I.A. Ibragimov
Y.A. Rozanov

Gaussian Random Processes

Translated by A.B. Aries

Applications of Mathematics

9

I.A. Ibragimov
Y.A. Rozanov

Gaussian Random Processes

Translated by A.B. Aries

Springer-Verlag
New York Heidelberg Berlin

I.A. Ibragimov
Lomi
Fontanka 25
Leningrad 191011
U.S.S.R.

Y.A. Rozanov
V.A. Steklov Mathematics Institute
Zazilov St. 42
Moscow 3-333
U.S.S.R.

AMS Subject Classifications: 60G10, 60G15, 60G35

Library of Congress Cataloging in Publication Data
Ibragimov, Il'dar Abdulovich.
Gaussian random processes.
(Applications of mathematics ; 9)
Translation of Gaussovskie sluchainye protsessy.
Bibliography: p.
1. Stochastic processes. I. Rozanov, Iurii
Antol'evich, joint author. II. Title.
QA274.4.I2613 519.2 78-6705

The original Russian edition GAUSSOVSKIE SLUCHAINYE PROTSESSY
was published in 1970 by Nauka.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-90302-x Springer-Verlag New York
ISBN 3-540-90302-x Springer-Verlag Berlin Heidelberg

Preface

The book deals mainly with three problems involving Gaussian stationary processes. The first problem consists of clarifying the conditions for mutual absolute continuity (equivalence) of probability distributions of a "random process segment" and of finding effective formulas for densities of the equivalent distributions. Our second problem is to describe the classes of spectral measures corresponding in some sense to regular stationary processes (in particular, satisfying the well-known "strong mixing condition") as well as to describe the subclasses associated with "mixing rate". The third problem involves estimation of an unknown mean value of a random process, this random process being stationary except for its mean, i.e., it is the problem of "distinguishing a signal from stationary noise". Furthermore, we give here auxiliary information (on distributions in Hilbert spaces, properties of sample functions, theorems on functions of a complex variable, etc.).

Since 1958 many mathematicians have studied the problem of equivalence of various infinite-dimensional Gaussian distributions (detailed and systematic presentation of the basic results can be found, for instance, in [23]). In this book we have considered Gaussian stationary processes and arrived, we believe, at rather definite solutions.

The second problem mentioned above is closely related with problems involving ergodic theory of Gaussian dynamic systems as well as prediction theory of stationary processes. From a probabilistic point of view, this problem involves the conditions for weak dependence of the "future" of the process on its "past". The employment of these conditions has resulted in a fruitful theory of limit theorems for weakly dependent variables (see, for instance, [14], [22]); the best known condition of this kind is obviously the so-called condition of "strong mixing". The problems arising in considering regularity conditions reduce in the case of Gaussian processes to a peculiar approximation problem related to linear spectral theory. The book contains the results of investigations of this problem which helped solve it almost completely.

The problem of estimating the mean is perhaps the oldest and most widely known in mathematical statistics. There are two approaches to the solution of this problem: first, the best unbiased estimates can be constructed on the basis of the spectral density of stationary noise; otherwise the least squares method can be applied.

We suggest one common class of "pseudobest" estimates to include best unbiased estimates as well as classical least squares estimates. For these "pseudobest" estimates explicit expressions are given, consistency conditions are found, asymptotic formulas are derived for the error correlation matrix, and conditions for asymptotic effectiveness are defined. It should be mentioned that the results relevant to regularity conditions and the mean estimation are formulated in spectral terms and can automatically be carried over (within the "linear theory") to arbitrary wide-sense stationary processes.

Each chapter has its own numbering of formulas, theorems, etc. For example, formula (4.21) means formula 21 of Section 4 of the same chapter where the reference is made. For the convenience of the reader we provide references to textbooks or reference books. The references are listed at the end of the book.

Contents

CHAPTER I
Preliminaries 1

I.1 Gaussian Probability Distribution in a Euclidean Space 1
I.2 Gaussian Random Functions with Prescribed Probability Measure 2
I.3 Lemmas on the Convergence of Gaussian Variables 5
I.4 Gaussian Variables in a Hilbert Space 7
I.5 Conditional Probability Distributions and Conditional Expectations 13
I.6 Gaussian Stationary Processes and the Spectral Representation 16

CHAPTER II
The Structures of the Spaces $H(T)$ and $L_T(F)$ 28

II.1 Preliminaries 28
II.2 The Spaces $L^+(F)$ and $L^-(F)$ 35
II.3 The Construction of Spaces $L_T(F)$ When T Is a Finite Interval 38
II.4 The Projection of $L^+(F)$ on $L^-(F)$ 45
II.5 The Structure of the σ-algebra of Events $\mathfrak{U}(T)$ 55

CHAPTER III
Equivalent Gaussian Distributions and their Densities 63

III.1 Preliminaries 63
III.2 Some Conditions for Gaussian Measures to be Equivalent 74
III.3 General Conditions for Equivalence and Formulas for Density of Equivalent Distributions 85
III.4 Further Investigation of Equivalence Conditions 90

CHAPTER IV
Conditions for Regularity of Stationary Random Processes 108

IV.1 Preliminaries 108
IV.2 Regularity Conditions and Operators B_τ 114
IV.3 Conditions for Information Regularity 125
IV.4 Conditions for Absolute Regularity and Processes with Discrete Time 129
IV.5 Conditions for Absolute Regularity and Processes with Continuous Time 138

CHAPTER V
Complete Regularity and Processes with Discrete Time 144

V.1 Definitions and Preliminary Constructions with Examples 144
V.2 The First Method of Study: Helson—Sarason's Theorem 147
V.3 The Second Method of Study: Local Conditions 153
V.4 Local Conditions (continued) 163
V.5 Corollaries to the Basic Theorems with Examples 177
V.6 Intensive Mixing 181

CHAPTER VI
Complete Regularity and Processes with Continuous Time 191

VI.1 Introduction 191
VI.2 The Investigation of a Particular Function $\gamma(T; \mu)$ 195
VI.3 The Proof of the Basic Theorem on Necessity 200
VI.4 The Behavior of the Spectral Density on the Entire Line 207
VI.5 Sufficiency 212
VI.6 A Special Class of Stationary Processes 217

CHAPTER VII
Filtering and Estimation of the Mean 224

VII.1 Unbiased Estimates 224
VII.2 Estimation of the Mean Value and the Method of Least Squares 237
VII.3 Consistent Pseudo-Best Estimates 244
VII.4 Estimation of Regression Coefficients 251

References 274

CHAPTER I

Preliminaries

I.1 Gaussian Probability Distributions in a Euclidean Space

A probability distribution P in an n-dimensional vector space $\mathbb{R}^n$ is said to be *Gaussian* if the characteristic function

$$\varphi(u) = \int_{\mathbb{R}^n} e^{i(u,x)} \mathsf{P}(dx), \qquad u \in \mathbb{R}^n$$

(here $(u, x) = \sum u_k x_k$ denotes the scalar product of vectors $u = (u_1, \ldots, u_n)$ and $x = (x_1, \ldots, x_n)$) has the form

$$\varphi(u) = \exp\left\{ i(u, a) - \frac{1}{2}(Bu, u) \right\}, \qquad u \in \mathbb{R}^n, \tag{1.1}$$

where $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ is the *mean* and B is a linear self-adjoint non-negative definite operator called a *correlation operator*; the matrix $\{B_{kj}\}$ defining B is said to be a *correlation matrix*. In this case

$$(u, a) = \int_{\mathbb{R}^n} (u, x) \mathsf{P}(dx),$$

$$(Bu, v) = \int_{\mathbb{R}^n} [(u, x) - (u, a)][(v, x) - (v, a)] \mathsf{P}(dx), \tag{1.2}$$

$$u, v \in \mathbb{R}^n.$$

The distribution P with mean value a and correlation operator B is concentrated in an m-dimensional hyperplane L of $\mathbb{R}^n$ (m being the rank of the correlation matrix), which can be expressed as

$$L = a + B\mathbb{R}^n$$

(L being the totality of all vectors $y \in \mathbb{R}^n$ of the form $y = a + Bx$, $x \in \mathbb{R}^n$).

In fact,

$$\mathsf{P}(\mathbb{R}^n \backslash L) = 0,$$

the distribution P being absolutely continuous with respect to Lebesgue measure dy in the hyperplane L, so that

$$\mathsf{P}(\Gamma) = \int_{\Gamma \cap L} p(y)\,dy, \tag{1.3}$$

where the distribution density $p(y)$, $y \in L$, has the form

$$p(y) = \frac{1}{(2\pi)^{m/2} \det B} \exp\left\{-\frac{1}{2}(B^{-1}(y-a), (y-a))\right\}. \tag{1.4}$$

Here $\det B$ denotes the determinant of the matrix that prescribes the operator B in the subspace $\mathbb{R}^m = B\mathbb{R}^n$, and B^{-1} is the inverse on this subspace.

I.2 Gaussian Random Functions with Prescribed Probability Measure

Let $(\Omega, \mathfrak{A}, \mathsf{P})$ be a probability space, i.e., a measurable space of elements $\omega \in \Omega$ with probability measure P on a σ-algebra $\mathfrak{A}$ of sets $A \subseteq \Omega$.

Any real measurable function $\xi = \xi(\omega)$ on a space Ω is said to be a random variable. The totality of random variables $\xi(t) = \xi(\omega, t)$ (parameter t runs through a set T) is said to be a random function of parameter $t \in T$. The random variables $\xi(t)$ themselves are said to be values of this random function $\xi = \xi(t)$; for fixed $\omega \in \Omega$ the real function $\xi(\omega, \cdot) = \xi(\omega, t)$ of $t \in T$ is said to be a *sample function* or a *trajectory* of the random function $\xi = \xi(t)$.

We shall consider another space X of real functions $x = x(t)$ of $t \in T$, which includes all trajectories $\xi = \xi(\omega, t)$, $t \in T$, of the random function $\xi = \xi(t)$. (For instance, the space $X = \mathbb{R}^T$ of all real functions $x = x(t)$, $t \in T$, possesses this property.) Denote by $\mathfrak{B}$ the minimal σ-algebra of sets of X containing all cylinder sets of this space, i.e., sets of the form

$$[x(t_1), \ldots, x(t_n)] \in \Gamma \tag{2.1}$$

(the set indicated by (2.1) consists of the functions $x = x(t)$ for which the values $[x(t_1), \ldots, x(t_n)]$ at the points $t_1, \ldots, t_n \in T$ prescribe a vector that belongs to a Borel set Γ in an n-dimensional vector space $\mathbb{R}^n$). The mapping $\xi = \xi(\omega)$ under which each $\omega \in \Omega$ corresponds to a pertinent sample function $\xi(\omega, \cdot) = \xi(\omega, t)$ of $t \in T$—an element of the space X—is a measurable mapping in a probability space $(\Omega, \mathfrak{A}, \mathsf{P})$ onto a measurable space $(X, \mathfrak{B})$. The sets $A \in \mathfrak{A}$ of the form $A = \{\xi \in B\}$—the preimages of sets $B \in \mathfrak{B}$ under the mapping $\xi = \xi(\omega, \cdot)$ indicated—form (in the aggregate) a σ-algebra. This σ-algebra $\mathfrak{A}_\xi$ is minimal among σ-algebras of the sets containing all sets of the form

$$[\xi(t_1), \ldots, \xi(t_n)] \in \Gamma \tag{2.2}$$

(the set indicated consists of the elements $\omega \in \Omega$ for which the values $[\xi(\omega, t_1), \ldots, \xi(\omega, t_n)]$ prescribe a vector belonging to a Borel set Γ of an n-dimensional vector space $\mathbb{R}^n$), or, in other words, the σ-algebra $\mathfrak{A}_\xi$ is to be generated by values $\xi(t)$, $t \in T$. Probability measure P_ξ defined on a σ-algebra $\mathfrak{B}$ by the relation

$$\mathsf{P}_\xi(B) = \mathsf{P}\{\xi \in B\}, \qquad B \in \mathfrak{B}, \tag{2.3}$$

is said to be a *probability distribution of the random function* $\xi = \xi(t)$ (on the pertinent function space X).

We shall discuss next the question: When is the family of real variables $\xi(t) = \xi(\omega, t)$ given on a space Ω (parameter t runs through a set T) a random function with the given probability distribution P_ξ? More precisely, when does there exist probability measure P in the space Ω related with the given distribution P_ξ by means of (2.3)? We assume in this case that the set $\xi(\Omega)$ of all sample functions $\xi(\omega, \cdot) = \xi(\omega, t)$ of $t \in T$ belongs to the space X.

It is readily seen that *such a probability measure* P *exists if and only if the set* $\xi(\Omega)$ *has a complete exterior measure*, i.e.,

$$\mathsf{P}_\xi(B) = 1 \quad \text{for } B \supseteq \xi(\Omega) \tag{2.4}$$

for any measurable set $B \in X$.

In fact, if P_ξ is the probability distribution of the random function $\xi = \xi(t)$, for any set $\bar{B} \in X$ in the complement of the set $\xi(\Omega)$ the set $\{\xi \in \bar{B}\}$ is empty and

$$\mathsf{P}_\xi(\bar{B}) = \mathsf{P}\{\xi \in \bar{B}\} = 0.$$

On the other hand, for any sets $B_1, B_2 \in \mathfrak{B}$, such that $\{\xi \in B_1\} = \{\xi \in B_2\}$, the symmetric difference $B_1 \circ B_2 = (B_1 \backslash B_2) \cup (B_2 \backslash B_1)$ is contained in the complement of the set $\xi(\Omega)$, and, under the condition (2.4), $\mathsf{P}_\xi(B_1 \circ B_2) = 0$, $\mathsf{P}_\xi(B_1) = \mathsf{P}_\xi(B_2)$. Hence the relation

$$\mathsf{P}\{\xi \in B\} = \mathsf{P}_\xi(B), \qquad B \in \mathfrak{B}, \tag{2.5}$$

defines the single-valued function $\mathsf{P} = \mathsf{P}(A)$ on the σ-algebra $\mathfrak{A}_\xi$ of all sets of the form $A = \{\xi \in B\}$, $B \in \mathfrak{B}$, generated by $\xi(t)$, $t \in T$. Obviously, P is a probability measure and $\xi = \xi(t)$ is a random function on a probability space $(\Omega, \mathfrak{A}, \mathsf{P})$ with the given probability distribution P_ξ.

The measure P on the σ-algebra $\mathfrak{A}_\xi$ generated by the variables $\xi(t)$, $t \in T$, can be defined uniquely by finite-dimensional distributions $\mathsf{P}_{t_1, \ldots, t_n}$ of which each is a Borel measure on $\mathbb{R}^n$ defined by

$$\mathsf{P}_{t_1, \ldots, t_n}(\Gamma) = \mathsf{P}\{[\xi(t_1), \ldots, \xi(t_n)] \in \Gamma\}, \tag{2.6}$$

$\mathsf{P}_{t_1, \ldots, t_n}$ being the probability distribution of the random vector $[\xi(t_1), \ldots, \xi(t_n)]$. In fact,

$$\mathsf{P}(A) = \inf \sum_k \mathsf{P}(A_k), \tag{2.7}$$

where the lower bound is taken over all sets A_k of the form (2.2), whose union covers the set $A \in \mathfrak{A}$. In particular, this fact refers to the probability distribution P_ξ on the corresponding function space X—it is the probability measure on the σ-algebra $\mathfrak{B}$ generated by the given variables $\xi(t) = \xi(x, t)$ on the space X, i.e., the variables of the form

$$\xi(t, x) = x(t), \qquad x \in X \tag{2.8}$$

(where the parameter t, fixed for each functional $\xi(x, t) = x(t)$ of $x \in X$, runs through the set T).

Denote by $\Gamma \times \mathbb{R}^{n-m}$ the Borel set in an n-dimensional space of vectors $[x(t_1), \ldots, x(t_n)]$ such that $[x(t_{i_1}), \ldots, x(t_{i_m})] \in \Gamma$ (Γ is a Borel set in an m-dimensional subspace $\mathbb{R}^m \subseteq \mathbb{R}^n$) with the remaining coordinates $x(t_i)$ arbitrary. Finite-dimensional distributions are "compatible" in the sense that

$$\mathsf{P}_{t_1, \ldots, t_n}(\Gamma \times \mathbb{R}^{n-m}) = \mathsf{P}_{t_{i_1}, \ldots, t_{i_m}}(\Gamma) \tag{2.9}$$

for all sets of the above type.

Let $X = \mathbb{R}^T$ be the space of all real functions $x = x(t)$, $t \in T$. According to a well-known theorem due to Kolmogorov,* any given family of distributions $\mathsf{P}_{t_1, \ldots, t_n}$ prescribes a continuous additive function P (defined by (2.6), where the variables have the explicit form (2.8)), on the algebra of all cylinder sets (2.1). This function extends uniquely to a probability measure P on the σ-algebra $\mathfrak{B}$. The random function $\xi = \xi(t)$ with values $\xi(t) = \xi(x, t)$ in the probability space $(X, \mathfrak{B}, \mathsf{P})$ has finite-dimensional distributions coinciding with the initial compatible distributions $\mathsf{P}_{t_1, \ldots, t_n}$.

Starting from probability distribution $\mathsf{P} = \mathsf{P}_\xi$ on the function space X, under the condition (2.4) we can define (see (2.5)) a probability measure on the corresponding space Ω.

The random functions $\xi = \xi(t)$ and $\tilde{\xi} = \tilde{\xi}(t)$ with values in the same space are said to be *equivalent* if with probability one (for almost all $\omega \in \Omega$)

$$\xi(\omega, t) = \tilde{\xi}(\omega, t)$$

for each fixed $t \in T$. Obviously, the finite-dimensional distributions of equivalent random functions coincide. Taking an equivalent random function $\xi = \xi(t)$ with the trajectories in any function space X, we can define (see (2.3)) a probability measure in this space as well.

Random variables are said to be *Gaussian* if their finite-dimensional distributions are Gaussian. More precisely, (when we deal with a random function $\xi = \xi(t)$ with parameter $t \in T$ under some parametrization), the values $\xi(t) = \xi(\omega, t)$ and the function $\xi = \xi(t)$ itself are said to be *Gaussian* if all finite-dimensional distributions $\mathsf{P}_{t_1, \ldots, t_n}$ are Gaussian. Probability measure P on a σ-algebra $\mathfrak{A}_\xi$ generated by all $\xi(t)$ is also said to be *Gaussian*.

Each of the finite-dimensional distributions $\mathsf{P}_{t_1, \ldots, t_n}$ of the Gaussian random function $\xi = \xi(t)$ has mean value $[a(t_1), \ldots, a(t_n)]$ and correlation

* See [10], p. 150.

matrix $\{B(t_k, t_j)\}$ where $a(t)$, $t \in T$, is *the mean value of* the function $\xi = \xi(t)$, and $B(s, t)$, s, $t \in T$, is *its correlation function*:*

$$a(t) = \mathsf{M}\xi(t),$$
$$B(s, t) = \mathsf{M}[\xi(s) - a(s)][\xi(t) - a(t)], \qquad s, t \in T. \tag{2.10}$$

Therefore, the Gaussian measure P on a σ-algebra $\mathfrak{A}_\xi$ can be defined uniquely by means of its mean value $a(t)$, $t \in T$, and its correlation function $B(s, t)$, $s, t \in T$.

The mean value $a(t)$, $t \in T$, can be arbitrary, and the correlation function $B(s, t)$, $s, t \in T$, need only satisfy the positive definiteness condition

$$\sum_{k, j=1}^{n} c_k c_j B(t_k, t_j) \geqslant 0 \tag{2.11}$$

for any $t_1, \ldots, t_n \in T$ and real $c_1, \ldots, c_n$.

For any function $a(t)$, $t \in T$, and a positive definite correlation function $B(s, t)$, $s, t \in T$, there exists a Gaussian random function with the mean $a(t)$, $t \in T$, and a correlation function $B(s, t)$, $s, t \in T$. Actually, Gaussian distributions $\mathsf{P}_{t_1, \ldots, t_n}$ with the mean $[a(t_1), \ldots, a(t_n)]$ and correlation matrices $\{B(t_k, t_j)\}$ are compatible distributions, and define a Gaussian measure P in the space $X = \mathbb{R}^T$ of all real functions $x = x(t)$ of $t \in T$, on the σ-algebra $\mathfrak{B} = \mathfrak{A}_\xi$, which can be generated by the given values $\xi(t) = \xi(x, t)$ on X of the form (2.8) (parameter t runs through the set T).

I.3 Lemmas on the Convergence of Gaussian Variables

Let $\xi_n = \xi_n(\omega)$, $n = 1, 2, \ldots$, be a sequence of random variables on a probability space $(\Omega, \mathfrak{A}, \mathsf{P})$. The sequence ξ_n, $n = 1, 2, \ldots$, is said to be *convergent in probability on a set* $A \in \mathfrak{A}$ *to some variable* $\xi = \xi(\omega)$ if for any $\varepsilon > 0$

$$\lim_{n\to\infty} \mathsf{P}(\{|\xi_n - \xi| > \varepsilon\} \cap A) = 0. \tag{3.1}$$

Let us recall that a sequence ξ_n, $n = 1, 2, \ldots$, converges in probability if and only if this sequence is Cauchy,† i.e., on the same set A the sequence $\Delta_{nm} = \xi_n - \xi_m$, $n, m = 1, 2, \ldots$, converges to zero in probability.

Lemma 1. *If a sequence of Gaussian variables* ξ_n, $n = 1, 2, \ldots$, *converges in probability on a set* $A \in \mathfrak{A}$ *of positive measure* ($\mathsf{P}(A) > 0$), *it is convergent in the mean*:

$$\lim_{n\to\infty} \mathsf{M}[\xi_n - \xi]^2 = 0. \tag{3.2}$$

* $\mathsf{M}\xi$ denotes the expectation of a random variable $\xi = \xi(\omega)$ on a probability space $(\Omega, \mathfrak{A}, P)$: $\mathsf{M}\xi = \int_\Omega \xi(\omega)\mathsf{P}(d\omega)$.

† See, for example, [10], p. 90.

Proof. We shall consider Gaussian variables $\Delta_{nm} = \xi_n - \xi_m$. For any $\varepsilon > 0$

$$\mathsf{P}\{|\Delta_{nm}| > \varepsilon\} = 2\int_\varepsilon^\infty \frac{1}{\sqrt{2\pi}\sigma_{nm}} \exp\left\{-\frac{(x-a_{nm})^2}{2\sigma_{nm}^2}\right\} dx,$$

where $a_{nm} = \mathsf{M}\Delta_{nm}$, $\sigma_{nm}^2 = \mathsf{P}(\Delta_{nm} - a_{nm})^2$. Suppose that the sequence ξ_n, $n = 1, 2, \ldots$, is not convergent in the mean; this is equivalent to

$$\overline{\lim_{n,m\to\infty}} (a_{nm}^2 + \sigma_{nm}^2) > 0.$$

It can be easily seen that under this condition for a positive ε we have

$$\overline{\lim_{n,m\to\infty}} \mathsf{P}\{|\Delta_{nm}| > \varepsilon\} \geqslant 1 - p/2,$$

where $p = \mathsf{P}(A) > 0$. But then

$$\overline{\lim_{n,m\to\infty}} \mathsf{P}(\{|\Delta_{nm}| > \varepsilon\} \cap A) \geqslant p/2,$$

which fact contradicts (3.1). Hence

$$\lim_{n,m\to\infty} \mathsf{M}\Delta_{nm}^2 = \lim_{n,m\to\infty} (a_{nm}^2 + \sigma_{nm}^2) = 0,$$

i.e., the sequence ξ_n, $n = 1, 2, \ldots$, is Cauchy (a fundamental) in the mean and, therefore, is convergent in the mean.

In particular, if a sequence of Gaussian variables ξ_n, $n = 1, 2, \ldots$, is convergent with positive probability (i.e., convergent for all ω from a set $A \in \mathfrak{A}$ of positive measure), it is convergent in the mean. □

Let us consider a sequence of independent Gaussian variables ξ_n, $n = 1, 2, \ldots$.

Lemma 2. *The series $\sum_{n=1}^\infty \xi_n^2$ is convergent with positive probability if and only if the series $\sum_{n=1}^\infty \mathsf{M}\xi_n^2$ is convergent.*

Proof. Obviously,

$$\sum_{n=1}^\infty \mathsf{M}\xi_n^2 = \mathsf{M}\sum_{n=1}^\infty \xi_n^2,$$

and hence the convergence of the series $\sum_{n=1}^\infty \mathsf{M}\xi_n^2$ implies that the variable $\xi^2 = \sum_{n=1}^\infty \xi_n^2(\omega)$ is finite for almost all $\omega \in \Omega$, i.e., the series $\sum_{n=1}^\infty \xi_n^2$ is convergent with probability one. Let the series $\sum_{n=1}^\infty \xi_n^2$ be convergent with positive probability (by the well-known zero–one law* this series is convergent with probability one as well). Then the sequence ξ_n, $n = 1, 2, \ldots$, converges to 0 in the mean: $\mathsf{M}\xi_n^2 \to 0$ for $n \to \infty$ (see Lemma 1). Let $a_n =$

* See, for example, [10], p. 157.

$\mathsf{M}\xi_n$, $\sigma_n^2 = \mathsf{M}(\xi_n - a_n)^2$. Then

$$a_n^2 + \sigma_n^2 = \mathsf{M}(\xi_n')^2 + \int_{|x|>1} x^2 \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left\{-\frac{(x-a_n)^2}{2\sigma_n^2}\right\} dx,$$

where the random variables $\xi_n' = \xi_n'(\omega)$ are defined as

$$\xi_n'(\omega) = \begin{cases} \xi_n(\omega) & \text{for } |\xi_n| \leq 1 \\ 0 & \text{for } |\xi_n| > 1. \end{cases}$$

For $a_n^2 + \sigma_n^2 \to 0$ we have

$$\int_{|x|>1} x^2 \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left\{-\frac{(x-a_n)^2}{2\sigma_n^2}\right\} dx = o(a_n^2 + \sigma_n^2),$$

so that

$$\mathsf{M}(\xi_n')^2 \sim a_n^2 + \sigma_n^2.$$

By the well-known three-series theorem* a necessary condition for the series $\sum_{n=1}^{\infty} \xi_n^2$ of independent variables ξ_n^2, $= 1, 2, \ldots$, to be convergent is that $\sum_{n=1}^{\infty} \mathsf{M}(\xi_n')^2 < \infty$. But $\mathsf{M}(\xi_n')^2 \sim a_n^2 + \sigma_n^2$, and, consequently, it follows from the convergence of the series $\sum_{n=1}^{\infty} \xi_n^2$ that $\sum_{n=1}^{\infty} (a_n^2 + \sigma_n^2) < \infty$. □

I.4 Gaussian Variables in a Hilbert Space

A random variable ξ in a Euclidean n-dimensional space $\mathbb{R}^n$ is said to be *Gaussian* if its probability distribution is Gaussian.

The random variable $\xi \in \mathbb{R}^n$ is Gaussian if and only if the real variable $\xi(u) = (u, \xi)$ (equal to the scalar product of the elements $u, \xi \in \mathbb{R}^n$) is Gaussian for each $u \in \mathbb{R}^n$.

In fact, the value at a point $u \in \mathbb{R}^n$ of the characteristic function $\varphi(u)$ of the random variable $\xi \in \mathbb{R}^n$ coincides with the value of the characteristic function of the real random variable $\xi(u) = (u, \xi)$ at the point 1 and has the form

$$\varphi(u) = \mathsf{M}e^{i(u, \xi)} = \exp\left\{i(u, a) - \frac{1}{2}(Bu, u)\right\}, \qquad u \in \mathbb{R}^n$$

(see (1.1), where (u, a) is the mean and (Bu, u) is the variance of the Gaussian variable $\xi(u) = (u, \xi)$).

It is clear that the random variable $\xi \in \mathbb{R}^n$ is Gaussian if and only if the random function of the form $\xi(u) = (u, \xi)$ of $u \in \mathbb{R}^n$ is Gaussian.

Let U be a complete separable Hilbert space and let $\xi = \xi(\omega)$ be a function on a probability space $(\Omega, \mathfrak{A}, \mathsf{P})$ with the values in U. The random element ξ of a Hilbert space U is said to be a random variable in U if the scalar product (u, ξ) for each $u \in U$ is a real random variable, i.e., it is a measurable function on the probability space $(\Omega, \mathfrak{A}, \mathsf{P})$.

* See, for example, [10], p. 166.

The random variable ξ in a Hilbert space U is said to be *Gaussian* if the real random variable $\xi(u) = (u, \xi)$ is Gaussian for each $u \in U$. This fact is equivalent to the fact that the random function $\xi(u) = (u, \xi)$ of $u \in U$ is Gaussian since values $\xi(u) = (u, \xi)$ as well as any vector values $[\xi(u_1), \ldots, \xi(u_n)]$ are Gaussian.

In fact, for any vector $\lambda = [\lambda_1, \ldots, \lambda_n]$ in $\mathbb{R}^n$ the scalar product $\sum_{k=1}^n \lambda_k \xi(u_k)$ is equal to

$$\sum_{k=1}^{n} \lambda_k \xi(u_k) = \left(\sum_{k=1}^{n} \lambda_k u_k, \xi \right) = (u, \xi),$$

where $u = \sum_{k=1}^n \lambda_k u_k \in U$; by hypothesis, the variable $\xi(u) = (u, \xi)$ is Gaussian.

Obviously, the mean

$$a(u) = \mathsf{M}(u, \xi), \qquad u \in U,$$

of the random function $\xi(u) = (u, \xi)$, $u \in U$, is a linear functional, and the correlation function

$$B(u, v) = \mathsf{M}[(u, \xi) - a(u)][(v, \xi) - a(v)], \qquad u, v \in U,$$

is a bilinear positive functional on the Hilbert space U. In this case, since the scalar product (u, ξ) is a continuous function of $u \in U$ for each fixed $\omega \in \Omega$, a Gaussian function $\xi(u) = (u, \xi)$ of $u \in U$ must be continuous in the mean (see Lemma 1):

$$\lim_{\|u - v\| \to 0} \mathsf{M}[(u, \xi) - (v, \xi)]^2 = 0 \tag{4.1}$$

($\|u\|$ denotes the norm of the element $u \in U$). But

$$\mathsf{M}[(u, \xi) - (v, \xi)]^2 = a(u - v)^2 + B(u - v, u - v)$$

and (4.1) implies that the functionals $a(u)$ and $B(u, v)$ are continuous.

Being a linear continuous functional, the mean $a(u)$ can be expressed as

$$a(u) = (u, a), \qquad u \in U, \tag{4.2}$$

for some element a in U. Any element $a \in U$ having the property that

$$(u, a) = \int_{\Omega} (u, \xi(\omega)) \mathsf{P}(d\omega) \tag{4.3}$$

for all $u \in U$ is said to be *the mean** of a random variable $\xi \in U$. Being a continuous positive bilinear functional, the correlation function $B(u, v)$ can be expressed

$$B(u, v) = (Bu, v), \qquad u, v \in U, \tag{4.4}$$

where B is a linear positive (i.e., nonnegative self-adjoint) operator in a Hilbert space U called a *correlation operator.*

Let us show that the correlation operator B is completely continuous.

* For the integrability of functions with values in a Hilbert space, see, for example, [12], p. 59.

In fact, any orthonormalized sequence $v_1, v_2, \ldots$ goes to zero weakly, so that the Gaussian variables $\xi_n = (v_n, \xi)$, $n = 1, 2, \ldots$, where $\xi = \xi(\omega) \in U$, goes to zero as $n \to \infty$ for all $\omega \in \Omega$. Therefore (see Lemma 1), they are convergent in the mean, i.e.,

$$\mathsf{M}\xi_n^2 = (Bv_n, v_n) \to 0$$

(here and further on we assume for simplicity of notation that the mean $a \in U$ is 0). If the operator B was not assumed completely continuous, outside some ε-neighborhood of zero there would be an infinite number of spectral points (taking into account the multiplicity), and, therefore, an infinite number of invariant orthogonal subspaces for each element of which

$$(Bu, u) = \int_{|\lambda| > \varepsilon} \lambda d(E_\lambda u, u) \geqslant \varepsilon \|u\|^2,$$

where $B = \int \lambda \, dE_\lambda$ is the spectral representation of the continuous self-adjoint operator B.

Further, we shall choose a complete orthonormalized basis of eigen-elements $v_1, v_2, \ldots$ of this completely continuous symmetric positive operator B corresponding to eigenvalues $\sigma_1^2, \sigma_2^2, \ldots$. The corresponding variables $\xi_k = (v_k, \xi)$, $k = 1, 2, \ldots$, are uncorrelated:

$$\mathsf{M}\xi_k \xi_j = (Bv_k, v_j) = \begin{cases} \sigma_k^2 & \text{for } j = k, \\ 0 & \text{for } j \neq k. \end{cases}$$

In this case

$$\sum_1^\infty \xi_k^2(\omega) = \sum_1^\infty (v_k, \xi(\omega))^2 = \|\xi(\omega)\|^2.$$

As is well known, uncorrelated Gaussian variables are independent and the convergence of the series $\sum_1^\infty \xi_k^2(\omega)$ (for all ω) implies the convergence of the series $\sum_1^\infty \mathsf{M}\xi_k^2$ (see Lemma 2). Consequently,

$$\sum_1^\infty (Bv_k, v_k) = \sum_1^\infty \mathsf{M}\xi_k^2 = \sum_1^\infty \sigma_k^2 < \infty,$$

i.e., the correlation operator B is a nuclear operator:* for any orthonormal system $u_1, u_2, \ldots, \in U$,

$$\sum_1^\infty (Bu_k, u_k) < \infty. \tag{4.5}$$

Therefore, if we have a Gaussian random variable $\xi \in U$, the random function $\xi(u) = (u, \xi)$ of parameter $u \in U$ has a mean of the form (4.2) and a correlation function of the form (4.4) where the correlation operator B is a nuclear operator on the Hilbert space U.

Next, let $\xi(u)$, $u \in U$, be an arbitrary Gaussian random function with a mean of the form (4.2) and a correlation function of the form (4.4), where B

* See, for example, [7], p. 55.

is a nuclear operator on the Hilbert space U. Then there exists an equivalent random function $\xi(u)$, $u \in U$, and a Gaussian random variable $\xi = \xi(\omega)$ in U such that

$$\xi(u) = (u, \xi), \qquad u \in U. \tag{4.6}$$

The variable $\xi \in U$ indicated can be defined for almost all elementary outcomes ω by the formula

$$\xi(\omega) = \sum_{k=1}^{\infty} \xi(v_k)v_k, \tag{4.7}$$

where $v_1, v_2, \ldots$ is the complete orthonormal system of eigenelements of the nuclear operator B, and, by virtue of the relation

$$\mathsf{M} \sum_{k=1}^{\infty} \xi(v_k)^2 = \sum_{k=1}^{\infty} B(v_k, v_k) < \infty$$

for independent Gaussian variables $\xi(v_1), \xi(v_2), \ldots$, the series $\sum_{k=1}^{\infty} \xi(v_k)^2$ is convergent with probability one. In fact, $\xi(u)$, $u \in U$, is a random linear functional in the sense that with probability one

$$\xi(\lambda_1 u_1 + \lambda_2 u_2) = \lambda_1 \xi(u_1) + \lambda_2 \xi(u_2)$$

for any real λ_1, λ_2 and any elements $u_1, u_2 \in U$ since, as we can easily verify,

$$\mathsf{M}[\xi(\lambda_1 u_1 + \lambda_2 u_2) - \lambda_1 \xi(u_1) - \lambda_2 \xi(u_2)]^2 = 0.$$

Furthermore, the random functional $\xi(u)$ is continuous in the mean (see (4.1) and below), and, since

$$u = \lim_{n \to \infty} \sum_{k=1}^{n} (u, v_k)v_k,$$

we have

$$\xi(u) = \lim_{n \to \infty} \xi\left(\sum_{k=1}^{n} (u, v)v_k\right) = \lim_{n \to \infty} \sum_{k=1}^{n} (u, v_k)\xi(v_k)$$

(in the sense of convergence in the mean); at the same time with probability one

$$(u, \xi) = \lim_{n \to \infty} \sum_{k=1}^{n} \left(u, \sum_{k=1}^{n} \xi(v_k)v_k\right) = \lim_{n \to \infty} \sum_{k=1}^{n} (u, v_k)\xi(v_k),$$

so that we have the equality (4.6) with probability one for each value $\xi(u)$ of the primary random function $\xi(u)$, $u \in U$.

Thus, we have arrived at the following result.*

Theorem 1. *The Gaussian functional $\xi(u)$, $u \in U$, on a Hilbert space U can be represented by* (4.6) *if and only if the mean $a(u)$, $u \in U$, is a continuous*

* A survey of results related to distributions in linear spaces can be found, for example, in Yu. V. Prokhorov, "The method of characteristic functionals," *Proceedings of the 4th Berkeley Symposium, Vol. 2*, 1961, pp. 403–419.