad \text{for} \quad k \neq r,$$

$$\chi_{A \cap B}(\omega)=\chi_A(\omega)\,\chi_B(\omega),$$

$$\chi_{A \setminus B}(\omega)=\chi_A(\omega)-\chi_B(\omega), \quad \text{if} \quad B \subset A,$$

$$\chi_{\overline{\lim} A_n}(\omega)=\overline{\lim}\, \chi_{A_n}(\omega), \qquad \chi_{\underline{\lim} A_n}(\omega)=\underline{\lim}\, \chi_{A_n}(\omega).$$

A random variable ξ is called *discrete* if it admits only a finite or countable number of distinct values. Such a variable can be expressed as $\xi=\sum_k c_k \chi_{A_k}(\omega)$, where A_k are $\mathfrak{S}$-measurable sets pairwise disjoint and $\bigcup_k A_k=\Omega$. For each ω only one summand is nonzero in the r.h.s. of the last equality and $\xi=c_k$ if $\omega \in A_k$. For an arbitrary random variable ξ one can always construct a sequence ξ_n of discrete random variables taking on only a finite number of possible values and converging to ξ for each ω. To prove this assertion it is sufficient to set

$$\xi_n=\sum_{j=-n}^{n-1} \sum_{k=1}^{n} \left(j+\frac{k-1}{n}\right) \chi_{A_{jk}},$$

where

$$A_{jk}=\left\{\omega: j+\frac{k-1}{n} \leqslant \xi < j+\frac{k}{n}\right\}.$$

It then follows that $|\xi-\xi_n|<\frac{1}{n}$, if $|\xi|<n$.

It is easy to verify that for a non-negative ξ one can construct a monotonically increasing sequence of non-negative discrete random variables (taking on a countable number of values) uniformly converging to ξ. Indeed, in this case we set

$$\xi_n=\sum_{k=0}^{\infty} \frac{k}{2^n} \chi_{A_{kn}}, \quad \text{where} \quad A_{kn}=\left\{\omega: \frac{k}{2^n} \leqslant \xi < \frac{k+1}{2^n}\right\}.$$

Then $0 \leqslant \xi-\xi_n<2^{-n}$ for all ω.

Random elements. The notion of a random variable can be generalized to the notion of a random element with the values in an arbitrary measurable space $\{\mathscr{X}, \mathfrak{B}\}$. Let $\{\Omega, \mathfrak{S}\}$ and $\{\mathscr{X}, \mathfrak{B}\}$ be two measurable spaces. The mapping $g: \omega \rightarrow x$ $(x \in \mathscr{X})$ is called a measurable mapping of $\{\Omega, \mathfrak{S}\}$ into $\{\mathscr{X}, \mathfrak{B}\}$ if $g^{-1}(B)=\{\omega: g(\omega) \in B\} \in \mathfrak{S}$ for an arbitrary $B \in \mathfrak{B}$.

Definition 7. *A random element ξ with values in a measurable space* $\{\mathscr{X}, \mathfrak{B}\}$ is a measurable mapping of $\{\Omega, \mathfrak{S}, \mathsf{P}\}$ into $\{\mathscr{X}, \mathfrak{B}\}$.

If $\mathscr{X}$ is a metric space then $\mathfrak{B}$ is always assumed to be a σ-algebra of Borel sets (unless stipulated otherwise). If $\mathscr{X}$ is a vector space, then ξ is called a *random vector*.

Let a sequence of random elements $\{\xi_k; k=1, 2, \dots, n\}$ be given, defined on a fixed probability space $\{\Omega, \mathfrak{S}, \mathsf{P}\}$ with values in the spaces $\{\mathscr{X}_k, \mathfrak{B}_k\}$ correspondingly. This sequence can be considered as a single random element ζ, which will be called the direct product of random elements $\xi_1, \dots, \xi_n$, with values in a measurable space $\{\mathscr{Y}, \mathfrak{B}\}$ where $\mathscr{Y} = \prod_{k=1}^{n} \mathscr{X}_k$ is the product of the spaces $\mathscr{X}_1, \mathscr{X}_2, \dots, \mathscr{X}_n$ and $\mathfrak{B} = \prod_1^n \mathfrak{B}_k$ is the product of the σ-algebras $\mathfrak{B}_1, \mathfrak{B}_2, \dots, \mathfrak{B}_n$.

The last remark is also valid in the more general case of an arbitrary set of random elements ξ_α, $\alpha \in A$, with the values in $\{\mathscr{X}_\alpha, \mathfrak{B}_\alpha\}$ where A is a set of indices. Here the product $\mathscr{Y} = \prod_{\alpha \in A} \mathscr{X}_\alpha$ represents the space of all the mappings $y = y(\alpha)$: $\alpha \to x_\alpha$; $x_\alpha \in \mathscr{X}_\alpha$, $\alpha \in A$, i.e. the space of all functions defined on A admitting a value in $\mathscr{X}_\alpha$ for each $\alpha \in A$.

A cylindrical set in $\mathscr{Y}$ is called a set C of all $y \in \mathscr{Y}$ satisfying the relations of the type

$$y(\alpha_k) \in B_{\alpha_k}, \; k=1, \dots, n, \; B_{\alpha_k} \in \mathfrak{B}_{\alpha_k}.$$

Here n is an arbitrary integer and α_k are arbitrary elements of A. More precisely, we call $C = C_{\alpha_1 \dots \alpha_n}(B_{\alpha_1} \times \dots \times B_{\alpha_n})$ a cylindrical set with the bases $B_{\alpha_1} \times B_{\alpha_2} \times \dots \times B_{\alpha_n}$ over the coordinates $\alpha_1, \alpha_2, \dots, \alpha_n$. The minimal σ-algebra containing all the cylindrical sets is denoted by $\mathfrak{B}$ and is called the product of σ-algebras $\mathfrak{B}_\alpha$, $\mathfrak{B} = \prod_{\alpha \in A} \mathfrak{B}_\alpha$. It is easy to observe that the mapping $g: \omega \to y(\alpha)$ defined by the relations $g(\omega) = g(\omega, \alpha) = f_\alpha(\omega)$ where $f_\alpha(\omega) = \xi_\alpha$ is a measurable mapping of $\{\Omega, \mathfrak{S}\}$ into $\{\mathscr{Y}, \mathfrak{B}\}$. If all $\mathscr{X}_\alpha$ are the same, $\mathscr{X}_\alpha = \mathscr{X}$, then $\mathscr{Y} = \mathscr{X}^A$ represents the space of all functions with values in $\mathscr{X}$ defined on A and the mapping $g(\omega)$ associates a function from $\mathscr{X}^A$ with each elementary event ω; in other words the mapping $g(\omega)$ is a random function. Thus, the family of random variables $\{\xi_\alpha, \alpha \in A\}$ may be regarded as a random function.

Let $\xi = f(\omega)$ be a random element with the values in $\{\mathscr{Y}, \mathfrak{B}\}$.

Definition 8. *A σ-algebra generated by a random element* ξ is a σ-algebra σ_ξ or $\sigma(\xi)$ consisting of all sets of the form $\{f^{-1}(B); B \in \mathfrak{B}\}$.

Clearly the class of sets $\{f^{-1}(B); B \in \mathfrak{B}\}$ is a σ-algebra.

The following statement is an equivalent formulation of the above: the σ-algebra σ_ξ is the minimal σ-algebra in Ω with respect to which the random element ξ is measurable.

It is intuitively clear that measurability of a certain random variable η with respect to σ_ξ means that η is a function of ξ.

Lemma 1. *Let* $\xi=f(\omega)$ *be a random element on* $(\Omega, \mathfrak{S}, \mathsf{P})$ *with values in* $\{\mathscr{X}, \mathfrak{B}\}$ *and* η *be a* σ_ξ*-measurable random variable. Then there exists a* $\mathfrak{B}$*-measurable real valued function* $g(x)$ *such that* $\eta=g(\xi)$.

Proof. Assume that η is a discrete random variable admitting values a_n, $n=1,2,\dots$. Let $A_n=\{\omega:\eta=a_n\}$. Then there exists $B_n\in\mathfrak{B}$ such that $f^{-1}(B_n)=A_n$. Put $B_n'=B_n\setminus\bigcup_{k=1}^{n-1} B_k$. The sets $B_n'\in\mathfrak{B}$ are disjoint, $f^{-1}(B_n')=$

$$=A_n\setminus\bigcup_{k=1}^{n-1} A_k=A_n, \text{ and } f^{-1}\left(\bigcup_1^\infty B_n'\right)=\bigcup_1^\infty A_n=\Omega, \text{ i.e. } f(\Omega)\subset\bigcup_1^\infty B_n'. \text{ Now}$$

put $g(x)=a_n$ if $x\in B_n'$. Then $\eta=g(\xi)$.

We now consider the general case. There exists a sequence of discrete σ_ξ-measurable random variables η_n, convergent to η for each ω. Therefore $\eta_n=g_n(\xi)$, where $g_n(x)$ is $\mathfrak{B}$-measurable. The set of points S on which the functions $g_n(x)$ converge to a certain point is $\mathfrak{B}$-measurable, it contains $f(\Omega)$ and $\lim g_n(x)=\lim\eta_n=\eta$ for $x\in f(\Omega)$. Putting $g(x)=\lim g_n(x)$ for $x\in S$ and $g(x)=0$ for $x\notin S$ we obtain $\eta=g(\xi)$. □

Mathematical expectation. The mathematical expectation of a random variable is its most important numerical characteristic. This notion corresponds to the intuitive notion of the value of the arithmetic mean of observations on a random variable in a long sequence of identical stochastic experiments.

By definition the *mathematical expectation* of a random variable $\xi=f(\omega)$ is equal to the integral of $f(\omega)$ with respect to the measure P. We denote it as

$$\mathsf{E}\xi=\int_\Omega f(\omega)\,\mathsf{P}(d\omega)=\int_\Omega \xi\,d\mathsf{P}.$$

Often the designation Ω of the region of integration is omitted. Mathematical expectation possesses a number of properties which are well known from the theory of abstract integration.

Convergence in probability. Various types of convergence of sequences of random variables play an important role in probability theory. The definition of convergence with probability 1 (almost surely) was presented earlier.

Definition 9. If there exists a random variable ξ such that for any $\varepsilon>0$

$$\mathsf{P}\{|\xi_n-\xi|>\varepsilon\}\to 0 \quad \text{as} \quad n\to\infty,$$

we say that the *sequence* $\{\xi_n; n=1, 2, \ldots\}$ *converges in probability to the random variable* ξ and denote

$$\xi = \text{P-lim}\, \xi_n .$$

In measure theory convergence in probability corresponds to convergence in measure. The following corollaries follow from the general results of measure theory:

a) *If a sequence* $\{\xi_n; n=1, 2, \ldots\}$ *converges almost surely it converges in probability. The converse is generally not true. However, a subsequence which converges almost surely can be selected from a sequence of random variables convergent in probability.*

b) *A necessary and sufficient condition for convergence in probability of a sequence of random variables is as follows*: for arbitrary $\varepsilon > 0$ and $\delta > 0$ an $n_0 = n(\varepsilon, \delta)$ can be found such that for n and $n' > n_0$

$$\mathsf{P}\{|\xi_{n'} - \xi_n| > \varepsilon\} < \delta .$$

This condition is called the *condition of fundamentality in probability of the sequence* $\{\xi_n, n=1, 2, \ldots\}$.

c) If $\xi = \text{P-lim}\, \xi_n$ and $\eta = \text{P-lim}\, \xi_n$ then $\xi = \eta \pmod{\mathsf{P}}$.

d) *Let* $\eta_k = \text{P-lim}\, \xi_{kn} (k = 1, 2, \ldots, m)$ *and let the function* $\varphi(t_1, t_2, \ldots, t_m)$ *be everywhere continuous in the m-dimensional Euclidean space* $\mathscr{R}^m$, *except possibly on a Borel set* $D (D \subset \mathscr{R}^m)$ *such that*

$$\mathsf{P}\{(\eta_1, \eta_2, \ldots, \eta_m) \in D\} = 0 .$$

Then the sequence $\zeta_n = \varphi(\xi_{1n}, \xi_{2n}, \ldots, \xi_{mn})$ *converges in probability to* $\eta = \varphi(\eta_1, \eta_2, \ldots, \eta_m)$. *In particular, if the sequences* ξ_{kn} *are convergent in probability, so are the sequences* $\xi_{1n} + \xi_{2n}$, $\xi_{1n}\xi_{2n}$ *and* ξ_{1n}/ξ_{2n}, *the latter under the assumption that* $\mathsf{P}\{\text{P-lim}\, \xi_{2n} = 0\} = 0$ *and, moreover*

$$\text{P-lim}(\xi_{1n} + \xi_{2n}) = \text{P-lim}\, \xi_{1n} + \text{P-lim}\, \xi_{2n}, \qquad \text{P-lim}\, \frac{\xi_{1n}}{\xi_{2n}} = \frac{\text{P-lim}\, \xi_{1n}}{\text{P-lim}\, \xi_{2n}} .$$
$$\text{P-lim}(\xi_{1n} \cdot \xi_{2n}) = \text{P-lim}\, \xi_{1n} \text{P-lim}\, \xi_{2n},$$

A sufficient condition for convergence with probability 1 as stated below is useful in various specific problems:

Lemma 2. If there exists a sequence $\varepsilon_n > 0$, such that

$$\sum_{n=1}^{\infty} \mathsf{P}\{|\xi_{n+1} - \xi_n| > \varepsilon_n\} < \infty, \quad \sum_{n=1}^{\infty} \varepsilon_n < \infty,$$

then ξ_n *converges with probability* 1 *to a certain random variable* ξ. *If for any* $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} \mathsf{P}\{|\xi - \xi_n| > \varepsilon\} < \infty ,$$

then ξ_n converges to ξ with probability 1.

Proof. Let A_n denote the event $|\xi_{n+1}-\xi_n|>\varepsilon_n$. Then

$$\mathsf{P}(\overline{\lim}\, A_n)=\mathsf{P}\left(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} A_n\right)\leqslant \lim_{m\to\infty}\sum_{m}^{\infty}\mathsf{P}(A_n)=0.$$

Therefore, the terms of the series $\xi_1+\sum_{1}^{\infty}(\xi_{n+1}-\xi_n)$ starting with some index $m=m(\omega)$ are dominated with probability 1 by the terms of the convergent series $\sum_{n=1}^{\infty}\varepsilon_n$. This proves the first assertion. Next, let

$$B_{Nn}=\left\{|\xi-\xi_n|>\frac{1}{N}\right\}.$$

Then

$$\mathsf{P}\{\lim|\xi-\xi_n|>0\}=\mathsf{P}\left\{\bigcup_{N=1}^{\infty}\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} B_{Nn}\right\}\leqslant \lim_{N\to\infty}\lim_{m\to\infty}\sum_{n=m}^{\infty}\mathsf{P}(B_{Nn})=0,$$

which proves the second assertion. □

$\mathscr{L}_p$-spaces. By $\mathscr{L}_p=\mathscr{L}_p(\Omega,\mathfrak{S},\mathsf{P})$ $(p\geqslant 1)$ we denote a linear normed space of random variables ξ on $(\Omega,\mathfrak{S},\mathsf{P})$ satisfying $\mathsf{E}|\xi|^p<\infty$. The norm in $\mathscr{L}_p$ is defined by

$$\|\xi\|=\{\mathsf{E}|\xi|^p\}^{1/p}.$$

The convergence of the sequence ξ_n to its limit ξ in $\mathscr{L}_p$ (the $\mathscr{L}_p$-convergence) signifies that

$$\mathsf{E}|\xi-\xi_n|^p\to 0 \quad \text{as} \quad n\to\infty.$$

The $\mathscr{L}_p$-convergence implies convergence in probability. This fact follows directly from Chebyshev's inequality

$$\mathsf{P}\{|\xi_n-\xi|>\varepsilon\}\leqslant\frac{\mathsf{E}|\xi-\xi_n|^p}{\varepsilon^p}.$$

The space $\mathscr{L}_p$ is complete. The most important $\mathscr{L}_p$-spaces are $\mathscr{L}_1=\mathscr{L}$ and $\mathscr{L}_2$. We shall now discuss $\mathscr{L}_2$ in some detail. Note that all the definitions above and the theorems in this section are valid with no modifications for the complex-valued random variables.

The space $\mathscr{L}_2=\mathscr{L}_2(\Omega,\mathfrak{S},\mathsf{P})$ of complex-valued random variables becomes a Hilbert space if we define in $\mathscr{L}_2$, for each pair of random variables ξ and η, their scalar product putting it equal to $\mathsf{E}\xi\bar{\eta}$.

Two random variables ζ and η are called orthogonal if $\mathsf{E}\zeta\bar{\eta}=0$. In the case when ζ and η are real and $\mathsf{E}\zeta=\mathsf{E}\eta=0$, orthogonality is equivalent to the property that variables are uncorrelated. Convergence of the