A.M. Yaglom

Correlation Theory of Stationary and Related Random Functions
Volume I: Basic Results

With 49 Illustrations

A.M. Yaglom
Institute of Atmospheric Physics
Academy of Sciences
109017 Moscow
U.S.S.R.

To the memory of my beloved father,
Moisei (Moses) Akivovich Yaglom,
who taught me everything and made
me enthusiastic about mathematics,
and to my dear Professor Andrei Nikolaevich Kolmogorov,
who acquainted me with the fascinating world
of random functions and the richness of
their applications.

Preface

The theory of random functions is a very important and advanced part of modern probability theory, which is very interesting from the mathematical point of view and has many practical applications. In applications, one has to deal particularly often with the special case of *stationary random functions*. Such functions naturally arise when one considers a series of observations $x(t)$ which depend on the real-valued or integer-valued argument t ("time") and do not undergo any systematic changes, but only fluctuate in a disordered manner about some constant mean level. Such a time series $x(t)$ must naturally be described statistically, and in that case the stationary random function is the most appropriate statistical model.

Stationary time series constantly occur in nearly all the areas of modern technology (in particular, in electrical and radio engineering, electronics, and automatic control) as well as in all the physical and geophysical sciences, mechanics, economics, biology and medicine, and also in many other applied fields. One of the important trends in the recent development of science and engineering is the ever-increasing role of the fluctuation phenomena associated with the stationary disordered time series. Moreover, at present, more general classes of random functions related to a class of stationary random functions have also been appearing quite often in various applied studies and hence have acquired great practical importance.

The widespread usefulness of the concept of a stationary random function has naturally resulted in a swift increase in the last few years in literature on the applied theory of such functions. Simultaneously the mathematical and statistical literature on the subject has been expanding from year to year to meet the pressing needs of practice. As a result, the available scientific literature (of both purely theoretical and applied nature) dealing with stationary random functions has attained such enormous pro-

portions that even a specialist has a hard time groping his way through it, to say nothing of a layman, who gets completely lost in this jungle.

The situation is somewhat different as regards the study of various classes of random functions more general than ordinary stationary functions (e.g., homogeneous and isotropic random fields or processes with stationary increments of some order). The literature devoted to such functions is not too voluminous as yet, but still it comprises many dozens of books and papers.

Along with the vigorous growth, the literature on stationary random functions is being differentiated into more and more distinct parts having little in common and meant for different groups of readers. This refers primarily to the division of the entire presentation of the material into literature for mathematicians and applied literature for workers in various fields in which stationary series of observations arise (first of all for engineers and physicists). The enormous body of applied literature contains, as a rule, very few references (which are usually of minor importance) to purely mathematical works, while books and papers on the mathematical theory of stationary random functions rarely mention the works of an applied nature. The extensive statistical literature on analysis of time series occupies from this point of view an intermediate place. Almost all this literature is in fact based on strict mathematical theory, but the more advanced parts of this theory are nevertheless usually ignored here, while most of the attention is devoted to an important, albeit a rather specialized problem of digital processing of time series data. The small number of references to more sophisticated mathematical results in applied texts may create an impression that an advanced mathematical theory of random functions is practically useless and can only confuse a reader who is not a pure mathematician. The author hopes that this book will prove the groundlessness of such a viewpoint.

The present book comprises an elementary introduction to the most important part of the theory of stationary and related random functions dealing only with first and second moments of these functions. (This part of general theory is usually called *correlation theory*, or *second-order theory*; the present book uses the former term.) The presentation is intended for readers without any special mathematical background. They are only required to know elementary calculus. Formally speaking, they may not even be familiar with the rudiments of probability theory, because all the necessary information relating to this discipline is given in the Introduction to the book. However, a knowledge of the basic concepts of probability theory and its simplest applications will greatly facilitate the understanding of the material presented in the book, and it can hardly be of any use without this knowledge. The principal definitions and formulae from probability theory included in the Introduction are therefore meant primarily to refresh in the reader's memory the facts formerly studied by him and widely used in the book.

It is assumed that most of the readers are particularly interested in the

applied aspects of the theory of stationary and related random functions. With this in mind, many examples illustrating the considered concepts and relations and facilitating applied work are included in the book, while the mathematical proofs are often only briefly outlined and sometimes even omitted altogether. (Mathematical proofs are in many cases presented more comprehensively, although often without lingering over technical details and not under the most general conditions, in Volume II to the book, which will be discussed in more detail in what follows.) The author, however, always tried to pinpoint all facts whose proofs are omitted in the book, even if these facts are of a rather subtle nature and require a level of sophistication far beyond the scope of the knowledge expected of most of the readers. Among other things, the book does not employ a nonrigorous (but usually leading to correct results and therefore widely employed in engineering and physical texts) method where the derivatives and integrals of the realizations of random functions actually used in applications are replaced, without any reservations, by theoretically simpler derivatives and integrals in the mean-square sense. On the whole, the level of mathematical rigor of the presentation adopted in the book is somewhat higher than in most of the books and papers intended for students of applied science or workers in various applied fields. Therefore it is believed that the book will also come in handy for readers who are not quite happy with the heuristic level of rigor of most of the engineering texts and who wish to get more precise information on the fundamentals of the mathematical theory of stationary and related random functions without too much effort or wasted time. The intention of satisfying this group of readers was another reason for discussing, in comparative detail, some rather important theoretical topics traditionally excluded from books intended for those specializing in applied science (e.g., questions of strict definition and domain of applicability of the concept of a random function, of the mathematical and physical ergodicity conditions, and of the exact meaning of the spectral representations). It is also hoped that the method of presentation adopted in the book will make it accessible to students of mathematics not yet versed in the theory of stationary random functions and who are interested not in logical subtleties of this theory but in the underlying fundamental facts and their possible applications.

When writing this book, the author encountered numerous difficulties in choosing symbols and terms. At present, very many symbols and terms adopted in the mathematical literature on the random function theory do not coincide with those used in the applied literature, and different applied fields sometimes diverge drastically in this respect. Specific choice of symbols and terms is also inherent in the enormous statistical literature on time series analysis which is closely related to the contents of many parts of this book. In an attempt to facilitate the mutual understanding between mathematicians, statisticians, and workers in various applied fields, the author decided to resort to a compromise without giving preference to any one of the parties. For instance, the angular frequency is

denoted by ω, as is customary in the applied literature, and not by λ used in most of the purely mathematical books and papers. Also, probability averaging is symbolized by angular brackets which are widely used by physicists and engineers for this purpose but are almost never found in the mathematical and statistical literature. At the same time, in accordance with usage in mathematical texts, the spectral density is denoted, not by some upper-case letter, but by the lower-case f and is not referred to as a power spectrum (the last term is most common to physicists and engineers), while the capital F is used in the book to denote the spectral distribution function, which is rarely mentioned in the applied literature (and whenever it is, the term ''integrated spectrum'' is used instead of ''spectral distribution function''). At the same time, the terms ''correlation function'' and ''multidimensional random function,'' which are customary for engineers and physicists, are used here in place of the more specialized mathematical (and statistical) terms ''covariance function'' and ''multivariate random function.''

This book consists of an introduction and four chapters. It ends with a comparatively short list of basic literature on the subject which includes all the books and papers directly referred to in the text. (The references are indicated by the name of the author supplemented, in parentheses or after a comma, by the date of publication of the work. The initials of the author are given only in cases where there are several authors with the same surname.) The Introduction recapitulates the principal facts of the classical probability theory and presents an heuristic outline of the concept of a random function accompanied by a number of examples.

Chapter 1 contains the strict definitions of a general random function and of a stationary random function, the explanation of the essence of the correlation theory of such functions, and the simplest properties of stationary random functions and stationary correlation functions.

Chapter 2 discusses the basic facts of the very important spectral theory of stationary random functions and some of its simple applications (in particular, applications to the theory of linear time-invariant transformations of stationary functions, which play a crucial part in modern electrical and radio engineering).

Chapter 3 deals with methods for determining the main statistical characteristics of a stationary random function from its single observed realization (i.e., with the statistical analysis of stationary time series). This topic is treated in the enormous statistical literature. Its presentation in this book takes into account all the important recent developments, but differs in style and content from all other presentations of the same subject known to the author.

Finally, Chapter 4 covers a number of generalizations of the concept of a stationary random function which are of great mathematical interest and are at present rather widely employed in many applications. The material of this chapter has never been presented systematically in any book.

Some parts of the book contain more specialized or detailed material

which can be omitted in the first reading without loss of continuity. These parts are indicated by narrower-than-usual margins on both sides and all of them begin and end with a special mark* (see, e.g., pp. 132–133 and 133–140). In addition, many superscripted numbers (which begin with number 1 in the Introduction and in each of the four chapters) can be found in different places in the book. Any of these numbers indicates that a note is prepared by the author commenting on the corresponding passage of the text.

All the notes are collected in Volume II which is of a more special nature than the main body of the book. The readers previously uninitiated into the theory of stationary and related random functions are advised to skip all the notes in the first reading, i.e., to read the present book without opening Volume II and to disregard all the references to the notes. Later on, however, they may wish to gain additional information on the topics considered in the book (perhaps only on some of these topics) and find out more data on the relevant literature. They will then have to resort to Volume II.

The notes in this volume are very diverse in their contents: some give a more detailed derivation of results whose proofs are only briefly outlined (or omitted altogether) in this book, while many others contain more specialized additional material (in particular, they often give a lot more sophisticated mathematical results which are never mentioned in the applied literature, or contain some remarks about additional interesting applications of the conclusions from this book). Some notes deal with the history of the question at hand or discuss the connection of the results considered in this book, with certain results relating to other branches of mathematics and physics. Finally, the notes include nearly all the references, which the author has intentionally made very numerous.

It was noted at the beginning of this preface that the literature on stationary and related random functions is enormous and that books and papers of an applied nature usually cite only a small portion devoted to closely related applied problems, while mathematicians, as a rule, refer almost exclusively to the mathematical literature. By contrast, the Bibliography at the end of Volume II includes a great number of books and papers of varied degrees of complexity which are purely mathematical, or deal with statistical problems, or are devoted to various applications (of a very diverse nature). These references are intended to give the reader a reasonably complete picture of the present state of investigations in the area considered, to aid him in finding the necessary information on matters of special interest, and help him in finding his way in the tremendous amount of books and papers dealing with stationary and related random functions. Naturally, even an extensive list of references such as the one contained in Volume II is insufficient to cover a significant portion of all the available literature on the subject which numbers many tens of thousands of papers and books. The reader should also keep in mind that a number of not very important works have been included in the bibliography

only as examples illustrating the general trends of research in the field discussed in this book, while there is no doubt that many interesting works have been unjustly omitted in the list of references and many authors ignored without any reason. Nevertheless, the notes and bibliography collected in Volume II embrace a very wide range of investigations on the theory of stationary and related functions. Therefore, the present Volume I and Volume II taken together can very well be used as a sufficiently complete reference book, while the present book alone can reasonably be regarded as an introductory textbook for students and research workers who have never been initiated into the theory of random functions.

In conclusion, I wish to note that the writing of this book was originally intended as a revision of my old, long survey paper first published in Russian in 1952 and then issued twice in the United States as a slightly enlarged separate book (see Yaglom, 1962). The revision process dragged on and on, ultimately leading to a completely new book, which differs drastically from the previously one (primarily owing to the inclusion of extensive fresh material and elimination of the whole theory of linear extrapolation and filtering of stationary random functions, which constituted the bulk of the older book). Part of the material of Chapters 1–3 of this book was also included in my comparatively new book (which is much smaller than the present one) published in Leningrad a few years ago (Yaglom, 1981).

Contents

INTRODUCTION

The theory of probability studies those experiments (observation, trials) whose results cannot be exactly predicted, since they depend on some circumstances of which we are ignorant or whose effect we cannot determine beforehand. However, not all such experiments can serve as objects of probabilistic studies. First of all, it is necessary that the experiments (observations, trials) concerned belong to the category of *mass phenomena*, i.e. that they could be repeated many times under practically identical conditions. It is further required that the experimental results show a definite *statistical stability*, namely that after numerous repetitions of the experiment the frequency of appearance of any prescribed result remains approximately the same almost all the time, only slightly fluctuating about some constant number p. The requirement of such a statistical stability (or, to put it differently, of a *statistical law* relating to a given experiment) may seem very restrictive at first glance. In actuality it is surprisingly often satisfied both in natural and man-made conditions.* The experimental result under study is then called a *random event*, and the number p, the *probability* of this event.

---

*Note, in particular, that all the quantum-mechanical laws are statistical. Therefore many prominent contemporary physicists even tend to believe that some statistical laws lie at the basis of all the laws of nature.

Random events are the very first object of study in the theory of probability, but not the main object of investigation. The principal place in probability theory is occupied by the study of numerical characteristics of random events, i.e. of quantities taking different numerical values depending on the results of a certain random experiment (or random event).   Such quantities are usually called *random variables*. Typical examples are: (a) the number of points showing after tossing a die, (b) the number of calls coming through to a telephone exchange within a certain period of time, (c) the lifetime of an electric bulb from the moment of its manufacture to the moment of its burnout, and (d) the error of measurement of some physical quantity by a given instrument. All these examples refer to different areas of human activity; the possibility of numerous repetitions of related experiments is quite obvious here, and the statistical stability of the experimental outcomes is confirmed by an analysis of a large number of observations. Examples of random variables of natural origin that are unrelated to human activity are (e) the temperature at a fixed place at noon on one of the days of January, and (f) the total area of sunspots recorded on a given day by a solar observatory. Here, too, experiment confirms the feasibility of the conditions required for these variables to be considered random in the probabilistic sense.

Random variables will usually be denoted by capital characters, and the numerical (i.e. experimental) values taken by them (which are generally called *sample values, observed values,* or *realizations* of the corresponding random variable), by the corresponding lower-case characters. If X is a random variable, then the occurrence of the value of X which is less than a fixed number x will be a random event. Its probability

(0.1) $\mathbf{P}\{X < x\} = F(x)$

is a function of x, which is called the *distribution function* of the random variable X. (The symbol $\mathbf{P}\{....\}$ will always stand for the probability of fulfillment of the relations between the braces.) In probability theory, a random variable X is assumed to be given if its distribution function is known; then it is also said that the *probability distribution* of X is known. It follows from (0.1) that any distribution function

$F(x)$ is a monotone nondecreasing function of x tending to zero as $x \to -\infty$ and to unity as $x \to \infty$, i.e. it satisfies the conditions

(0.2)

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1,$$

$$F(x + a) - F(x) \geqslant 0 \text{ for } a > 0.$$

Figure 1 shows schematically the distribution functions of the random variables mentioned in Examples (a) - (d) on p. 2. We can see that these functions differ noticeably. In Examples (a) and (b), the function $F(x)$ is a "step function", which increases only at discrete discontinuity points ("jumps") and remains constant in the intervals between the jumps.

(a)

(b)

(c)

(d)
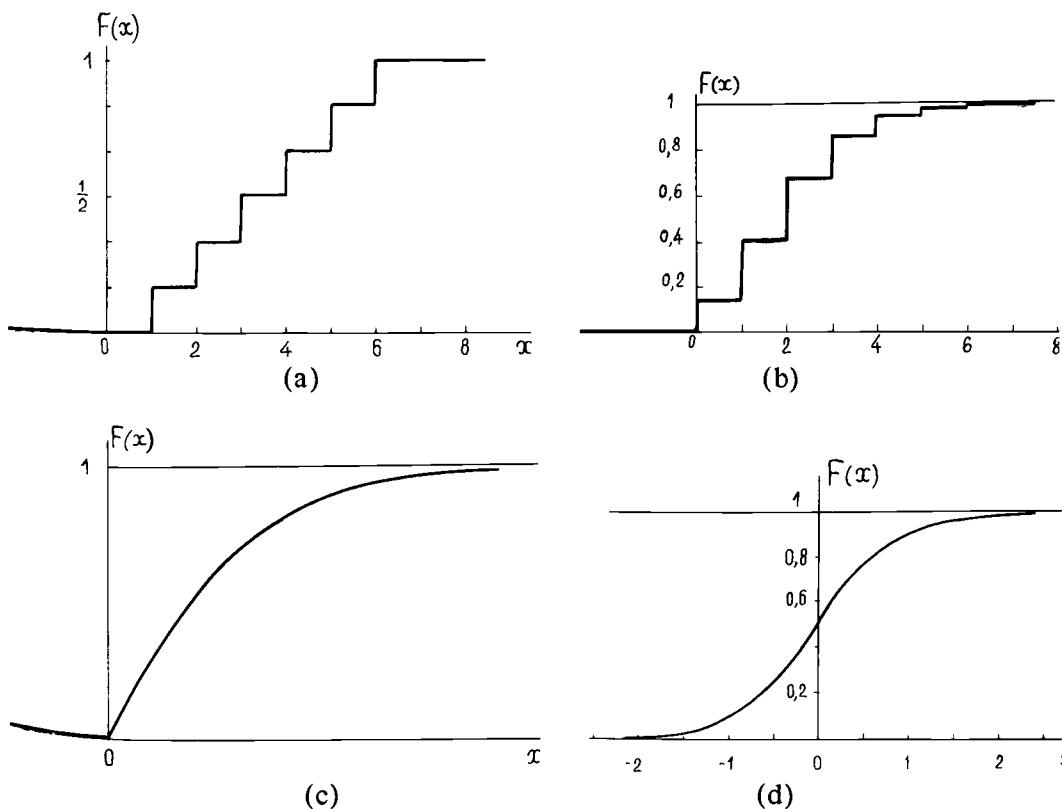
Fig. 1. (a) - (d). Four examples of probability distribution functions.

Such functions $F(x)$ indicate that the random variable X is
discrete, i.e. that it can take only discrete values x_1, x_2, ...
(coinciding with the points of jumps of $F(x)$) with
probabilities p_1, p_2, ... equal to the values of the
corresponding jumps. In Examples (c) and (d), $F(x)$ is a
continuous differentiable function of x, which is equal to,
plus or minus constant, an indefinite integral of its
derivative $dF(x)/dx$. (In mathematical analysis, functions
possessing such a property are called *absolutely continuous.*) It
is clear, therefore, that in cases (c) and (d) the assignment of
values to the function $F(x)$ is equivalent to the assignment of
values to its derivative − the *probability density* $p(x)$:

$$(0.3) \qquad F(x)=\int_{-\infty}^{x} p(x')dx'; \quad p(x) = \frac{dF(x)}{dx}.$$

Equations (0.3) indicate that the random variable X is
continuous. For such a variable the probability distribution
of X can be represented by its probability density $p(x)$.
(Schematic graphs of the densities $p(x)$ corresponding to the
above Examples (c) and (d) are given in Fig. 2(a) and (b).
The conditions (0.2), as applied to a continuous random
variable, take the form
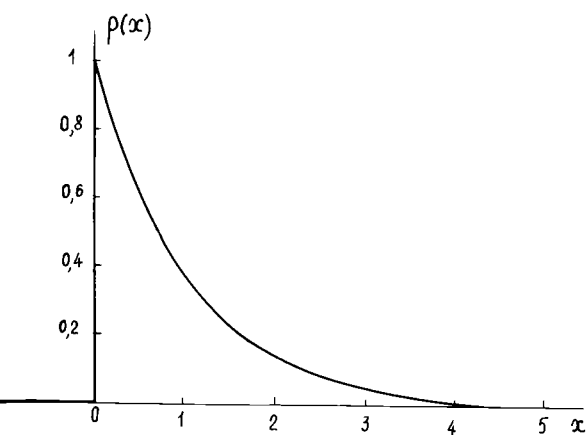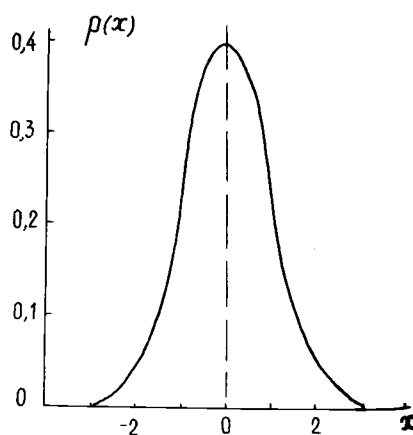
(a)   (b)

Fig. 2(a) and (b). Two examples of probability densities.

(0.4) $\int_{-\infty}^{\infty} p(x)dx = 1, \quad p(x) \geqslant 0$ for all x.

Apart from purely discrete and purely continuous random variables, there may exist *mixed* random variables associated with the distribution function $F(x)$, which experience jumps, but increases smoothly in the intervals between them (see Fig. 3). A function of this type can be written as the sum of the continuous function $F_I(x)$ and the step function $F_{II}(x)$. In such cases it is also possible to use formally the concept of probability density $p(x) = dF(x)/dx$. However, here $p(x)$ will no longer be an ordinary function, but it will now include generalized Dirac δ-functions:[1]

(0.5) $p(x) = p_0(x) + \sum_k p_k \delta(x - x_k),$

where $p_0(x)$ is an ordinary non-negative function. In the particular case of a purely discrete random variable the probability density is simply a linear combination of δ-functions: $p(x) = \sum_k p_k \delta(x - x_k)$. In principle, there may also exist still more complex random variables, whose distribution function contains a term which is a continuous (jumpless), but not an absolutely continuous function (i.e. it is not representable as an integral with respect to its derivative).[2] However, such distribution functions are never encountered in physical and engineering applications, so that in applied problems one may act as if only discrete, continuous, and mixed random variables exist.
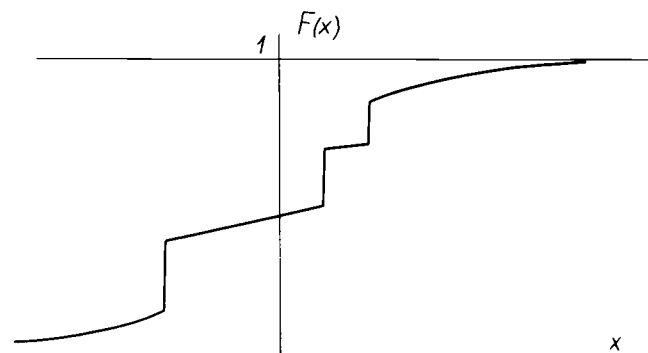
Fig. 3. An example of the mixed distribution function $F(x)$.

The distribution function $F(x)$ (and the probability density $p(x)$) provides a complete specification of the random variable X. (Another important method to specify the random variable completely consists in assigning a so-called *characteristic function* of X. This method will be briefly considered in Note 3 in Volume II.) In describing the particular properties of the variable X, however, it is often much more convenient to use simpler numerical statistical characteristics, determined by the probability distribution. The *moments* of the variable X

$$(0.6) \qquad \mu^{(n)} = \langle X^n \rangle = \int_{-\infty}^{\infty} x^n dF(x)$$

are the most important numerical characteristics. The angular brackets in the middle part of (0.6) symbolize the *probabilistic averaging* (i.e. averaging over the statistical ensemble of all possible realizations),[4] while the right-hand side gives the analytical formulation of this averaging operation. The integral in the right-hand side of (0.6) is the so-called *Stieltjes integral*, which, for finite limits of integration, is defined by the equation

$$(0.7) \qquad \int_a^b f(x)dF(x) = \lim_{\max|x_k - x_{k-1}| \to 0} \sum_{k=1}^{N} f(x_k')[F(x_k) - F(x_{k-1})],$$

where $a = x_0 < x_1 < \dots < x_{N-1} < x_N = b$, $x_{k-1} \leqslant x_k' \leqslant x_k$. (As to the improper Stieltjes integral $\int_{-\infty}^{\infty} f(x)dF(x)$ it is defined in the usual manner as the limit $\int_a^b f(x)dF(x)$ as $a \to -\infty$ and $b \to \infty$.)* For a discrete random variable, the integral on the right-side of (0.6) reduces to a sum: here

$$(0.6a) \qquad \mu^{(n)} = \sum_k x_k^n p_k.$$

If a probability density $p(x)$ exists, then

$$(0.6b) \qquad \mu^{(n)} = \int_{-\infty}^{\infty} x^n p(x)dx.$$

---

*Formula (0.6) is a particular case of the general formula for the mean value $\langle f(X) \rangle$ of the arbitrary function $f(X)$ of the random variable X: $\langle f(X) \rangle = \int_{-\infty}^{\infty} f(x)dF(x)$.

For a discrete or mixed random variable it is also possible to use formally (0.6b), but the function $p(x)$ will now contain δ-functions, and therefore the integral will actually include a sum of the type (0.6a). From now on we shall, for simplicity, systematically use formulae of the type (0.6b), which enable us to embrace all the classes of random variables of practical interest, provided that δ-functions are also used.

The simplest of the moments $\mu^{(n)}$ is the first moment $\mu^{(1)} = m_X$, where

$$(0.8) \qquad m_X = \langle X \rangle = \int_{-\infty}^{\infty} x p(x) dx.$$

The moment m_X is called the *mean value* (or the *mathematical expectation*) of the random variable X. (Below we shall always assume that $\int_{-\infty}^{\infty} |x| \, p(x) dx < \infty$ and hence X has a finite mean value.) The value of m_X determines the "distribution center" around which all the realizations x of the random variable X are grouped.

A mere knowledge of the mean value is usually of little use if we do not have an idea of the amount of spread of the observed values of X about its mean value. There are several distinct methods for evaluating the spread,[5] but the *variance* of the random variable X

$$(0.9) \qquad \text{var } X = \sigma_X^2 = \langle (X - \langle X \rangle)^2 \rangle = \int_{-\infty}^{\infty} (x - m_X)^2 p(x) dx$$

is most convenient for use in mathematical theory, and is therefore the most important numerical characteristic of the spread. The positive square root of the variance σ_X^2 is called the *standard* (or *root-mean-square*) *deviation* (of the variable X from m_X);[*] it determines the order of magnitude of the spread of the observed values of X about the "distribution center" m_X. The variance σ_X^2, which is later also always assumed finite, is expressed in a very simple way via the first and second moments:

$$(0.10) \qquad \sigma^2(x) = \langle X^2 \rangle - \langle X \rangle^2 = \mu^{(2)} - m_X^2.$$

---

[*]In the following text the symbols σ_X^2 and σ_X will sometimes be replaced by $\sigma^2(X)$ and $\sigma(X)$.

Moreover, if a is an arbitrary number, it is easy to see that

$$
\begin{aligned}
\langle (X - a)^2 \rangle &= \langle [(X - m_X) + (m_X - a)]^2 \rangle \\
&= \sigma_X^2 + (m_X - a)^2 .
\end{aligned}
$$
(0.11)

Thus, *the mean square of the deviation of* X *from* a *will be the least precisely when* a *coincides with the mean value* $m_X = \langle X \rangle$.

Moments of a higher order of the variable X (the third and fourth orders at any rate) also describe quite definite properties of the corresponding probability distributions and may sometimes be of applied value; however, we shall not dwell on them here.

To illustrate the foregoing, recall the above-mentioned random variables (a) - (d), whose probability distributions were depicted in Fig. 1. The experiments determining these variables were previously described only in very general terms. Now we note that both the observed data and some theoretical (but not quite rigorous) considerations show that the probability distribution of these variables can often be described accurately enough by formulae of the type:

case (a) $p_k = P\{X = k\} = 1/6$ for $k = 1,2,3,4,5,$ and 6 (this is an example of a *discrete uniform probability distribution*);

case (b) $p_k = P\{X = k\} = e^{-\lambda} \dfrac{\lambda^k}{k!}, \quad k = 0, 1, 2, ...$

(the *Poisson probability distribution* with parameter λ, where $\lambda > 0$);

case (d) $p(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{for } x \geqslant 0, \\ 0 & \text{for } x < 0 \end{cases}$

(the *exponential distribution* with parameter λ, where again $\lambda > 0$);

case (d) $p(x) = \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left[-\dfrac{(x - m)^2}{2\sigma^2} \right]$

(the *normal probability distribution* or *Gaussian distribution* with parameters m and σ, where m is arbitrary, and $\sigma > 0$). The formulae (0.8) and (0.9) (where in (a) and (b) the

integrals must naturally be replaced by the corresponding sums) make it possible to readily calculate[6] that

$$m_X = 3.5, \quad \sigma_X^2 = 78.75 \text{ for (a)};$$

$$m_X = \lambda, \quad \sigma_X^2 = \lambda \quad \text{for (b)};$$

$$m_X = 1/\lambda, \quad \sigma_X^2 = 1/\lambda^2 \text{ for (c); and}$$

$$m_X = m, \quad \sigma_X^2 = \sigma^2 \quad \text{for (d)}.$$

Suppose now that X and Y are two numerical characteristics of some random event. (For instance, X and Y could be two coordinates of the point of explosion of a shell ejected during shooting training; or the mean temperature of a day in May at some fixed place and the total precipitation during the same day.) The pair of random variables X and Y can be regarded as a single *two-dimensional* (or *bivariate*) *random variable* $\mathbf{X} = (X,Y)$. In assigning such a random variable $\mathbf{X}$ (i.e. in assigning its probability distribution), use is made of the *two-dimensional (bivariate) distribution function*

$$(0.12) \quad F(x,y) = \mathbf{P}\{X < x, Y < y\},$$

depending on two variables x and y. It is easy to see that the function $F(x,y)$ must, by definition, tend to zero as $x \to -\infty$ or $y \to -\infty$, tend to one as $x \to \infty$ and $y \to \infty$, and be a monotone nondecreasing function of both variables x and y. Moreover, this function must also satisfy the following condition:

$$(0.13) \quad F(x + a, y + b) - F(x + a, y) - F(x, y + b)$$

$$+ F(x, y) \geqslant 0 \text{ for all } a > 0, b > 0,$$

which refers to both variables simultaneously and means that $\mathbf{P}\{x \leqslant X < x + a, y \leqslant Y < y + b\} \geqslant 0$ for all $a > 0, b > 0$. Consider now the particular case of a continuous two-dimensional variable $\mathbf{X} = (X,Y)$. In such a case $\mathbf{P}\{x \leqslant X < x + \Delta x, y \leqslant Y < y + \Delta y\} \approx p(x,y)\Delta x\Delta y$ for small Δx and Δy and, hence, there exists a *two-dimensional probability density* $p(x,y)$ such that

$$(0.14) \quad F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p(x',y')dx'dy', \quad p(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}.$$

Here, all the conditions on $F(x,y)$ reduce to two conditions

(0.15) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)dxdy = 1$, $p(x,y) \geqslant 0$ for all x,y.

The simplest numerical statistical characteristics of the two-dimensional random variable $\mathbf{X} = (X,Y)$ are its *moments*

(0.16) $\mu^{(n,m)} = \langle X^n Y^m \rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^n y^m d^2 F(x,y)$.

The integral on the right-hand side of (0.16) is a two-dimensional Stieltjes integral, which can be defined quite similarly to the definition of a one-dimensional integral of the same type (cf. (0.7)). In this book, however, it is sufficient to consider only the simplest case of the continuous variable $\mathbf{X}$, which has a two-dimensional probability density $p(x,y)$ (i.e. when $\mathbf{X}$ is such that (0.14) is fulfilled). Here, the integral on the right-hand side of (0.16) reduces to an ordinary double integral:

(0.17) $\mu^{(n,m)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^n y^m p(x,y)dxdy$.

The moments $\mu^{(n,0)} = \langle X^n \rangle$ and $\mu^{(0,m)} = \langle Y^m \rangle$ obviously coincide with the ordinary moments of one-dimensional random variables X and Y. If $n > 0$ and $m > 0$, the moment $\mu^{(n,m)}$ is called the *joint moment* of the variables X and Y. The simplest (and the most important) joint moment is the second (sometimes called the *product moment*)

(0.18) $\mu^{(1,1)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy p(x,y)dxdy$.

It is often more convenient to use, instead of $\mu^{(1,1)}$, the corresponding *central joint moment*, i.e. the moment of the variables $\mathring{X} = X - \langle X \rangle$, $\mathring{Y} = Y - \langle Y \rangle$:

(0.19)
$$b_{XY} = \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle = \langle XY \rangle - \langle X \rangle \langle Y \rangle$$
$$= \mu^{(1,1)} - m_x m_y.$$

The second central moment is often called the *covariance* of the random variables X and Y and is denoted by $\mathrm{cov}(X,Y)$. The value of the covariance normalized by the product of the corresponding standard deviations, i.e.

(0.20) $r = \dfrac{b_{XY}}{\sigma_X \sigma_Y} = \dfrac{\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle}{\langle (X - \langle X \rangle)^2 \rangle^{1/2} \langle (Y - \langle Y \rangle)^2 \rangle^{1/2}}$

is called the *correlation coefficient* of the random variables X and Y.

The range of possible values of the correlation coefficient r is restricted by the inequalities

(0.21) $-1 \leqslant r \leqslant 1.$

In fact, the mean value of the non-negative random variable $(\mathring{X} - a\mathring{Y})^2$, where $\mathring{X} = X - \langle X \rangle$, $\mathring{Y} = Y - \langle Y \rangle$, clearly cannot be negative. Since $\langle (\mathring{X} - a\mathring{Y})^2 \rangle = \sigma_X^2 - 2b_{XY}a + \sigma_Y^2 a^2$, we obtain

(0.22) $\sigma_X^2 - 2b_{XY}a + \sigma_Y^2 a^2 \geqslant 0$

for any a. By virtue of the known criterion of the non-negativity of the quadratic binominal, it follows from (0.22) that

(0.23) $|b_{XY}| \leqslant \sigma_X \sigma_Y.$

This inequality is obviously equivalent to (0.21); it is sometimes called the *Cauchy inequality*, or the *Buniakovsky inequality*, or the *Schwarz inequality* (after the scientists who proved some related inequalities in other branches of mathematics). It should also be noted that if $r = 1$ (or $r = -1$) then $\langle (\mathring{X} - a\mathring{Y})^2 \rangle = 0$ for some positive (negative) a, i.e. X, coincides here with probability one, is equal to a linear combination of the form $aY + b$.

As a particular example of a two-dimensional probability distribution we consider the very important case of the *two-dimensional normal (or Gaussian) distribution*. This distribution is determined by a two-dimensional probability density of the form

(0.24)
$$p(x,y) = \dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\, \exp\left\{ -\dfrac{1}{2(1-r^2)} \left[\dfrac{(x-m_1)^2}{\sigma_1^2} \right.\right.$$
$$\left.\left. - 2r\dfrac{(x-m_1)(y-m_2)}{\sigma_1\sigma_2} + \dfrac{(y-m_2)^2}{\sigma_2^2} \right]\right\},$$

where m_1 and m_2 are arbitrary numbers, $\sigma_1 > 0$, $\sigma_2 > 0$, and $-1 < r < 1$. It is easy to verify that the function (0.24) satisfies the conditions (0.15), i.e. it is indeed the probability density of a two-dimensional random variable. In three-dimensional space the surface $z = p(x,y)$ is bell-shaped; on the plane (x,y) it is associated with equiprobability curves having the shape of similar ellipses. If $r = 1$ or $r = -1$, (0.24) is meaningless. As $r \to \pm 1$ all the ellipses $p(x,y) = $ const. tend to segments of the

straight line $x - m_1 = \pm \dfrac{\sigma_1}{\sigma_2} (y - m_2)$. In other words, as $r \to \pm 1$

the distribution with a density (0.24) tends to a singular distribution concentrated on this line, and both X and Y have a one-dimensional normal distribution. The distribution of the pair (X,Y) is then called a *degenerate* (or *improper) normal distribution*. It is easy to verify that for the two-dimensional density (0.24) the one-dimensional densities of the components X and Y are equal to

$$p_1(x) = \int_{-\infty}^{\infty} p(x,y)dy = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-(x-m_1)^2/2\sigma_1^2},$$

(0.25)

$$p_2(y) = \int_{-\infty}^{\infty} p(x,y)dx = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-(x-m_2)^2/2\sigma_2^2},$$

i.e. they are both one-dimensional normal densities. From these formulae it follows at once that m_1 and σ_1^2 are equal to the mean value and variance of the random variable X, while m_2 and σ_2^2 are equal to the mean value and variance of the variable Y. The probabilistic meaning of the parameter r in (0.24) can be illustrated by evaluation of the second joint moment $\mu^{(1,1)}$ with the use of (0.18). It turns out that $\mu^{(1,1)} = m_1 m_2 + r\sigma_1\sigma_2$ and, hence, $r = \dfrac{\mu^{(1,1)} - m_1 m_2}{\sigma_1 \sigma_2}$ is the correlation coefficient of the variables X and Y. Note that when $r = 0$ (0.24) takes the form

(0.26) $p(x,y) = p_1(x)p_2(y),$

where $p_1(x)$ and $p_2(x)$ are one-dimensional probability densities of the components X and Y, respectively. Hence, *for*

normally distributed variables **X** = (X,Y) *the non-correlatedness of* **X** *and* **Y** *implies their mutual independence.* For other two-dimensional distributions this statement does not, in general, hold. Of course, from the independence of the variables *X* and *Y* it always follows that they are uncorrelated (since if (0.26) holds, then $\mu^{(1,1)} = m_X m_Y$, i.e. $b_{XY} = 0$ and $r = 0$), but the lack of correlation by no means proves the independence of *X* and *Y*.[7]

When we have to deal simultaneously with *n* different random variables $X_1, X_2, \ldots, X_n$, all these variables taken together can be regarded as a single *n-dimensional* (or *n-variate*) random variable **X** = $(X_1, X_2, \ldots, X_n)$. In assigning the probability distribution of the variable **X** one can use the *n-dimensional distribution function*

(0.27) $F(x_1, x_2, \ldots, x_n) = \mathbf{P}\{X_1 < x_1, X_2 < x_2, \ldots, X_n < x_n\}.$

The function $F(x_1, x_2, \ldots, x_n)$ must tend to zero when at least one of the variables $x_1, x_2, \ldots, x_n$ tends to $-\infty$. Further, it must tend to unity when all the variables $x_1, x_2, \ldots, x_n$ simultaneously tend to $+\infty$, and be a monotone nondecreasing function of each of the variables x_i. Moreover, it must satisfy the following condition generalizing the condition (0.13): a combination of the values of this function expressing the probability

(0.28) $\mathbf{P}\{x_1 \leqslant X_1 < x_1 + a_1, x_2 \leqslant X_2 < x_2 + a_2, \ldots, x_n \leqslant X_n < x_n + a_n\}$

cannot be negative for any $x_1, x_2, \ldots, x_n, a_1 \geqslant 0, a_2 \geqslant 0, \ldots, a_n \geqslant 0$.

In the particular case of the existence of the *n-dimensional probability density* $p(x_1, x_2, \ldots, x_n)$ we have:

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} p(x_1', \ldots, x_n') dx_1' \ldots dx_n',$$

(0.29)

$$p(x_1, \ldots, x_n) = \frac{\partial^n F(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n}.$$

Here, all the conditions imposed on $F(x_1, \ldots, x_n)$ reduce to the conditions

$$(0.30) \quad \int_{-\infty}^{\infty}...\int_{-\infty}^{\infty} p(x_1,...,x_n)dx_1...dx_n = 1,$$

$$p(x_1,...,x_n) \geqslant 0 \quad \text{for all } x_1,...,x_n.$$

The moments of the n-dimensional random variable $\mathbf{X}$ are defined by the equations

$$(0.31) \quad \mu^{(j_1,j_2,...,j_n)} = \langle X_1^{j_1} X_2^{j_2} ... X_n^{j_n} \rangle$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty} x_1^{j_1} x_2^{j_2}...x_n^{j_n} d^n F(x_1,x_2,...,x_n),$$

where the right-hand side contains an n-dimensional Stieltjes integral, which is a natural generalization of a one-dimensional integral. In this book we can, however, restrict ourselves to the simplest case of continuous variables $\mathbf{X}$, having a probability density $p(x_1,x_2, ..., x_n)$. In this case (0.31) can be replaced by the equation of the form

$$(0.32) \quad \mu^{(j_1,j_2,...,j_n)}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty} x_1^{j_1} x_2^{j_2}...x_n^{j_n} p(x_1,x_2,...,x_n)dx_1 dx_2...dx_n,$$

containing only an ordinary n-dimensional integral.

The simplest moments are those of the first order

$$(0.33) \quad \mu_k^{(1)} = m_{X_k} = \langle X_k \rangle, \ k = 1, ..., n,$$

which, taken together, form the n-dimensional *mean value vector* $\mathbf{m} = (m_{X_1}, ..., m_{X_n})$. Also of great importance are the second moments

$$(0.34) \quad \mu_{jk}^{(1,1)} = B_{jk} = \langle X_j X_k \rangle, \ j,k = 1,...,n,$$

forming the $n \times n$ matrix $\mathsf{B} = \|B_{jk}\|_{j,k=1,...,n}$. Instead of the second moment B_{jk} it is often more convenient to consider the second central moments

$$(0.34a) \quad b_{jk} = \langle (X_j - \langle X_j \rangle)(X_k - \langle X_k \rangle) \rangle = B_{jk} - m_{X_j} m_{X_k},$$

which are, at $j = k$, *variances of the variables* X_j and, at $j \neq k$, *covariances of* X_j *and* X_k. In some cases the normed second central moments (i.e. correlation coefficients) $r_{jk} = b_{jk}/\sigma_j \sigma_k$ are

also useful. The matrix of the second central moments $\overset{\circ}{\mathcal{B}} = \|b_{jk}\|$ is sometimes called the *covariance matrix* of the multidimensional random variable X, and the matrix of the normed moments $R = \|r_{jk}\|$ is called the *correlation matrix* of X. However the term *correlation matrix* is sometimes also applied to the matrix $\mathcal{B}$ or $\mathcal{B}$. We shall seldom use the matrix in this book, and for brevity we shall also apply the term correlation matrix to both matrices $\mathcal{B}$ and $\overset{\circ}{\mathcal{B}}$ (but we shall occasionally call $\overset{\circ}{\mathcal{B}}$ the *centered correlation matrix* to avoid confusion).

Both correlation matrices $\mathcal{B}$ and $\overset{\circ}{\mathcal{B}}$ are symmetric by definition: $B_{jk} = B_{kj}$ and $b_{jk} = b_{kj}$. Moreover, these matrices are also *positive definite: for any real numbers* $c_1, ..., c_n$

(0.35) $\displaystyle\sum_{j=1}^{n} \sum_{k=1}^{n} B_{jk} c_j c_k \geqslant 0$

and

(0.36) $\displaystyle\sum_{j=1}^{n} \sum_{k=1}^{n} b_{jk} c_j c_k \geqslant 0.$

This is clear, since the left-hand sides of (0.35) and (0.36) are equal to the mean values of non-negative quantities $(c_1 X_1 + ...$

$+ c_n X_n)^2$, and, respectively, $(c_1 \overset{\circ}{X}_1 + ... + c_n \overset{\circ}{X}_n)^2$, where $\overset{\circ}{X}_j = X_j - \langle X_j \rangle$. Note also that the inequality (0.36) is a particular case of (0.35), because the moments B_{jk} transform to b_{jk} when X_j, $j = 1, ..., n$ are replaced by $\overset{\circ}{X}_j$. In the particular case of $n = 2$ the conditions for positive definiteness of $\mathcal{B}$ and $\overset{\circ}{\mathcal{B}}$ reduce to the single inequality $|B_{jk}| \leqslant (B_{jj} B_{kk})^{1/2}$ and, respectively, $|b_{jk}| \leqslant (b_{jj} b_{kk})^{1/2}$; the second of these inequalities has already been discussed (see (0.23)). It is also easy to show that for any symmetric positive definite matrix $\mathcal{B}$ there exists a multidimensional probability distribution with mean values equal to zero and the matrix of second moments equal to . This follows from the description of multidimensional normal probability distributions, which will be given below.

Let us now consider the general *multidimensional normal* (or *Gaussian*) *probability distribution*. This is the distribution with a probability density of the form

(0.37) $p(x_1, x_2, ..., x_n) = A \exp\left\{ -\dfrac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} g_{jk} (x_j - m_j)(x_k - m_k) \right\}$

Here, m_j, $j = 1, ... \ n$, are arbitrary real constants; g_{jk}, $j, k + 1, ...,$

n, are real constants such that $g_{jk} = g_{kj}$, and $G = \|g_{jk}\|$ is a strictly positive definite matrix (i.e. $\sum_{j=1}^{n} \sum_{k=1}^{n} g_{jk}c_jc_k > 0$ for any real c_1, ..., c_n not all of which are equal to zero). Moreover, A is a constant determined by the normalization condition

$$(0.38) \quad A\int_{-\infty}^{\infty}....\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}g_{jk}(x_j - m_j)(x_k - m_k)\right\}dx_1...dx_n = 1$$

Integrating the left-hand side of (0.38) (for which it is convenient to make the coordinate transformation reducing the quadratic form in the exponent to a sum of squares), it is not difficult to verify that

$$(0.39) \quad A = \frac{G^{1/2}}{(2\pi)^{n/2}},$$

where $G = |g_{jk}|$ is the determinant of $G = \|g_{jk}\|$. From this it follows that if   is only a positive definite (but not a strictly positive definite) matrix, so that $\sum_{j=1}^{n} \sum_{k=1}^{n} g_{jk}c_jc_k$ can vanish at some nonzero values of c_1, ..., c_n, then $A = 0$ (since here $G = 0$). Thus in this case the function (0.37) is no longer the probability density. Nevertheless, the multidimensional normal distribution can, in fact, be determined when $G = \|g_{jk}\|$ is a positive definite (but not strictly positive definite) matrix.   In such a case it should only be assumed that the random variables X_1, ..., X_n satisfy, with probability one, one or more linear equations, so that the entire probability distribution is concentrated on some ($n -$ m)-dimensional linear subspace (where $m > 0$ is the number of linear equations) of the n-dimensional space of points $\mathbf{x} = (x_1, ..., x_n)$. If the distribution concentrated on this subspace is an ($n -m$)-dimensional normal distribution, the initial distribution of n variables X_1, ..., X_n will also be called a normal (but only a *degenerate*, or *improper*, *normal*) distribution.   From now on we shall consider, for simplicity, only nondegenerate (i.e. proper) normal distributions with a probability density of the form (0.37).   However, it can be easily shown that all the facts given below are also valid for degenerate normal distributions.[8]

The constants m_j and g_{jk} in (0.37) are simply related to the first and second moments of the distribution under consideration.   In fact, substituting (0.37) into the general

equations for the first and second central moments and integrating over all the variables (for which it is again convenient to transform the quadratic form on the right-hand side of (0.37) to a sum of squares) it is easy to show that

$$(0.40) \qquad m_{X_j} = \langle X_j \rangle = m_j, \quad b_{jk} = \langle (X_j - \langle X_j \rangle)(X_k - \langle X_k \rangle) \rangle = \frac{G_{jk}}{G} .$$

Here, $G = \det G = |G_{jk}|$, as above, and $G_{jk} = \dfrac{\partial G}{\partial g_{jk}}$ is the cofactor of the element g_{jk} in the determinant G (so that the matrices $B = \|b_{jk}\|$ and $G = \|g_{jk}\|$ are mutually inverse). When $n = 2$, eqs. (0.37), (0.39), and (0.40) evidently agree with (0.24). From (0.40) we also obtain an expression for the ordinary (noncentral) second moments of the normal distribution with a probability density (0.37):

$$(0.41) \qquad B_{jk} = \langle X_j X_k \rangle = \frac{G_{jk}}{G} + m_j m_k.$$

Note that, according to (0.40), (0.41), and (0.37), *the probability density of the normal probability distribution is fully determined by its first and second moments.* Therefore, in this case the values of all the statistical characteristics of the variables X_1, ..., X_n are functions of their first and second moments.

Let $\mathbf{X} = (X_1, ..., X_n)$ have a normal probability distribution with a density (0.37). Consider a subset X_{i_1}, ..., X_{i_m} containing $m < n$ variables of the X_j, $j = 1, ... n$. To determine the probability distribution of this subset one must integrate the density (0.37) over all the other variables except x_{i_1}, ..., x_{i_m} from $-\infty$ to ∞. It is readily seen that the resulting function of x_{i_1}, ..., x_{i_m} will again have the form of a constant multiplied by

$$\exp\left\{ -\frac{1}{2} \sum_{j=1}^{m} \sum_{k=1}^{m} h_{jk}(x_{i_j} - m_{i_j})(x_{i_k} - m_{i_k}) \right\},$$

where $\|h_{jk}\|$ is some strictly positive definite m-dimensional matrix. Therefore, if the *random variables* X_1, ..., X_n *have an n-dimensional normal probability distribution, then the probability distribution of any* $m < n$ *of these variables also is normal.* Note now that under any nondegenerate linear

transformation of the coordinates x_1, ..., x_n the strictly positive quadratic form

$$\sum_{j=1}^{n} \sum_{k=1}^{n} g_{jk}(x_j - m_j)(x_k - m_k)$$

transforms to some strictly positive quadratic form of the new variables. Since the Jacobian of the linear transformation is equal to a constant, the probability density (0.37) will again transform to a normal probability density. It follows therefore that *if the variables X_1, ..., X_n have a normal probability distribution, the probability distribution of any n linearly independent combinations Y_1, ..., Y_n of X_1, ..., X_n is also normal.* Hence *any number m of such linear combinations Y_1, ..., Y_m have also a joint normal distribution.* (In particular, each

separate linear combination $Y = \sum_{j=1}^{n} c_j X_j + c_0$ has a

one-dimensional normal distribution.)

The study of one-dimensional (scalar) and multidimensional (vector) random variables constitutes the principal content of classical probability theory. In recent years, however, it has been found useful (and often necessary) to consider more complex probabilistic objects, namely *random functions*. This term denotes functions of some argument t (mostly, although not always, time) which are specified by the results of some experiment or observation and which can take different values (characterized by certain statistical stability) when the experiment or observation is repeated many times. (The precise meaning of the statement concerning statistical stability will be discussed below.) The study of some particular classes of such random functions is the subject of the present book.

The simplest situation where the concept of the random function is useful is when one has to deal with the data of some observations repeated at regular time intervals. Here, of course, the researcher has at his disposal only a certain finite number N of observations which can be regarded as the value of some N-dimensional random variable $\mathbf{X} = (X_1, ..., X_N)$. However, if N is very large (for instance, of the order of some hundreds or thousands), it is often more convenient to regard the values X_1, ..., X_N as a part of an infinite time series $...X_{-1}$, X_0, X_1, ..., X_N, X_{N+1} ... of all the possible

observations. Such a time series is a random function of the discrete argument t (the observation number), which runs through all possible integer values. The random functions of the integral argument t are often called *random sequences*. More complicated experiments, from the standpoint of theory, are those in which observations are conducted continuously, i.e. which register the value of some function of the argument t running through all real values of a certain, possibly infinite, interval. Random functions of such a continuous argument t are very often called *random processes*. Finally, in the case of the experimental (or observational) function $X(t_1, ..., t_k)$ of several real arguments (which usually have the meaning of the coordinates of a point in some space), the term *random field* is often used.[9]

Historically, one of the first examples of experiments producing random functions were the observations of the *Brownian motion* of small particles suspended in a fluid (a liquid or a gas). Such observations were first described over a century and a half ago.[10] The observed particles always move extremely irregularly (see Fig. 4(a)). When the values of a coordinate of one of these particles are recorded at definite time intervals we obtain one *realization* (or *observed value*, or *sample value*, or *trajectory*) of some random sequence $X(t)$ (see the schematic diagram of Fig. 4(b),(c)). Similarly, the values of the coordinate $x(t)$ of a Brownian particle at all t form a *realization* of the random process $X(t)$. Brownian motion is also closely related to the current and voltage fluctuations in electrical circuits. Strictly speaking, the voltage $V = V(t)$ across any conductor and the current $I = I(t)$ in the conductor are always random functions of time, since the thermal motion of the electrons (which is similar to the Brownian motion of suspended particles) causes uncontrollable fluctuations in the voltage and current, known as *thermal noise.* The role of electrical fluctuations became particularly important when vacuum tubes and semiconductor circuits came into wide use. A vacuum tube and a semiconductor are always sources of considerable noise (i.e. of strong fluctuations) arising from a number of physical causes. (For example, inevitable fluctuations in the number of electrons passing through the tube during identical time intervals produce the so-called *shot noise* while the fluctuations in cathode emission intensity cause the *flicker noise.*) In radio receivers, apart from the noise arising in the

electrical circuit, one also always observes "fadings". These
fadings are a random fluctuation of the power of the signal
received due to scattering of the radio waves from
inhomogeneities in the refractive index of the atmosphere and
to the influence of extraneous electrical discharges
(meteorological and industrial noises). These noises are of
greatest importance in the operation of radar and space
communication systems, where the "useful signal" is usually
rather weak, and the distorting noises are always strong.
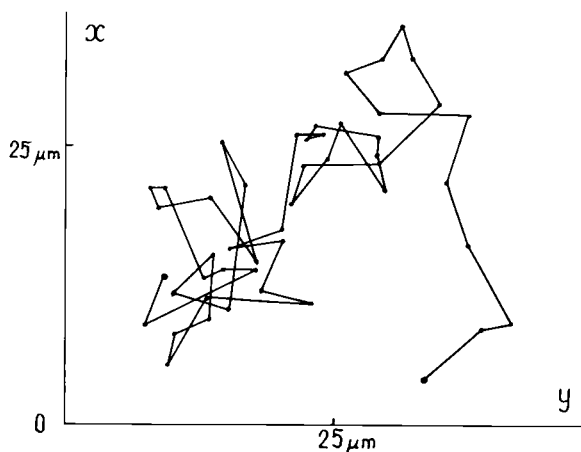
(a)
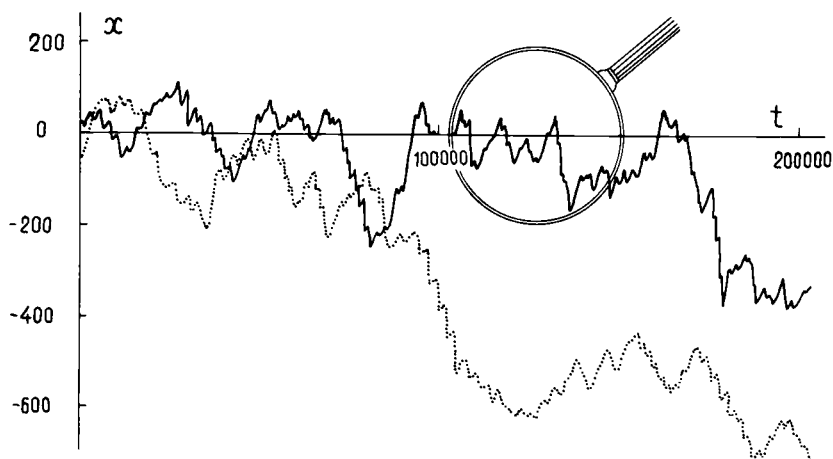
(b)
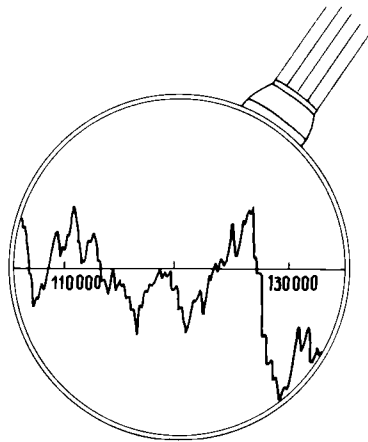
(c)

Fig. 4. (a) Microscopic observations of a Brownian particle at times separated by 30 sec intervals (Perrin, 1916). (b) Two schematic models of the dependence of x-coordinate on time for a Brownian trajectory (Wold, 1965). (c) Twelve times enlarged portion of one of the curves in (b) (Wold, 1965).

Electrical circuits and devices are essential parts in most servomechanisms and automatic control systems; on the other hand, the operation of many automatic control systems is also affected by irregular noise of non-electrical origin. (For example, the control system of aircraft flying in a turbulent atmosphere is affected by disordered gusts of wind.) Therefore it is not surprising that the wide development of electrical and radio engineering, electronics, radar techniques, and automatic control instruments became a powerful stimulus for a vigorous development of random function theory. As a result, a very important new applied discipline emerged, which is now called "random noise theory", or "statistical radio engineering", or "statistical radiophysics", or "statistical communication theory", or "statistical automatic control theory", depending on the principal interests of a given author or authors.[11]
    Above, speaking of the control system of aircraft, we mentioned *turbulence*. This word denotes irregular time and space variations of the velocity, pressure, temperature, and

other fluid mechanical quantities, which are involved in almost any flow of a liquid or a gas. (So-called laminar flows, which are not accompanied by irregular fluctuations, are in fact a rare exception.) Owing to turbulence the values of the three velocity components, pressure, temperature, and other flow parameters must usually be regarded as realizations of certain random fields — random functions of time and the three space coordinates. This is true for both the flows occurring in nature (in the earth's and other planets' atmospheres, rivers, lakes, seas, and oceans) and for flows encountered in engineering or created in physical laboratories. The effect of turbulent gusts of wind on plants, buildings, and other structures is one of the sources of *microseisms* — small random vibrations of the earth's surface recorded by sensitive seismometers and making the level of the earth's surface at a fixed point a random function of time. A similar "random" nature is exhibited by so-called *geomagnetic variations*, i.e. disordered fluctuations of the earth's magnetic field. Many time functions of biological origin can also naturally be regarded as realizations of random processes; a good example are oscillations in electric brain potentials recorded on electroencephalograms (EEG). Very often, random functions arise in diverse branches of technology: if some product is manufactured continuously at a factory, its parameters always fluctuate owing to some uncontrollable factors, and the time variation of each of these parameters can be regarded as the realization of a random process. Consider, for example, an ordinary weaver's thread: it is never completely uniform and its thickness can be regarded as a random function of the distance from the beginning of the thread. Similarly, the vertical coordinate of any rough horizontal surface is a random function of two horizontal coordinates. An important example of a rough surface is the sea surface; since sea waves are mobile, the height of the sea surface is a random function of three variables: two space coordinates and time. The specific examples of random functions listed here are the subject of a huge literature, differing greatly in style and level.[12] In the following text we shall also encounter other examples of random functions of applied origin; now, however, it hardly makes sense to continue citing them.
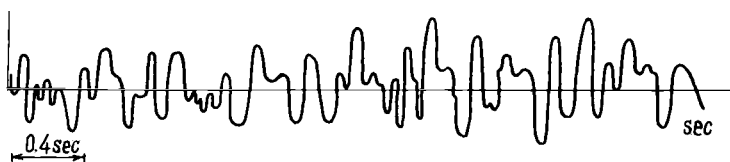
    As examples of observed values of random functions, Fig. 5 shows curves resulting from the measurement of (a) fadings of a

radio signal received by a radar; (b) the temperature at a fixed point of the surface layer of the atmosphere; (c) fluctuations of the earth's level (microseisms) recorded in 1964 by a seismograph station in East Kazakhstan; (d) the time variation of sea wave height at a given point of the sea; (e) electroencephalogram of a healthy man during his sleep; and (f) variations in the diameter of a manufactured thread along its length.   We can see that all the curves of Fig. 5 are similar to each other although they refer to entirely different phenomena.   Curves of this type, which fluctuate in a disordered manner with sharp and irregular variations are very often encountered in diverse practical activities; their external appearance alone usually suggests that we are dealing with a realization of some random function, which requires statistical treatment.   Sometimes, however, this first impression may nevertheless prove erroneous.   Therefore, we shall give more attention to the applicability of the theory of random functions to actual observations.

From the foregoing it follows that the main feature of the random function is by no means the involved shape and irregularity of its plot, but the presence of a whole "statistical ensemble" of similar realizations $x_1(t)$, $x_2(t)$ ... etc., each of which corresponds to one of a set of possible experiments (or observations) of the same type and which, taken together, possess some "statistical stability". The presence of only one extremely irregular curve (such as those depicted in Fig. 5) only indicates that we are dealing with a phenomenon for which we can hardly hope to create a full-fledged deterministic ("dynamic") theory enabling us to describe all the observed bends of the curve. (It is also clear that such a dynamic theory, if possible at all, must be very complicated and hardly can be useful, because exact values of an extremely involved function are usually useless in practice.) Therefore, it seems quite natural to apply the statistical approach here and try to find an ensemble of curves which would contain our curve and which would, on the whole, correspond to a single random function $X(t)$ with comparatively simple statistical properties.

All the curves depicted in Fig. 5 represent observations admitting of many repetitions under similar conditions. Therefore, the problem of determining a random function corresponding to these observations seems to be relatively simple.   Note, however, that the "statistical ensemble" of all

possible realizations $x(t)$ can often be chosen in different ways. For instance, in the case of the curve (b) in Fig. 5 we may include in the ensemble only measurements at the same point, at the same time of the day, and in approximately the same weather, or we may consider all the possible measurements at different points, which begin at arbitrary time moments throughout the year. Similarly, in the case of curve (e) we may consider only electroencephalograms of healthy people of about the same age and in an appropriately

(a)
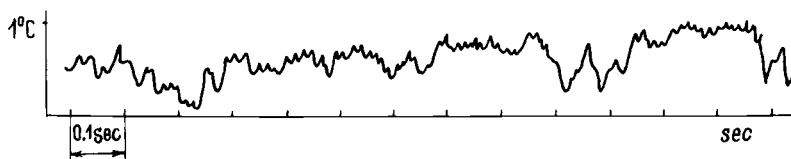
(b)
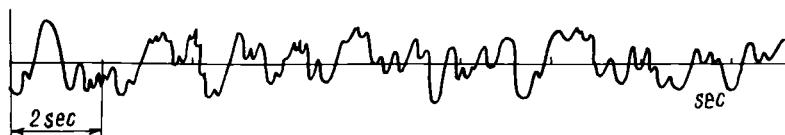
(c)
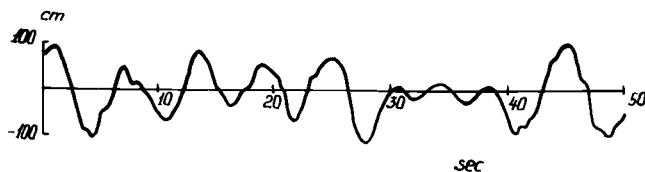
(d)

(e)
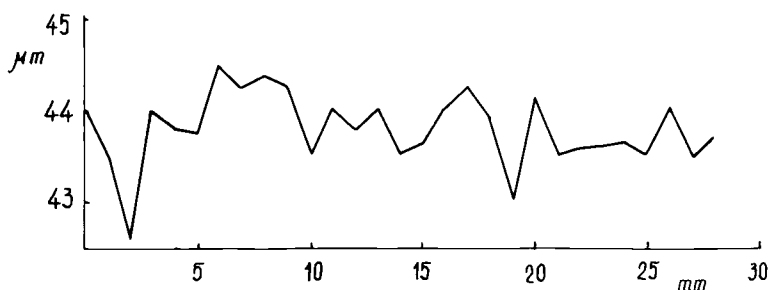
(f)

Fig. 5. Observed values of random functions. (a) Fading of the intensity of radio signals received in a radar (James et al., 1947; the units of intensity are not indicated). (b) Fluctuations of air temperature at a point in the atmosphere (Monin and Yaglom, 1971). (c) Earth level fluctuations recorded in East Kazakhstan (Vinnik, 1968; the units of the level fluctuations are not indicated). (d) Fluctuations of the sea height at a given point produced by waves (Taira et al., 1971). (e) Electroencephalogram of a healthy sleeping man (Kozhevnikov and Meshcherskii, 1963). (f) Variation in the diameter of a nylon thread (Cramer and Leadbetter, 1967).

identical physical and psychic state, or we may include in the "statistical ensemble" all the recorded electroencephalograms (most of which are made in hospitals and refer to sick people). The same is true for curve (f) where we may consider only the threads spun by one loom (or a bank of similar looms), or we may include in the number of realizations the measurements referring to any and all weaver's threads put out at all the possible factories with very different equipment, and so on. Hence the same observed curve $x(t)$ may sometimes be regarded as the

realization of several different random functions with different statistical characteristics. In some cases, when solving different problems referring to one and the same curve, such a curve may sometimes usefully be regarded as the realization of a random function, and sometimes as a curve determined, at least approximately, by certain dynamic equations.

The concept of random functions is also widely applied to many functions $x(t)$ arising in certain geophysical, astrophysical, economic, and related problems, in which it seems difficult to determine the "statistical ensemble of similar experiments or observations" producing a set of similar realizations. Of course, curves (b), (c), and (d) in Fig. 5 are also of geophysical origin, but they are related to small-scale fluctuations, which naturally fit into the scheme of "mass phenomena". However, curves of the same type often appear when we plot data representing much slower large-scale variations of geophysical parameters. As an example, Fig. 6(a)-(c) shows the variations in the mean daily wind velocity, mean monthly (January) air temperature, and mean annual temperature at some fixed observation points. All these three curves are rather similar to curve (b) in Fig. 5, although the scales at the time axis are quite different here. The difference in scale makes the determination of the appropriate "statistical ensemble of similar observation series" a rather difficult task (at least in cases of curves (b) and (c) in Fig. 6). Another, appreciably more exotic, example of the same type is provided by the palaeoclimatic time series in Fig. 6(d) showing the fluctuations in the Earth's glaciation (i.e. in the global ice-volume) during the last million years, as recorded by changes in isotopic composition of the fossil plankton and representing the long-period variations in the mean Earth's temperature. A similar example from the field of economics is given by the famous Beveridge series of wheat prices in Western and Central Europe from 1500 to 1869 plotted in Fig. 6(e). This series was made up of recorded prices in nearly 50 places in various countries. To exclude the effect of systematic price changes, Beveridge calculated for each year the ratio (in percent) of the average price for that year to the mean price of the 31 years for which it is a center; the variations of precisely this index are presented in Fig. 6(e). Finally, the curve of Fig. 6(f) exhibits the astrophysical series of average annual values of

Wolfer's sunspot numbers characterizing the solar activity.[13] To be definite, we shall discuss just the last curve in Fig. 6.

We can, of course, assume that the curve of Fig. 6(f) is included in an ensemble of similar curves referring to all possible stars which have all their principal parameters close to the corresponding parameters of our Sun and are in approximately the same stage of evolution. However, since we cannot precisely indicate any one of such stars, the assumption formulated here can hardly be considered fruitful. Let us examine, instead, some segment $x(t)$, $x(t + 1)$, ..., $x(t + N - 1)$ of the time series depicted in Fig. 6(f). Then we can regard as all its possible "realizations" simply all the possible sets of N successive observations of Wolfer's number; the observed data give us ground to believe that the statistical stability of the results will actually take place for such an "ensemble of realizations". This approach means in fact that we accept as an "ensemble of all the possible realizations" of the random function $X(t)$ the set of all the possible time shifts (i.e. transformations of the type $x(t) \rightarrow x(t + a)$, where a is fixed) of one and the same function $x(t)$. It is important to stress here that the results of this book will often be applicable to the obtained "random function" $X(t)$, provided that the probabilistic averaging will be replaced by time averaging (which evidently coincides with averaging over the introduced "set of realizations").[14] In particular, it is natural to think that the same approach can be applied to curves (c), (d), and (e) in Fig. 6. On the other hand, for instance, in the case of the data plotted in Fig. 7(a) on variations of the average annual level of the Caspian Sea or the data in Fig. 7(b) on the monthly total numbers of internal airline passengers, the efficiency of statistical methods seems doubtful. In fact, it is quite difficult to indicate a reasonable "statistical ensemble of similar observations", which includes the data presented in Fig. 7(a) or 7(b), but at the same time the set of all the possible time shifts has apparently no statistical stability any more. Nevertheless, methods of the theory of random functions were also applied by some investigators to both series depicted in Fig. 7. In fact, there exist several particular techniques of statistical analysis suitable for treatment of such nonstationary time series.[15] It should be borne in mind, however, that the practical expediency of the application of random function

(a)

(b)

(c)

(d)

(e)

years
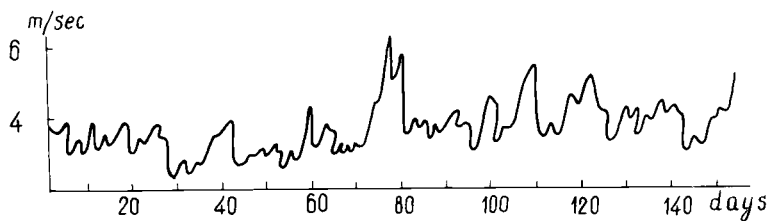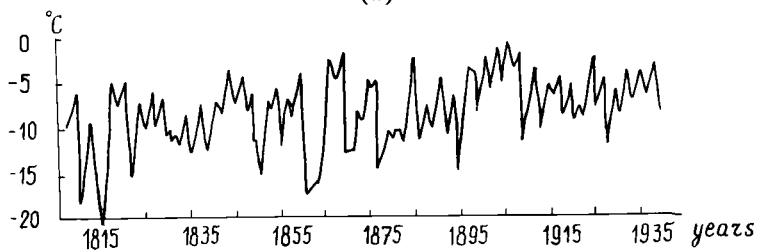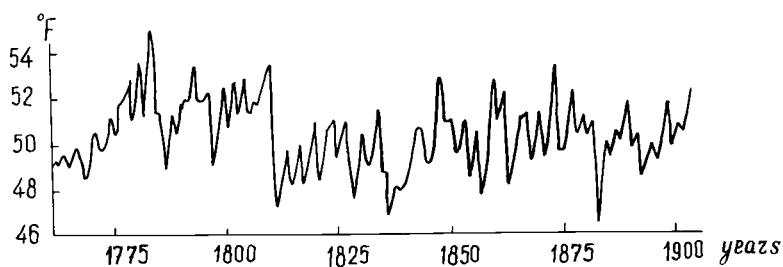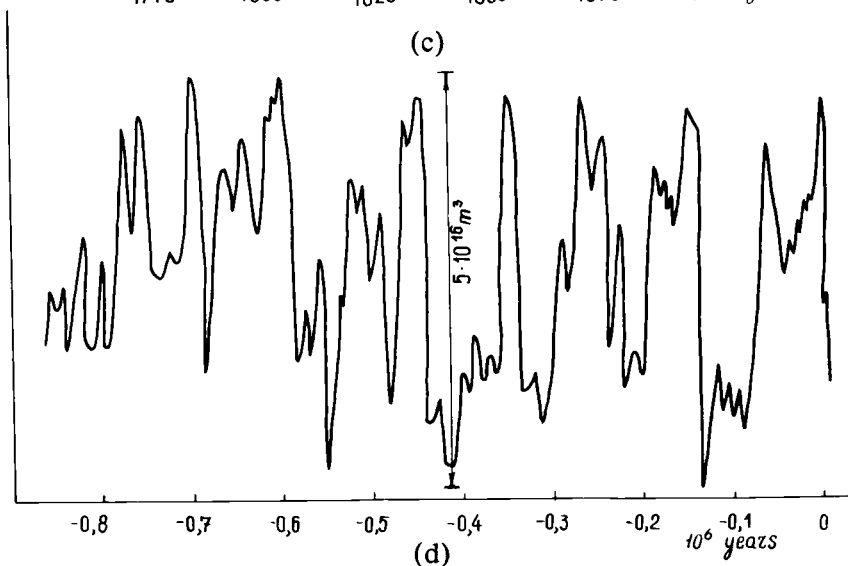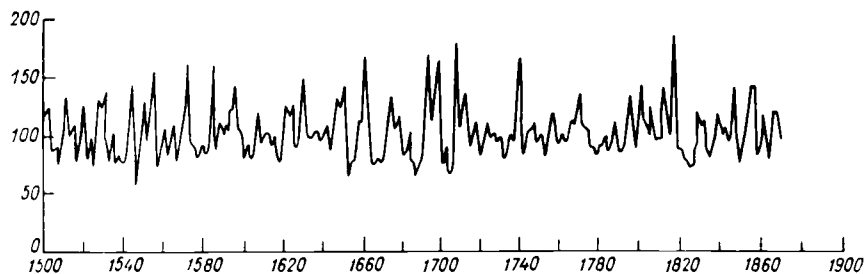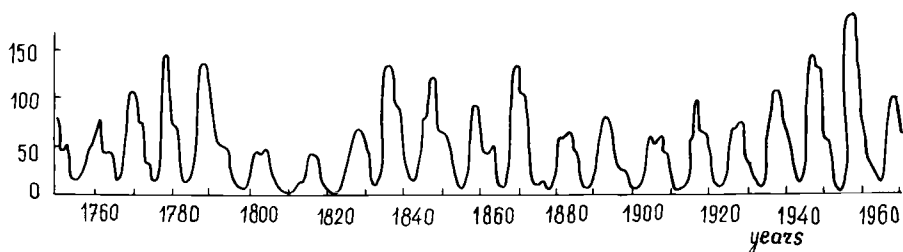
(f)

Fig. 6. Observed values of random functions. (a) Variations of daily mean wind velocity registered at Kew Observatory (Sheppard, 1951). (b) Monthly mean temperatures for January in Leningrad (Alisov et al., 1952). (c) London yearly mean temperatures from 1763 to 1900 (Cramér and Leadbetter, 1967). (d) Fluctuations in the Earth's glaciation for the last million years (GARP Publications Series, 1975). (e) The Beveridge series describing wheat price fluctuations in Western and Central Europe from 1500 to 1869 (Beveridge, 1921). (f) Yearly mean values of Wolfer's sunspot number (Waldmeir, 1961).

theory to such situations can hardly be elucidated from general considerations, but it requires a special investigation each time.

While the applicability of random function theory to the intricate and clearly irregular curves shown in Fig. 7 requires special investigation, the possibility of applying the same theory to many much simpler and "regular" functions is absolutely indisputable. Let us consider, for instance, the
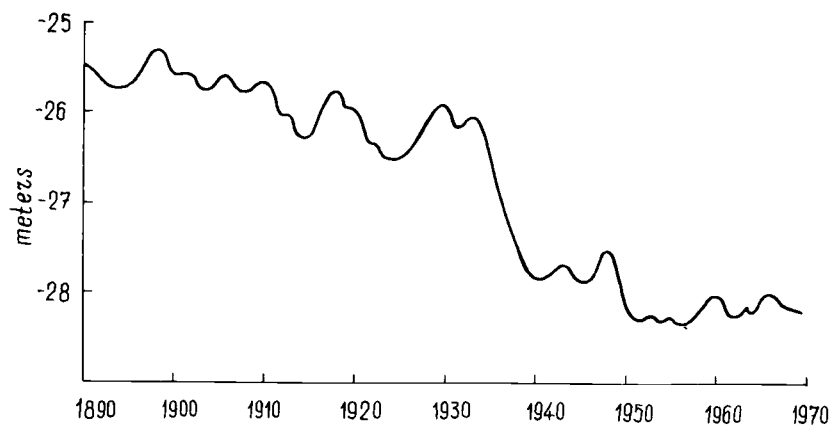
(a)
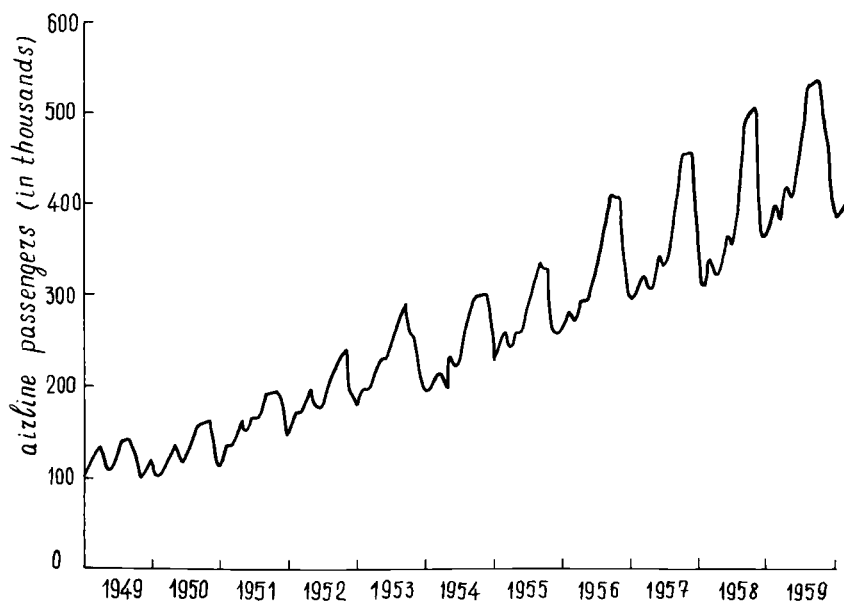
(b)

Fig. 7. (a) Long-term variation in the Caspian Sea level
(Privalsky, 1973). (b) Monthly mean numbers of
international airline passengers (Box and Jenkins,
1970).
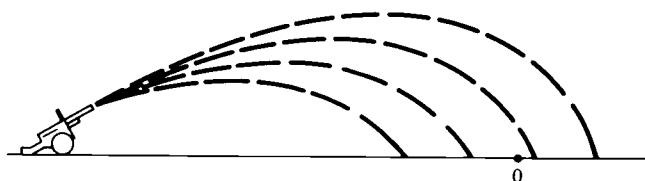
Fig. 8. Schematic graph of shell trajectories by firing at target O from a single cannon.

data on shooting at a fixed target from a single cannon (see Fig. 8). The "statistical ensemble" of the possible shell trajectory can naturally be regarded as the realization of a random function (which is approximately representable by the equation for a parabola with random parameters). As another example of the same type we consider an harmonic oscillation of high frequency ω_0 modulated in amplitude or in phase (or both) by a slowly varying process. Here, we deal with a function of the form $x(t) = A(t)\cos[\omega_0 t + \varphi(t)]$, where at least one of the functions $A(t)$ and $\varphi(t)$ is nonconstant, but changes noticeably only within a time containing many periods $T_0 = 2\pi/\omega_0$. In this case, for time intervals exceeding T_0 only by a few times, our function will be representable rather accurately by a section of a cosinusoid, but its amplitude and phase will be different for different intervals. In electrical and radioengineering, modulated oscillations are often used for transmitting information; then the changes in $A(t)$ and $\varphi(t)$ are usually very irregular, although slow, and therefore different sections of the oscillation process $x(t)$ can conveniently be regarded as the realizations of an harmonic
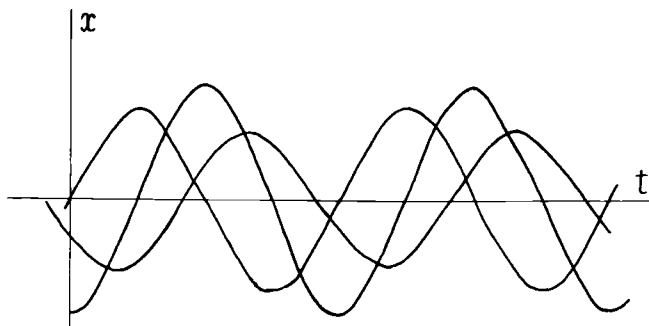
Fig. 9. Three realizations of a random harmonic oscillation of fixed frequency.

oscillation of fixed frequency ω_0 with a random amplitude and phase (see Fig. 9). Such a *random harmonic oscillation* is unquestionably a random function, but its realizations differ sharply from the complicated curves depicted in Figs. 5 and 6.

Of course, the above illustrations cover by no means all the types of random functions encountered in practice. In the probabilistic theory of queues, which has many important applications, the so-called *Poisson process* $N(t)$ is quite useful. This process represents, for example, the number of calls arriving at a telephone exchange during a time interval of length t. All the realizations of $N(t)$ have the form portrayed in Fig. 10(a): they are "step-functions", which rise by a unit at each of their discontinuities. The set of the discontinuity points ("jumps") of the Poisson process $N(t)$ constitutes the so-called *Poisson point system,* or *Poisson point process* (of fixed "intensity" $\lambda > 0$). For such a point process, all the distances ℓ_k between neighboring points are realizations of independent non-negative random variables having an identical exponential probability distribution (i.e. probability density $p(\ell) = \lambda\exp(-\lambda\ell)$ for $\ell \geqslant 0$). The numbers of points on any two nonintersecting time intervals are also represented by independent random variables; the number $N(s)$ of points on any interval of length s has a Poisson distribution with parameter λs (so that $P\{N(s) = n\} = \exp(-\lambda s)(\lambda s)^n/n!$, $n = 0, 1, \dots$ and $\langle N(s)\rangle = \lambda s$). The Poisson process $N(t)$ can also be conveniently extended to negative values of t by assuming that the Poisson system of discontinuity points $\{t_k\}$ is assigned on the entire real axis $-\infty < t < \infty$, and for $t < 0$ the value of $-N(t)$ is equal to the number of points t_k such that $t \leqslant t_k < 0$. In this case the realization of $N(t)$ will have the form shown in Fig. 10(b).

The Poisson distribution of the discontinuity points $\{t_k\}$ is also characteristic for the so-called *random telegraph signal* $X(t)$, which is encountered in some problems of the statistical theory of communication. In this case $X(t)$ takes, in turn, only one of the two values $+a$ and $-a$ in the intervals between its points of discontinuity (see Fig. 11). Of a similar nature are the so-called *point processes with adjoined random variables,* which correspond to the Poisson sequences of discontinuity points. These processes differ from random telegraph signals in that their values in the intervals between the discontinuity points are the realization of some sequence
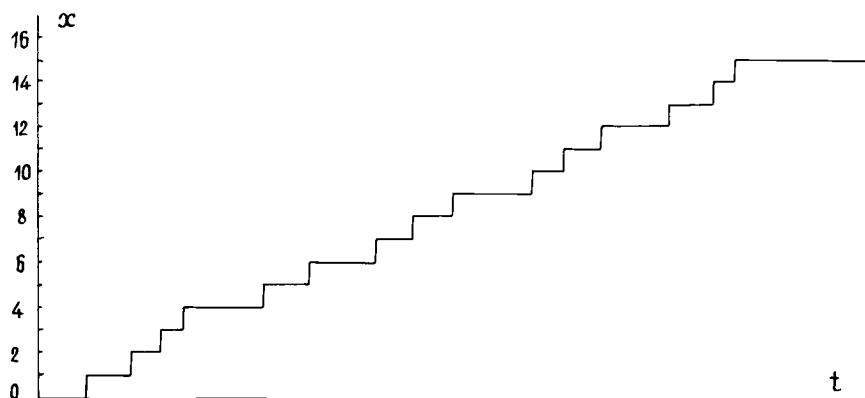
(a)
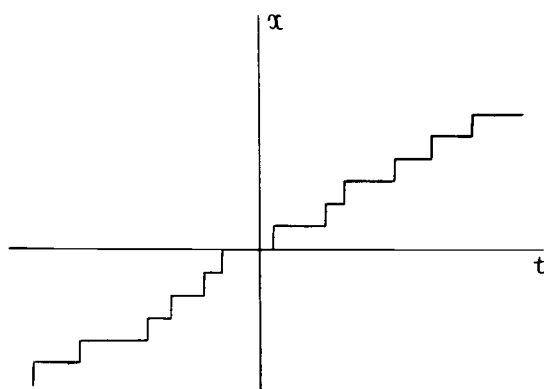
(b)

Fig. 10. (a) Realization of the Poisson process for $t \geqslant 0$.
(b) The Poisson process on the whole line $-\infty < t < \infty$.

of independent, identically distributed random variables having the given distribution function $F(x)$. (In other words, the constant value X_n of the process $X(t)$ for $t_n < t \leqslant t_{n+1}$ is selected at random so that $P\{X_n < x\} = F(x)$ for all n; see Fig. 12.) Finally, one more important class of random processes constructed with the aid of the Poisson point process $\{t_n\}$ consists of the *Poisson pulse processes* represented by the formula
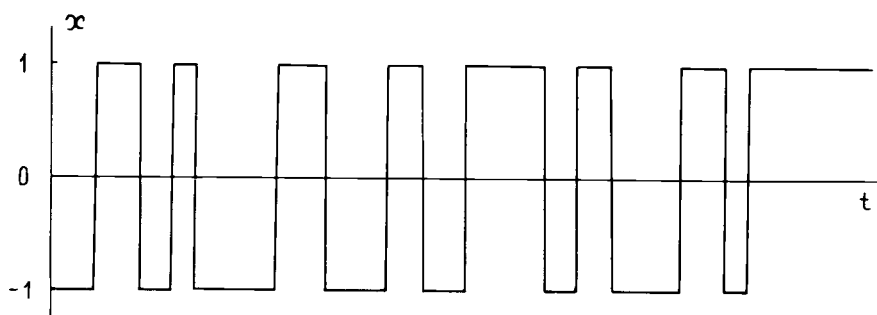
Fig. 11. Observed values of a random telegraph signal.

(0.42) $X(t) = \sum_n E_n \Gamma(t - t_n)$.

Here, $\Gamma(t)$ is a fixed function (which determines the "pulse shape"), and the "amplitudes" E_n form a sequence of independent (of each other and of points t_n), identically distributed random variables with a known distribution function (see Fig. 13, in which all the pulses are taken to be rectangular). An important particular case of the pulse process (0.42) is obtained when all the amplitudes E_n in (0.42) are not random variables but simply identical numbers (then $E_n \Gamma_n(t - t_n)$ can be denoted as $\Gamma_n(t - t_n)$, i.e. it can be assumed that $E_n = 1$ for all n). This particular process can be used as a model of shot noise and therefore it is often called the *shot noise process*.[16]

Assuming further that the point system $\{t_n\}$ is not necessarily a Poisson system but is simply some random point
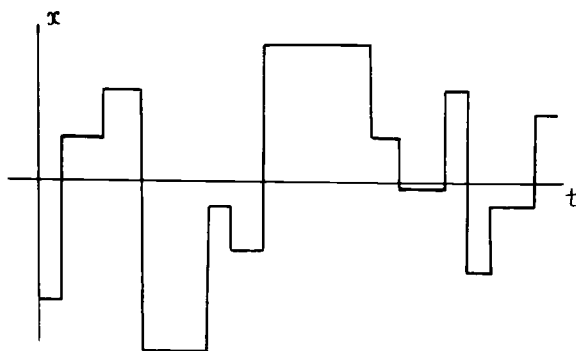
Fig. 12. Observed values of a point process with adjoined random variables.
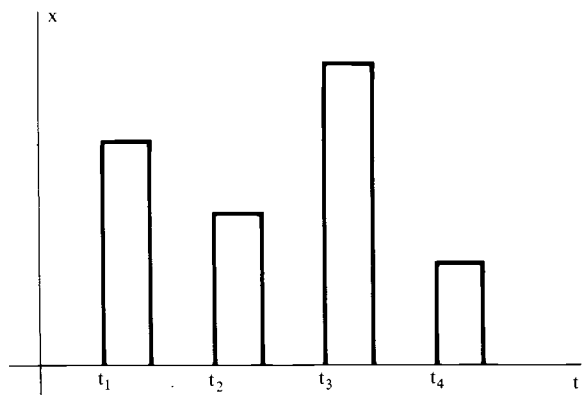
Fig. 13. The Poisson pulse process in the case of rectangular pulses.

process (i.e. a random sequence of points obeying definite statistical laws), we can substantially generalize the classes of the random processes schematically represented in Figs. 10-13. The new statistical models thus obtained also include a number of examples of great applied value.   For instance, by assuming that all the distances $\ell_n = t_{n+1} - t_n$ between the discontinuity points in Fig. 10 are realizations of mutually independent random variables with an identical (but not necessarily exponential) probability distribution, we arrive at the so-called *renewal processes*, which describe, e.g., the number of replacements of a certain machine part within a time t, provided that the time of operation of the part from the moment of its installation to the moment of its failure has a fixed probability distribution.   If, however, the points ..., t_n, t_{n+1}, ... form an arbitrary random point process, Fig. 10 is the realization of a general random *counting process* representing the counting of observations in a series of random events.   If we assume that in Fig. 11 the sequence of discontinuity points t_n is described by the formula $t_n = nT_0 + \Phi$, where T_0 is a fixed number ("period"), and Φ is a random variable, we shall obtain a rectangular *wave with a random phase* (Fig. 14).   By assuming that the sequence $\{t_n\}$ in (0.42) is nonrandom and periodic (i.e. $t_n = nT_0$) and $\{E_n\}$ is a random sequence, we arrive at the *pulse-amplitude modulation* model, which is sometimes used for transmitting information. If, however, all the E_n are constant, but $t_n = nT_0 + \tau_n$, where τ_n is a random sequence, then the process $X(t)$ will describe the *pulse-phase modulation* or, what is the same, *pulse-position*
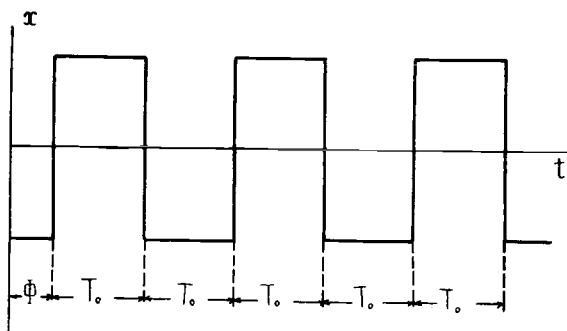
Fig. 14. The rectangular wave with a random phase.

modulation (see Fig. 15), etc.[17] In the available enormous
literature on applied random processes theory one can also
find a variety of other examples of random function, on
which we shall not dwell here.

The concept of a random function does not fit into the
framework of classical probability theory and requires the use
of a new mathematical apparatus. We have already remarked
that investigation of one-dimensional random variables can be
reduced to the study of probability distributions on the straight
line $-\infty < x < \infty$. Such distributions are represented by
monotone nondecreasing functions $F(x)$ which can be studied
with the aid of conventional methods of mathematical analysis,
a branch of mathematics dealing with functions of a real
variable x. To study multidimensional random variables $\mathbf{X} =
(X_1, ..., X_n)$ one has to deal with multidimensional distribution
functions $F(x_1, ..., x_n)$ describing probability distributions in an
n-dimensional space of points $\mathbf{x} = (x_1, ..., x_n)$. These functions
are much more complicated than one-dimensional distribution
functions, but they, too, do not yet extend beyond the scope of
traditional calculus. Moreover, in practice one can often
restrict oneself to n-dimensional distributions determined by a
continuous probability density $p(x_1, ..., x_n)$, which allows one to
simplify the relevant theory still further and to consider only
ordinary multiple integrals of continuous functions. When
studying random functions $X(t)$, however, one already has to use
a much more complicated notion of a *probability distribution in
an infinite dimensional (functional) space* of all the possible
realizations of the function concerned. (For instance, in the case
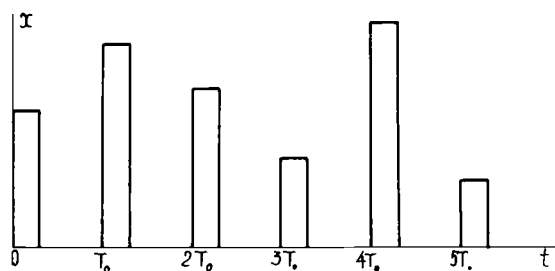
(a)
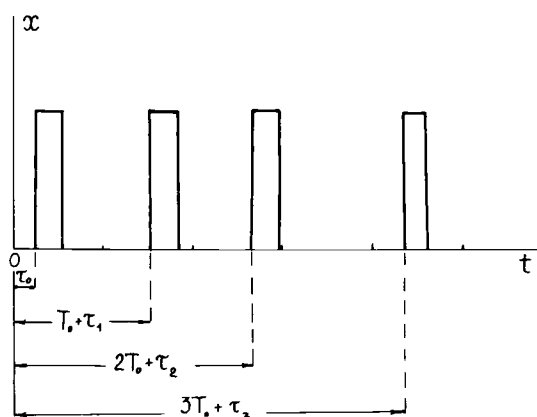
(b)

Fig. 15. (a) Pulse-amplitude modulated signal in the case of rectangular pulses. (b) Pulse-phase modulated signal (rectangular pulses).

of a random process $X(t)$ in continuous time t, one has to speak of the probability distribution in the space of all the possible trajectories of our process.) In an infinite-dimensional space there is no natural "volume element" and, hence, the concept of "probability density" is meaningless; the notion of "distribution function" also cannot be defined in a simple way here. That is why a rigorous mathematical theory of random functions cannot be developed without the use of comparatively complex and non-elementary concepts and methods relating to advanced modern mathematics, unknown to the majority of applied scientists.

The first attempts at a mathematical investigation of

probabilistic schemes leading to the notion of a random function appeared as early as the beginning of this century. Moreover, the first physical theories dealing with random processes sprang up at about the same time. However, the formulation of the corresponding general theory was the work of the 1920s, '30s, and '40s.   (The development of this theory did not stop there, of course, but has been going on at a high pace till today.) In the 1930s foundations were also laid for the theory of   the most important special classes of random functions, namely *processes with independent increments, Markov random processes,* and *stationary random processes.*   The theory of these classes of random functions continued to develop vigorously in subsequent years along with the theory of several more general (or, on the contrary, more particular) new classes of random functions (*branching processes, martingales,* general *Gaussian random functions, processes with stationary increments* of some finite order, *homogeneous random fields,* etc.).   In the late 1940s and early 1950s the first large survey papers and special monographs on the theory of random functions appeared; since then the number of books on the subject has been steadily increasing, and is now running into the many hundreds.[17]

This book covers only a small, but in practice very important, part of the theory of random functions. Its main part is devoted to an elementary introduction to the theory of stationary random sequences and processes -- special classes of random functions, which are very often encountered in various applied problems. Of the whole theory of stationary random functions $X(t)$ this book deals almost exclusively with the *correlation* (or the *second-order*) *theory,* i.e. with problems whose study involves only first- and second-order moments of random variables $X(t)$. This selection can be justified by the fact that the correlation theory actually constitutes the central (the most advanced and widely used) part of the entire modern theory of stationary random functions.   Moreover, the adopted limitation also enables us to dispense with the non-elementary niceties of the general theory of infinite-dimensional probability distributions and to make the book accessible to students of applied sciences or workers in applied fields having no special mathematical training. Note, in conclusion, that in the last chapter of this book a number of generalizations of the concept of a stationary random function are studied. These generalizations are rather important since they are often useful in dealing with applied problems.

Chapter 1
BASIC PROPERTIES OF STATIONARY
RANDOM FUNCTIONS

1. Definition of a Random Function

Let T be an arbitrary set of elements t, s, ... By a *random function* of an argument t in T (synonymously, a random function *on T*), we mean a function $X(t)$ whose values are random variables. Thus, a random function on T is a family of random variables $X(t),X(s)$, ... , corresponding to all elements t, s, ... of the set T. If the set T contains only a finite number of elements, then the random function $X(t)$ is a finite family of random variables. In this case we can regard $X(t)$ as a single *multidimensional random variable*. Such random variables are studied in classical probability theory, and they are specified by using multidimensional distribution functions $F(x_1,x_2, ..., x_n)$ or multidimensional probability densities $p(x_1,x_2, ..., x_n)$. If, however, the set T is infinite, then $X(t)$ is an infinite family of random variables. Such families are not studied in classical probability theory, and hence the specification of $X(t)$ requires a special definition.

We have already mentioned at the end of the Introduction that both the probability density concept and the distribution function concept cannot be simply extended to the infinite-dimensional case. Therefore, to specify the probability distribution of an infinite family of random variables $X(t)$ we must attack the problem from a different angle. Namely, *we shall regard the random function $X(t)$ as*

being specified if for each element t in the set T there is defined the distribution function

(1.1) $F_t(x) = P\{X(t) < x\}$

of the random variable X(t), if for each pair of elements t_1, t_2 in the set T we are given the distribution function

(1.2) $F_{t_1,t_2}(x_1,x_2) = P\{X(t_1) < x_1, X(t_2) < x_2\}$

of the two-dimensional random variable $X(t_1,t_2) = \{X(t_1),X(t_2)\}$ and so on, and, in general, if for any finite number n of elements $t_1,t_2, ..., t_n$ in the set T we are given the distribution function

(1.3) $F_{t_1,t_2,...,t_n}(x_1,x_2,...,x_n) = P\{X(t_1) < x_1, X(t_2) < x_2,...,X(t_n) < x_n\}$

of the n-dimensional random variable $X(t_1, ..., t_n) = \{X(t_1), ..., X(t_n)\}$. In many (although not all) cases of practical interest the distribution functions (1.1) $-$ (1.3) can be replaced by the *probability densities*

(1.4) $p_t(x), \quad p_{t_1,t_2}(x_1,x_2), ..., p_{t_1,t_2,...,t_n}(x_1,x_2,...,x_n), ...$.

Thus, the random function X(t) on the infinite set T is specified by a whole hierarchy of multidimensional probability distributions of different orders $n = 1, 2, 3, ...$. It is the existence of all these probability distributions that points to the *statistical stability* of the outcomes of the experiment generating the function X(t), which was mentioned in the Introduction when we first spoke of random functions.

The functions (1.1) $-$ (1.3), and (1.4) cannot be quite arbitrary. For instance, all the functions (1.4) must be non-negative and the integral of each of them with respect to all the variables must be equal to unity. Similarly, the functions (1.1) -- (1.3) must satisfy conditions characterizing multidimensional distribution functions. Moreover, these functions must also satisfy certain symmetry and compatibility conditions. Specifically, according to the definition of the densities $p_{t_1,t_2,...,t_n}(x_1, x_2, ..., x_n)$ they must satisfy the relationship

(1.5) $\qquad p_{t_{i_1}, t_{i_2}, \ldots, t_{i_n}}(x_{i_1}, x_{i_2}, \ldots, x_{i_n}) = p_{t_1, t_2, \ldots, t_n}(x_1, x_2, \ldots, x_n),$

where $i_1, i_2, \ldots, i_n$ is any permutation of the indices $1, 2, \ldots, n$. Furthermore, if $n > m$, then for any n points $t_1, t_2, \ldots, t_m, t_{m+1}, \ldots, t_n$ in T the following relation must hold:

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p_{t_1, t_2, \ldots, t_m, t_{m+1}, \ldots, t_n}(x_1, x_2, \ldots, x_m, x_{m+1}, \ldots, x_n) dx_{m+1} \ldots dx_n$$

(1.6) $\qquad\qquad = p_{t_1, t_2, \ldots, t_m}(x_1, x_2, \ldots, x_m).$

The *symmetry condition* (1.5) and the *compatibility condition* (1.6), as applied to the distribution functions $F_{t_1, t_2, \ldots, t_n}(x_1, x_2, \ldots, x_n)$, take the following forms:

(1.5a) $\qquad F_{t_{i_1}, t_{i_2}, \ldots, t_{i_n}}(x_{i_1}, x_{i_2}, \ldots, x_{i_n}) = F_{t_1, t_2, \ldots, t_n}(x_1, x_2, \ldots, x_n),$

$$F_{t_1, t_2, \ldots, t_m, t_{m+1}, \ldots, t_n}(x_1, x_2, \ldots, x_m, \infty, \ldots, \infty)$$

(1.6a) $\qquad\qquad = F_{t_1, t_2, \ldots, t_m}(x_1, x_2, \ldots, x_m).$

These symmetry and compatibility conditions are not only necessary, but also sufficient in the sense that any system of multidimensional distribution functions $(1.1) - (1.3)$ for which the conditions (1.5a) and (1.6a) are fulfilled can be regarded as defining some random function $X(t)$ on the set T. In particular, any system of probability densities (1.4) satisfying (1.5) and (1.6) defines such a function.[1]

The specification of a random function $X(t)$ by a hierarchy of finite-dimensional probability distributions $(1.1) - (1.3)$ or (1.4) is the most general and, therefore, theoretically the basic method of its specification. However, this method is extremely cumbersome and inconvenient for practical applications. Therefore, in real practice, other more particular methods of specifying random functions $X(t)$ are often used. These methods also permit one, in principle, to determine all the finite-dimensional probability distributions (1.3) (or densities (1.4)), but they have a simpler form. Thus, in a number of problems it is convenient to define a random function $X(t)$ by an analytic formula, containing parameters $Y_1, \ldots, Y_k$, which are random variables

(1.7) $X(t) = f(t, Y_1, ..., Y_k)$.

Some simple examples of such random functions are represented in Figs. 8, 9, and 14; other examples are provided by polynomials (ordinary or trigonometric) with random coefficients.   In principle, functions of the form (1.7) can be studied in the framework of classical probability theory, because all the statistical properties of $X(t)$ (i.e.   all the distributions (1.3)) are completely specified in this case by the probability distribution of the random vector $\mathbf{Y} = (Y_1, ..., Y_k)$. This does not mean, of course, that in practice the investigation of any random function of the form (1.7) should always be replaced by the study of the properties of the random vector $\mathbf{Y}$; but from the standpoint of the theory, such a replacement is always possible.

As regards the general specification of $X(t)$ by an infinite hierarchy of probability distributions of various orders, it appears to be practical only for some special classes of random functions for which the distributions (1.3) possess particular   simple   properties.   For   instance,   such   a specification is often convenient in cases where all the distributions (1.3) are completely defined by the values of the distributions of only a few lower orders.   The most important class of random functions $X(t)$ of a discrete argument $t = 0, \pm1, \pm2, ...$ or $t = 0, 1, 2, ...$, completely defined by their one-dimensional probability distributions (1.1), are sequences of independent random variables, for which

(1.8) $F_{t_1,t_2,...,t_n}(x_1,x_2,...,x_n) = F_{t_1}(x_1)F_{t_2}(x_2)...F_{t_n}(x_n)$

(and $p_{t_1,t_2,...,t_n}(x_1,x_2,...,x_n) = p_{t_1}(x_1)p_{t_2}(x_2)...p_{t_n}(x_n)$).   If the

random variables $X(t)$ are also identically distributed, then all the statistical properties of the function $X(t)$ will be defined by   a   single   one-dimensional   probability   distribution. Sequences of independent random variables are studied comprehensively in classical probability theory (primarily in connection with the investigation of various limit theorems). However, from the standpoint of random function theory such   sequences   seem   to   be   too   special,   simple,   and uninteresting. Nevertheless, many important types of random functions can be specified by relationships reducing the random character of the quantities $X(t)$ to their dependence

on some sequence of mutually independent random variables. For instance, it would be appropriate to consider random functions of the type

$$(1.9) \qquad X(t) = \sum_{n=0}^{\infty} a_n E_{t-n}, \quad t = 0, \pm 1, \dots,$$

or

$$(1.10) \qquad X(t) = \sum_{n=0}^{\infty} E_n \varphi_n(t), \quad a \leqslant t \leqslant b,$$

where a_n and $\varphi_n(t)$ are numerical coefficients and functions, while E_n, $n = 0, \pm 1, \pm 2, \dots$, or $n = 0, 1, 2, \dots$ are identically distributed independent random variables. (We shall encounter functions of these two types later in this book.) Formulae (1.9) and (1.10) offer examples of the specification of a random function $X(t)$ by an equation which defines $X(t)$ as a given function of infinitely many variables E_n. In general, specification of $X(t)$ as an ordinary (i.e. non-random) function of t and of some random function with relatively simple statistical properties provides a very important practical method of effective description of random functions. This method, for example, has already been used in the Introduction to determine the Poisson process, a random telegraph signal, and other functions $X(t)$ depending on the Poisson point process $\{t_n\}$ — a simple random function of an integer-valued argument n.

Passing over to random functions $X(t)$, which have the property that their finite-dimensional distributions (1.3) are determined by the values of one-dimensional distributions (1.1) and two-dimensional distributions (1.2), we can indicate several important classes of such functions. It is easy to see that the indicated property is inherent in all the *random processes with independent increments* — random functions of the real argument t, for which the differences $X(t_2) - X(t_1)$ and $X(t_4) - X(t_3)$ are independent random variables, provided that $t_1 < t_2 \leqslant t_3 < t_4$. Furthermore, the same property is possessed by all the *Markov random processes*, for which the conditional distribution of the quantity $X(t_k)$ at known values of $X(t_{k-1}), X(t_{k-2}), \dots, X(t_{k-m})$, where $t_{k-m} < \dots < t_{k-2} < t_{k-1} < t_k$ depends only on the "last value" $X(t_{k-1})$. Indeed, if a Markov process $X(t)$ has probability densities (1.4), then for $t_1 < t_2 < t_3 < \dots < t_n$ the following relation will evidently hold:

$$p_{t_1,t_2,t_3,\ldots,t_n}(x_1,x_2,x_3,\ldots,x_n)$$

(1.11)
$$= p_{t_1}(x_1)p_{t_2,t_1}(x_2|x_1)p_{t_3,t_2}(x_3|x_2)\ldots p_{t_n,t_{n-1}}(x_n|x_{n-1}),$$

where $p_{t,s}(x|y) = p_{t,s}(x,y)/p_s(y)$ is the conditional probability density of $X(t)$, provided that $X(s) = y$. In the same way, for general Markov processes, the distribution function $F_{t_1,t_2,\ldots,t_n}$

$(x_1,x_2,\ldots,x_n)$ can, in principle, be determined if the functions $F_t(x)$ and $F_{t,s}(x,y)$ are known. (However, in practice another method of specifying a Markov process $X(t)$ is generally used.) Finally, the values of one-dimensional and two-dimensional probability distributions completely specify the random function $X(t)$ on the arbitrary set T in the important case of a *Gaussian random function* X(t), i.e. such $X(t)$ for which all its finite-dimensional probability distributions (1.3) are normal (Gaussian) distributions. Indeed, we have already remarked in the Introduction that normal probability distributions are completely determined by their first and second moments, and the knowledge of all the one-dimensional distributions of $X(t)$ and the two-dimensional distributions of the pairs $\{X(t),X(s)\}$ clearly allow one to evaluate the first and second moments of the function $X(t)$.

There also exist additional methods of specifying a random function. However, we shall not discuss them here.[2]

2. Moments of a Random Function. Correlation Theory

To give a complete description of a random function $X(t)$ it is necessary, as we know, to specify all its finite-dimensional distribution functions (1.3) or probability densities (1.4). However, in most applied problems it is excessively complicated and costly to measure the multidimensional (and even one-dimensional) probability distributions experimentally, and they are generally too cumbersome to be used in practice. Moreover, the theoretical determination of probability distributions (1.3) is possible only in a few exceptional cases (the most important of which relate to Gaussian functions $X(t)$). Therefore, in developing the theory of random functions it is natural to restrict oneself (at least, in the beginning) to a study of the properties determined by just the simplest numerical characteristics of multidimensional distributions (1.3). It is well

known that the simplest and most widely used numerical characteristics of probability distributions are their *moments*. In the case of the distributions (1.3), the moments take the form

$$\mu^{(j_1,j_2,\ldots,j_n)}(t_1,t_2, \ldots, t_n) = \langle X(t_1)^{j_1}X(t_2)^{j_2}\ldots X(t_n)^{j_n}\rangle$$

(1.12)
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} x_1^{j_1}x_2^{j_2}\ldots x_n^{j_n}\, d^nF_{t_1,t_2\ldots,t_n}(x_1,x_2, \ldots, x_n),$$

where the integral on the right-hand side is an *n*-fold Stieltjes integral mentioned in the Introduction. In the particular case of the existence of probability densities (1.4), to which we shall usually restrict ourselves for simplicity, the right-hand side of (1.12) can be written in the form of an ordinary multiple integral:

(1.12a)
$$\mu^{(j_1,j_2,\ldots,j_n)}(t_1,t_2, \ldots, t_n) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \ldots$$
$$\int_{-\infty}^{\infty}x_1^{j_1}x_2^{j_2}\ldots x_n^{j_n}\, p_{t_1,t_2,\ldots,t_n}(x_1,x_2, \ldots, x_n)dx_1dx_2 \ldots dx_n.$$

The simplest of the moments (1.12) is, of course, the first moment

(1.13) $$\mu^{(1)}(t) = m(t) = \langle X(t)\rangle = \int_{-\infty}^{\infty}xp_t(x)dx$$

i.e. the *mean value* of the random function $X(t)$. The function $m(t)$ may have any form; when it is not constant, its dependence on t describes the *systematic variation* (or *trend*) of the random function $X(t)$. The mean value is a very important characteristic of $X(t)$, but it describes only the coarsest properties of the function $X(t)$ and does not even reflect the random character of $X(t)$. (Hence the term "systematic variation", which is naturally opposed to "random variation".) Much more subtle properties of $X(t)$ are described by its *second moment*

(1.14) $$\mu^{(1,1)}(t,s) = B(t,s) = \langle X(t)X(s)\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyp_{t,s}(x,y)dxdy$$

or by the *central second moment* − the *covariance* between the variables $X(t)$ and $X(s)$:

(1.15) $$b(t, s) = \langle (X(t) - \langle X(t)\rangle)(X(s) - \langle X(s)\rangle)\rangle = B(t, s) - m(t)m(s).$$

The values of the function (1.15) characterize the random deviations of the values $X(t)$ from the respective mean values $m(t)$. In particular, $b(t,t) = \sigma^2[X(t)] = \sigma^2(t)$ is the variance of $X(t)$, $[b(t,t)]^{1/2} = \sigma(t)$ is the corresponding standard deviation, and

$$(1.16) \qquad \frac{b(t,s)}{\sigma(t)\sigma(s)} = R(t,s)$$

is the correlation coefficient between $X(t)$ and $X(s)$. The function $B(t,s)$ is often called the *correlation function* of the random function $X(t)$. (The same name is also applied to functions $b(t,s)$ or $R(t,s)^*$, so, where this will cause no confusion, we shall call both the functions $B(t,s)$ and $b(t,s)$ correlation functions. However, whenever confusion may arise, the function $b(t,s)$ will be called the *centered correlation function*, and the correlation coefficient $R(t,s)$ will be called the *normalized correlation function* of the random function $X(t)$.)

It is clear that all three functions $B(t,s)$, $b(t,s)$, and $R(t,s)$ are symmetric by definition:

$$(1.17) \qquad B(t,s) = B(s,t), \quad b(t,s) = b(s,t), \quad R(t,s) = R(s,t).$$

Moreover, all of them are *positive definite kernels* on T, i.e. they possess the property that

$$(1.18) \qquad \sum_{j,k=1}^{n} B(t_j,t_k)c_j c_k \geqslant 0,$$

$$(1.18a) \qquad \sum_{j,k=1}^{n} b(t_j,t_k)c_j c_k \geqslant 0,$$

$$(1.18b) \qquad \sum_{j,k=1}^{n} R(t_j,t_k)\, c_j c_k \geqslant 0$$

for any positive integer n, points t_1, ..., t_n of the set T and real numbers c_1, ..., c_n. Indeed, the left-hand side of (1.18) is equal to $\langle (c_1 X(t_1) + ... + c_n X(t_n))^2 \rangle$ and hence cannot be negative. As to the functions $b(t,s)$ and $R(t,s)$, they are particular cases of the function $B(t,s)$; they correspond to the

---

*In general, the terminology is not completely established here. For example, in the literature the function B(t,s) is sometimes called the <u>autocorrelation function</u>, or the <u>covariance</u> (function), or the <u>autocovariance</u> (function).

random functions $\overset{\circ}{X}(t) = X(t) - \langle X(t) \rangle$ (*fluctuation* of the
function $X(t)$) and $X_0(t) = \overset{\circ}{X}(t)/\sigma(t)$ (*normed fluctuation* of $X(t)$).
When $n = 1$, condition (1.18b) is evidently fulfilled (since $R(t,t) = 1$), and the conditions (1.18) and (1.18a) reduce to the obvious
requirements that

(1.19) $B(t,t) \geqslant 0,$ $b(t,t) \geqslant 0.$

When $n = 2$, from (1.18) $-$ (1.18b) it follows that

(1.19a) $|B(t,s)| \leqslant [B(t,t)B(s,s)]^{1/2},$ $|b(t,s)| \leqslant \sigma(t)\sigma(s),$ $|R(t,s)| \leqslant 1$

(cf. the inequalities (0.21) and (0.23) in the Introduction).

According to (1.18) and (1.18a), at a given $\langle X(t) \rangle = m(t)$
both the functions $B(t,s)$ and $B(t,s) - m(t)m(s)$ must be positive
definite kernels. It is also easy to show that *any positive
definite kernel $B(t,s)$ on T is the correlation function* (and
simultaneously the centered correlation function) *of some
random function $X(t)$ with a zero mean value.* Moreover, *if
the difference $B(t,s) - m(t)m(s)$ is also a positive definite kernel,
then the function $X(t)$ can be chosen so that $\langle X(t) \rangle = m(t)$.* To
prove these statements it is sufficient to consider families of
normal probability densities $p_{t_1, t_2, ..., t_n}(x_1, x_2, ..., x_n)$ of the

form (0.37) with the second moment matrix $(t_1, ..., t_n) = \|B(t_j, t_k)\|, j, k = 1, ... n$, and zero mean values or, respectively,
the mean value vector $\mathbf{m}(t_1, ..., t_n) = (m(t_1), ..., m(t_n))$, and then
prove the validity, for these densities, of the symmetry and
compatibility conditions (1.15) and (1.16).[3]

The first and second moments of the random function $X(t)$
obviously do not specify this function uniquely: there are
many quite dissimilar functions $X(t)$ having the same mean
value $m(t)$ and correlation function $B(t,s)$. Consider, for
instance, the Poisson process $X(t) = N(t)$, represented in Fig.
10(a). It is easy to show that if λ is the parameter of the
Poisson point process $\{t_k\}$ (i.e. the mean rate of points t_k per
unit time), then $m(t) = \lambda t$, and $B(t,s) = \lambda s + \lambda^2 ts$ for $t \geqslant s$
while $B(t,s) = \lambda t + \lambda^2 ts$ for $t < s$.[4] Hence in the case of the
centered Poisson process $\overset{\circ}{X}(t) = X(t) - \lambda t$ (schematically plotted
in Fig. 16)

(1.20) $m(t) = \langle \overset{\circ}{X}(t) \rangle = 0,$ $B(t,s) = \langle \overset{\circ}{X}(t)\overset{\circ}{X}(s) \rangle = \lambda \min(t,s),$

where min(t,s) is the smaller of the two numbers t and s. Consider now the *Brownian motion process* $W(t)$, which is sketched in Fig. 4(b). In Sec. 12 we shall show that it is a Gaussian random process with a zero mean value and a correlation function of the form $B(t,s) = \sigma^2 \min(t,s)$. Therefore, if $\sigma^2 = \lambda$ then the centered Poisson process and the Brownian motion process (which is also often called the *Wiener process*) have the same first and second moments, although their realizations are of an entirely different nature.

As still another example, we consider a *random telegraph signal* (see Fig. 11) corresponding to the Poisson point sequence $\{t_k\}$, with parameter λ, and a *point process with adjoined random variables* (Fig. 12) corresponding to the Poisson point sequence $\{t_k'\}$ with parameter λ' and the adjoined variables X_k with zero mean and variance σ^2. It is clear that the mean value of both these processes is zero. Moreover, the correlation function of a random telegraph signal has the form

(1.21) $$B(t, s) = \langle X(t)X(s) \rangle = a^2 e^{-2\lambda |t-s|},$$

while

(1.22) $$B(t, s) = \langle X(t)X(s) \rangle = \sigma^2 e^{-\lambda' |t-s|}$$

for a point process with adjoined random variables.[4] Hence, by setting $\sigma^2 = a^2$ and $\lambda' = 2\lambda$ we again arrive at a pair of very different random processes with identical first and second moments. Below, in Secs. 10 and 12, we shall also encounter the so-called *Ornstein-Uhlenbeck process* — a Gaussian random process with zero mean and correlation
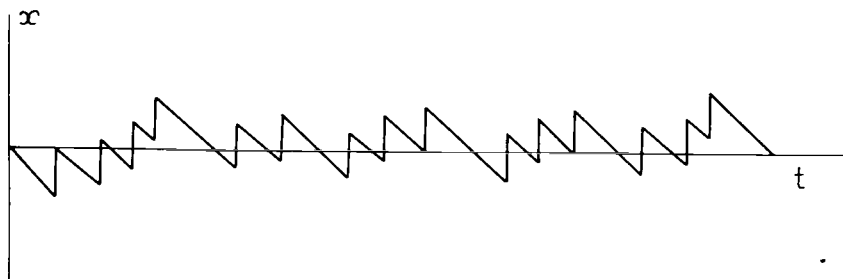
Fig. 16. The centered Poisson process $N(t) - \langle N(t) \rangle$.

function $B(t,s) = \sigma^2\exp(-\lambda'|t - s|)$. This process is quite unlike the processes depicted in Figs. 11 and 12, but it also has the same first and second moments. Moreover, we shall show in Sec. 6, Example 5, that it is possible to construct a random process with the same moments of the first two orders, all of whose realizations are exact sinusoids. Finally, Example 4 of the same section and Note 9 to Chap. 2 supply constructions of two distinct random pulse processes for which deviations from the mean value have, once again, first and second moments of precisely the same form.

There are, however, many properties of a random function $X(t)$, which are completely determined by its mean value $m(t)$ and by its correlation function $B(t,s)$. Such properties are particularly important in applications, because they depend only on the two simplest statistical characteristics of the function $X(t)$, which can often be measured fairly accurately with little effort.   Therefore, it was even thought expedient to give a special name to that part of the general theory of random functions which studies only those properties of such functions which are determined by their first and second moments; it is called *correlation theory* (of random functions).[5] It is the correlation theory that will be principally considered in this book. It should also be emphasized that correlation theory always assumes that all the random variables under consideration (i.e. all the $X(t)$ at any value of t) have a finite mean value and a finite variance; this assumption will always be considered valid in what follows.   Under this assumption the correlation function $B(t,s)$ will be finite for any t and s by virtue of inequality (1.19a).

The practical value of correlation theory is substantially enhanced by the fact that random functions encountered in many situations very often turn out to be either precisely Gaussian (i.e. all of their finite-dimensional probability distributions (1.3) are multidimensional Gaussian normal distributions) or at least approximately Gaussian, with an accuracy sufficient for many practical purposes. Since multidimensional Gaussian distributions are completely specified by their moments of the first two orders, it is clear that for Gaussian random functions $X(t)$ the mean value $m(t)$ and the correlation function $B(t,s)$ uniquely specify all the finite-dimensional distributions (1.3), and so they completely specify the function $X(t)$.   It follows that the answer to any

meaningful question concerning Gaussian random function $X(t)$ can, in principle, be obtained from the values of $m(t)$ and $B(t,s)$; in other words, correlation theory does not differ here from complete theory.   It should also be noted that for any random function $X(t)$ with finite first and second moments it is always possible to construct a Gaussian function $X^{(0)}(t)$ with the same mean value and the same correlation function as $X(t)$ (cf.   Note 3 to this chapter in Vol. II).   Therefore, within the framework of correlation theory one can assume, without loss of generality, that the function $X(t)$ under investigation is Gaussian.

It is significant that in a number of physical and engineering problems the finite-dimensional probability distributions of the function $X(t)$ can legitimately be regarded as Gaussian, either by virtue of the so-called *central limit theorem* of probability theory (this will be the case when $X(t)$ arises from the superposition of a large number of comparatively weak random effects — for example, in studying phenomena associated with the electrical "shot noise"), or because of the normality of the Maxwellian distribution of thermal molecular velocities (here, an example is the "thermal noise" theory).   We note in passing that in applied investigations the assumption of normal distributions (1.3) is sometimes taken for granted without adequate justification, even in cases where this is in fact unnecessary. Indeed, if only the results from correlation theory are used, we have no need to mention the relevant probability distributions.

3. Stationarity

From now until Chapter 4, the variable t will always take only real values and will usually be interpreted as time. As regards the set T, it will sometimes contain all the real numbers $-\infty < t < \infty$ (so that, in accordance with the terminology indicated in the Introduction, $X(t)$ will be a *random process*), and sometimes will consist only of integers ... -2, -1, 0, 1, 2, ... (in this case $X(t)$ will be a *random sequence*). As far as the term *random function* is concerned, we shall use it only when we find it convenient to consider random processes and random sequences simultaneously.

The random function $X(t)$ will be called stationary if all its finite-dimensional distribution functions (1.3) (or probability densities (1.4)) remain the same when shifted along the time axis, i.e. when for any n, t_1, t_2, ..., t_n and τ

(1.23) $F_{t_1+\tau, t_2+\tau, ..., t_n+\tau}(x_1, x_2, ..., x_n) = F_{t_1, t_2, ..., t_n}(x_1, x_2, ..., x_n)$

or, accordingly,

(1.24) $p_{t_1+\tau, t_2+\tau, ..., t_n+\tau}(x_1, x_2, ..., x_n) = p_{t_1, t_2, ..., t_n}(x_1, x_2, ..., x_n)$.

It follows, in particular, that for a stationary random function all the one-dimensional probability distributions have to be identical (i.e. $F_t(x) = F(x)$ and $p_t(x) = p(x)$ cannot depend on t), all the two-dimensional distribution functions and probability densities $F_{t_1, t_2}(x_1, x_2)$ and $p_{t_1, t_2}(x_1, x_2)$ can depend only on the difference $t_2 - t_1$, and so on.   As to the functions $F_{t_1 t_2, ..., t_n}(x_1, x_2, ..., x_n)$ and $p_{t_1, t_2, ..., t_n}(x_1, x_2, ..., x_n)$, they can depend only on the $n - 1$ differences $t_2 - t_1$, ..., $t_n - t_1$.   Assuming $\tau = - t_1$ in (1.23) and (1.24), we find that these relations can also be written as

(1.23a) $F_{t_1, t_2, ..., t_n}(x_1, x_2, ..., x_n) = F_{0, t_2-t_1, ..., t_n-t_1}(x_1, x_2, ..., x_n)$

and

(1.24a) $p_{t_1, t_2, ..., t_n}(x_1, x_2, ..., x_n) = p_{0, t_2-t_1, ..., t_n-t_1}(x_1, x_2, ..., x_n)$.

Suppose now that $y = g(x)$ is an arbitrary function of a single variable.   It is easy to see that the distribution function $F(x_1, ..., x_n) = P\{X_1 < x, ..., X_n < x_n\}$ of the random variables X_1, X_2, ..., X_n, uniquely determines the value of the distribution function $F^{(g)}(y_1, ..., y_n) = P\{Y_1 < y_1, ..., Y_n < y_n\}$ of the variables $Y_1 = g(X_1)$, $Y_2 = g(X_2)$, ..., $Y_n = g(X_n)$. Consequently, if the distribution function of the variables $X_1 = X(t_1)$, $X_2 = X(t_2)$, ..., $X_n = X(t_n)$ depends only on the differences $t_2 - t_1$, ..., $t_n - t_1$, then the distribution function of the random variables $Y = g(X(t_1))$, $Y_2 = g(X(t_2))$, ..., $Y_n = g(X(t_n))$ will also depend only on these differences. Therefore, *if $X(t)$ is a stationary random function, any function $Y(t) = g(X(t))$ of $X(t)$ will also be a stationary random function.* This

important property distinguishes the class of stationary random functions, for instance, from the class of Gaussian random functions $X(t)$ or from the class of Markov random processes $X(t)$, for which, the function $g(X(t))$ in general, no longer belongs to the class of random functions under consideration.

The physical meaning of the stationarity condition is quite clear. It means that a phenomenon whose numerical characteristic is the random function $X(t)$ is stationary in the sense that none of the observed macroscopic factors influencing this phenomenon change in time. To put it differently, $X(t)$ describes the time variation of some characteristic of a steady-state phenomenon, for which no choice of the time has any advantage over any other choice. For instance, the current and voltage fluctuations in an electrical circuit ("noise") represent stationary processes, provided that the circuit has a "stationary regime", i.e. that none of its parameters (including the temperature of the conductors and all the external electromotive forces) change in time. Similarly, the velocity component of a turbulent flow will be a random stationary time function if the flow is steady, i.e. all its averaged characteristics remain constant; the diameter of a thread will be a homogeneous function of the cross section coordinate* if the conditions under which the thread is manufactured do not change in time, and so on.

In general, the random functions $X(t)$ encountered in practical applications can very often be regarded as stationary to a high degree of accuracy. Such a possibility is usually manifested vividly enough on visual inspection of the observed values of functions $X(t)$. Thus, when looking at the curves in Figs. 5 and 6 we notice that the general outlines of the recorded irregular fluctuations remain the same throughout the observation period: the curves have no systematic trend, the fluctuation amplitude does not increase or decrease appreciably along the curve, and the typical

---

*If the variable t, on which the random function depends, is other than the time, then the expression "stationary random function" is often replaced by "homogeneous random function". Moreover, the expression "time-homogeneous (or homogeneous in time) function" has precisely the same meaning as "stationary function".

periods of the observed fluctuations also show no perceptible changes. All this gives us sufficient ground to believe that the curves under review are realizations of some stationary random processes. In this connection it should be emphasized that, in practice, in order to regard a random function $X(t)$ as stationary it is not necessary that all the external characteristics of the phenomenon generating it be strictly invariant in time. For practical purposes, it is sufficient that these internal characteristics do not change appreciably during the time of observation of $X(t)$. This last remark has a direct bearing, for instance, on the curve of Fig. 5(b), characterizing the atmospheric turbulence, which is undoubtedly affected by the daily and annual variations of the average air temperature. However, in observations for comparatively short periods (e.g., one hour or less) the corresponding random functions can usually be confidently regarded as stationary.

Passing on to the other examples of random functions, which were discussed in the Introduction, we note that none of the very different curves given in Figs. 4(b), 7, 8, and 10 seem to arise from stationary random functions. However, in the case of the curve of Fig. 7(a), for instance, one can still assume that during separate, comparatively short time intervals, it can be regarded as approximately stationary. Also, one cannot be sure that after a considerable lengthening of the observation period it will not again resume the shape of a realization of a stationary process containing some long-period oscillations.[6] An amplitude-modulated harmonic oscillation with a fixed phase is not a stationary random process, because here the zeros, minima and maxima of $X(t)$, correspond to some strictly definite time instants. For a similar reason, pulse-amplitude modulation processes with a fixed phase will also not be stationary (see Fig. 15(a)). A phase-modulated harmonic oscillation $X(t) = A\cos(\omega_0 t + \Phi)$, where Φ is a random variable taking values from the interval $0 \leqslant \Phi < 2\pi$, is also a nonstationary process if the probability distribution of Φ is not uniform (for instance, if the probability density of Φ has a sharp peak). But if Φ is distributed uniformly on $[0, 2\pi]$ with a constant density $p = 1/2\pi$, then the oscillation $X(t)$ will be a stationary random process for any probability distribution of the amplitude A (under the assumption that A is independent of Φ). Indeed, since a time shift is equivalent to some phase shift, no time

shift can affect the probability distributions of $X(t)$. For the same reason, the "square wave with a random phase" of Fig. 14 will also be a stationary process if the phase Φ is distributed uniformly on the interval $[0, T_0]$. The sum of harmonic oscillations

$$(1.25) \qquad X(t) = \sum_{k=1}^{n} A_k \cos(\omega_k t + \Phi_k),$$

is also a stationary process if all the phases Φ_1, ..., Φ_n are independent random variables uniformly distributed on the interval $[0, 2\pi]$, and the amplitudes A_1, ..., A_n are independent of each other and of the phases Φ_1, ..., Φ_n. (When A_1, ..., A_n are fixed constants, the process (1.25) is called a *random phase process*.) Similarly, pulse-amplitude and pulse-phase modulation processes will also be stationary, provided we assume that the "time datum point" in Figs. 15(a) and (b) was not fixed beforehand, but has a uniform probability distribution on an interval of length T_0.[7]

A random telegraph signal and a point process with adjoined random variables are both stationary processes, provided that the corresponding Poisson point sequence $\{t_k\}$ has a constant intensity λ, and the time datum point is chosen "at random", i.e. is in no way related to the points t_k (this last assumption was in fact adopted in calculations of the correlation functions of these processes described in Sec. 2). However, if some point t_0 of the Poisson sequence $\{t_k\}$ is taken to be the time datum point, then the process $X(t)$ will no longer be stationary (since here the moment $t = 0$ is distinguished by the condition that it coincides with the discontinuity point of $X(t)$). A modified random telegraph signal will not be stationary, either, if it is constructed as follows: segments $\boldsymbol{l}_1 = t_1$, $\boldsymbol{l}_2 = t_2 - t_1$, $\boldsymbol{l}_3 = t_3 - t_2$, $\boldsymbol{l}_4 = t_4 - t_3$, which are realizations of independent random variables with a given probability distribution, are placed successively along the positive semiaxis, beginning from the point $t = 0$. Then it is assumed that $X(t) = a$ for $0 \leqslant t = t_1$, $X(t) = -a$ for $t_1 \leqslant t < t_2$, $X(t) = a$ for $t_2 \leqslant t < t_3$, $X(t) = -a$ for $t_3 \leqslant t < t_4$, and so on. Indeed, here $X(t)$ is defined only for $t \geqslant 0$, and at small (compared to the average length λ^{-1} of the segments $\boldsymbol{l}_i$) values of t the equality $X(t) = a$ holds true with a probability close to one, whereas at $\lambda t \gg 1$ the values of $X(t) = a$ and $X(t) = -a$ are approximately equiprobable. It is easy to see, however, that at $t \gg \lambda^{-1}$ the special choice of the time datum

point $t = 0$ adopted in our construction will actually cease to affect the values of $X(t)$. Therefore any segment of realization of $X(t)$ between the points t' and t'', where $t' \gg \lambda^{-1}$, and $t'' > t'$, can be regarded as a segment of realization of the stationary process $X^{(0)}(t)$ determined for $-\infty < t < \infty$.

The behavior of the modified random telegraph signal $X(t)$ is very typical of a wide class of functions $X(t)$ arising in the study of particular manifestations of human activity. Namely, the initial period of "small values of t" is very often characterized by the presence of some nonstationary "transient process", which essentially depends on the choice of some "initial conditions". Further on, however, this "transient process" gradually dies out, and the entire system passes into the "equilibrium state" corresponding to the steady-state regime of its functioning. Thus any physical device or industrial apparatus can be considered to be in the steady-state regime after a complete dying out of the "transient process" which was triggered by turning it on. Naturally, there is no clearly defined time boundary separating the "transient process" from the "steady-state regime"; the choice of such a boundary depends exclusively on the required degree of accuracy. From this point of view the concept of a stationary random function, which assumes that our function is defined for all values of t and has precisely identical statistical properties for all t, is always only a mathematical idealization, i.e. it describes some simplified model, which only approximately reflects the properties of certain real processes. But this model has a very wide field of application, within which it is extremely useful for solving many important problems. This last circumstance ultimately determines the high practical value of the mathematical theory of stationary random functions.

Since the one-dimensional distribution function $F_t(x) = F(x)$ of a stationary function $X(t)$ is independent of t, it is clear that the mean value $m(t) = \langle X(t) \rangle$ must be constant:

(1.26) $\langle X(t) \rangle = m.$

Similarly, in the stationary case a two-dimensional distribution function $\{X(t), X(s)\}$ depends only on the difference $t - s$; therefore, the correlation function $B(t,s) = \langle X(t)X(s) \rangle$ has the form

(1.27) $\langle X(t)X(s) \rangle = B(t - s).$

It is also clear that if $X(t)$ is stationary then $b(t,s) = B(t-s) - m^2 = b(t-s)$ differs from $B(t-s)$ only by a constant summand, and

$$R(t,s) = b(t - s)/b(0) = R(t - s)$$

differs from $b(t - s)$ only by a constant factor. Thus, in the correlation theory of stationary random functions these functions are completely characterized by a constant m and a function $B(\tau)$ (or $b(\tau)$) of a single variable $\tau = t - s$ (integral for random sequences and arbitrary real for random processes).

Since in the correlation theory of random functions multidimensional probability distributions do not appear at all, the very definition of stationarity in this theory should be changed.   In the correlation theory it is natural to call a random function $X(t)$ stationary provided that *its mean value* $\langle X(t) \rangle$ *is a constant and its correlation function* $\langle X(t)X(s) \rangle$ *depends only on* t − s.   Random functions satisfying these two conditions are said to be *stationary in the wide sense* or *widely stationary*.[8] These will be the main functions studied in Chaps. 1 − 3 of this book (and hence they will simply be called stationary for brevity).   It is clear that the general condition (1.23) or (1.24) of ordinary stationarity (which is also called *stationarity in the strict sense* or *strict stationarity*) may not hold, generally speaking, for random functions, which are stationary only in the wide sense.   We note, however, that in practice the fulfillment of the wide sense stationarity conditions (1.26) and (1.27) almost always means that the function $X(t)$ under investigation is generated by some steady phenomenon and is therefore also strictly stationary. (Indeed, what other factors can ensure the exact fulfillment of these two conditions?) Thus, functions $X(t)$, which are stationary in the wide sense, but are not strictly stationary, are almost never encountered in practical applications. Nevertheless, this does not mean that the concept of wide sense stationarity is of no practical interest at all.   The point is that the verification of the general conditions (1.23) or (1.24) of strict stationarity is very difficult in many real situations and sometimes altogether impossible, while conditions (1.26) and (1.27) are much easier to verify since they relate only to the simplest statistical characteristics $m(t)$ and $B(t,s)$ of the function $X(t)$, which can often be measured rather accurately

with little effort. If it turns out that the conditions (1.26) and (1.27) are fulfilled for a function $X(t)$ with a high degree of accuracy, this will be sufficient for applying to it the results of the correlation theory of stationary random functions, neglecting the fact that the accuracy of the conditions (1.23) and (1.24) may, in principle, be unsatisfactory in this case.

In conclusion it should be noted that for Gaussian random functions, for which the mean value $m(t)$ and the correlation function $B(t,s)$ (or $b(t,s)$) completely specify all the multidimensional probability distributions, the concept of stationarity in the wide sense and of stationarity in the strict sense are exactly the same. In fact, a Gaussian function always has finite first and second moments; moreover, if $m(t)$ is independent of t and $B(t,s)$ depends only on $t - s$, it can be easily shown that all the corresponding multidimensional (Gaussian) probability distributions satisfy the condition (1.23).

4. Properties of Correlation Functions.   Derivative and Integral of a Random Process

It will be shown in Sec. 16 that under very general conditions the mean value m of a stationary random function $X(t)$ can be measured rather easily and with a good degree of accuracy. This often allows one to consider, instead of the function $X(t)$, its "fluctuation" $\overset{\circ}{X}(t) = X(t) - m$, i.e. the deviation of $X(t)$ from its mean value. The function $\overset{\circ}{X}(t)$ has a zero mean value, and its correlation function $B(\tau)$, which is the only characteristic of $\overset{\circ}{X}(t)$ used in correlation theory, coincides with the centered correlation function $b(\tau)$ of the initial function $X(t)$. Since even when m is unknown, addition of a constant to $X(t)$ hardly changes most of the theoretical considerations and can usually be easily taken into account, we shall assume, as a rule, that $\langle X(t) \rangle = 0$. If the situation is different, however, this will usually be specifically indicated.

By virtue of the general properties (1.17), (1.19), and (1.19a) of the correlation function $B(t,s)$ the function $B(\tau)$ must be such that

$$(1.28) \qquad B(0) > 0, \quad B(-\tau) = B(\tau), \quad |B(\tau)| \leqslant B(0).$$

Moreover, by (1.18), for any positive integer n, any n numbers (integer-valued in the case of stationary sequences and arbitrary real in the case of stationary processes) τ_1, ..., τ_n and any n real numbers c_1, ..., c_n the following inequality must hold:

(1.29) $\sum\limits_{j,k=1}^{n} B(\tau_j - \tau_k) \, c_j c_k \geqslant 0.$

The function $B(\tau)$, such that $B(-\tau) = B(\tau)$ and for which condition (1.29) is fulfilled at all n, τ_1, ..., τ_n and c_1, ..., c_n, is called a *positive definite function* of the argument τ (integral or arbitrary real).* Thus, *the correlation function of a stationary random function is always a positive definite function.* The converse is also true: *any positive definite function of an integer-valued or, respectively, a real argument τ is a stationary correlation function,* i.e. a correlation function of some random stationary sequence or, respectively, process (see Sec. 2, where a similar statement was formulated for general random functions on an arbitrary set T).

It follows that *the class of all stationary correlation functions coincides with the class of positive definite functions* (of an integer-valued or arbitrary real argument). We also note that the class of centered stationary correlation functions $b(\tau)$ coincides with the class of ordinary (noncentered) stationary correlation functions $B(\tau)$ (since both of them coincide with the class of positive definite functions). Moreover, the class of normalized stationary correlation functions $R(\tau)$ coincides with the class of positive definite functions equal to unity at the point $\tau = 0$.

By using the coincidence of the class of stationary correlation functions and the class of positive definite functions or by proceeding directly from the definition of the correlation function $B(\tau)$, one can obtain many important general properties of such functions. For example, it is easy to see that *if $B_1(\tau)$ and $B_2(\tau)$ are two stationary correlation functions while a_1 and a_2 are non-negative constants, then*

(1.30) $B(\tau) = a_1 B_1(\tau) + a_2 B_2(\tau)$

---

*The positive definite functions..., B(-1), B(0), B(1),... of an integer-valued argument τ are also called <u>positive definite sequences</u>.

is also a stationary correlation function, since from the validity of inequalities (1.29) for the functions $B_1(\tau)$ and $B_2(\tau)$ it follows at once that (1.29) will hold true for the function (1.30) as well. [It is also possible to base this deduction on the obvious fact that $B(\tau)$ is a correlation function of the random function $X(t) = \sqrt{a_1}X_1(t) + \sqrt{a_2}X_2(t)$, where $X_1(t)$ and $X_2(t)$ are uncorrelated stationary random functions with correlation functions $B_1(\tau)$ and $B_2(\tau)$.] The simplest linear combination (1.30) can be also replaced by a more general linear combination

$$(1.31) \qquad B_1(\tau) = \sum_{i=1}^{n} a_i B_i(\tau)$$

of n stationary correlation functions $B_1(\tau)$, ..., $B_n(\tau)$ with non-negative coefficients a_1, ..., a_n. It is clear that such a combination will also always be a stationary correlation function. Moreover, the arbitrary integer n in (1.31) can even be set equal to infinity if the corresponding infinite sum converges. Indeed, *if $B^{(n)}(\tau)$, $n = 1, 2, ...,$ is any sequence of stationary correlation functions having a limit*

$$(1.32) \qquad B(\tau) = \lim_{n \to \infty} B^{(n)}(\tau),$$

then $B(\tau)$ is also a stationary correlation function (since if condition (1.29) is valid for $B^{(n)}(\tau)$ for all $n = 1, 2, ...,$ then it is clearly also valid for $B(\tau)$). Since $\sum_{i=1}^{\infty} a_i B_i(\tau) = \lim_{n \to \infty} \sum_{i=1}^{n} a_i B_i(\tau)$, the convergent infinite sum of stationary correlation functions multiplied by non-negative constants also represents a stationary correlation function.

Let us suppose now that *$B(\tau;a)$ is a stationary correlation function which depends on the real parameter* a, *h(a) is a non-negative function of* a, *and* H(a) *is a monotone nondecreasing function of* a. *Then both the functions*

$$(1.33) \qquad B_h(\tau) = \int_A B(\tau; a)h(a)da$$

and

$$(1.34) \qquad B_H(\tau) = \int_A B(\tau; a)dH(a)$$

(*where* A *is an axis* $-\infty < a < \infty$ *or any part of it*) *are stationary correlation functions.*   (This statement is clearly a simple consequence of the result related to the function $B(\tau)$ on the left-hand side of (1.32).) Equations (1.33) and (1.34) can also be easily generalized to the multidimensional case concerning the stationary correlation function $B(\tau; \mathbf{a}) = B(\tau; a_1, ..., a_k)$, which depends on the vector parameter $\mathbf{a} = (a_1, ..., a_k)$.

Some classes of the non-stationary correlation functions will be considered in Chap. 4 of this book, but in Chaps. 1 − 3 such functions will be mentioned only in a few cases. Therefore in Chaps. 1 − 3 we shall not use the adjective "stationary" when speaking of functions $B(\tau)$ and $b(\tau)$, but shall mean by a correlation function a stationary correlation function unless otherwise specified.   It is easy to show that *the product of any two correlation functions* $B_1(\tau)$ *and* $B_2(\tau)$ *is also a correlation function.*   In proving this, it is easier to proceed, not from the condition (1.29), but from the definition of the correlation function as the second moment of a random function.   Let $X_1(\tau)$ and $X_2(\tau)$ be stationary random functions, whose correlation functions are equal to $B_1(\tau)$ and $B_2(t)$, respectively.   The functions $X_1(\tau)$ and $X_2(\tau)$ are specified by the sets of their finite-dimensional probability distributions.   In order to specify the pair $(X_1(t), X_2(t))$ we must give the joint probability distributions of all possible groups $\{X_1(t_1), ..., X_1(t_n), X_2(t_1'), ..., X_2(t_m')\}$, which include the values of both the first and second functions.   It is always possible to specify the joint distributions of the two functions by the relation

$$P\{X_1(t_1) < x_1, ..., X_1(t_n) < x_n, \; X_2(t_1') < x_1', ..., X_2(t_m') < x_m'\}$$

$$= P\{X_1(t_1) < x_1, ..., X_1(t_n) < x_n\} \, P\{X_2(t_1') < x_1', ..., X_2(t_m') < x_m'\},$$

without changing the distributions of any one of the functions $X_1(t)$ and $X_2(t)$ separately (i.e. to choose $X_1(t)$ and $X_2(t)$ to be mutually independent).   Let us now examine a new random function $X(t) = X_1(t)X_2(t)$. Since the mean value of the product of independent random variables is equal to that of the corresponding mean values,

$$(1.35) \quad B(\tau) = \langle X_1(t + \tau)X_2(t + \tau) \ X_1(t)X_2(t) \rangle$$

$$= \langle X_1(t + \tau)X_1(t) \rangle \ \langle X_2(t + \tau)X_2(t) \rangle = B_1(\tau)B_2(\tau).$$

This shows that $B_1(\tau)B_2(\tau)$ is also a correlation function.

If we have some examples of correlation functions, it would be possible to obtain a large number of new examples of such functions by using the above reasoning. At first glance it would seem, however, that it is very hard to verify the validity of the condition (1.29) for any given function, and therefore it is not clear where we can obtain the initial examples of correlation functions $B(\tau)$. We shall see below that this first impression is false and that there are comparatively simple methods of constructing positive definite functions $B(\tau)$. The verification of the validity of condition (1.29) for a given function $B(\tau)$ is also not as complicated as it may seem at first; see Secs. 10 and 15 of Chap. 2, where numerous specific examples of correlation (i.e. positive definite) functions $B(\tau)$ are given.

Note also that $R(\tau) = \dfrac{b(\tau)}{b(0)}$ (where $b(\tau) = B(\tau) - m^2$, and $m =$ $\langle X(t) \rangle$) is the correlation coefficient between the random variables $X(t + \tau)$ and $X(t)$, which depends only on τ in the stationary case. In most practical applications of the theory of stationary random functions it is natural to assume that the statistical relationship between the quantities $X(t)$ and $X(t + \tau)$ must be very weak if τ is large enough. More precisely, it is natural to assume that

$$(1.36) \quad b(\tau) \to 0, \quad B(\tau) \to m^2 = \text{const for } |\tau| \to \infty,$$

where $m = \langle X(t) \rangle$ is no longer considered necessarily equal to zero. The condition (1.36), unlike the above-discussed conditions referring to $B(\tau)$, is not strictly necessary (below we shall also encounter examples of functions $B(\tau)$ for which it fails). However, in applications it can confidently be considered valid in almost all cases.

In the remainder of this section we consider only stationary random processes, i.e. functions $X(t)$ of a real argument t. We shall need the concept of the limit of a sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$. In probability theory

theory there are several "reasonable" definitions of such a limit, and the choice of the most convenient one depends on the problem at hand.   In the correlation theory of random functions the following definition is the most appropriate: *The random variable* X *is the limit of the sequence of random variables* $X_1, X_2, ..., X_n, ...$ *(i.e.* $X = \lim\limits_{n \to \infty} X_n$*) if*

(1.37) $\lim\limits_{n \to \infty} \langle (X - X_n)^2 \rangle = 0,$

i.e. if for any $\epsilon > 0$ *there exists* $N = N(\epsilon)$ *such that*

(1.38) $\langle (X - X_n)^2 \rangle < \epsilon \quad if \quad n > N.$

A limit defined in this way is usually called a *mean square limit,* or a *limit in the mean (square)* of the sequence X_n, n = 1, 2, ... and X_n is said to *converge to* X *in the mean (square).* In the rest of this book the limit of a sequence of random variables will always be understood as a limit of this kind unless otherwise stipulated. Using the well-known Chebyshev inequality [9]

(1.39) $P\{ |X - X_n| > \epsilon \} \leqslant \dfrac{\langle (X - X_n)^2 \rangle}{\epsilon^2},$

it is easy to show that (1.37) implies that the relation

(1.40) $\lim\limits_{n \to \infty} P\{ |X - X_n| > \epsilon \} = 0$

holds for any $\epsilon > 0$. Equation (1.40) means that given any $\epsilon > 0$ and $\eta > 0$, there exists $N_0 = N_0(\epsilon, \eta)$ such that

(1.41) $P\{ |X - X_n| < \epsilon \} > 1 - \eta \quad if \quad n > N_0.$

The relation (1.41) makes it clear why the random variable X satisfying (1.37) can appropriately be called the *limit* of the sequence $X_1, X_2, ..., X_n, ...$.   It is also clear that a variable X satisfying (1.40) for all $\epsilon > 0$ can naturally be considered as the limit of the sequence $X_1, X_2, ..., X_n, ...$ even when the condition (1.37) does not hold; such an X is usually called the *limit in probability* of the sequence $X_1, X_2, ..., X_n, ...$.   The result just proved can, therefore, be stated as follows: *If* X *is the limit in the mean of the sequence* $X_1, X_2, ..., X_n, ...,$ *then* X *is also the limit in probability of the same sequence.*

Now let $x, x_1, x_2, ..., x_n ...$ be some realization of random

variables $X, X_1, X_2, ..., X_n, ...$. According to what has been said above, if X is the limit in probability of the sequence $X_1, X_2, ...,$ $X_n, ...$, and if n is sufficiently large, then the difference $x - x_n$ will be very small in absolute value with a near-unity probability. This does not mean, however, that $|x - x_n|$ will necessarily be very small at *all* sufficiently large values of n at once. In other words, the convergence in probability of the sequence of random variables $X_1, X_2, ..., X_n, ...$ to X, as $n \to \infty$, does not yet imply that $x_n \to x$, as $n \to \infty$, for each realization $x, x_1, x_2,$ ..., $x_n, ...$. If, however, $x_n \to x$, as $n \to \infty$, for all realizations (except maybe some exceptional set of "pathological realizations", which are so improbable that the total probability of their appearance is zero), the sequence $X_1, X_2, ..., X_n, ...$ is said *to converge to the random variable* X *almost surely* (or *with probability one*). Accordingly, X is called in this case the *almost-sure limit* (or the *limit with probability one*) of the sequence $X_1, X_2, ..., X_n, ...$. It is not hard to show that *from almost-sure convergence of the sequence* $X_1, X_2, ..., X_n, ...$ *to* X *it follows that* X_n *also converges to* X *in probability* (although it may not converge to X in the mean); but the convergence of X_n to X in probability (or even in the mean) does not imply that X_n also converges to X almost surely.[10]

Let us now continue the consideration of the limits in the mean (square) which are of greatest interest to us. We assume as usual that all the considered random variables have finite variances. Using the Cauchy–Buniakovsky–Schwarz inequality

$$(1.42) \qquad |\langle UV \rangle| \leqslant [\langle U^2 \rangle \langle V^2 \rangle]^{1/2} \text{ for all } U, V$$

(see above (0.23) and (1.19a)), it is easy to show that *if* $X = \lim\limits_{n \to \infty} X_n$ *and* $Y = \lim\limits_{m \to \infty} X_m$, *then*

$$(1.43) \qquad \lim\limits_{n \to \infty, m \to \infty} \langle X_n Y_m \rangle = \langle XY \rangle.$$

Indeed,

$$|\langle X_n Y_m \rangle - \langle XY \rangle| = |\langle [(X_n - X) + X][(Y_m - Y) + Y] \rangle$$

$$- \langle XY \rangle| \leqslant |\langle (X_n - X)(Y_m - Y) \rangle| + |\langle (X_n - X)Y \rangle| + |\langle X(Y_m - Y) \rangle|.$$

Applying (1.42) to all the terms on the right-hand side, we

find that it tends to zero, as $n \to \infty$, $m \to \infty$, i.e. that (1.43) holds true. It is still easier to prove that *if* $X = \lim_{n \to \infty} X_n$, *then*

(1.44) $\lim_{n \to \infty} \langle X_n Y \rangle = \langle XY \rangle$ *for any* Y.

[(1.44) is in fact a particular case of (1.43) corresponding to the sequence $(Y_1, Y_2, ..., Y_m, ...) = (Y, Y, ... Y, ...)$ of identical terms]. Noting now that

$$(X_n - X_m)^2 = [(X_n - X) - (X_m - X)]^2 = (X_n - X)^2$$

$$-2 (X_n - X)(X_m - X) + (X_m - X)^2,$$

and using (1.42) again, it is easy to show that *if the sequence* $X_1, X_2, ..., X_n, ...$ *converges (i.e. if* $\lim_{n \to \infty} X_n = X$ *exists), then*

(1.45) $\lim_{n \to \infty, m \to \infty} \langle X_n - X_m)^2 \rangle = 0.$

It is more difficult to prove the converse statement that *if* (1.45) *holds for some sequence of random variables* $X_1, X_2, ..., X_n, ...,$ *then this sequence necessarily converges, i.e.* $\lim_{n \to \infty} X_n = X$ *exists*;

nevertheless, this statement is also true.[11] After removing the parentheses in (1.45) we find that *the sequence* $X_1, X_2, ..., X_n, ...$ *will converge if the following limit exists:*

(1.46) $\lim_{n \to \infty, m \to \infty} \langle X_n X_m \rangle$

(because $\lim_{m \to \infty} \langle X_n^2 \rangle$ and $\lim \langle X_m^2 \rangle$ are only particular cases of

(1.46)). It is also clear that for any convergent sequence X_1, $X_2, ..., X_n, ...$ the limit (1.46) exists and is equal to $\langle X^2 \rangle$, where $X = \lim_{n \to \infty} X_n$.

We now return to the study of stationary random processes $X(t)$ and their correlation functions $B(\tau)$. It is clear that

$$\langle [X(t + h) - X(t)]^2 \rangle = 2[B(0) - B(\tau)].$$

Therefore, *if the function* $B(\tau)$ *is continuous at the point* $\tau = 0$ (i.e. $\lim_{h \to 0} B(h) = B(0)$), *the process* $X(t)$ *will be mean square*

continuous (or *continuous in the mean*) *in the sense that*

(1.47) $\lim\limits_{h \to 0} \langle [X(t + h) - X(t)]^2 \rangle = 0$, i.e. $\lim\limits_{h \to 0} X(t + h) = X(t)$

for all t (and, hence, $\lim\limits_{h \to 0} P\{|X(t + h) - X(t)| > \epsilon\} = 0$ for all *t*

and any $\epsilon > 0$).

If, however, the function $B(\tau)$ is not continuous at zero, the process $X(t)$ will not be mean square continuous at any point (i.e. the condition $\lim\limits_{h \to 0} \langle [X(t + h) - X(t)]^2 \rangle = 0$ will fail at any

t). Such discontinuous random processes must naturally behave extremely irregularly and can hardly serve as a suitable model for any reasonable applied problem. Therefore such processes will not be considered at all in the subsequent text, i.e. all the considered correlation functions $B(\tau)$ will be assumed continuous at the point $\tau = 0$. It is easy to show that in this case the function $B(\tau)$ will also be everywhere continuous. In fact, if we let $U = X(t + h) - X(t)$ and $V = X(t - \tau)$ in inequality (1.42), we find that

$$|B(\tau + h) - B(\tau)| \leqslant \sqrt{2} \, [B(0) - B(h)]^{1/2} \, [B(0)]^{1/2}.$$

Hence, *if* $\lim\limits_{h \to 0} B(h) = B(0)$, *then also* $\lim\limits_{h \to 0} B(\tau + h) = B(\tau)$

for all τ.

Note that the mean square continuity of the process $X(t)$ for all *t* does not mean that all the realizations $x(t)$ of this process are continuous functions. Indeed, we have already seen in Sec. 2 that, e.g., $B(\tau) = \langle X(t + \tau)X(t) \rangle = a^2 \exp(-2\lambda|\tau|)$ for a random telegraph signal $X(t)$, i.e. the function $B(\tau)$ is continuous here. However, the realizations $x(t)$ of this process $X(t)$ are discontinuous (see Fig. 11). There is no contradiction here, of course: although $x(t)$ has discontinuity points, for each fixed *t* the probability of discontinuity between the instants *t* and $t + h$ for a small *h* is very low, and therefore the mean square of $X(t + h) - X(t)$ is also small. To find conditions guaranteeing that the realizations $x(t)$ will be continuous at all the points at once, one evidently needs quite a different approach, no longer based on consideration of the mean square limits alone; we shall not pursue this topic here.[12]

If $X(t)$ is a mean square continuous random process, then it is also natural to consider a sequence of random variables

$X(t, h_k) = \dfrac{X(t + h_k) - X(t)}{h_k}$, which correspond to the given sequence

of numbers $h_1, h_2, ..., h_k, ...$ converging to zero.   If, for any such sequence $h_1, h_2, ..., h_k, ...$, the sequence of random variables $X(t, h_k)$ converges in the mean to a unique random variable, then the process $X(t)$ is said to be *mean square differentiable* at a point t, and the quantity

$$(1.48) \qquad X'(t) = \lim_{h \to 0} \frac{X(t + h) - X(t)}{h}$$

is called the *(mean square) derivative* of the process $X(t)$ at the point t. It is easy to see that the stationary process $X(t)$ is either everywhere differentiable or nondifferentiable for any t. The process $X(t)$ is differentiable (i.e. the derivative (1.48) exists) provided that the following limit exists:

$$
(1.49) \quad
\begin{aligned}
&\lim_{h \to 0, h' \to 0} \left\langle \frac{X(t + h) - X(t)}{h} \; \frac{X(t + h') - X(t)}{h'} \right\rangle \\
&= \lim_{h \to 0, h' \to 0} \frac{B(h - h') - B(h) - B(h') + B(0)}{h\,h'} .
\end{aligned}
$$

It is easy to see that if the correlation function $B(\tau)$ has a second derivative $B''(0)$ at the point $\tau = 0$, the limit (1.49) will be finite and equal to $- B''(0)$. Therefore, in this case the process $X(t)$ will be differentiable at any t.   Since

$$
\left\langle \frac{X(t + \tau + h) - X(t + \tau)}{h} \; \frac{X(t + h) - X(t)}{h} \right\rangle
$$

$$
= \frac{2B(\tau) - B(\tau + h) - B(\tau - h)}{h^2} ,
$$

it is clear that the second derivative $B''(\tau)$ exists for any τ if the process $X(t)$ is differentiable and

$$(1.50) \qquad \langle X'(t + \tau) X'(t) \rangle = - B''(\tau).$$

Hence *if the correlation function* $B(\tau)$ *is twice differentiable at the point* $\tau = 0$, *then it is also twice differentiable everywhere.* The corresponding stationary process $X(t)$ is then differentiable and its derivative $X'(t)$ has the correlation function equal to $-B''(\tau)$

(i.e. $-B''(\tau)$ is also a correlation function). It is also clear that

(1.51) $\langle X'(t + \tau)X(t) \rangle = B'(\tau), \quad \langle X(t + \tau)X'(t) \rangle = - B'(\tau).$

Since $B(-\tau) = B(\tau)$, if $B(\tau)$ is a differentiable function, then
necessarily $B'(0) = 0$; therefore, $\langle X'(t)X(t) \rangle = 0$ *for any*
differentiable $X(t)$. It should also be remarked that the function
$B''(\tau)$ is necessarily continuous at all points τ, provided it is
continuous at the point $\tau = 0$ (since the function $-B''(\tau)$ is a
correlation function). If, however, the function $B(\tau)$ has no
finite second derivative at the point $\tau = 0$, the limit (1.49) does
not exist and the process $X(t)$ is everywhere not mean square
differentiable.
    Processes $X(t)$, for which $B''(\tau)$ is a continuous function
and, hence, for which there exists a mean square continuous
process $X'(t)$, will   from   now   on   be   simply   called
differentiable. From the foregoing discussion of continuous
processes it follows that the realization $x'(t)$ of the
derivative $X'(t)$ of a differentiable process $X(t)$ may be a
discontinuous function of t.   On the other hand, the process
$X(t)$ may also be nondifferentiable when all its realizations
$x(t)$ have continuous derivatives $x'(t)$.   Indeed, it has already
been remarked in Sec. 2 that the function $B(\tau) = \sigma^2 \exp(-$
$\lambda'|\tau|)$ (for which the derivative $B'(\tau)$ is discontinuous at the
point $\tau = 0$, and $B''(0)$ does not exist) may be a correlation
function of a process $X(t)$ having only sinusoidal realizations.
It should, however, be noted that for all the differentiable
stationary processes $X(t)$ encountered in practice, realizations
$x(t)$ also turn out to be differentiable, and the functions $x'(t)$
coincide with the realizations of the process $X'(t)$.[13] This
rarely noted fact is of considerable applied value, because it
indicates   that   the   mean   square   derivative $X'(t)$ (more
precisely, its realization $x'(t)$) can usually be determined
from one realization $x(t)$ of the differentiable process $X(t)$ by
differentiating this realization.

    Let us now consider *integrals* of the type

(1.52) $\int_a^b f(t)X(t)dt,$

where $f(t)$ is a numerical function, and $X(t)$ is a stationary
process. Such an integral can be appropriately defined as the
limit (in the mean) of the corresponding integral sum, i.e.

$$(1.53) \qquad \int_a^b f(t)X(t)dt = \lim_{\max|t_k - t_{k-1}| \to 0} \sum_{k=1}^{n} f(t'_k)X(t'_k)(t_k - t_{k-1}),$$

where $a = t_0 < t_1 < ... < t_{n-1} < t_n = b$, $t_{k-1} \leqslant t'_k \leqslant t_k$. It is easy to deduce from this that a sufficient condition for the existence of the integral (1.52) is that the double integral

$$(1.54) \qquad \int_a^b \int_a^b B(t - s)f(t)f(s)ds dt < \infty,$$

exists and that under this condition

$$(1.55) \qquad \langle (\int_a^b f(t)X(t)dt)^2 \rangle = \int_a^b \int_a^b B(t - s)f(t)f(s)ds dt.$$

Similarly

$$(1.56) \qquad \langle \int_a^b f(t) X(t)dt \int_c^d g(s)X(s)ds \rangle = \int_a^b \int_c^d B(t - s)f(t)g(s)ds dt.$$

The arguments required to prove condition (1.54) and relations (1.55) and (1.56) are very close to those given above in connection with the existence of the derivative (1.48). Since we have agreed that the function $B(\tau)$ is considered to be continuous, the condition (1.54) will definitely hold, for instance, for any bounded piece-wise continuous function $f(t)$.

The improper integrals

$$\int_a^\infty f(t)X(t)dt, \quad \int_{-\infty}^b f(t)X(t)dt \text{ and } \int_{-\infty}^\infty f(t)X(t)dt$$

are defined in the usual way as the (mean square) limits of (1.52) as $b \to \infty$, or $a \to -\infty$, or $a \to -\infty$ and $b \to \infty$. The condition (1.54) for the existence of the integral and the relation (1.55) and (1.56) are also valid for improper integrals, provided we assume that the limits of integration in (1.54) -- (1.56) can take infinite values.

It should be emphasized, however, that although the above results relating to the integrals (1.52) are very simple, these results can hardly satisfy practical workers. In applications we often encounter only one realization $x(t)$ of the process $X(t)$ (more rarely, a finite number of such realizations) and evaluate the integral of the process by integrating this realization (numerically or with the aid of an integrator of some kind). Then we naturally obtain the quantity

(1.52a) $\int_a^b f(t)x(t)dt$

and not the mean square limit (1.53), which includes averaging over the whole set of realizations. Therefore, it is extremely important in applications that mean square continuous stationary processes $X(t)$ always have integrable (with probability one) realizations, and the integral (1.52) coincides here with a random variable, the sampling values of which are equal to the integrals (1.52a). However, this fact is seldom noted in any book except advanced mathematical treatises. Since its proof requires the use of rather subtle mathematical concepts and theorems, we shall not pursue it here.[14]

5. Complex Random Functions. Spaces of Random Variables

When considering examples of random functions in the Introduction we mentioned, in particular, "random harmonic oscillation" (see Fig. 9). It will be seen later that this example is basic to the theory of stationary random functions in the following sense: Any stationary function $X(t)$ can be represented as a superposition of random harmonic oscillations, i.e. it can be represented by some modification of the conventional Fourier integral.

The following remark may be quite useful in this connection. So far we have been speaking exclusively of real random variables and real random functions. However, it is known that harmonic oscillations can be described very conveniently by complex functions of the form $A\exp(i\omega t)$, and the complex form of the Fourier integral is also simpler in some respect than its real form. Therefore, in theoretical considerations it is often convenient to consider more general complex random variables and complex random functions. A complex random variable X can be defined as $X = X_1 + iX_2$, where X_1 and X_2 are two real random variables. Similarly, a complex random function $X(t)$ is, by definition, equal to the sum $X_1(t) + iX_2(t)$ where $X_1(t)$ and $X_2(t)$ are real random functions.

The complex random function $X(t) = X_1(t) + iX_2(t)$ is specified by the set of distribution functions (or probability densities) of all groups of the $2n$ real random variables $X_1(t_1)$, $X_1(t_2)$, ..., $X_1(t_n), X_2(t_1), X_2(t_2)$, ..., $X_2(t_n)$ for all possible values of

n, t_1, t_2, ..., t_n. The function $X(t)$ is said to be *stationary* if all these probability distributions are invariant under arbitrary time shifts, i.e. if they remain unaltered when an arbitrary number τ is added to all the arguments t_1, t_2, ..., t_n. It is clear that in this case the real and imaginary parts of $X(t)$ — the functions $X_1(t)$ and $X_2(t)$ — are real-valued stationary random functions.

The mean value $\langle X(t)\rangle = \langle X_1(t) + iX_2(t)\rangle$ of a complex stationary random function is obviously a (complex) constant $m = m_1 + im_2$. Further on we assume, as a rule, that $\langle X(t)\rangle = m = 0$; this will not lead to a loss of generality. As for the second moments of the function $X(t)$, they include the following four real functions of the argument τ: $\langle X_1(t + \tau)X_1(t)\rangle$, $\langle X_1(t + \tau)X_2(t)\rangle$, $\langle X_2(t + \tau)X_1(t)\rangle$, and $\langle X_2(t + \tau)X_2(t)\rangle$. However, only the following combination of them will be of importance to us:

$$B(\tau) = \langle X(t + \tau)\overline{X(t)}\rangle = \langle (X_1(t + \tau) + iX_2(t + \tau))(X_1(t) - iX_2(t))\rangle$$

$$(1.57) \quad = \langle X_1(t + \tau)X_1(t)\rangle + \langle X_2(t + \tau)X_2(t)\rangle$$

$$+ i[\langle X_2(t + \tau)X_1(t)\rangle - \langle X_1(t + \tau)X_2(t)\rangle].$$

(The overbar denotes, as usual, the complex conjugate.) The function $B(\tau)$ will be called the *correlation function* of the complex stationary random function $X(t)$. Accordingly, if $\langle X(t)\rangle = m \neq 0$, then

$$b(\tau) = \langle (X(t + \tau) - m)(\overline{X(t) - m})\rangle = B(\tau) - |m|^2$$

is called the *centered correlation function* of $X(t)$, and $R(\tau) = b(\tau)/b(0)$, the *normalized correlation function*. (If $m = 0$ then obviously $b(\tau) = B(\tau)$.) In particular, if $X(t)$ is real (i.e. $X_2(t) = 0$), the above definitions coincide with those used before. Complex random variables X and Y will be called *uncorrelated* if $\langle (X - \langle X\rangle)(\overline{Y - \langle Y\rangle})\rangle = 0$ (in particular, if $\langle X\overline{Y}\rangle = 0$ when $\langle X\rangle = \langle Y\rangle = 0$).

The correlation function $B(\tau)$ is a complex function satisfying the first and third of the relations (1.28). However, the second relation (1.28) assumes the following form in the complex case:

$$(1.58) \quad B(-\tau) = \overline{B(\tau)}.$$

Similarly, the inequality (1.29) now turns into the following statement: *for any positive integer* n, *arbitrary integral or real (depending on whether* X(t) *is a sequence or a process),* $\tau_1, \tau_2, ...,$ τ_n *and complex numbers* $c_1, c_2, ..., c_n$

$$(1.59) \qquad \sum_{j,k=1}^{n} B(\tau_j - \tau_k) c_j \overline{c}_k \geq 0.$$

Actually, the proof of these facts does not differ from the corresponding proofs in the real case. (For instance, the left-hand part of (1.59) is equal to $\langle |\sum_{j=1}^{n} X(\tau_j) c_j|^2 \rangle \geq 0$.) The complex function $B(\tau)$ for which (1.59) holds is called the *positive definite* complex function of the argument τ. (It is easy to show that the first and third of the relations (1.28) and condition (1.58) follow from (1.59); cf. Note 2 to Chap. 1.) Hence, in the complex case, also, *the correlation function of a stationary random function is necessarily positive definite.* Quite analogously to the real case, it is also possible to prove the converse, namely that *any positive definite function of an integer-valued or, respectively, a real argument* τ *is the correlation function of some stationary sequence or, respectively, of a stationary process.*[15]

In the following chapter the stationary random functions $X(t)$ under consideration will always be assumed complex, with zero mean values. As regards examples, we shall, as a rule, choose them to involve only real random functions, since just such functions are actually encountered in most of the applications.

It is useful to conclude the chapter with one more remark on the use of geometric language, which is often convenient in investigation of random variables and random functions. We have already noted at the very beginning of the Introduction that random variables X are always associated with an experiment, which may have different outcomes. To each of these outcomes corresponds its own numerical value x of the quantity X. Let us first consider the simplest situation arising when we deal with an experiment which only has a finite number k of different outcomes (e.g., the toss of a die). Let $p_1, p_2, ..., p_k$ denote the probabilities of these k outcomes. The random variable X associated with our experiment takes certain numerical values $x_1, x_2, ..., x_k$ depending on the outcome of the experiment. (The indicated values are all possible realizations of the variable X.) Quantities X of this

kind can be regarded as k-dimensional vectors with components $(x_1, x_2, ..., x_k)$. The set of all random variables associated with our experiment is then a k-dimensional vector space H_k (real or complex, depending on whether we consider real or complex random variables).

To make this geometric model more meaningful, we must introduce a metric in the space H_k, i.e. define a "length" (or *norm*) $\|X\|$ of the vector X. The most appropriate definition of the norm $\|X\|$ is

$$(1.60) \qquad \|X\|^2 = \sum_{i=1}^{k} p_i x_i^2$$

(or, in the complex case, $\|X\|^2 = \sum_{i=1}^{k} p_i |x_i|^2$).   In other words, we

assume that the probabilities p_i define the length scales along k coordinate axes in H_k.   Then the *scalar product* (X, Y) of two vectors $X = (x_1, x_2, ..., x_k)$ and $Y = (y_1, y_2, ..., y_k)$ (which

can be expressed as $(1/2)(\|X + Y\|^2 - \|X\|^2 - \|Y\|^2)$ in the real case) should be defined by the formula

$$(1.61) \qquad (X,Y) = \sum_{i=1}^{k} p_i x_i y_i$$

(or in the complex case, by $(X, Y) = \sum_{i=1}^{k} p_i x_i \overline{y}_i$).   In probability

theory language, (1.60) and (1.61) mean that the square of the norm of a vector and the scalar product of two vectors are given by the following mean values:

$$(1.62) \qquad \|X\|^2 = \langle X^2 \rangle, \quad (X, Y) = \langle X Y \rangle$$

or, in the complex case,

$$(1.62a) \qquad \|X\|^2 = \langle |X|^2 \rangle, \quad (X, Y) = \langle X \overline{Y} \rangle.$$

Thus, the metric just introduced has a simple probabilistic meaning.

Let us now pass on to the general case of an experiment, the set of possible outcomes of which is no longer assumed finite.   The random variable X will now be a numerical function on the entire set of outcomes (because to each of them corresponds a certain value x of the quantity X).

Generally speaking, a set of such variables X will not form a finite-dimensional space.   Nevertheless the variables X can again be thought of as vectors, since they form a *linear vector space* in the sense that the elements of this space (i.e. random variables) can be added together (i.e. $X + Y$ can be determined for any X and Y) and multiplied by numbers (if c is a number and X is a random variable, then cX is also a random variable). We shall consider only complex random variables X with a bounded mean square $\langle |X|^2 \rangle$ (or, what is the same, with a finite variance) and introduce the space H of all such random variables.   In the space H we define the square of the norm (the "square of the length of a vector") and the scalar product of the two vectors with the aid of formulae (1.62a).   Note that our new space H is also a linear vector space. In fact, if $\langle |X| \rangle^2 < \infty$ and $\langle |Y| \rangle^2 < \infty$, then also $\langle |cX|^2 \rangle = |c|^2 \langle |X|^2 \rangle < \infty$ and, as can readily be shown by means of the Cauchy–Buniakovsky–Schwarz inequality $\langle |X + Y|^2 \rangle < \infty$, i.e. X and $X + Y$ are also elements of H.[16] Moreover, it is easy to see that the square of the length of a vector and the scalar product of the two vectors, defined by (1.62a), have the usual properties of the same quantities in a finite-dimensional vector space.   Under this condition the fact that the space H is, generally speaking, infinite-dimensional (i.e. is a *Hilbert space*) only slightly restricts the applicability of the conventional geometric concepts to H.[17]

From our new standpoint, the random sequence $X(t)$ is a denumerable sequence of elements ("vectors" or "points") of the Hilbert space H, whereas the random process $X(t)$ is a one-parameter family of points of H, i.e. a *curve* in this space.[18] The correlation function $B(t,s) = (X(t),X(s))$ is a scalar product of the vectors $X(t)$ and $X(s)$, i.e. the basic geometric characteristic of the corresponding sequence of points or curves.   Thus, in a certain sense, the correlation theory of random processes is equivalent to the theory of curves in a Hilbert space H.   In particular, stationary random processes correspond to curves $X(t)$ in H for which the scalar product $(X(t),X(s))$ depends only on $t - s$.   It is not hard to see that geometrically this means that there exists a one-parameter group of rotations in H which transforms the "point" $X(0)$ into an arbitrary point $X(t)$, and the whole curve into itself. (In three dimensions, only circles have this property.)

In conclusion, let us remark that when investigating a random function $X(t)$ it is unnecessary to take into consideration the entire Hilbert space H of random variables

X with a finite variance. One can restrict oneself to the consideration of a narrower Hilbert space H_X (which is a *linear subspace* of the space H) consisting of all possible finite linear combinations $\Sigma_{k=1}^m c_k X(t_k)$ and all the limits in the mean of the sequences of such linear combinations. The norm and scalar product in H_X are defined exactly in the same way as in the whole space H. The space H_X, however, has the advantage that it is directly related to the function $X(t)$ and contains no "extra" elements (because H_X is the smallest linear subspace of H which contains all the elements $X(t)$).

Chapter 2
EXAMPLES OF STATIONARY RANDOM FUNCTIONS. SPECTRAL REPRESENTATIONS

A number of examples of stationary random functions have already been considered in the Introduction and in Sec. 3. Below we give several additional examples of this type; some of them will naturally lead us to the important spectral representation of an arbitrary stationary function in the form of an appropriately modified Fourier integral.

6. Examples of Stationary Random Sequences

Example 1. The simplest example of a stationary random sequence is a sequence $E(t)$, $t = 0, \pm 1, \pm 2, ...$, of (real) independent random variables with identical probability distributions. We assume that $\langle E(t) \rangle = 0$ and $\sigma_E^2 = \langle E^2(t) \rangle = 1$. Then, the correlation function of the sequence $E(t)$ equals

$$(2.1) \qquad B(t) = \langle E(t + \tau)E(t) \rangle = \begin{cases} 1 & \text{for } \tau = 0, \\ 0 & \text{for } \tau \neq 0. \end{cases}$$

In the general case, the elements $E(t)$ of a wide sense stationary sequence with mean value zero and correlation function (2.1) may be neither identically distributed nor mutually independent. It only follows from (2.1) that the

variables $E(t)$ have identical (unit) variances and are
uncorrelated. For this reason, such sequences will be called
*sequences of uncorrelated random variables.** (In this book, $E(t)$,
where t is discrete, will always denote such a sequence, unless
otherwise specified.) From the standpoint of the Hilbert space
H of Sec. 5, a sequence of uncorrelated random variables is
an orthonormal sequence of vectors (i.e. a sequence of
mutually orthogonal vectors of unit length).

Example 2. Let

(2.2) $$X(t) = \frac{E(t) + E(t - 1) + ... + E(t - m + 1)}{\sqrt{m}} .$$

It is clear that $X(t)$ is also a stationary random sequence
with mean value zero. It is easily verified that

(2.3) $$B(\tau) = \langle X(t + \tau)X(t)\rangle = \begin{cases} 1 - \dfrac{|\tau|}{m} & \text{for } |\tau| < m, \\ 0 & \text{for } |\tau| \geqslant m, \end{cases}$$

i.e., as τ increases, $B(\tau)$ falls off linearly from 1 to 0 and
then remains 0.

Example 3. A generalization of the stationary sequence
of Example 2 is given by the sequence

(2.4) $$X(t) = \sum_{k=0}^{h} b_k E(t - k), \qquad t = ..., -2, -1, 0, 1, 2, ...,$$

where b_k, $k = 0, 1, ..., n$, are given numbers (we consider them
real for simplicity). In this case we obviously have

(2.5) $$B(\tau) = \begin{cases} \sum_{k=0}^{n-|\tau|} b_{k+|\tau|} \, b_k & \text{for } |\tau| \leqslant n, \\ 0 & \text{for } |\tau| > n. \end{cases}$$

A sequence $X(t)$ of the form (2.4) is called a *moving*

---

*Such sequences are also sometimes called <u>purely random sequences</u> or <u>discrete</u>
<u>white noises</u> .   Note that in the special case of <u>Gaussian</u> stationary
sequences, lack of correlation implies independence, and the identity of
variances (at zero mean values) implies the identity of probability
distributions.

average sequence of order *n* (or a *sequence of moving averages* of order *n*).   It can be shown that if for a stationary sequence $X(t)$ the equality $B(\tau) = 0$ holds for $|\tau| > n$, but $B(n) \neq 0$, then $X(t)$ can be represented as a moving average sequence of order *n*. [1]

Example 4.   Example 3 can be generalized by assuming that the number *n* in (2.4) is infinite. For the infinite sum

(2.6) $$X(t) = \sum_{k=0}^{\infty} b_k E(t - k)$$

to be convergent, the following inequality must hold:

(2.7) $$\sum_{k=0}^{\infty} |b_k|^2 < \infty$$

(and then the sum (2.7) equals $B(0) = \langle |X(t)|^2 \rangle$).   One   can also assume that the index *k* in (2.6) takes both positive and negative values, i.e.

(2.8) $$X(t) = \sum_{k=-\infty}^{\infty} b_k E(t - k), \quad \text{where} \sum_{k=-\infty}^{\infty} |b_k|^2 < \infty$$

(and the numbers b_k may   even   be   complex).   Then,

(2.9) $$B(\tau) = \langle X(t + \tau)\overline{X(t)} \rangle = \sum_{k=-\infty}^{\infty} b_{k+\tau} \overline{b}_k.$$

Stationary sequences of the form (2.6) or (2.8) are also often called *moving average sequences* (of infinite order). When one wishes to emphasize that one deals specifically with the form (2.6) or (2.8), the term *one-sided* or, respectively, *two-sided moving averages* is used.   It is clear that (2.6) and, especially, (2.8) describe a comparatively wide class of stationary random sequences. [2]

Example 5.   Suppose that

(2.10) $$X(t) = aX(t - 1) + cE(t),$$

where $|a| < 1$ and, for simplicity, we assume *a* and *c* to be real.   According to (2.10) each term of the sequence $X(t)$ is obtained by adding to the preceding term $X(t - 1)$, multiplied by *a*, a term which is independent of all the past values of $X(t)$ and has a fixed variance $\sigma^2 = |c|^2$.   In other words, the function $X(t)$ partially damps out within a unit time (because $|a| < 1$), but at the same time it is subjected to an "external

disturbance", which amounts to the addition of a "random impact" $cE(t)$.   If we specify a sequence $E(t)$ and choose, arbitrarily, the "initial value" $X(t_0) = x_{t_0}$, it is easy to

determine from formula (2.10) the values of $X(t_0 + 1)$, $X(t_0 + 2)$, $X(t_0 + 3)$, etc., successively. It is clear that at first the value x_{t_0} will significantly affect the values of $X(t_0 + n)$, but

as n increases, the role of x_{t_0} will decrease progressively.

This reasoning suggests that there must exist a stationary sequence $X(t)$ satisfying (2.10), and it can be obtained by choosing x_{t_0} arbitrarily and then assuming that $t_0 \to -\infty$.

It is easy to see that

$$X(t_0 + n) = a^n x_{t_0} + c \sum_{k=1}^{n} a^{n-k} E(t_0 + k)$$
$$= a^n x_{t_0} + c \sum_{j=0}^{n-1} a^j E(t_0 + n - j)$$

is the solution of (2.10) with the initial condition $X(t_0) = x_{t_0}$.   Hence it is clear that the stationary solution of this

equation has the form

(2.11) $X(t) = c \sum_{j=0}^{\infty} a^j E(t - j)$, $t = ..., -2, -1, 0, 1, 2, ...,$

i.e. it is a special sequence of one-sided moving averages of infinite order.
       A stationary sequence $X(t)$ that satisfies (2.10) is called a *first-order autoregressive sequence.*   With the aid of (2.9) and (2.11) it is easy to show that its correlation function has the form

(2.12) $B(\tau) = \langle X(t + \tau) X(t) \rangle = C a^{|\tau|}$, $C = \dfrac{|c|^2}{1 - a^2}$.

       Example 6.   The generalization of the sequence of Example 5 is an *m-th order autoregressive sequence* defined as the stationary solution of the difference equation of a form

(2.13) $X(t) + a_1 X(t - 1) + ... + a_m X(t - m) = cE(t)$,

where $a_m \neq 0$. Using (2.13) we can find successively all the values of $X(t)$ with $t > t_0$ if the values of $E(t)$ and m "initial

conditions" $X(t_0) = x_{t_0}$, $X(t_0 - 1) = x_{t_0-1}$, ..., $X(t_0 - m + 1) = x_{t_0-m+1}$

are given. Such a solution of (2.13) will approach stationarity as $t - t_0 \to \infty$ (in particular as $t_0 \to -\infty$ and t is fixed) if all the roots of the algebraic equation

$$(2.14) \qquad 1 + a_1 z + ... + a_m z^m = 0$$

lie *outside* the unit circle, i.e. are larger than 1 in absolute value. Another way of stating the condition for stationarity is that all the roots α_1, α_2, ..., α_m of the equation $z^m + a_1 z^{m-1} + ... + a_m = 0$ must lie *inside* the unit circle, i.e. must be less than 1 in absolute value. It can be shown that (2.13) will then have a unique stationary solution, which is a special sequence of one-sided moving averages constructed from $E(t)$.[3]

Let us assume that we have at our disposal a sequence of "random numbers" $e(1)$, $e(2)$, ..., simulating the realization of a sequence of independent random variables $E(1)$, $E(2)$, ... with $\langle E(t) \rangle = 0$, $\langle E(t)E(s) \rangle = \delta_{t-s}$* and a specified probability distribution. (The numbers $e(1)$, $e(2)$, ... may, for instance, be taken from some "table of random numbers" or calculated on a computer, as is usually done when Monte Carlo methods are used.)[4] Then it is quite easy to obtain the realization $x(t)$ of the autoregressive sequence $X(t)$ with $|\alpha_j| < 1$, $j = 1, ..., m$. For this we simply need to choose some initial conditions

$$x(0) = x_0, x(-1) = x_{-1}, ..., x(-m + 1) = x_{-m+1}$$

and determine successively the values of $x(t)$ with $t > 0$ from the equation $x(t) = -a_1 x(t - 1) - ... - a_m x(t - m) + ce(t)$. Discarding some number of initial terms of the obtained time series $x(t)$, which is sufficient for a practically complete dying out of the "transient process" affected by the choice of the initial conditions (cf. above Sec. 3, p.55), we obtain the desired realization. Some examples of such simulated realizations of autoregressive stationary sequences will be used in Chap. 3.

Representation of the sequence $X(t)$ in the form (2.6)

---

*δ_m is the Kronecker delta, i.e. $\delta_0 = 1$ and $\delta_m = 0$ when m ≠ 0.

can, in particular, be used for evaluation of the correlation function $B(\tau) = \langle X(t + \tau)\overline{X(t)}\rangle$. (Another, simpler method for determining the function $B(\tau)$ for autoregressive sequences will be considered in Sec. 15.) For instance, if $m = 2$ and the roots α_1 and α_2 of the equation $z^2 + a_1 z + a_2 = 0$ are distinct (and both are less than unity in absolute value), then it can be shown that

$$(2.15) \qquad B(\tau) = \frac{|c|^2}{(\alpha_1 - \alpha_2)(1 - \alpha_1\alpha_2)}\left[\frac{\alpha_1^{|\tau|+1}}{1 - \alpha_1^2} - \frac{\alpha_2^{|\tau|+1}}{1 - \alpha_2^2}\right].$$

In the case of multiple roots $\alpha_1 = \alpha_2 = \alpha$ we must only pass to the limit $\alpha_1 \to \alpha_2 = \alpha$ in (2.15) to obtain the formula

$$(2.15a) \qquad B(\tau) = \frac{|c|^2}{(1 - \alpha^2)^2}\left[|\tau| + \frac{1 + \alpha^2}{1 - \alpha^2}\right]\alpha^{|\tau|}.$$

Equation (2.15) is convenient only when the roots α_1 and $\alpha_2 \neq \alpha_1$ are real. If, however, the roots $\alpha_1 = \alpha e^{i\theta}$ and $\alpha_2 = \alpha e^{-i\theta}$ are complex conjugates, then this equation is more conveniently rewritten as

$$(2.15b) \qquad B(\tau) = C\alpha^{|\tau|}\cos(\theta |\tau| - \psi),$$

where

$$C = \frac{|c|^2}{(1 - \alpha^2)(1 - 2\alpha^2\cos2\theta + \alpha^4)^{1/2}\sin\theta},$$

$$\psi = \tan^{-1}\frac{(1 - \alpha^2)\text{ctn}\theta}{1 + \alpha^2}.$$

Hence in this case $B(\tau)$ has the shape of a damped harmonic oscillation.[5]

Note also that since the autoregressive sequence $X(t)$ can be represented in the form (2.6), then $\langle E(t)X(t - k)\rangle = 0$ for $k > 0$.   Hence, multiplying both sides of the complex conjugate equation (2.13) by $E(t)$ and taking mean values of the products, we find that $\langle E(t)X(t)\rangle = c\langle|E(t)|^2\rangle = c$. Let us now multiply all the terms of (2.13) by $X(t - k)$ and form the mean values again.   Then we arrive at the equations

(2.16) $B(k) + a_1 B(k - 1) + ... + a_m B(k - m) = 0$

for $k = 1, 2, ...,$

and

(2.16a) $B(k) + a_1 B(k - 1) + ... + a_m B(k - m) = |c|^2$

for $k = 0$.

Equations (2.16) and (2.16a) are often called the *Yule–Walker equations*; it can readily be verified that the functions (2.12), (2.15), (2.15a), and (2.15b) indeed satisfy such equations.

Example 7. A generalization of the sequences of both Examples 6 and 3 is a stationary sequence $X(t)$ satisfying the difference equation

(2.17) $X(t) + a_1 X(t - 1) + ... + a_m X(t - m)$

$$= c[E(t) + b_1 E(t - 1) + ... + b_n E(t - n)].$$

Such a sequence is usually called an *autoregressive – moving average sequence* (or *ARMA-sequence*).[*] We shall consider only the case where all the roots of (2.14) are larger than unity in absolute value; then sequence $X(t)$ will again be representable as a sum of the form (2.6) with special coefficients b_k.[6] Proceeding from this, one can often easily obtain an explicit equation for $B(\tau)$. (The particular case of $m = n = 1$ and of real $a_1 = \alpha$ and $b_1 = \beta$ less than unity will be considered again in Sec. 15.) Since it is possible to represent $X(t)$ in the form (2.6), it follows again that $\langle E(t)X(t - k) \rangle = 0$ for $k > 0$. Therefore (2.17) implies that the correlation function $B(\tau)$ satisfies the following *generalized Yule – Walker equations*:

$$B(k) + a_1 B(k - 1) + ... + a_m B(k - m) = 0$$
(2.18)

for $k = n + 1, n + 2, ...$.

---

[*]Another term to denote such a sequence is an <u>autoregressive sequence with moving average residuals</u>.

It is easy to indicate many other examples of stationary random sequences $X(t)$. In particular, we can use methods quite similar to those applied in the next section to the case where t is continuous. However, we shall not consider such examples in order to avoid unnecessary repetition.

7. Examples of Stationary Random Processes

In this and the following five sections we consider only random processes $X(t)$ depending on a continuously varying argument t. Unfortunately, in the case of random processes, we do not have a simple analog for a sequence of uncorrelated random variables $E(t)$, $t = ..., -1, 0, 1, ...$. This circumstance complicates the construction of examples of stationary random processes analogous to the examples of stationary sequences discussed in the previous Sec. 6. In fact, such examples of random processes nevertheless do exist and even play an important role in many applications, but they are not so simple as examples in Sec. 6. Therefore, we postpone their consideration until Sec. 12 and meanwhile shall use other methods to construct examples of stationary processes. Namely, we shall discuss a number of processes $X(t)$ specified by simple analytic expressions containing random parameters. We begin with the simplest expression given by the formula $X(t) = X\Phi(t)$, where X is a random variable and $\Phi(t)$ is a numerical function of t (where $-\infty < t < \infty$).

Example 1. Let

(2.19) $X(t) = X\Phi(t),$

where X is a random variable, and $\Phi(t)$ is a complex-valued function of the real argument t. Since $\langle X\Phi(t)\rangle = \Phi(t)\langle X\rangle$, it is clear that if $\Phi(t)$ is not a constant (i.e. the process (2.19) does not reduce to a constant random variable), then $X(t)$ can be a stationary process only when

(2.20) $\langle X\rangle = 0.$

If condition (2.20) holds, then for $X(t)$ to be stationary it is necessary that the function

$$\langle X(t + \tau)\overline{X(t)} \rangle = \Phi(t + \tau)\overline{\Phi(t)}\langle|X|^2\rangle$$

be independent of t, i.e. that $\Phi(t + \tau)\overline{\Phi(t)}$ be independent of t. Thus setting $\tau = 0$, we find that

(2.21) $|\Phi(t)|^2 = r^2 = \text{const}, \quad \Phi(t) = re^{i\varphi(t)}$,

where r is a real number, and $\varphi(t)$ is a real function of t. Substituting (2.21) into the product $\Phi(t + \tau)\overline{\Phi(t)}$, we find that the difference $\varphi(t + \tau) - \varphi(t)$ must be independent of t. If we assume, for simplicity, that the function $\varphi(t)$ is differentiable,* we immediately get

$$\frac{d}{dt}[\varphi(t + \tau) - \varphi(t)] = 0, \text{ i.e. } \varphi'(t + \tau) = \varphi'(t).$$

Since τ is arbitrary, this means that $\varphi'(t) = \omega = \text{const}$, i.e.

(2.22) $\varphi(t) = \omega t + \theta, \quad \Phi(t) = re^{i(\omega t+\theta)}$.

The numerical factor $re^{i\theta}$ can be included in the random variable X (i.e. the product $Xre^{i\theta}$ can be denoted simply by the single symbol X). Thus, a nonconstant random process (2.19) will be stationary if and only if it has the form

(2.23) $X(t) = Xe^{i\omega t}$,

where X is a complex random variable with mean value zero, and ω is a real constant.

The random process (2.23) clearly describes a periodic oscillation of angular frequency ω, with random amplitude and random phase. The real (and also the imaginary) part of each of its realizations is a cosinusoidal oscillation of the form $x^{(i)}(t) = a\cos(\omega t + \varphi)$, where a and φ vary from one realization to another (cf. Fig. 9 in the Introduction). The corresponding correlation function has the form

(2.24) $B(\tau) = \langle X(t + \tau)X(\overline{t}) \rangle = \langle|X|^2\rangle e^{i\omega\tau} = fe^{i\omega\tau}$

---

*It is not hard to see that this condition can, in fact, be weakened considerably. For instance, it is sufficient to assume that the function $\varphi(t)$ is only continuous (i.e. to consider an arbitrary continuous process of the form (2.19)). We shall not, however, pursue this here.

where $f = \langle |X|^2 \rangle > 0$ is the mean value of the square of the oscillation amplitude, which is proportional to the averaged energy (or energy per unit time, i.e. power) of the oscillation.* Note that the correlation function (2.24) does not depend at all on the statistical characteristics of the phase φ of the oscillation $X(t)$.

Example 2.   Further examples of stationary random processes can be obtained by forming linear combinations of processes of the type (2.23). For example, consider the process

$$(2.25) \qquad X(t) = X_1 e^{i\omega_1 t} + X_2 e^{i\omega_2 t},$$

where X_1 and X_2 are random variables with mean zero. It is clear that in this case $\langle X(t) \rangle = 0$ and

$$\langle X(t + \tau)\overline{X(t)} \rangle$$
$$= \langle (X_1 e^{i\omega_1(t+\tau)} + X_2 e^{i\omega_2(t+\tau)})(\overline{X}_1 e^{-i\omega_1 t} + \overline{X}_2 e^{-i\omega_2 t}) \rangle$$
$$= \langle |X_1|^2 \rangle e^{i\omega_1 \tau} + \langle X_1 \overline{X}_2 \rangle e^{i(\omega_1 - \omega_2)t + i\omega_1 \tau}$$
$$+ \langle \overline{X}_1 X_2 \rangle e^{-i(\omega_1 - \omega_2)t + i\omega_2 \tau} + \langle |X_2|^2 \rangle e^{i\omega_2 \tau}.$$

If the process (2.25) is to be stationary, the last expression must be independent of t. Because of the linear independence of the functions $e^{i(\omega_1 - \omega_2)t}$, $e^{-i(\omega_1 - \omega_2)t}$, and 1, this expression, where $\omega_2 \neq \omega_1$, can be independent of t only if

$$(2.26) \qquad \langle X_1 \overline{X}_2 \rangle = \langle \overline{X}_1 X_2 \rangle = 0.$$

Thus, the random process (2.25) is stationary if and only if X_1 and X_2 are uncorrelated complex random variables with mean values zero. Then $X(t)$ is a superposition of two uncorrelated (possibly independent) oscillations of frequencies

---

*If $X(t)$ is the velocity of some mechanical vibration, then f is proportional to its average kinetic energy. If $X(t)$ is a fluctuating current or voltage, then f is proportional to the average electrical power, etc.   Below we shall usually refer to f and similar quantities as the power.

ω_1 and ω_2, with random amplitudes and phases. The correlation function of (2.25) is

$$(2.27) \quad \begin{aligned} B(\tau) &= \langle |X_1|^2 \rangle e^{i\omega_1 \tau} + \langle |X_2|^2 \rangle e^{i\omega_2 \tau} \\ &= f_1 e^{i\omega_1 \tau} + f_2 e^{i\omega_2 \tau}, \end{aligned}$$

where f_1 and f_2 are the average powers of the individual oscillations. Note that $B(\tau)$ is completely independent of the phases of these oscillations.

It should be remarked that the process (2.25) can be real whereas the process (2.23) is always complex. For (2.25) to be real, it is only necessary that $\omega_2 = -\omega_1 = -\omega$, and $X_2 = \bar{X}_1$ (i.e. $X_1 = (U - iV)/2$, $X_2 = (U + iV)/2$). Then (2.25) can be written in the form

$$(2.28) \quad X(t) = U\cos\omega t + V\sin\omega t,$$

where U and V are real random variables. Conditions (2.26) reduce here to conditions

$$(2.29) \quad \langle U^2 \rangle = \langle V^2 \rangle = g, \quad \langle UV \rangle = 0,$$

where $g = 2\langle X_1^2 \rangle = 2\langle X_2^2 \rangle$, and (2.27) takes the form

$$(2.30) \quad B(\tau) = g \cos\omega\tau.$$

Example 3. In the same way, we can construct processes of the form

$$(2.31) \quad X(t) = \sum_{k=1}^{n} X_k e^{i\omega_k t},$$

where either $\langle X_1 \rangle = \langle X_2 \rangle = \dots \langle X_n \rangle = 0$ or $\langle X_2 \rangle = \dots = \langle X_n \rangle = 0$ and $\omega_1 = 0$. Just as in Example 2, it can be shown that the process (2.31) is stationary if and only if

$$(2.32) \quad \langle X_i \bar{X}_j \rangle = 0 \quad \text{for } i \neq j.$$

The correlation function of $X(t)$ has the form

$$(2.33) \quad B(\tau) = \sum_{k=1}^{n} \langle |X_k|^2 \rangle e^{i\omega_k \tau} = \sum_{k=1}^{n} f_k e^{i\omega_k \tau},$$

where $f_k > 0$ for all k. The process (2.31) is a superposition of n periodic oscillations of different frequencies and the coefficients f_k specify the average powers of these oscillations. Setting $\tau = 0$ in (2.33), we obtain

$$(2.34) \qquad B(0) = \langle |X(t)|^2 \rangle = \sum_{k=1}^{n} f_k.$$

This formula shows that in a superposition of uncorrelated periodic oscillations, the average power of the composite oscillation equals the sum of the average powers of its separate harmonic components.

Let us assume that $\omega_k \neq 0$ for all k (so that $\langle X_k \rangle = 0$ for all k and $\langle X(t) \rangle = 0$). Then, for the process (2.31) to be real, the number n must be even (equal to $2m$, say) and the terms of the sum (2.31) must separate into m pairs of complex conjugate terms. In this case, the process $X(t)$ can be rewritten in the form

$$(2.35) \qquad X(t) = \sum_{k=1}^{m} (U_k \cos\omega_k t + V_k \sin\omega_k t),$$

where $\omega_k > 0$ for all k, U_k and V_k are real random variables with $\langle U_k \rangle = \langle V_k \rangle = 0$, and, according to (2.32),

$$\langle U_k V_j \rangle = 0 \text{ for all } k \text{ and } j,$$

$$\langle U_k U_j \rangle = \langle V_k V_j \rangle = 0 \text{ for } k \neq j,$$

$$\langle U_k^2 \rangle = \langle V_k^2 \rangle = g_k = 2f_k, \; k = 1, ..., m$$

(cf. (2.29)). If $X(t)$ is real but the frequencies ω_k include the zero frequency $\omega_1 = 0$ ($\langle X(t) \rangle$ may differ from zero in this case), then (2.35) holds true anyway, and all the other above-mentioned relations are also valid. (The only difference is that now $\omega_k \geqslant 0$, and $\langle U_1 \rangle$ may be nonzero.) The correlation function (2.33), along with the term $f_k e^{i\omega_k \tau}$, where $\omega_k \neq 0$, includes, in the real case, the complex conjugate term $f_k e^{-i\omega_k \tau}$ and can thus be rewritten as

$$(2.36) \qquad B(\tau) = \sum_{k=1}^{m} g_k \cos\omega_k \tau.$$

In formula (2.31) we can also set $n = \infty$. Then, however, for the infinite sum on the right-hand side of (2.31) to be convergent to a limit that is a random variable of finite variance, the following condition must hold:

$$(2.37) \qquad \sum_{k=1}^{\infty} \langle |X_k|^2 \rangle = \sum_{k=1}^{\infty} f_k < \infty.$$

If this is true, then the series

$$(2.38) \qquad X(t) = \sum_{k=1}^{\infty} X_k e^{i\omega_k \tau},$$

where $\langle X_k \rangle = 0$ for all k and $\langle X_k \bar{X}_j \rangle = 0$ for $k \neq j$, will converge (in the usual sense of the mean square convergence). The correlation function $B(\tau)$ of the process (2.38) is

$$(2.39) \qquad B(\tau) = \sum_{k=1}^{\infty} f_k e^{i\omega_k \tau}.$$

In the real case we can write the formulae (2.38) and (2.39) in a form similar to (2.35) and (2.36).

A stationary random process of the form (2.31) or (2.38) is called a *process with a discrete spectrum*, and the set of numbers ω_1, ω_2, ... is called the *spectrum* of the process.[7] From (2.39) it is easy to deduce that in this case the spectrum can always be determined from the correlation function of the process. In fact, the spectrum consists of the numbers ω for which

$$(2.40) \qquad \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} B(\tau) e^{-i\omega\tau} d\tau \neq 0.$$

It is also easy to see that at $\omega = \omega_k$ the limit (2.40) is equal to f_k. Thus, the correlation function $B(\tau)$ also determines the mean values f_k of the squares of the amplitudes X_k of the separate harmonic components $X_k e^{i\omega_k t}$ of the process $X(t)$. In the special case where it is known that a real stationary process $X(t)$ with a discrete spectrum is Gaussian (normal), it can be shown that the real and imaginary part of each of the complex amplitudes X_k have identical normal probability distributions. Thus, all the real amplitudes $|X_k|$ have a so-called *Rayleigh probability distribution* with a probability density of the form $p(x) = 2x f_k^{-1} \exp(-x^2/f_k)$ for $x > 0$, while all the phases $\varphi_k = \arg X_k$ are uniformly distributed from 0 to 2π. Therefore, in the Gaussian case the correlation function

$B(\tau)$ uniquely determines the process $X(t)$.

As the next, slightly more complicated, example, we consider the formula for $X(t)$, which contains an infinite number of random parameters.

Example 4. Suppose $X(t)$ is a *Poisson pulse process* of the form

$$(2.41) \qquad X(t) = \sum_n E_n \Gamma(t - t_n).$$

Here, as in (0.42), $\Gamma(t)$ is a fixed (nonrandom) function, which we assume to be bounded and tending to zero as $t \to \pm \infty$ (rapidly enough to guarantee the convergence of all the integrals and sums encountered below). Moreover, the sequence ... t_{n-1}, t_n, t_{n+1}, ... is a Poisson point system of given intensity λ (i.e. the mean number of points t_n in a time interval of length s is equal to λs), and ..., E_{n-1}, E_n, E_{n+1}, ... is a sequence of independent (of each other and of all the points t_n), identically distributed random variables with mean values $\langle E_n \rangle = m_E$ and variances $\langle (E_n - \langle E_n \rangle)^2 \rangle = \sigma_E^2$. Thus, the random quantities in (2.41) are the variables E_n and t_n.

We begin by calculating the mean value of $X(t)$. Since the random variables E_n and t_n (and, hence, also E_n and $\Gamma(t - t_n)$) are independent, we have

$$\langle X(t) \rangle = \sum_n \langle E_n \rangle \langle \Gamma(t - t_n) \rangle = m_E \langle \sum_n \Gamma(t - t_n) \rangle.$$

Now let $N(t)$, $-\infty < t < \infty$, be a Poisson process corresponding to the point system $\{t_n\}$ (this process is defined on p. 32 of the Introduction and is sketched in Fig. 10b). Then it is easy to see that

$$(2.42) \qquad \sum_n \Gamma(t - t_n) = \int_{-\infty}^{\infty} \Gamma(t - s) dN(s),$$

where the integral on the right-hand side is a Stieltjes integral, defined in the Introduction (see, in particular, (0.7) where, of course, the limit must now be understood as the limit of a sequence of random variables). Since $N(t_2) - N(t_1)$, where $t_2 > t_1$, is equal to the number of points t_n on the interval $[t_1, t_2]$, $\langle N(t_2) - N(t_1) \rangle = \lambda(t_2 - t_1)$ and hence $\langle dN(s) \rangle = \lambda ds$. Thus

$$\langle \sum_n \Gamma(t - t_n)\rangle = \langle \int_{-\infty}^{\infty} \Gamma(t - s)dN(s)\rangle$$

$$= \int_{-\infty}^{\infty} \Gamma(t - s) \langle dN(s)\rangle$$

$$= \lambda \int_{-\infty}^{\infty} \Gamma(t - s)ds = \lambda \int_{-\infty}^{\infty} \Gamma(u)du$$

and

(2.43) $m = \langle X(t)\rangle = \lambda m_E \int_{-\infty}^{\infty} \Gamma(u)du.$

The calculation of the correlation function of the process $X(t)$ (i.e., of $B(\tau) = \langle X(t+\tau)X(t)\rangle$ is similar to the calculation of $m = \langle X(t)\rangle$ but is more complicated.[8] We find that

$$B(\tau) = \langle X(t + \tau)X(t)\rangle$$

(2.44)
$$= \lambda(m_E^2 + \sigma_E^2)\int_{-\infty}^{\infty} \Gamma(u + \tau)\Gamma(u)du$$

$$+ \lambda^2 m_E^2 \left[\int_{-\infty}^{\infty} \Gamma(u)du\right]^2,$$

and therefore

(2.45) $b(\tau) = B(\tau) - m^2 = \lambda(m_E^2 + \sigma_E^2)\int_{-\infty}^{\infty} \Gamma(u + \tau)\Gamma(u)du,$

(2.46) $\sigma_X^2 = b(0) = \lambda(m_E^2 + \sigma_E^2)\int_{-\infty}^{\infty} \Gamma^2(u)du.$

In the case where the pulse amplitude is constant (i.e. $\sigma_E^2 = 0$ and $X(t)$ is a *shot noise process*) the relations (2.43) and (2.46) are known as *Campbell's theorem* (cf. Note 8).

Consider now two particular cases of equations (2.43) and (2.45). If all the pulses are square wave-forms of fixed width T_0 (see Fig. 17a), then it is easy to see that

$$\int_{-\infty}^{\infty} \Gamma(u)du = T_0,$$

$$\int_{-\infty}^{\infty} \Gamma(u + \tau)\Gamma(u)du = \begin{cases} T_0 - |\tau| & \text{for } |\tau| \leqslant T_0, \\ 0 & \text{for } |\tau| > T_0. \end{cases}$$

Thus in this case

$$m = \lambda m_E \, T_0,$$

(2.47)
$$b(\tau) = \begin{cases} \lambda(m_E^2 + \sigma_E^2)(T_0 - |\tau|) & \text{for } |\tau| \leqslant T_0, \\ 0 & \text{for } |\tau| > T_0. \end{cases}$$

The second equation (2.47) implies, in particular, that the "triangular function" $b(\tau) = \max\{T - |\tau|, 0\}$, depicted in Fig. 17(b), is a positive definite function (since otherwise it could not be the correlation function of a stationary process).

If, however, $\Gamma(u) = e^{-\alpha u}$ for $u \geqslant 0$ and $\Gamma(u) = 0$ for $u < 0$ (see Fig. 18(a)), then

$$\int_{-\infty}^{\infty} \Gamma(u)\,du = \frac{1}{\alpha}, \quad \int_{-\infty}^{\infty} \Gamma(u + \tau)\Gamma(u)\,du = \frac{1}{2\alpha}\, e^{-\alpha|\tau|}.$$

Hence in this case

(2.48) $m = \lambda m_E/\alpha, \quad b(\tau) = \lambda(m_E^2 + \sigma_E^2)e^{-\alpha|\tau|}/2\alpha$

(see Fig. 18(b)).   In this case we see that the process $\overset{\circ}{X}(t) = X(t) - \langle X(t)\rangle$ has the first and second moments of the same form as the moments of a random telegraph signal or of a point process with adjoined random variables having a zero mean (see Sec. 2).

The number of such examples can, of course, easily be increased.[9]
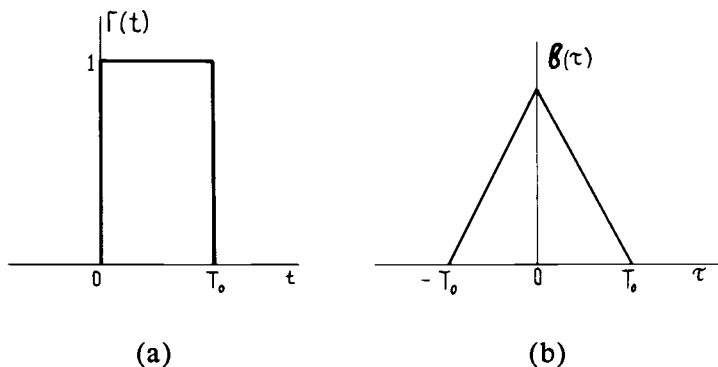

Fig. 17.   (a) Rectangular pulse $\Gamma(t)$. (b) The corresponding correlation function $b(\tau)$.
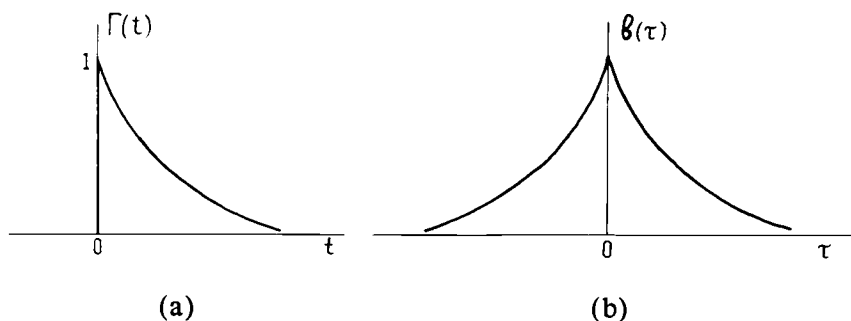
Fig. 18. (a) Exponentially attenuating pulse $\Gamma(t)$. (b) The corresponding correlation function $b(\tau)$.

In conclusion we consider one more example of a stationary random process specified by an analytic formula containing a few random parameters, but differing sharply from the formulae considered in Examples 1–3.

Example 5. Assume that

$$(2.49) \qquad X(t) = Y\, e^{it\Omega},$$

where Y is a random variable (real or complex) such that $\langle Y \rangle = 0$, $\langle |Y|^2 \rangle = a^2$, and Ω is a real random variable independent of Y and characterized by the probability distribution function

$$(2.50) \qquad P\{\Omega < \omega\} = F_\Omega(\omega).$$

Then both the random variables Y and $|Y|^2$ will be independent of the variable $e^{it\Omega}$. Since the mean value of the product of independent random variables is equal to the product of the mean values of the factors,

$$(2.51) \qquad \begin{aligned} \langle X(t) \rangle &= \langle Y \rangle \langle e^{it\Omega} \rangle = 0, \\ \langle X(t+\tau)\overline{X(t)} \rangle &= \langle |Y|^2 \rangle \langle e^{i\tau\Omega} \rangle. \end{aligned}$$

We see that the random process $X(t)$ is wide sense stationary, with mean value zero, while its correlation function can be represented as

(2.52) $B(\tau) = \int_{-\infty}^{\infty} e^{i\tau\omega} \, dF(\omega),$

where $F(\omega) = a^2 F_\Omega(\omega)$ (cf. the footnote following (0.7) in the Introduction). Here, $F_\Omega(\omega)$ is an arbitrary distribution function and a^2 is an arbitrary positive number; therefore $F(\omega)$ is an arbitrary nondecreasing bounded function of ω such that $\lim_{\omega \to -\infty} F(\omega) = 0$. (The last condition is obviously not a restriction, because the integral (2.52) does not change at all on addition of an arbitrary constant to $F(\omega)$.)

The above example shows that any function representable in the form (2.52), where $F(\omega)$ is a real nondecreasing bounded function, may be a correlation function of a stationary random process.[10] It should be observed that all the realizations of the process (2.49) have a special form of complex functions of the form $x(t) = ye^{i\omega t}$. If, however, the probability distribution of the random variable Ω is symmetric about the point $\omega = 0$ (i.e. $P\{a_1 \leqslant \Omega < a_2\} = P\{-a_2 < \Omega \leqslant -a_1\}$ for any a_1 and $a_2 > a_1$, so that, in particular, $dF_\Omega(\omega) = dF_\Omega(-\omega)$), then (2.52) can also be rewritten in the real form:

(2.53) $B(\tau) = \int_{-\infty}^{\infty} \cos\omega\tau \, dF(\omega) = \int_0^{\infty} \cos\omega\tau \, dG(\omega),$

where $dG(\omega) = 2F(\omega)$ for $\omega > 0$. The example of the process $X(t)$ can also be so modified in this case so that all its realizations $x(t)$ become real. For this, it is sufficient to consider, in place of (2.49), the random process

(2.54) $X_1(t) = \sqrt{2} \, Y\cos(t\Omega + \Phi),$

where Y, Ω, and Φ are three independent real random variables. (The variable Y can be replaced, if desirable, by the real number a.) Let $\langle Y^2 \rangle = a^2$, Ω has an arbitrary probability distribution symmetric about the point $\omega = 0$, while Φ is uniformly distributed from 0 to 2π. It is easy to see that in this case $\langle X_1(t) \rangle = 0$ (irrespective of the probability distributions of Y and Ω), and

(2.55) $B(\tau) = \langle X_1(t + \tau)X_1(t) \rangle = a^2 \int_{-\infty}^{\infty} \cos\omega\tau \, dF_\Omega(\omega)$

(where $F_\Omega(\omega)$ is the distribution function of Ω) which

coincides with (2.53).[11] All the realizations of the process (2.54) are obviously sine curves of the form $x(t) = \sqrt{2}y\cos(\omega t + \varphi)$ with a random phase φ, random frequency ω, and random amplitude $\sqrt{2}y$ (but if we take $Y = a$, a constant, then the amplitudes of all the realizations will be identical). Thus, for any real function $B(\tau)$ of the form (2.53) it is possible to construct a stationary random process $X(t)$ having $B(\tau)$ as its correlation function and such that all its realizations are sinusoids.

Note that in Examples 4 and 5 the class of obtained correlation functions is rather wide. Indeed, in Example 4 it depends on the real function $\Gamma(u)$, of which is only required to be square integrable (i.e. that formula (2.46) be meaningful), and in Example 5 it depends on the arbitrary nondecreasing bounded function $F(\omega)$, $-\infty < \omega < \infty$ (or, in the real case, on the arbitrary nondecreasing bounded function $G(\omega)$, $0 \leqslant \omega < \infty$). It is clear that formulae (2.45), (2.52), and (2.53) embrace a great number of particular examples of correlation functions. It is also not hard to show that all the functions (2.45) can be represented in the form (2.53), i.e. that the class of functions of the form (2.45) is a part of the class of functions of the form (2.53).[12] Moreover, it was proved by A.Ya.Khinchin in 1934 that the correlation function of *any* stationary random processes can be represented in the form (2.52) (and, in a real case, also in the form (2.53)), where $F(\omega)$ (or $G(\omega)$) is a real nondecreasing bounded function.[13] This *theorem of Khinchin*, which is sometimes also called the *Wiener–Khinchin theorem* for the reason explained in the next section, plays a very important role in the general theory of stationary random processes, and it will be discussed in detail in Sec. 9.[13a] However, to get a clearer idea of the physical meaning of formulae (2.52) and (2.53), it is convenient to consider first the spectral representation of stationary random processes themselves, although historically this representation was proved only after the appearance of Khinchin's result (and in fact its proof was based on this result).

To conclude this section, we wish to emphasize that, by virtue of Khinchin's theorem, Example 5 shows that for any function $B(\tau)$, which may be a correlation function of a stationary process, there exists a stationary process $X(t)$ having this correlation function and such that all its realizations $x(t)$ have the form of complex harmonics $ye^{i\omega t}$.

If, however, the function $B(\tau)$ is real, then it is also possible to construct $X(t)$ in such a way that all its realizations $x(t)$ will be real sinusoids. It follows from this that there exists a process $X(t)$ with a correlation function of the form $B(\tau) = \sigma^2 e^{-\lambda|\tau|}$, all of whose realizations are sinusoids. This fact has already been mentioned (without proof) in Sec. 2.

8. Spectral Representation of Stationary Random Processes

It is well known that in studying continuous oscillations it is often useful to accomplish a harmonic analysis, i.e. to represent the oscillation under consideration as a Fourier series or integral. It should, however, be borne in mind that representation in the form of a Fourier series is possible only for a very special class of periodic functions. (If we do not require that all the frequencies be multiples of a single "basic frequency" ω_0, then such a representation is also possible for a more general, but still rather restricted class of almost periodic functions.[14]) Moreover, representation in the form of a Fourier integral is possible only for functions which decay to zero rapidly enough at infinity.* At the same time, undamped (i.e. not decaying to zero) but also nonperiodic oscillations appear quite often in many applications and they cannot be represented either as a Fourier series or as a Fourier integral. The inapplicability of the customary techniques of harmonic analysis to such oscillations greatly hinders their study.

In applied work, this difficulty is often obviated in a very simple way: it is nevertheless assumed that Fourier representation of the studied function is still possible, and then conventional formulae are applied, which have been rigorously proved only under some severe restrictions. It is remarkable that despite the obvious mathematical looseness of this approach, it has been applied successfully by a number

---

*Instead of Fourier series and integrals it is also possible to consider more general Fourier-Stieltjes integrals, which we have already encountered (see, for instance, (2.52)) and which play an important role later on in this book. However, an ordinary (nonrandom) function can be represented in the form of the Fourier-Stieltjes integral only in some special cases (usually only when it is the sum of a function decaying to zero at infinity and a periodic or an almost periodic function).

of prominent physicists and has repeatedly yielded correct results. This shows clearly that harmonic analysis has, in fact, a wider region of application than indicated in standard mathematical texts.

In the special mathematical literature one can find several attempts to extend the scope of harmonic analysis to include some classes of undamped (and nonperiodic) functions, but none of them has yielded fully satisfactory results. The typical (and best-known) example of such a theory is given by Wiener's "generalized harmonic analysis", which stemmed directly from the author's attempt to attach a rigorous meaning to the physical concept of a spectrum of an undamped oscillation (see, e.g., Weiner, 1933). Wiener investigated a class of (nonrandom and, generally speaking, complex valued) functions $x(t)$, $-\infty < t < \infty$, for which, at any τ,

$$(2.56) \quad \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} x(t + \tau)\overline{x(t)}dt = B(\tau)$$

exists and is a continuous function of τ.* He proved a number of important mathematical theorems related to such functions. In particular, he showed that in this case $B(\tau)$ is always representable in the form (2.52) or (in the real case) (2.53) where the function $F(\omega)$ is nondecreasing and possesses many properties which agree with the physical concept of the "spectrum of an oscillation $x(t)$". Therefore Wiener named $F(\omega)$ the "spectrum" of the function $x(t)$. However, the "harmonic analysis" of the functions $x(t)$ (i.e. their representation as a superposition of complex harmonics $e^{it\omega}$) has a cumbersome and uncustomary form in Wiener's theory, and also, the relationship of this analysis of $x(t)$ with the "spectrum" $F(\omega)$ is rather complicated.[15]

A considerable further step toward substantiation of harmonic analysis of undamped functions was made in the spectral theory of stationary random processes developed during 1934–1942 in the works of Khinchin, Kolmogorov, and Cramér. Appreciably simpler and more complete results were achieved by passing from consideration of individual function $x(t)$ to simultaneous consideration of a set of such functions with a specified probability distribution, i.e.   to

---

*These conditions can be weakened slightly, but we shall not discuss these refinements here.

consideration of a random process $X(t)$.   The requirement of
the existence of the limit (2.56) was then naturally replaced
by the requirement of wide sense stationarity of $X(t)$.   Let us
now assume that the stationary process $X(t)$ has the property
that the time average

$$\frac{1}{T} \int_0^T x(t + \tau)\overline{x(t)}dt$$

of the product $x(t + \tau)\overline{x(t)}$ (where $x(t)$ is a realization of $X(t)$
and τ is an arbitrary number) converges almost surely (i.e. with
probability one) to the mean value $\langle X(t + \tau)\overline{X(t)} \rangle = B(\tau)$ as $t \to \infty$.
The formulated property is a well-known consequence of the
ergodicity of the process $X(t)$ and it is valid under very general
conditions (see Sec. 17 below). For the sake of brevity we shall
call (in this, and only in this, section) stationary processes $X(t)$
having this property, *stationary ergodic processes.*   Then a
realization $x(t)$ of the stationary ergodic process $X(t)$ will, with
probability one, be such a function that the limit (2.56) exists
and is equal to the correlation function $\langle X(t + \tau)\overline{X(t)} \rangle$ of $X(t)$.
Thus, with reference to the particular class of functions $x(t)$
possessing the time average (2.56) − the class of realizations of
ergodic stationary random processes − Wiener's analytical
theorem on the representability of the function (2.56) in the
form of (2.52) or (2.53) is a simple corollary of Khinchin's
theorem related to the theory of random processes.   The
considered class includes the most important examples of
functions for which the limit (2.56) exists; cf. Note 14 to the
Introduction.) On the other hand, it will be explained in Sec. 17
that stationary random processes encountered in practice are
almost always ergodic; therefore, for them Khinchin's theorem
follows at once from Wiener's theorem.   It is for the last-
mentioned   reason   that   Khinchin's   theorem   on   the
representability of a correlation function of any stationary
random process in the form of (2.52) (or (2.53)) is sometimes
called the Wiener−Khinchin theorem (cf. also Note 13a).
    The main advantage of passing from individual functions
$x(t)$ to stationary random processes $X(t)$ consists in a much
simpler harmonic analysis of such processes.   We have
already seen that in some cases the spectral representation of
the stationary random process $X(t)$ has a very simple form: If
$X(t)$ is a process with a discrete spectrum, then this process
can be represented as a simple sum of uncorrelated harmonic

oscillations, i.e.

$$X(t) = \sum_{k=1}^{\infty} X_k e^{i\omega_k t}.$$

Of course, stationary random processes with a discrete spectrum are far from being the only kind of stationary processes. (This follows, in particular, from the fact that the correlation function $B(\tau)$ of processes with a discrete spectrum has the form (2.39) and never decays to zero at infinity, whereas in practice the correlation functions of stationary processes with mean value zero usually tend to zero as $\tau \to \infty$.) However, it turns out that any stationary random process $X(t)$ can be obtained as the limit of a sequence of processes with discrete spectra, where these latter processes can even be chosen to have the special form (2.31). Namely, it can be shown that given any stationary process $X(t)$, any $\epsilon > 0$ (however small), and any T (however large), there exist complex random variables X_1, X_2, ..., X_n which are pairwise uncorrelated (the number n depends on ϵ and T, of course) and real numbers ω_1, ..., ω_n such that

$$(2.57) \quad \left\langle \left| X(t) - \sum_{k=1}^{n} X_k e^{i\omega_k t} \right|^2 \right\rangle < \epsilon$$

for any t in the interval $-T \leqslant t \leqslant T$.   This also implies that given any stationary process $X(t)$ with mean zero, and any $\epsilon > 0$, $\eta > 0$, and $T > 0$, there exist random variables X_1, X_2, ..., X_n and real numbers ω_1, ω_2, ..., ω_n such that $\langle X_k \rangle = 0$ for all k, $\langle X_k X_l \rangle = 0$ for $k \neq l$ and

$$(2.57a) \quad P\left\{ \left| X(t) - \sum_{k=1}^{n} X_k e^{i\omega_k t} \right| < \epsilon \right\} > 1 - \eta$$

for any t for which $|t| \leqslant T$ (cf. the transition from (1.38) to (1.41) in Sec. 4).

Formulae (2.57) and (2.57a) show that every stationary random process $X(t)$ can be approximated arbitrarily closely by a linear combination of pairwise uncorrelated harmonic oscillations with random amplitudes and phases.   These formulae follow from the general theorem on the spectral representation of stationary random processes, which is apparently the most important of all the theorems on such processes.   In addition to the statement on the existence of sums of harmonic oscillations satisfying the relations (2.57) and (2.57a), the theorem includes some statements on the limiting behavior of these sums, as $\epsilon \to 0$ (or $\epsilon \to 0$ and $\eta \to 0$).

It implies that in order to improve the accuracy and reliability of the approximation of the process $X(t)$ by the sum of harmonic oscillations (i.e. to decrease ϵ and n in (2.57a) or ϵ in (2.57)) and to increase the time during which the approximation holds (i.e. to increase T), we must, in general, increase n, the number of harmonic components, and decrease all the differences $\omega_{k+1} - \omega_k$ between neighboring frequencies. If we make ϵ arbitrarily small in formula (2.57) (or if we make ϵ and n arbitrarily small in formula (2.57a)) and if, at the same time, we make T arbitrarily large, then in the general case the number of frequencies ω_k within any given interval $\Delta\omega$ of ω-axis will become arbitrarily large. It turns out that when this limiting process is performed, the sum ΣX_k of uncorrelated random variables X_k corresponding to frequencies ω_k belonging to $\Delta\omega$ converges to some random variable (depending on $\Delta\omega$, of course), which we denote by $Z(\Delta\omega)$. This fact is the essence of the spectral representation theorem for stationary processes.   It allows us to represent the stationary process $X(t)$ in the form of the following Fourier–Stieltjes integral:

$$(2.58) \qquad X(t) = \int_{-\infty}^{\infty} e^{i\omega t} Z(d\omega).$$

The improper integral (2.58) means the limit

$$(2.59) \qquad \int_{-\infty}^{\infty} e^{i\omega t} Z(d\omega) = \lim_{a\to-\infty, b\to\infty} \int_{a}^{b} e^{i\omega t} Z(d\omega),$$

where

$$(2.60) \qquad \int_{a}^{b} e^{i\omega t} Z(d\omega) = \lim_{\max|\Delta_k\omega|\to 0} \sum_{k} e^{i\omega'_k t} Z(\Delta_k\omega),$$

and, on the right-hand side of (2.60), the summation is over all subintervals $\Delta_k\omega = [\omega_{k-1}, \omega_k]$ appearing in the partition $a = \omega_0 < \omega_1 < \dots < \omega_{n-1} < \omega_n = b$ of the interval $[a,b]$, ω'_k is an arbitrary point in the subinterval $\Delta_k\omega$, and $|\Delta_k\omega|$ is its length.

The function $Z(\Delta\omega)$ associates a random variable with each interval $\Delta\omega$ of the ω-axis, i.e., $Z(\Delta\omega)$ is a *random interval function*.   From the very definition of $Z(\Delta\omega)$, it is clear that if $\langle X(t) \rangle = 0$, i.e. $\langle X_k \rangle = 0$ for all k , $Z(\Delta\omega)$ has the following properties:

(a) $\langle Z(\Delta\omega)\rangle = 0$ for all $\Delta\omega$;

(b) $\langle Z(\Delta_1\omega)\overline{Z(\Delta_2\omega)}\rangle = 0$ if $\Delta_1\omega$ and $\Delta_2\omega$ are disjoint (nonintersecting) intervals;

(c) $Z(\Delta_1\omega) + Z(\Delta_2\omega) = Z(\Delta_1\omega + \Delta_2\omega)$, if $\Delta_1\omega$ and $\Delta_2\omega$ are adjacent intervals defined by the inequalities $\omega_1 < \omega \leqslant \omega_2$ and $\omega_2 < \omega \leqslant \omega_3$, respectively, and $\Delta_1\omega + \Delta_2\omega$ is the combined interval $\omega_1 < \omega \leqslant \omega_3$.

When $\langle X(t)\rangle = m \neq 0$, we need only add the constant m to all the approximating sums of the form $\Sigma X_n e^{i\omega_k t}$ corresponding to the process $X(t) - m$ with mean value zero. Therefore in this case $\langle X_0\rangle = m \neq 0$, where X_0 is the amplitude corresponding to the zero frequency $\omega_0 = 0$, but $\langle X_k\rangle = 0$ for $\omega_k \neq 0$. Consequently, when $\langle X(t)\rangle = m \neq 0$ condition (a) must be replaced by

(a₁) $\langle Z(\Delta\omega)\rangle = 0$ if $\Delta\omega$ does not contain the point $\omega = 0$, and $\langle Z(\Delta\omega)\rangle = m \neq 0$ if $\Delta\omega$ contains the point $\omega = 0$.

In the remainder of this section, however, we shall not consider processes $X(t)$ with $\langle X(t)\rangle \neq 0$, but shall always assume that $\langle X(t)\rangle = 0$.

Instead of the random interval function $Z(\Delta\omega)$ one can also consider the random point function $Z(\omega_1) = Z((-\infty, \omega_1])$, where $(-\infty, \omega_1]$ is the infinite half-line consisting of the points ω' for which $-\infty < \omega' \leqslant \omega_1$.* Then, the integral representation (2.58) can be written in the more usual form

$$(2.61) \qquad X(t) = \int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega),$$

---

*The definition of the function $Z(\Delta\omega)$ clearly makes it meaningful for infinite intervals $\Delta\omega$. Moreover, this function can also be determined for a wide class of point sets $\Delta\omega$ (not necessarily having the form of intervals), i.e. it can be assumed to be a <u>random set function</u>. The condition (c) will evidently transform then to the following general <u>additivity condition</u>:   (c₁) $Z(\Delta_1\omega) + Z(\Delta_2\omega) = Z(\Delta_1\omega + \Delta_2\omega)$ for any disjoint sets $\Delta_1\omega$ and $\Delta_2\omega$, where $\Delta_1\omega + \Delta_2\omega$ is a union of sets $\Delta_1\omega$ and $\Delta_2\omega$.

standing for the same limit as (2.59) and (2.60), which can now be written as

(2.62)
$$\int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega)$$
$$= \lim_{a \to -\infty, b \to \infty} \left\{ \lim_{\max |\omega_k - \omega_{k-1}| \to 0} \sum_{k=1}^{n} e^{i\omega_k' t} [Z(\omega_k) - Z(\omega_{k-1})] \right\}$$

where $\omega_0 = a$, ω_1, ..., ω_{n-1}, $\omega_n = b$ and ω_k' have the same meaning as above (cf. the definition of the ordinary Stieltjes integral on p. 6 of this book). In the new notation conditions (a) and (b) become

(a') $\langle Z(\omega) \rangle = 0$ for all ω;

(b') $\langle [Z(\omega_1 + \Delta\omega_1) - Z(\omega_1)][\overline{Z(\omega_2 + \Delta\omega_2) - Z(\omega_2)}] \rangle = 0$

if the intervals $[\omega_1, \omega_1 + \Delta\omega_1]$ and $[\omega_2, \omega_2 + \Delta\omega_2]$ are disjoint.

Random functions satisfying the condition (b') are often called *random functions with uncorrelated increments*.

When the process $X(t)$ is real, the random interval function $Z(\Delta\omega)$ is such that $Z(\Delta^*\omega) = \overline{Z(\Delta\omega)}$, where $\Delta^*\omega$ denotes the interval symmetric to $\Delta\omega$ with respect to the point $\omega = 0$. Assuming that $Z(\omega) = [Z_1(\omega) - iZ_2(\omega)]/2$ for $\omega \geqslant 0$, (2.61) can be rewritten in this case in a form similar to (2.35):

(2.63) $X(t) = \int_0^{\infty} \cos\omega t \, dZ_1(\omega) + \int_0^{\infty} \sin\omega t \, dZ_2(\omega)$,

where $Z_1(\omega)$ and $Z_2(\omega)$ are real random functions of the variable ω (which takes only non-negative values) such that

(2.64) $\langle [Z_i(\omega_1 + \Delta\omega_1) - Z_i(\omega_1)][Z_j(\omega_2 + \Delta\omega_2) - Z_j(\omega_2)] \rangle = 0$,

if $i \neq j$ or $i = j$ and the intervals $[\omega_1, \omega_1 + \Delta\omega_1]$ and $[\omega_2, \omega_2 + \Delta\omega_2]$ are disjoint, and

(2.65) $\langle [Z_1(\omega + \Delta\omega) - Z_1(\omega)]^2 \rangle = \langle [Z_2(\omega + \Delta\omega) - Z_2(\omega)]^2 \rangle$.

However, the utilization of (2.61) is usually simpler than that of (2.63), even when $X(t)$ is real; therefore, (2.63) will not be used below.

The representation of a stationary random process $X(t)$

in the form of the integral (2.61), where $Z(\omega)$ possesses the properties (a') and (b'), is called the *spectral representation* of $X(t)$. (The same term is applied also to forms (2.58) and (2.63) of the representation (2.61).) The spectral representation describes the harmonic analysis of a general stationary process $X(t)$, i.e. its representation in a form of a superposition of harmonic oscillations. It shows, in particular, that the writing of a fluctuation process (a "noise") as a finite (or infinite) sum (2.31) (or (2.38)) of uncorrelated harmonic oscillations, which is common in applications[16], is not restrictive in the stationary case. Indeed, the spectral representation implies that any stationary process $X(t)$ can be arbitrarily closely approximated by such a sum on any finite segment of the time axis. However, it should also be remembered that the use of Fourier–Stieltjes integrals of the form (2.61), although uncommon for engineers and other workers in applied fields, is in fact not more complicated than the use of sums of the form (2.31) or (2.38). Therefore the use of the above-mentioned approximate sums gives no gain in simplicity over the exact form and is hardly expedient. As for representation of stationary fluctuations in the form of an ordinary Fourier integral, which is also sometimes encountered in applied studies, it is obviously nonrigorous: it will be explained in Sec. 9 that the function $Z(\omega)$ in (2.61) is usually nowhere differentiable, so that the transition from the Fourier–Stieltjes integral (2.61) to the ordinary Fourier integral is impossible.

In the discussion, at the end of Sec. 4 on mean square integrals of random processes $X(t)$ (i.e. mean square limits of the sequences of the corresponding integral sums), it was emphasized that the practical value of such integrals is due primarily to the fact that their realizations coincide, as a rule, with integrals of the corresponding realizations $x(t)$ of the process $X(t)$, which are usually the actual object of investigation. It is clear, however, that this remark no longer applies to the integral on the right-hand side of (2.61): this integral must necessarily be understood only as the mean square limit (2.62). Indeed, the realization $x(t)$ of the stationary random process $X(t)$ is usually a nonrandom function, for which Wiener's requirement of the existence of the limit (2.56) holds, and we have already seen that such functions cannot be expanded either into the Fourier integral

or into the Fourier—Stieltjes integral.* Thus, it is precisely the feasibility of a probabilistic interpretation of the integral on the right-hand side of (2.61) as some mean square limit that made it possible to obtain a simple spectral representation (2.61) for any stationary random process.

It is natural to pose the question of the practical value of the above-indicated spectral representation in the usual situations, when the researcher has only one realization $x(t)$ of the process $X(t)$ at his disposal. It will be shown below that in actuality the expansion (2.61) has a definite physical meaning, which can also be attached to each separate realization $x(t)$. Moreover, it will be explained in Sec. 11 that the usual applications of harmonic analysis of nonrandom functions are extended without any change to stationary random processes, provided that the expansion into harmonics is understood as the representation (2.61). It is this last circumstance, coupled with the fact that the random undamped oscillations encountered in practice are almost always of a statistical nature, that explains why the formal application of Fourier expansions to undamped oscillations, being mathematically nonrigorous, have nevertheless led to absolutely correct results in the past.

The possibility of a spectral representation of an arbitrary stationary random process $X(t)$ was first shown by A.N. Kolmogorov in the early 1940s; his results were formulated in terms of the geometry of the Hilbert space H (see Sec. 5) and were derived by using certain facts from the spectral theory of linear operators. Later, several other authors interpreted and substantiated the same representation in a probabilistic context. Now many different proofs of the spectral representation theorem for stationary processes can be found in the existing extensive literature.[17] We shall not pursue these proofs here; however, in Note 17 two such proofs are discussed and a third proof is briefly outlined at the end of Sec. 11.

---

*Formally this is also manifested in the fact that the realizations $z(\omega)$ of the random functions $Z(\omega)$ are usually extremely irregular functions of an "unbounded variation" for which a Fourier-Stieltjes integral of the type (2.61) simply does not exist.

9. Spectral Representation of the Correlation Function

Recall Khinchin's formula

$$B(\tau) = \int_{-\infty}^{\infty} e^{i\tau\omega} dF(\omega),$$

for the correlation function $B(\tau)$ of a stationary random process $X(t)$ (see (2.52) above). The function $F(\omega)$ in this formula is a bounded monotone nondecreasing function of ω. It is clear that this function is determined by the formula for $B(\tau)$ only to within an arbitrary additional constant, which can be chosen so that $F(-\infty) = 0$.

In the real case, as is well known, $B(-\tau) = B(\tau)$ for all τ and, hence, $B(\tau) = [B(\tau) + B(-\tau)]/2$. Therefore in this case (2.52) can be rewritten in the form

$$B(\tau) = \int_{-\infty}^{\infty} \cos\omega\tau\, dF(\omega) = \int_{0}^{\infty} \cos\omega\tau\, dG(\omega),$$

where $dG(\omega) = dF(\omega) + dF(-\omega)$ for $\omega > 0$ (see (2.53)). Moreover, it is easy to show that the relation $B(\tau) = B(-\tau)$ implies that the increments of $F(\omega)$ on intervals symmetric about the point $\omega = 0$ must coincide. Thus, $dG(\omega) = 2dF(\omega)$ for $\omega > 0$, and if the function $F(\omega)$ is continuous at the point $\omega = 0$, then $G(\omega) = 2F(\omega) + \text{const}$. (However, if $F(\omega)$ is discontinuous at the point $\omega = 0$, then $G(\omega)$ will have the same, and not a two-fold, jump at this point.) It is convenient to assume, when the function $G(\omega)$ is considered, that $G(-0) = 0$ (where $G(-0)$ is the value of $G(\omega)$ just before its possible jump at the point $\omega = 0$).

Note that if $\langle X(t) \rangle = m \neq 0$, then $B(\tau) = b(\tau) + m^2$, where $b(\tau)$ is a centered correlation function, which also permits representation in the form (2.52). Therefore, if $\langle X(t) \rangle = m$, then the function $F(\omega)$ in (2.52) for $B(\tau)$ will necessarily satisfy the relation $F(+0) - F(-0) \geq |m|^2$ (i.e. $\omega = 0$ will be the discontinuity point of $F(\omega)$, and its jump at $\omega = 0$ will not be less than $|m|^2$). Further in this section, we shall again assume that $\langle X(t) \rangle = 0$ (i.e. whenever $\langle X(t) \rangle = m \neq 0$, we shall assume that the value of m has already been subtracted from all the values of the process, so that $B(\tau)$ is actually a centered correlation function).

In practical applications of stationary random processes, the correlation function $B(\tau)$ usually tends to zero as $|\tau| \to \infty$ if $\langle X(t) \rangle = 0$ (cf. (1.36) in Sec. 4). Suppose that the absolute value of $B(\tau)$ falls off so rapidly as $|\tau| \to \infty$, that

(2.66) $\int_{-\infty}^{\infty}|B(\tau)|d\tau < \infty$

(this condition is also usually fulfilled for practical situations). Then the function $B(\tau)$ can be represented as the Fourier integral

(2.67) $B(\tau) = \int_{-\infty}^{\infty} e^{i\tau\omega} f(\omega)d\omega,$

where $f(\omega)$ is a bounded and continuous function of ω. The representation (2.67) is also possible under some other conditions imposed on $B(\tau)$. For instance, (2.66) can be replaced by a less restrictive condition

(2.68) $\int_{-\infty}^{\infty}|B(\tau)|^2 d\tau < \infty,$

but in this case the function $f(\omega)$ will no longer be necessarily continuous and bounded.[18] Formula (2.67) is a particular case of formula (2.52); it indicates that, subject to the condition (2.66) (or (2.68)),

(2. 69) $F(\omega) = \int_{-\infty}^{\omega} f(\omega')d\omega',$

where

(2.70) $f(\omega) = F'(\omega),$

so that $F(\omega)$ is a differentiable function. In the case of a real process $X(t)$, where the function $B(\tau)$ is even, the function $f(\omega)$ will also be even, and the Fourier integral representation (2.67) can be rewritten as

(2.71) $B(\tau) = \int_{0}^{\infty} \cos\tau\omega\, g(\omega)d\omega,$ $g(\omega) = 2f(\omega).$

It is clear that in this case $g(\omega) = G'(\omega)$ and $G(\omega) = \int_{0}^{\omega} g(\omega')d\omega'$, where $G(\omega)$ is a function appearing in (2.53).

The Fourier representation (2.52) or (2.67) of the correlation function $B(\tau)$ is called the *spectral representation of the correlation function*. The function $F(\omega)$ appearing in this representation is called the *spectral distribution function* of the stationary random process $X(t)$. If, however, the relations (2.69) and (2.67) hold, then $f(\omega)$ us called the *spectral*

density (function) of the process $X(t)$).*

By virtue of (2.52)

$$(2.72) \qquad \int_{-\infty}^{\infty} dF(\omega) = B(0) = \langle |X(t)|^2 \rangle,$$

so that

$$(2.73) \qquad \int_{-\infty}^{\infty} dF(\omega) = F(\infty) - F(-\infty) < \infty.$$

We see that the spectral distribution function is necessarily bounded. If there exists a spectral density $f(\omega)$, then obviously

$$(2.74) \qquad \int_{-\infty}^{\infty} f(\omega) d\omega = B(0) < \infty;$$

thus the spectral density is always integrable. Moreover, since $F(\omega)$ is a monotone nondecreasing function, the spectral density $f(\omega) = F'(\omega)$ is everywhere non-negative:

$$(2.75) \qquad f(\omega) \geqslant 0.$$

Conversely, if $f(\omega)$ is an integrable non-negative function, the function (2.69) will obviously be bounded and monotone nondecreasing, i.e. the function (2.67) will satisfy the conditions of Khinchin's theorem. Thus, Khinchin's theorem acquires a particularly simple form when applied to rapidly falling off functions $B(\tau)$ representable in the form of the

---

*The terminology relating to spectral representations has not yet been finally established. For instance, the terms "spectral distribution function" and "spectral density (function)" are often used in relation to the function $G(\omega)$ and $g(\omega) = 2f(\omega)$ of the non-negative frequency ω. Moreover, the Fourier expansions in harmonics $e^{i\omega\tau}$ or $\cos\omega\tau$, where ω is the angular frequency (measured in "radians per second"), is often replaced by expansions in functions $e^{2\pi in\tau}$ or $\cos2\pi n\tau$, where n is the ordinary frequency measured in Hertz (i.e. cycles per second). In this case the spectral density $f(\omega)$ must be replaced by the density $f_1(n) = 2\pi f(\omega/2\pi)$, which is 2π times higher. Note also that the term "power spectrum", or simply "spectrum", is widely used in all the applied literature instead of the term "spectral density" preferred by mathematicians. (The term "spectrum" has, in mathematics, another meaning, which will be explained below in this section.) As to the spectral distribution function, it is rarely used in the applied literature, but if used it is generally called the "integrated spectrum".

Fourier integral (e.g. to functions satisfying the condition (2.66) or (2.68)): *a rapidly falling off function* B(τ) *will be a correlation function if, and only if, its Fourier transform* f(ω^2) *is everywhere non-negative.*

So far, we have not yet explained in this section how one proves Khinchin's theorem, which is so important for the entire content of the section; but this is discussed in Note 13, referring to Sec. 7. As indicated in that Note, Khinchin's theorem is a simple consequence of the following two statements, taken together:

(a) The class of functions $B(\tau)$, which are correlation functions of stationary random processes, coincides with the class of positive definite functions of the variable τ (see above, Sec. 4 for a real case and Sec. 5 for a complex case).

(b) A continuous function $B(\tau)$ of the real variable τ is positive definite if, and only if, it can be represented in the form (2.52), where $F(\omega)$ is bounded and nondecreasing (this statement was proved independently by Bochner and Khinchin, but was first published by Bochner and therefore is known as *Bochner's theorem*; see, e.g., Bochner (1959) and also Note 3 to Introduction).

In the preceding section it was emphasized that Khinchin's theorem lies at the basis of almost all the proofs of the spectral representation theorem for stationary random processes. It is, however, obvious that if we proved the spectral representation theorem without using Khinchin's theorem, this would also clearly imply the possibility of representing $B(\tau)$ in the form (2.52).   Indeed, replacing $X(t + \tau)$ and $X(t)$ in the formula $B(\tau) = \langle X(t + \tau)X(t) \rangle$ by their spectral representation (2.61) and then using the definition (2.62) of the corresponding Fourier–Stieltjes integral and the property (b') of the random function $Z(\omega)$, we obtain at once (2.52), where

$$(2.76) \qquad F(\omega + \Delta\omega) - F(\omega) = \langle |Z(\omega + \Delta\omega) - Z(\omega)|^2 \rangle$$

so that $F(\omega)$ is clearly a nondecreasing function.   Formula (2.76) can also be written in the differential form:

(2.77) $\langle |dZ(\omega)|^2 \rangle = dF(\omega).$

Moreover, (2.77) can be combined with the property (b') of $Z(\omega)$ in the form of a single symbolic relation

(2.78) $\langle dZ(\omega)\overline{dZ(\omega')} \rangle = \delta(\omega - \omega')dF(\omega)d\omega',$

where $\delta(\omega)$ is the Dirac δ-function. It is easy to see that the substitution of (2.78) into the expression for the mean value of any double integral with respect to $dZ(\omega)$ and $dZ(\omega')$ gives the correct result. As the simplest example we consider the following derivation of Khinchin's formula (2.52):

$$\langle X(t + \tau)\overline{X(t)} \rangle = \langle \int_{-\infty}^{\infty} e^{i(t+\tau)\omega}dZ(\omega) \overline{\int_{-\infty}^{\infty} e^{it\omega'} dZ(\omega')} \rangle$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i[(t+\tau)\omega-t\omega']}\langle dZ(\omega)\overline{dZ(\omega')} \rangle$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i[(t+\tau)\omega-t\omega']} \delta(\omega - \omega')dF(\omega)d\omega'$$

$$= \int_{-\infty}^{\infty} e^{i\tau\omega}dF(\omega).$$

Quite similarly, the following more general result can be derived:

(2.79) $\langle \int_{-\infty}^{\infty} g(\omega)dZ(\omega) \overline{\int_{-\infty}^{\infty} h(\omega)dZ(\omega)} \rangle = \int_{-\infty}^{\infty} g(\omega)\overline{h(\omega)}dF(\omega)$

where $g(\omega)$ and $h(\omega)$ are any two complex functions whose squared absolute values are integrable with respect to $dF(\omega)$. Note also that if the spectral density $f(\omega)$ exists, then the relations (2.77) and (2.78) obviously take the form

(2.80) $\langle |dZ(\omega)|^2 \rangle = f(\omega)d\omega,$

(2.81) $\langle dZ(\omega)dZ(\overline{\omega'}) \rangle = \delta(\omega - \omega')f(\omega)d\omega d\omega'.$

Formulae (2.76)–(2.78) and (2.80)–(2.81) establish the

relationship between the spectral representation of the correlation function (determined by the functions $F(\omega)$ and $f(\omega)$) and the spectral representation of the stationary random process $X(t)$ itself, which includes the random point function $Z(\omega)$ or the random interval function $Z(\Delta\omega) = Z(\omega_2) - Z(\omega_1)$, where $\Delta\omega = [\omega_1, \omega_2]$. We shall see in Sec. 11 that this relationship gives physical meaning to Khinchin's mathematical theorem and permits one to verify it experimentally when the stationary process $X(t)$ is realized in the form of oscillations of some measurable physical quantity X. Now, however, we shall consider some more special questions bearing on the indicated relationship.

We begin with the particular case where the function $F(\omega)$ is a "step-function", which changes (namely, increases by a positive value) only at discrete discontinuity points but takes a constant value between any two discontinuities (cf. Figs. 1(a) and 1(b)). In this case (2.76) and (2.77) imply that the random interval function $Z(\Delta\omega)$ is concentrated entirely on a discrete set of discontinuity points ω_1, ω_2, ... of the function $F(\omega)$, i.e. $Z(\Delta\omega) = 0$ for all intervals $\Delta\omega$ that do not contain any of these points. It is clear that then the spectral representation (2.58) reduces to the representation (2.31) (or (2.38)) of the process $X(t)$ as a superposition of separate uncorrelated harmonic oscillations with random amplitudes and phases, i.e. $X(t)$ will be a process with a discrete spectrum in the sense explained in Sec. 7. Moreover, (2.52) shows that the corresponding correlation function $B(\tau)$ has the form (2.33) or, respectively, (2.39). We see that the representability of the correlation function $B(\tau)$ of the process $X(t)$ in the form of a sum of harmonics (or, in the real case, of cosine functions) implies that the corresponding random process $X(t)$ is a sum of uncorrelated random harmonic oscillations. This particular case of the general spectral representation theorem is due to Slutsky (see Note 7).

In the general case the nondecreasing function $F(\omega)$ can be represented as a sum of a continuous function $F_I(\omega)$ and a "step-function" $F_{II}(\omega)$ constructed from the "jumps" (discontinuities) of $F(\omega)$. By virtue of (2.76) and (2.77), to each discontinuity point of $F(\omega)$ corresponds the discontinuity point of the random function $Z(\omega)$, where it changes jumpwise by a random variable X_k such that $\langle |X_k|^2 \rangle = f_k$ (here f_k is the increment of the function $F(\omega)$

at the corresponding discontinuity point).   The jumps of $Z(\omega)$ generate discrete terms of the form $X_k e^{i\omega_k t}$ on the right-hand side of (2.61); similarly, the jumps of $F(\omega)$ generate discrete terms of the form $f_k e^{i\omega_k t}$ on the right-hand side of (2.52).   Thus, representation of $F(\omega)$ as a sum $F_{\mathrm{I}}(\omega) + F_{\mathrm{II}}(\omega)$ implies representation of the correlation function $B(\tau)$ and of the process $X(t)$ in the form

(2.82) $B(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau} dF_{\mathrm{I}}(\omega) + \sum_k f_k e^{i\omega_k\tau}$,

and

(2.83) $X(t) = \int_{-\infty}^{\infty} e^{i\omega t} dZ_{\mathrm{I}}(\omega) + \sum_k X_k e^{i\omega_k t}$,

where $F_{\mathrm{I}}(\omega)$ is a continuous function of ω and to each term of the sum on the right-hand side of (2.82) there corresponds a term on the right-hand side of (2.83).   It is clear that if the sum on the right-hand side of (2.82) is different from zero, then $B(\tau)$ contains a periodic component and therefore does not tend to zero as $|\tau| \to \infty$. Since in practice one can usually be sure that $B(\tau) \to 0$ as $|\tau| \to \infty$ (i.e. that the correlation between $X(t)$ and $X(t + \tau)$ vanishes as $\tau \to \infty$), in most of the applied problems the sums on the right-hand side of (2.82) and (2.83) are missing.   Note also that in principle the first term on the right-hand side of (2.82) can also represent a function which does not go to zero at infinity; this is possible, however, only in some exceptional cases that are never encountered in practice.[19] Therefore, from the standpoint of applications it is safe to assume that the first term on the right-hand side of (2.82) always describes the component $B_{\mathrm{I}}(\tau)$ of the function $B(\tau)$, which tends to zero as $|\tau| \to \infty$ and is representable in the form

$$B_{\mathrm{I}}(\tau) = \int_{-\infty}^{\infty} e^{i\tau\omega} f(\omega) d\omega,$$

where $f(\omega) = F_{\mathrm{I}}'(\omega) = F'(\omega)$ and

$$F_{\mathrm{I}}(\omega) = \int_{-\infty}^{\omega} f(\omega') d\omega'.$$

The representation (2.82) explains the dual nature of the spectrum of the stationary process $X(t)$. By the *spectrum* of the process $X(t)$ is usually meant the set of frequencies of all the harmonic components of $X(t)$, i.e. the set of all frequencies ω which contribute to the left-hand side of (2.72). The following definition makes this concept precise: *the point ω is said to belong to the spectrum of the process X(t) if $F(\omega + \epsilon) - F(\omega - \epsilon) > 0$ for any $\epsilon > 0$, where $F(\omega)$ is the spectral distribution function of X(t)*. If the spectral density $f(\omega)$ exists, the spectrum obviously includes all the points ω where $f(\omega) > 0$ and all the limits of sequences of such points (i.e. it consists of all those points ω which have no vicinity where the spectral density identically vanishes). In the case where $B(\tau)$ has the form (2.33) or (2.39), the spectrum consists of all the corresponding frequencies ω_k (this case has already been considered in Sec. 7). In the more general case, where $F(\omega)$ is a sum of an indefinite integral $F_I(\omega)$ of the function $f(\omega) = F'(\omega)$ and of a step-function $F_{II}(\omega)$, the spectrum is a union of all the points ω where $f(\omega) > 0$ and all the limits of such points (the set of all these points is called *continuous spectrum* of $X(t)$) and of all the discontinuity points of $F(\omega)$ (these points constitute the *discrete spectrum* of $X(t)$).[20]

Let us now make one more remark bearing on the relations (2.80)–(2.81). Since correlation functions $B(\tau)$ encountered in practice almost always fall off rapidly at infinity and have a Fourier transform, the use of the spectral densities enables one, in most applications, to do without Stieltjes integrals as long as one deals only with the spectral representation of the correlation functions.* In the preceding section, however, we already emphasized that if the spectral representation of the stationary process $X(t)$ itself is also of interest, then the use of Stieltjes integrals is unavoidable: the random function $Z(\omega)$ is not differentiable in any reasonable sense, and there is no possibility of replacing the Fourier–Stieltjes integral in (2.61) by the Fourier integral of ordinary type. It is now easy to see why the function $Z(\omega)$ cannot be differentiable. Indeed, in the most usual cases where a positive spectral density $f(\omega)$ of the process $X(t)$ exists, (2.80) implies that the

*It is easy to see that also when studying a more general case, where $F(\omega)$ is a sum of an indefinite integral of $f(\omega)$ and a step-function $F_{II}(\omega)$, the use of Stieltjes integrals is in fact not necessary.

mean square of the increment $\Delta Z(\omega) = Z(\omega + \Delta\omega) - Z(\omega)$ of the function $Z(\omega)$, where $\Delta\omega$ is small, is close to $f(\omega)\Delta\omega$, i.e. it is of the same order of magnitude as $\Delta\omega$. Therefore, $\Delta Z(\omega)$ itself is, as a rule, of the order of $(\Delta\omega)^{1/2}$, which is incompatible with the assumption of the differentiability of the function $Z(\Delta\omega)$ (i.e. of the existence of a limit of the ratio $\Delta Z(\omega)/\Delta\omega$ as $\Delta\omega \to 0$). This reasoning refers, generally speaking, to any point ω of the frequency axis. Thus, here we deal with a rather rare case where in a problem with a real physical meaning, there arise nowhere differentiable functions, i.e. functions which many applied scientists quite recently considered a far-fetched mathematical abstraction, without any possible applications.[21]

Formulae (2.52) and (2.67) show that the knowledge of the spectral distribution function $F(\omega)$ or the spectral density $f(\omega)$ of the stationary process $X(t)$ permits one to determine easily its correlation function $B(\tau)$ as well. Conversely, if the correlation function $B(\tau)$ is known, then the spectral distribution function $F(\omega)$ and the spectral density $f(\omega)$ can also be obtained. The situation is particularly simple when the condition (2.66) (or (2.68)) is fulfilled, and therefore a spectral density $f(\omega)$ exists. According to (2.67), $B(\tau)$ is a Fourier transform of the function $f(\omega)$. Hence the spectral density $f(\omega)$ can be evaluated by the known formula for the inversion of a Fourier integral:

$$(2.84) \qquad f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega\tau} B(\tau) d\tau.$$

Then we can use (2.69) to find the spectral distribution function $F(\omega)$ as well. According to (2.69) and (2.84), the difference $F(\omega_2) - F(\omega_1) = \int_{\omega_1}^{\omega_2} f(\omega) d\omega$, where $\omega_2 > \omega_1$, equals

$$(2.85) \qquad F(\omega_2) - F(\omega_1) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-i\omega_2\tau} - e^{-i\omega_1\tau}}{- i\tau} B(\tau) d\tau.$$

In the general case, where the spectral distribution function $F(\omega)$ is not representable as the indefinite integral (2.69), i.e. the spectral density does not exist, (2.84) has no meaning. It turns out, nevertheless, that (2.85) is valid in the general case, provided only that the integral from $-\infty$ to ∞ is understood as

$\lim\limits_{T\to\infty} \int_{-T}^{T}$, and the values $F(\omega_2)$ and $F(\omega_1)$ are, when we deal
with the discontinuity point of $F(\omega)$, replaced by the half-sum
of the corresponding "pre-jump" and "post-jump" values. This
follows at once from the known "inversion formula for
characteristic functions" (see Note 3 to the Introduction) by
virtue of Khinchin's theorem, implying that the class of
characteristic functions of probability distributions coincides
with that of normalized correlation functions $R(\tau) =
B(\tau)/B(0)$. (However, we shall never use the general formula
(2.85) in this book.)

In the real case, where formula (2.71) is applicable, the
formula for $g(\omega) = 2f(\omega)$ implied by (2.84) takes the form

$$(2.86) \qquad g(\omega) = \frac{2}{\pi} \int_0^\infty \cos\omega\tau B(\tau)d\tau.$$

Similarly, the formula for the corresponding bounded
nondecreasing function $G(\omega)$ can be written as

$$(2.87) \qquad G(\omega) = \frac{2}{\pi} \int_0^\infty \frac{\sin\omega\tau}{\tau} B(\tau)d\tau.$$

Below, even in the case of real random processes, we shall
usually employ the complex form of the spectral
representation (2.52) or (2.67), i.e. we shall use functions $F(\omega)$
and $f(\omega)$, rather than $G(\omega)$ and $g(\omega)$.

Since the specification of the correlation function $B(\tau)$ is
equivalent to the specification of the corresponding spectral
distribution function $F(\omega)$ or (if (2.69) holds) of the spectral
density $f(\omega)$, it is obvious that any expression involving the
function $B(\tau)$ can also be rewritten in terms of $F(\omega)$ or
(subject to the condition (2.68)) $f(\omega)$.   Let us consider now a
few specific examples which will be of use later in this book.
(The process $X(t)$ and its correlation function $B(\tau)$ will be
assumed real in all of these examples.) We shall begin with

$$(2.88) \qquad T_1 = \frac{1}{B(0)} \int_0^\infty B(\tau)d\tau = \int_0^\infty R(\tau)d\tau$$

which is a quantity with the dimension of time (i.e. a time
scale) determined by the normalized correlation function of
$X(t)$. The quantity T_1, if finite, characterizes the time during
which appreciable correlation between $X(t)$ and $X(t + \tau)$ is

preserved (or, what is the same, the time needed for any correlation between $X(t)$ and $X(t + \tau)$ to die out). Therefore, T_1 is often called the *correlation time* (or the *integral time scale*) of a stationary process $X(t)$. It is clear that $T_1 = \infty$ if the function $B(\tau)$ does not go to zero as $\tau \to \infty$. If, however, $B(\tau)$ tends to zero as $\tau \to \infty$ and has a bounded Fourier transform $f(\omega)$, then, obviously,

$$(2.89) \qquad T_1 = -\frac{\pi f(0)}{B(0)} = \frac{\pi f(0)}{2\int_0^\infty f(\omega)d\omega} \; .$$

The problem of spectral interpretation of the quantity

$$(2.90) \qquad \beta_0 = \lim_{T\to\infty} \frac{1}{T} \int_0^T B(\tau)d\tau$$

is more complicated, but it deserves attention since this quantity will play an important part in Sec. 16. It is easy to see that $\beta_0 = 0$ if $B(\tau) \to 0$ as $\tau \to \infty$; in particular, $\beta_0 = 0$ in all the cases where the spectral density $f(\omega)$ does exist.   If the discrete spectrum of $X(t)$ is not missing, i.e. $B(\tau)$ includes a component of the form $\sum_k g_k \cos\omega_k \tau$, then all the terms with

$\omega_k \neq 0$ will contribute nothing to β_0 since $T^{-1}\int_0^T \cos\omega_k \tau d\tau \to 0$ as

$T \to \infty$.   However, if the above-mentioned component of $B(\tau)$ includes a constant term g_0 (i.e. the discrete spectrum of $X(t)$ includes the point $\omega_0 = 0$), then this term will give a non-zero contribution to β_0 equal to g_0.   It therefore seems natural to expect that in all cases

$$(2.91) \qquad \lim_{T\to\infty} \frac{1}{T} \int_0^T B(\tau)d\tau = \beta_0 = \Delta F(0)$$

where $\Delta F(0) = F(+0) - F(-0)$ is the jump of the spectral distribution function $F(\omega)$ at the point $\omega = 0$. This is indeed so, and (2.91) can easily be proved quite rigorously.[22]

The last expression to be considered here is

$$(2.92) \qquad \beta_1 = \lim_{T\to\infty} \frac{1}{T}\int_0^T B^2(\tau)d\tau.$$

If $B(\tau) \to 0$ as $\tau \to \infty$, then, obviously, $\beta_1 = 0$; in particular, $\beta_1 = 0$ if there exists a spectral density $f(\omega)$. If, however, $X(t)$ is a process with a discrete spectrum, then $B(\tau) = g_0 + \sum_k g_k \cos\omega_k \tau$, where ω_k

> 0 for all k. In this case

$$\beta_1 = \lim_{T\to\infty} \frac{1}{T}\int_0^T [g_0^2 + 2\sum_k g_0 g_k \cos\omega_k\tau + \sum_k \sum_\ell g_k g_\ell \cos\omega_k\tau\cos\omega_\ell\tau]d\tau$$

$$(2.93) \qquad = g_0^2 + \sum_k g_k^2/2.$$

It can be shown that (2.93) holds also in the general case of an arbitrary stationary process $X(t)$ since the component $B_I(\tau)$

$= \int_0^\infty e^{i\tau\omega}dF_I(\omega)$ of $B(\tau)$ (see (2.82)) always contributes nothing

to the value of β_1. Therefore, β_1 is always determined by the discrete component

$$B_{II}(\tau) = \sum_j f_j e^{i\omega_j\tau} = g_0 + \sum_k g_k \cos\omega_k\tau$$

(i.e. by the discrete harmonic oscillations contained in $X(t)$).[23]

10.   Examples of Correlation Functions of Stationary Processes

In Secs. 2 and 7 we have already met with a few specific examples of the correlation functions $B(\tau)$ of stationary processes and in Sec. 4 we have listed a number of properties of such functions which enable one to construct many new examples from the available stock of functions $B(\tau)$. Now, however, we have at hand a very simple general method for constructing examples of correlation functions. This method is based on the version of Khinchin's theorem, related to rapidly falling off functions: According to this theorem, it is sufficient to choose any everywhere non-negative and integrable function $f(\omega)$, and then its Fourier transform $B(\tau)$ will necessarily be a correlation function. Similarly, the verification of whether or not the integrable (or only having an integrable square) function $B(\tau)$ belongs to the class of correlation functions is rather simple: it is only necessary to evaluate its Fourier transform $f(\omega)$ by (2.84) and to verify whether $f(\omega) \geq 0$ for all ω (then $B(\tau)$ belongs to the class of correlation functions) or whether there exist values of ω where $f(\omega) < 0$ (then it does not). We also recall that the real correlation function $B(\tau)$ must necessarily satisfy the conditions $B(0) > 0$, $B(-\tau) = B(\tau)$, and $|B(\tau)| \leq B(0)$ (see

(1.28)). Therefore, in constructing examples of correlation functions $B(\tau)$, functions for which these conditions are not fulfilled are of no interest.

Example 1. *Exponential correlation function.* We begin with consideration of a function

$$(2.94) \qquad B(\tau) = Ce^{-\alpha|\tau|},$$

where $C > 0$ and $\alpha > 0$. We already know that this function is the correlation function of a stationary random process, because in Secs. 2 and 7 we have encountered several examples of processes $X(t)$ having such a correlation function (cf. also Note 9 to this chapter). Let us now verify that criterion (2.75) leads to the same conclusion. Simultaneously we shall also determine the corresponding spectral density $f(\omega)$, which is often of considerable interest.

According to formula (2.84), in the present case

$$f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Ce^{-\alpha|\tau|-i\omega\tau}d\tau$$

$$= \frac{C}{2\pi} \left\{ \int_{-\infty}^{0} e^{(\alpha-i\omega)\tau}d\tau + \int_{0}^{\infty} e^{-(\alpha+i\omega)\tau}d\tau \right\}$$

$$= \frac{C}{2\pi} \left\{ \frac{1}{\alpha - i\omega} + \frac{1}{\alpha + i\omega} \right\} = \frac{C\alpha}{\pi} \frac{1}{\alpha^2 + \omega^2}.$$

Thus we have

$$(2.95) \qquad f(\omega) = \frac{A}{\alpha^2 + \omega^2}, \quad A = \frac{C\alpha}{\pi},$$

so that $f(\omega) > 0$ for all ω. Hence (2.94) is in fact a correlation function for any $C > 0$ and $\alpha > 0$.

The graph of this function and that of the relevant spectral density (2.95) are given in Fig. 19. We note that the curve (2.94) falls off rapidly (in fact, exponentially) with $|\tau|$ and can already be considered virtually zero when the distance from the origin is only a few multiples of α^{-1}. This agrees with the fact that the correlation time T_1 corresponding to the function (2.94) is equal to $1/\alpha$, according to both (2.88) and (2.89). Thus, the constant α^{-1} (with the dimension of time) characterizes the typical time of decay of the correlation between the quantities $X(t)$ and $X(t + \tau)$.
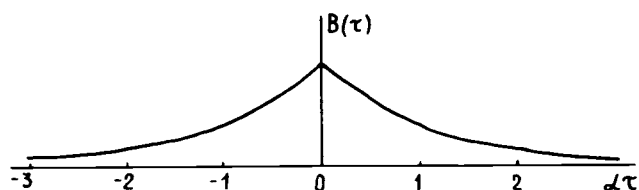
(a)
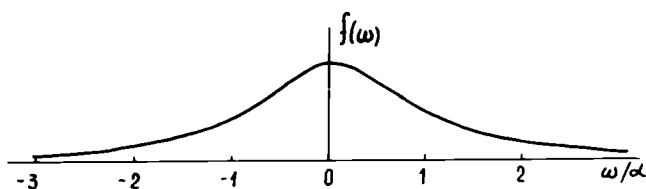
(b)

Fig. 19. (a) Graph of the correlation function (2.94).
          (b) Graph of the spectral density (2.95).

Therefore, in order to estimate the parameter α we only need to
know how long appreciable correlation is preserved between
$X(t)$ and the initial value $X(0)$. The spectral density $f(\omega)$ depicted
in Fig. 23(b) has a maximum at zero, remains almost constant
when ω is small compared to α, and then falls off rather slowly
(according to the power law with exponent -2). The value of
$f(\omega)$ is twice as small as its value at $\omega = 0$ when $\omega = \alpha$, five times
smaller when $\omega = 2\alpha$, ten times smaller when $\omega = 3\alpha$, and so on.
Thus, by increasing α we lengthen the interval in which $f(\omega)$ is
"approximately constant" and slow down the rate at which $f(\omega)$
falls off with ω, but simultaneously raise the rate at which $B(\tau)$
falls off with τ.   The fact that the increase of α leads to an
acceleration of the decrease in $B(\tau)$ with τ, but to the opposite
trend in the function $f(\omega)$ of ω, is associated with the known
"uncertainty principle" of function theory. According to this
principle the "effective width" of the function going to zero at
infinity (i.e. the length of the interval in which this function
differs appreciably from zero) is inversely proportional to the
"effective width" of its Fourier transform.[24]
   In addition to the above-discussed models, in Sec. 12 (see pp.
172–173) we shall encounter one more example of practical
importance of a random process $X(t)$, which has a correlation

function of the form (2.94). We also note that because of the great simplicity of (2.94) it is often used in applied studies even when the precise form of $B(\tau)$ is unknown or is found to be too complicated and inconvenient for practical use.

Example 1a. *White noise.* Of very great practical importance is the case where the correlation between the quantities $X(t)$ and $X(t + \tau)$ falls off extremely rapidly with τ, so that in using (2.94) one has to assume α to be very large. For example, consider the Brownian motion of a small particle immersed in a fluid, and let $X(t)$ be the fluctuation of the force acting on the particle. Variations of $X(t)$ depend on the collisions of molecules of the fluid with the particle. Since under normal conditions (at room temperature and with a real particle size) the particle in the fluid experiences about 10^{21} collisions per second, the values of $X(t)$ and $X(t + \tau)$ are practically independent for times τ of order 10^{-18} sec, say. Hence, if we approximate the correlation function of the process $X(t)$ by a function of the form (2.94), we must assume that α is considerably larger than 10^{18} $\sec^{-1}$. A similar sort of behavior is manifested by the fluctuations of the electromotive force in a conductor due to thermal motion of the electrons, or by the fluctuations of current flow in a vacuum tube due to the fluctuations in the number of electrons going from cathode to anode. In all these cases, if we use a correlation function of the form (2.94) to describe the phenomenon, we must choose α to be extremely large.

As already remarked, if ω is small compared to α , the spectral density $f(\omega)$ is practically constant, and equals $f(0) = f_0$. Consequently, if α is very large, $f(\omega)$ is constant over a very wide frequency range. In practice, however, we are almost never interested in all possible values of ω. In fact, in most practical problems, only a certain restricted frequency range is of any importance (this will be explained in more detail in the next section). Thus, if the spectral density is virtually constant over this frequency range, then in solving the given problem we can in general neglect the variation of $f(\omega)$ and assume that

(2.96) $f(\omega) = f_0 = \text{const.}$

It is clear that, strictly speaking, the spectral density (2.96) can

never exist, for otherwise $\langle X^2(t)\rangle = \int_{-\infty}^{\infty} f(\omega)d\omega$, the total (average)

power of the process $X(t)$, would be infinite (recall the condition (2.73) on p. 105).   Nevertheless, the concept of a stationary random process with a constant spectral density turns out to be a very useful mathematical idealization (similar to the concept of a point mass in mechanics, which contradicts the physically obvious fact that the density of any real object must be finite). The idealized random process with a constant spectral density is often called *white noise*, or it is said that it has a *white spectrum* (by analogy with white light in optics, which has a very wide optical spectrum).   The value of this idealized model of the "generalized random process" is primarily due to the fact that very often, if we carry out a calculation using the spectral density (2.95) (or, what is the same, the correlation function (2.94)) and then in the final result let A (or C) and α go to infinity in such a way that the ratio $A/\alpha^2 = C/\pi\alpha = f_0$ remains constant, we ultimately obtain a meaningful result (which, of course, depends only on f_0, and not on A and α separately).   In all such cases, this limiting result can be used with good accuracy for all sufficiently large α, and it is much simpler to obtain the final result by assuming from the very outset that the random process under consideration has a spectral density of the form (2.96).   Moreover, the same result can also be applied, with considerable accuracy, to many other forms of the correlation function $B(\tau)$ where $B(\tau)$ falls off rapidly from an appreciable value $B(0)$ to virtually zero at $|\tau| > \eta$ and η is very small in comparison with all typical time scales of the problem. (The spectral density $f(\omega)$ is then necessarily almost constant over the relevant frequency range.   Another justification of the utilization of the concept of a white noise will be discussed in Sec. 24, devoted to so-called "generalized stationary processes" which include also the white noise process.)

The spectral density (2.96) corresponds to the "improper" correlation function

(2.97) $B(\tau) = 2\pi f_0 \delta(\tau),$

which is proportional to the δ-function (see Note 1 to the Introduction).   The   form   (2.97)   of   the   function $B(\tau)$ (represented schematically together with the spectral density (2.96) in Fig. 20) implies that in fact we refer to a limiting situation where $B(\tau) \to 0$ for all $\tau \neq 0$, and $B(0) \to \infty$ so that
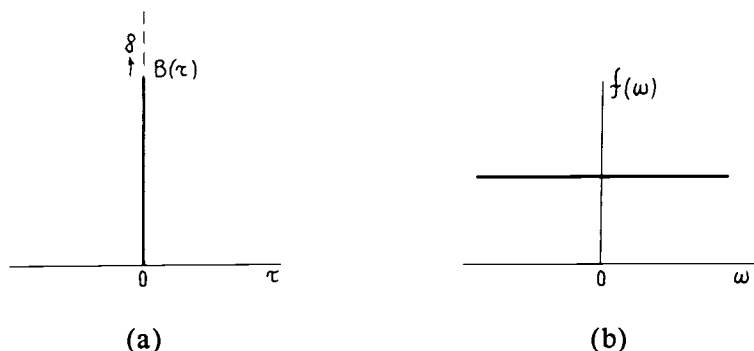
Fig. 20. (a) Schematic graph of the correlation function for "white noise". (b) Graph of the corresponding spectral density.

$\int_{-\infty}^{\infty} B(\tau)d\tau$ is constant. For instance, for any positive definite $B_0(\tau)$, we take $aB_0(a\tau)$ as the new correlation function, and then let $a \to \infty$ (the form of $B_0(\tau)$ is limited here only by the condition $\int_{-\infty}^{\infty} B_0(\tau)d\tau \neq 0$). Stated otherwise, it means that the values of $X(t)$ at any two instants of time (however close together) are assumed to be uncorrelated, i.e. that $X(t)$ is an "absolutely random" (or "purely random") process. Such a process is obviously quite analogous to a "stationary sequence of uncorrelated random variables" (also called a *discrete white noise*) $E(t)$, but refers to a continuous time t. Therefore, when t is continuous, $E(t)$ will always denote a "white noise" of unit intensity (i.e. such that $f_0 = 1/2\pi$, and $B(\tau) = \delta(\tau)$).

Note also that estimating the values of the parameters in the formula for $B(\tau)$ (e.g. the values of a in the expression $aB_0(a\tau)$), for which we can begin to use the limiting result obtained with the assumption (2.96) or (2.97), is equivalent, in a sense, to estimating the "spectral bandwidth" which is important in the given problem. Here we only indicate that in the three examples we mentioned of random processes of physical origin (i.e. fluctuation of the force acting on a Brownian particle, voltage fluctuation due to thermal noise in a conductor, and current fluctuation due to the shot effect) the approximation (2.96) - (2.97) is quite acceptable in almost

all problems of practical interest.

*It should be emphasized that when we say that only a restricted frequency range, where the spectral density remains virtually constant, is essential to the given problem, we thereby imply that the total power of the process $\langle X^2(t) \rangle = B(0)$, determined by the whole spectral distribution, does not play an important part in the problem. Therefore, the fact that the density (2.94) corresponds to an infinite total power should not bother us. The value $f_0 = f(0)$ is, naturally, the only essential spectral characteristic of the random process $X(t) = cE(t)$ for which the representation (2.94) of the spectral density is possible. Thus, the determination of this value is the first and foremost task of the theory of such a process.

In the case where the process $X(t)$ is the fluctuating force acting on a Brownian particle, statistical physics leads to the formula $f_0 = wkT/\pi$, where w is the proportionality constant between the friction force and the velocity ($w = 6\pi r\mu$ for a spherical particle of radius r immersed in a fluid of viscosity μ), T is the absolute temperature, and k is the Boltzmann constant (the derivation of this result will be considered in Sec. 12). In the case where $X(t)$ is the fluctuating electromotive force in a conductor, we need only replace the coefficient w in the preceding formula by R, the resistance of the conductor, i.e. in this case $f_0 = RkT/\pi$. (In the theory of electrical fluctuations this result is known as the *Nyquist formula*.) Finally, if $X(t)$ is the fluctuating current in a vacuum tube due to the shot effect, then under wide conditions the so-called *Schottky formula* holds, which states that $f_0 = \epsilon i_0/2\pi$, where $-\epsilon$ is the charge of the electron, and i_0 is the mean value of the current flowing through the tube.[25] *

Example 2. *Cosine curve.* It has already been shown in Sec. 7 (see (2.30)) that a function

(2.98) $B(\tau) = C \cos\omega_0\tau,$

where $C > 0$, $\omega_0 > 0$, is a correlation function of a stationary random process. Since $\cos\omega_0\tau = (e^{-i\omega_0\tau} + e^{i\omega_0\tau})/2$, to the function (2.98) corresponds the spectral distribution function

$$(2.99) \quad F(\omega) = \begin{cases} 0 & \text{for } \omega \leqslant -\omega_0, \\ C/2 & \text{for } -\omega_0 < \omega \leqslant \omega_0, \\ C & \text{for } \omega > \omega_0. \end{cases}$$
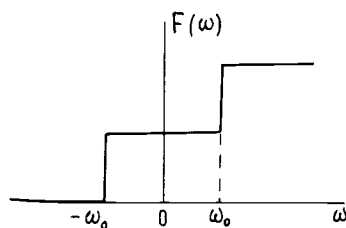
Formally, the same result can be written in the form of the following "spectral density formula":

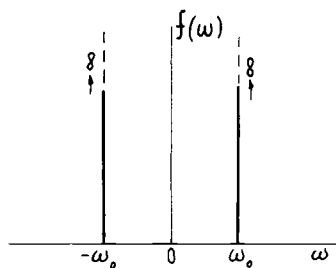$$(2.100) \quad f(\omega) = \frac{C}{2} [\delta(\omega + \omega_0) + \delta(\omega - \omega_0)]$$

(see Fig. 21).   The appearance of δ-functions on the right-hand side of (2.100) is due to the fact that the correlation function (2.98) does not go to zero at infinity.
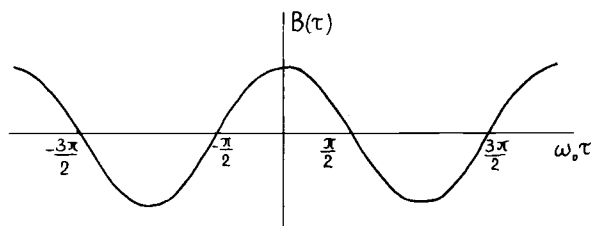
(a)

(b)

(c)

Fig. 21. (a) The cosinusoidal correlation function (2.98).
(b) The corresponding spectral distribution function $F(\omega)$. (c) The spectral density $f(\omega) = F'(\omega)$.

Example 3.   *Damped oscillation curve.*   The function

(2.101) $B(\tau) = Ce^{-\alpha|\tau|}\cos\omega_0\tau,$

where $C > 0$, $\alpha > 0$, and $\omega_0 > 0$, is often used to represent a
"damped oscillation". This function is the product of functions
of the forms (2.94) and (2.98) and hence it must also belong to
the class of correlation functions according to the result given
in Sec. 4 (see pp. 60–61). Let us determine the corresponding
spectral density $f(\omega)$:

$$f(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} Ce^{-\alpha|\tau|-i\omega\tau}\cos\omega_0\tau\,d\tau$$

$$= \frac{C}{4\pi}\left\{\int_{-\infty}^{\infty} e^{-\alpha|\tau|-i(\omega+\omega_0)\tau}d\tau\right.$$

(2.102)

$$\left.+ \int_{-\infty}^{\infty} e^{-\alpha|\tau|-i(\omega-\omega_0)\tau}d\tau\right\}$$

$$= \frac{C\alpha}{2\pi}\left\{\frac{1}{\alpha^2 + (\omega + \omega_0)^2} + \frac{1}{\alpha^2 + (\omega - \omega_0)^2}\right\}$$

(cf. the derivation of (2.95)).[26] To put it differently,

(2.103) $f(\omega) = \dfrac{A(\omega^2 + b^2)}{\omega^4 + 2a\omega^2 + b^4}$,

where

(2.104) $A = \dfrac{C\alpha}{\pi}$, $a = \alpha^2 - \omega_0^2,$ $b = (\alpha^2 + \omega_0^2)^{1/2}.$

It is clear that $f(\omega) > 0$ for all ω, and this confirms again that
(2.101) is a correlation function.

The correlation function (2.101) corresponds to some simple
models of stationary random processes encountered in
practice.[27] Moreover, it is often very suitable for approximating
certain empirical correlation functions which change their signs
several times (i.e. which alternate between positive and
negative values). As a typical example, consider the fading of
radio signals received by radar which is partially shown in Fig.
5(a). The fading is centered, i.e. $\langle X(t)\rangle = 0$; its correlation
function $B(\tau) = \langle X(t + \tau)X(t)\rangle$ can be calculated approximately

from a given (sufficiently lengthy) realization $x(t)$ of $X(t)$. To do this we need only to perform the time averaging of the values of $x(t + \tau)x(t)$, where τ is fixed, over the whole length of the realization, and to repeat this procedure many times for different values of τ. (We shall discuss this method of approximate determination of the values of the correlation function in Sec. 17 below.) The result of the application of this procedure to the fading of radio signals is shown in Fig. 37(a) (see p. 238). It is also explained in Sec. 17 that the unordered small waves seen in the graph of the empirical normalized correlation function $R(\tau) = B(\tau)/B(0)$, Fig. 37(a), at large values of τ are, in fact, highly unreliable and should be ignored. Therefore we reproduce in Fig. 22 only the initial enlarged portion of the curve depicted in Fig. 37(a). The dashed line in Fig. 22 indicates the curve

$$(2.105) \quad \tilde{R}(\tau) = e^{-24|\tau|}\cos40\tau$$

which has the form (2.101). We see that the curve (2.105) closely fits the experimental solid curve in the same figure over the whole region where the experimental values can be considered reliable. Thus, for theoretical purposes, we can replace the empirical correlation function shown in Fig. 37(a),
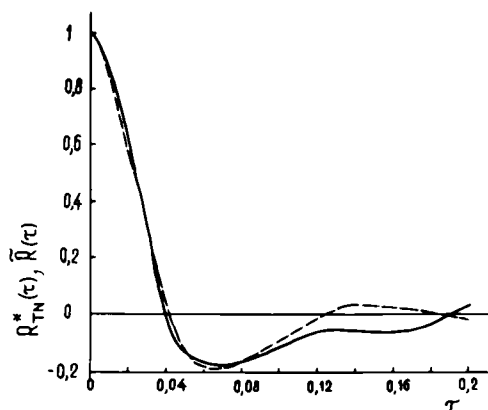
Fig. 22. Enlarged initial portion of the calculated normalized correlation function $R^{*}_{TN}(\tau)$ for fading of radio signals (the solid line) and its approximation by a damped cosinusoid $\tilde{R}(\tau)$ (the dashed line) (James et al., 1947).

by the function (2.105) given by a simple formula. In particular, formulae (2.103) − (2.104) now permit us to determine approximately the normalized spectral density for fading of radio signals (partly shown in Fig. 5(a). The result thus obtained is presented in Fig. 23.

According to Fig. 23 the spectral density $f(\omega)$ reaches its largest value when the frequency ω is approximately equal to $\omega_0 = 40$ rad/sec, the frequency of the periodic factor in formula (2.105). In the general case of the correlation function (2.101), an elementary calculation shows that if $\alpha < \sqrt{3}\omega_0$ (i.e. $a < b^2/2$), then the function (2.102) has its maximum at the point $\omega = \omega_1$, where

$$\omega_1 = b\{[2(1 - a/b^2)]^{1/2} - 1\}^{1/2}.$$

If α is appreciably less then ω_0, then the last formula implies that $\omega_1 \approx (1 - \alpha^4/8\omega_0^4)\omega_0 \approx \omega_0$. The sharpness of the maximum of the function (2.102) is clearly determined by the quantity α, i.e. by the rate at which the function (2.101) falls off with τ. For very small α the maximum is very sharp (and very close to the point $\omega = \omega_0$); hence the function (2.102) is close in shape to the "improper" spectral density (2.100) corresponding to the correlation function (2.98) (see Fig. 24(a). In this case, realizations of a random process with spectral density (2.102)
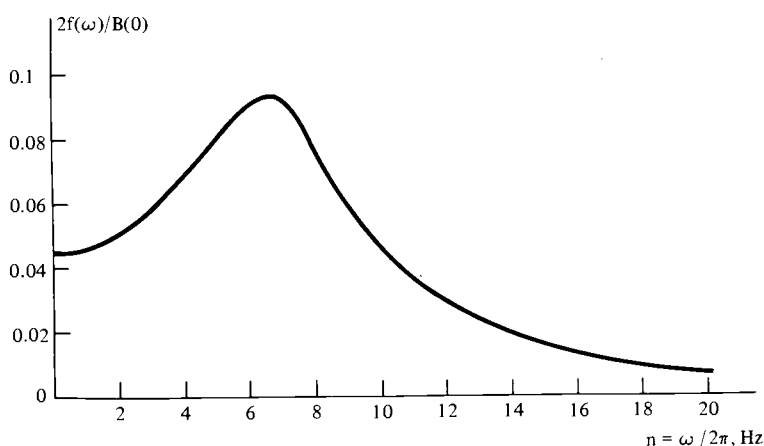
Fig. 23.   The normalized spectral density for fading of radio signals determined from the approximate correlation function (2.105) (James et al., 1947).

(a)

(b)
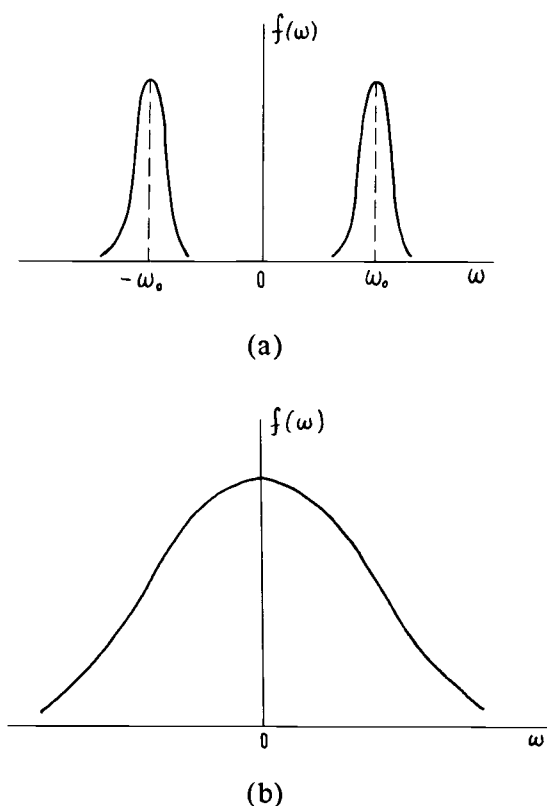
Fig. 24.   The shape of the spectral density (2.102); (a) for $\alpha \ll \omega_0$ and (b) for $\alpha > \sqrt{3}\omega_0$.

are close in their shape to realization of a random harmonic oscillation of frequency $\omega_0/2\pi$. More precisely, these realizations can be represented by the formula

$$(2.106) \qquad x(t) = a(t)\cos[\omega_0 t + \varphi(t)]$$

where $|a'(t)/a(t)| \ll \omega_0$ and $|\varphi'(t)/\varphi(t)| \ll \omega_0$ i.e., they can be regarded as sinusoidal oscillations of angular frequency ω_0 with slowly varying (relative to the time scale $1/\omega_0$) amplitude and phase.

If $\alpha \geqslant \sqrt{3}\omega_0$ (i.e. $a \geqslant b^2/2$), then the spectral density (1.102) has a single maximum at $\omega = 0$ and decreases monotonically for both positive and negative ω (see Fig. 24(b)). For very large α this maximum is quite broad, so that in many cases

one can even make the simple assumption that $f(\omega) = f_0 =$ const, thus regarding the process $X(t)$ as a white noise.

Example 4.   So far we have been discussing examples where the correlation function $B(\tau)$ was specified by some simple formula, which allowed us to calculate the corresponding spectral density $f(\omega)$ and to verify that this density was non-negative.   We now do the opposite; in other words, we begin with the formula for the spectral density $f(\omega)$. So, for our next example, we consider the correlation function corresponding to a spectral density of the form

$$(2.107) \quad f(\omega) = \frac{A}{\omega^4 + 2a\omega^2 + b},$$

where, A, a, and b are real constants.   It is easy to verify that in order for the density (2.107) to be everywhere non-negative and integrable, the numbers A, a, and b have to satisfy the inequalities

$$(2.108) \quad A > 0, \quad b > 0, \quad \sqrt{b} + a > 0.$$

If these inequalities hold, we can denote $\sqrt{b} = \omega_1^2$, $\sqrt{b} + a = 2\alpha^2$ and then rewrite (2.107) in the form

$$(2.109) \quad f(\omega) = \frac{A}{(\omega^2 - \omega_1^2)^2 + 4\alpha^2\omega^2} .$$

Formula (2.109) clearly implies that $f(\omega) > 0$ for all ω.

For any A, a, and b satisfying (2.108) or, what is the same, for any $A > 0$, $\alpha^2 > 0$, and $\omega_1^2 > 0$ the function (2.107) or, respectively, (2.109) is a spectral density corresponding to the correlation function $B(\tau)$ where

$$(2.110) \quad B(\tau) = \int_{-\infty}^{\infty} \frac{Ae^{i\omega\tau}d\omega}{(\omega^2 - \omega_1^2)^2 + 4\alpha^2\omega^2} .$$

The integral on the right-hand side of (2.110) can readily be evaluated by using the theory of residues.   The resulting formula for $B(\tau)$ depends on the sign of the difference $\omega_1^2 - \alpha^2 = (\sqrt{b} - a)/2$.   If $\omega_1^2 - \alpha^2 = \omega_0^2 > 0$, we obtain

$$(2.111) \quad B(\tau) = \frac{\pi A}{2\alpha\omega_1^2}e^{-\alpha|\tau|}(\cos\omega_0\tau + \frac{\alpha}{\omega_0} \sin\omega_0|\tau|).$$

In particular, this form (with $\omega_0 = \alpha$) has a correlation function corresponding to the spectral density

(2.112) $f(\omega) = \dfrac{A}{\omega^4 + 4\alpha^4}$.

If $\omega_1 = \alpha^2$, the spectral density (2.109) becomes

(2.113) $f(\omega) = \dfrac{A}{(\omega^2 + \alpha^2)^2}$,

and then

(2.114) $B(\tau) = \dfrac{\pi A}{2\alpha^3}\, e^{-\alpha|\tau|}(1 + \alpha\,|\tau|).$

Finally, if $\omega_1^2 - \alpha^2 = -\beta^2 < 0$, the parameter ω_0 in (2.111) is purely imaginary: $\omega_0 = i\beta$. Then (2.111) takes the form

(2.115) $B(\tau) = \dfrac{\pi A}{4\alpha\,\omega_1^2\beta}\left[(\alpha + \beta)e^{-(\alpha-\beta)|\tau|} - (\alpha - \beta)e^{-(\alpha+\beta)|\tau|}\right].$

The graph of the spectral density (2.109) has the same character as the graph of the function (2.102). It is symmetric about the ordinate axis and, if $\omega_1^2 > 2\alpha^2$ then, for $\omega \geqslant 0$, it is a "dome-shaped" curve (similar to the curve shown in Fig. 23), with its maximum at the point $\omega = (\omega_1^2 - 2\alpha^2)^{1/2}$ (close to ω_0 if $\alpha \ll \omega_1$). If, however, $\omega_1^2 \leqslant 2\alpha^2$, the density $f(\omega)$ has a single maximum at the point $\omega = 0$ and decreases monotonically on both sides of this point (i.e. it is similar to the curve given in Fig. 24(b). The correlation functions (2.111), (2.114), and (2.115) have the forms shown in Fig. 25. Formulae (2.109) and (2.111), (2.114), (2.115) are often used to approximate experimental spectral densities and correlation functions encountered in practical applications. Moreover, as will be shown in Sec. 12 (see (2.198) below), these formulae can sometimes also be justified on theoretical grounds.

Example 5. *General damped oscillation curve.* The following formula represents the most general damped oscillation, and is the generalization of both formulae (2.101) and (2.111):

(2.116) $B(\tau) = Ce^{-\alpha|\tau|}\cos(\omega_0|\tau| - \psi),$

where $C > 0$, $\alpha > 0$, and $\omega_0 > 0$. Since the condition $|B(\tau)| \leqslant B(0)$ must be fulfilled for the correlation function $B(\tau)$, it is clear that the function (2.116) can be a correlation function only when $B'(+0) \leqslant 0$, where, as usual, $B'(+0)$ is the limit of
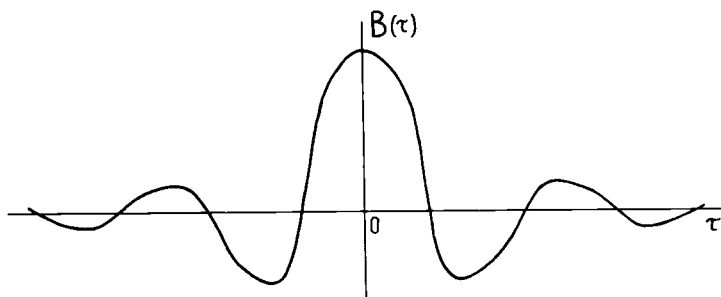
$[B(h) - B(0)]/h$, as $h \to 0$ through positive values. It readily follows from this that the phase ψ must satisfy the inequality

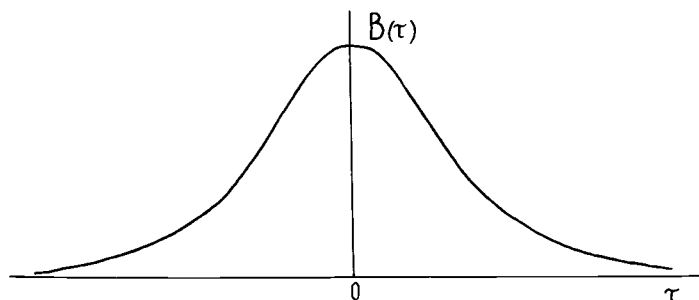(2.117) $|\psi| \leqslant \tan^{-1}(\alpha/\omega_0)$.

At $\psi = 0$ the function (2.116) obviously has the form (2.101), and at $\psi = \tan^{-1}(\alpha/\omega_0)$ the form (2.111).

It is not hard to see that the Fourier transform of the function (2.116) is

$$(2.118) \quad f(\omega) = \frac{c\omega^2 + d}{(\omega^2 + \alpha^2 - \omega_0^2)^2 + 4\alpha^2\omega_0^2}$$

(a)

(b)

Fig. 25. (a) The shape of the correlation function (2.111); (b) the general shape of the functions (2.114) and (2.115).

where

$$(2.119) \quad c = \frac{C}{\pi}(\alpha\cos\psi - \omega_0\sin\psi)$$

$$d = \frac{C(\alpha^2 + \omega_0^2)}{\pi}(\alpha\cos\psi + \omega_0\sin\psi).$$

Equations (2.118) and (2.119) imply that the function $f(\omega)$ is non-negative if, and only if, the condition (2.117) holds. Hence, this condition is necessary and sufficient for the function (2.116) to be a correlation function.

Functions of the form (2.116) can also often be conveniently used to approximate experimental correlation functions. As an example, the solid line in Fig. 26 indicated the empirical (normalized) correlation function $R(\tau) = B(\tau)/B(0)$ of the atmospheric pressure fluctuations averaged over the whole European part of the USSR. This correlation function was calculated by the author in the early 1950s by time averaging of the data of the semi-diurnal meteorological observations for the autumn of 1949. The dashed line in Fig. 26 indicates the curve

$$(2.120) \quad \tilde{R}(\tau) = C_1 e^{-0.06|\tau|}\cos(0.21|\tau| - 0.08),$$
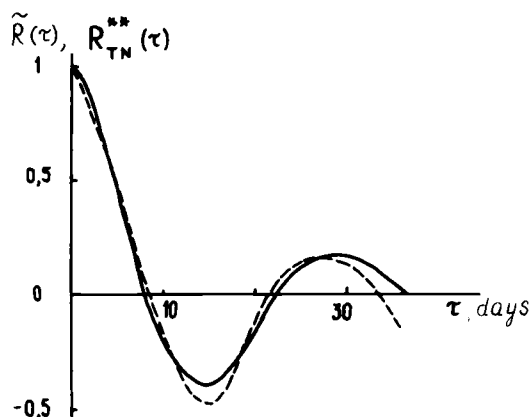
Fig. 26. Empirical normalized correlation function $R_{TN}^{**}(\tau)$ of the averaged atmospheric pressure and its approximation by the function (2.120) (Yaglom, 1955).

which has the form (2.116).[28] We can see that the curve
(2.120) fits very closely the experimental curve over the
region which includes several changes of sign of the
correlation function.

Example 6.   *Band limited white noise.*   A white noise with
a spectral density (2.96) is a mathematical idealization, which
does not fit into the conventional definition of the stationary
random process.   However, it is often reasonable to assume
in applications that the spectral density $f(\omega)$ of the process
$X(t)$ is constant (equal to f_0) only over a large, but *finite*,
range (namely, for $|\omega| \leqslant \Omega$, where Ω is some fixed "cut-off
frequency"), and is effectively zero outside this range (i.e. for
$|\omega| > \Omega$).   The function

$$(2.121) \quad f(\omega) = \begin{cases} f_0 & \text{for } |\omega| \leqslant \Omega, \\ 0 & \text{for } |\omega| > \Omega, \end{cases}$$

is a possible spectral density, and the correlation function
$B(\tau)$ corresponding to this density is of the form

$$(2.122) \quad B(\tau) = \int_{-\Omega}^{\Omega} f_0 e^{i\tau\omega} d\omega = 2f_0 \frac{\sin\Omega\tau}{\tau}$$

(see Fig. 27).   The ideal white noise of Example 1(a) can be
regarded as the limit of the "band limited white noise"
considered here, as $\Omega \to \infty$ (but f_0 is fixed).

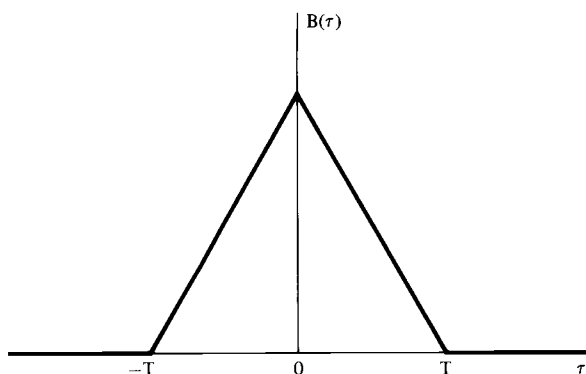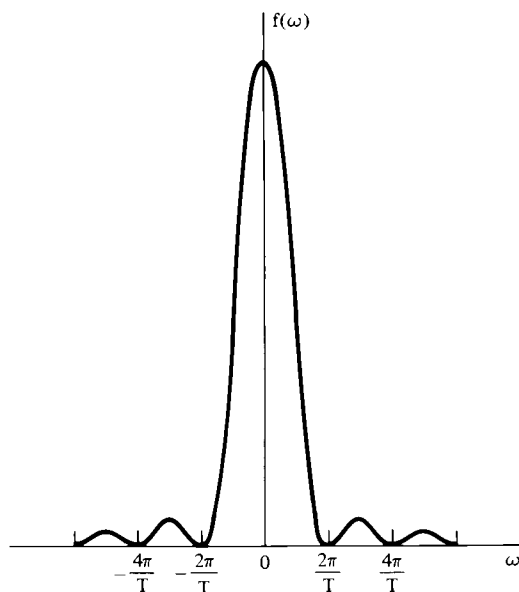(a)   (b)
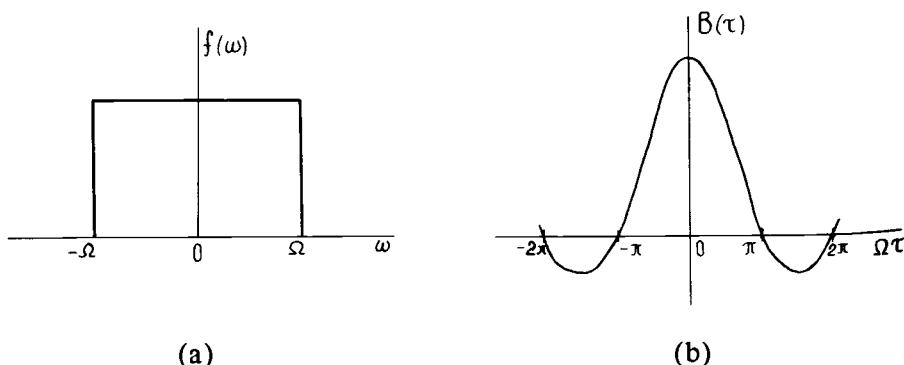
Fig. 27.   (a) The spectral density (2.121); (b) the correlation
         function (2.122).

Example 7. *Triangular correlation function.* We have already seen in Sec. 7 (see (2.47) and Fig. 17(b)) that the function

(a)

(b)

Fig. 28. (a) The triangular correlation function (2.123); (b) the spectral density (2.124).

(2.123) $B(\tau) = \begin{cases} C(T - |\tau|) & \text{for } |\tau| \leqslant T, \\ 0 & \text{for } |\tau| > T, \end{cases}$

where $C > 0$, $T > 0$, belongs to the class of correlation functions. The corresponding spectral density is given by the formula

(2.124) $f(\omega) = \frac{C}{2\pi} \int_{-T}^{T} (T - |\tau|) e^{-i\omega\tau} d\tau = \frac{2C}{\pi} \frac{\sin^2(\omega T/2)}{\omega^2}$

(see Fig. 28).   We see that indeed $f(\omega) \geqslant 0$ for all ω.

   *Consider now the function

(2.125) $B(\tau) = \begin{cases} C(T^2 - \tau^2) & \text{for } |\tau| \leqslant T, \\ 0 & \text{for } |\tau| > T, \end{cases}$

where $C > 0$, $T > 0$ (see Fig. 29).   This function is of the same kind as function (2.123); it satisfies conditions (1.28) and is also used sometimes by workers in applied fields as a model of the correlation function.[29] However, this function is not positive definite, and, hence, does not belong to the class of correlation functions. No calculations are needed to prove this.   In fact, it was shown in Sec. 4 (see p. 66) that if the correlation function is twice differentiable at the point $\tau = 0$, then it must be twice differentiable everywhere. However, function (2.125) has derivatives of all orders at $\tau = 0$ but no second derivative (and even no first derivative) at points $\tau = \pm T$. Therefore it cannot be a correlation function.
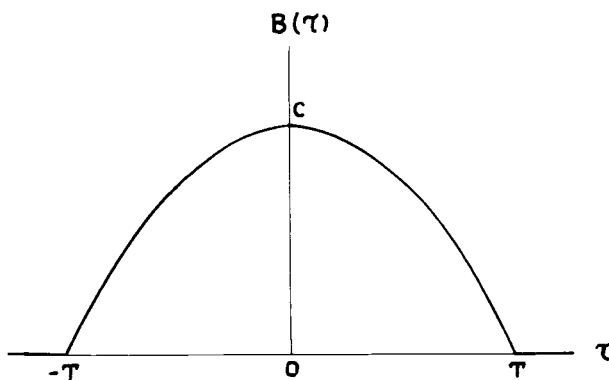

Fig. 29. The function (2.125) which is not a correlation function.

The family of functions

(2.125a) $B(\tau) = \begin{cases} C(T^m - |\tau|^m) & \text{for } |\tau| \leqslant T, \\ 0 & \text{for } |\tau| > T, \end{cases}$

where $C > 0$, $T > 0$, and $m > 0$, includes both the functions (2.123) and (2.125) as particular cases. All the functions (2.125a) clearly satisfy conditions (1.28); however, it is not hard to show that the function (2.125a) belongs to the class of correlation functions only if $0 < m \leqslant 1$, but it is not a correlation function if $m > 1$.[29a] *

Example 8. *Gaussian curve.* Let

(2.126) $B(\tau) = Ce^{-\alpha\tau^2}$,

i.e. assume that $B(\tau)$ has the form of the Gaussian probability density with mean value zero. The Fourier transform of such a function is well known (see Note 6 to the Introduction, eq. (0.11c')); it indicates that

(2.127) $f(\omega) = \dfrac{C}{2\sqrt{\pi\alpha}} e^{-\omega^2/4\alpha}$

(see Fig. 30). We see that $f(\omega) > 0$, and hence (2.126) is a correlation function.[30]

*Now we shall give a few more complicated examples of functions $B(\tau)$ and $f(\omega)$.

Example 9. *General rational spectral density.* Examples 1, 3, 4, and 5 above refer to the cases where the spectral density $f(\omega)$ is a rational function of ω. Stationary processes $X(t)$ having rational spectral densities are often encountered in applied problems and possess some important special properties, which greatly facilitate the solutions of many problems related to $X(t)$.[31] Now we consider the case of a general rational spectral density $f(\omega)$.
    Let

(2.128) $f(\omega) = C\,\dfrac{B(\omega)}{A(\omega)} = C\,\dfrac{(\omega - \beta_1)(\omega - \beta_2)...(\omega - \beta_q)}{(\omega - \alpha_1)(\omega - \alpha_2)...(\omega - \alpha_p)}$,
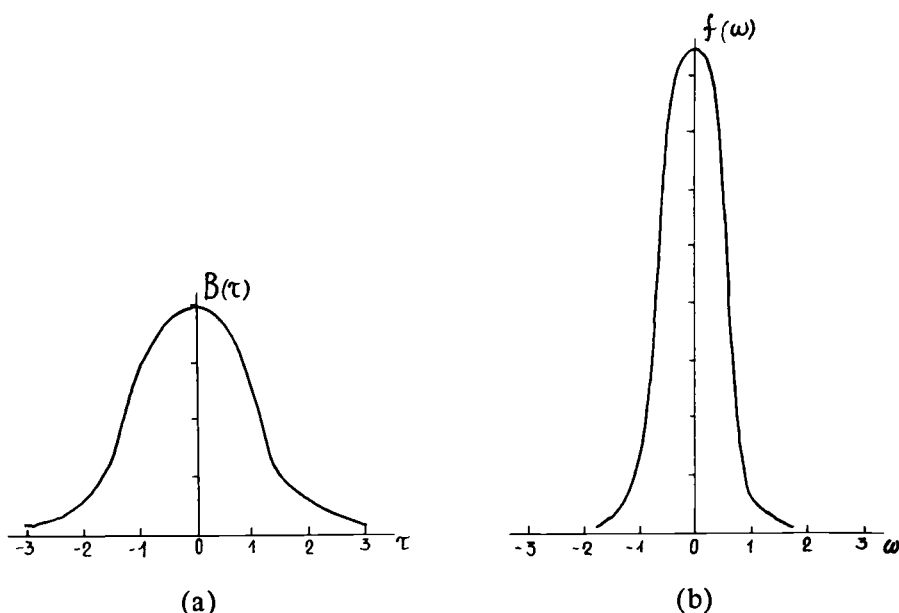
Fig. 30. (a) The "Gaussian correlation function" (2.126);
          (b) the corresponding spectral density $f(\omega)$.

where α_k, $k = 1, ..., p$, and β_j, $j = 1, ..., q$ are the roots of the polynomials $A(\omega)$ (of degree p) and $B(\omega)$ (of degree q). It is natural to suppose that the numerator and denominator in (2.128) have no common root, i.e. that none of the roots α_k coincides with any of the roots β_j (but, of course, multiple roots may be encountered both among $\alpha_1, ..., \alpha_p$ and $\beta_1, ..., \beta_q$). If $f(\omega)$ is a spectral density (i.e. a real, everywhere non-negative and integrable function of ω, where $-\infty < \omega < \infty$), then $p \geqslant q + 2$ and none of the roots α_k can be real. Moreover, in this case every real root β_j must be a root of even multiplicity (otherwise the function $f(\omega)$ would change sign in the vicinity of the point $\omega = \beta_j$). Besides,

$$C \frac{(\omega - \beta_1)...(\omega - \beta_q)}{(\omega - \alpha_1)...(\omega - \alpha_p)} = \bar{C} \frac{(\omega - \bar{\beta}_1)...(\omega - \bar{\beta}_q)}{(\omega - \bar{\alpha}_1)..(\omega - \bar{\alpha}_p)}$$

since $f(\omega)$ is real. Therefore all the roots α_k and all the complex roots β_j must be conjugate in pairs (i.e. must

include $\bar{\alpha}_k$ together with α_k and $\bar{\beta}_j$ together with β_j), and $p = 2m$ and $q = 2n$ must be even numbers. It is also clear that C is a real positive number (otherwise the condition $f(\omega) \geqslant 0$ would fail). Since $(\omega - \alpha_k)(\omega - \bar{\alpha}_k) = |\omega - \alpha_k|^2$ and $(\omega - \beta_j)(\omega - \bar{\beta}_j) = |\omega - \beta_j|^2$, (2.128) can be rewritten in the form

$$(2.129) \qquad f(\omega) = C \, \frac{|(\omega - \beta_1)...(\omega - \beta_n)|^2}{|(\omega - \alpha_1)...(\omega - \alpha_m)|^2} \, ,$$

where $m > n$, $C > 0$, and it can be assumed that all the roots α_1, ..., α_m and all the complex roots β_1, ..., β_n have positive imaginary part [for only one arbitrarily chosen root appears in (2.129) from each pair $(\alpha_k, \bar{\alpha}_k)$ and $(\beta_j, \bar{\beta}_j)$].

Formula (2.129), where C, m, n, α_1, ..., α_m, β_1, ..., β_n satisfy the indicated conditions, describes the general form of the rational spectral density $f(\omega)$.

If the process $X(t)$ is real, then $f(-\omega) = f(\omega)$ and the roots α_1, ..., $\alpha_p = \alpha_{2m}$ must include $-\alpha_k$, and $-\bar{\alpha}_k$ together with α_k and $\bar{\alpha}_k$ (the same is true also for the roots β_1, ..., $\beta_q = \beta_{2n}$). Therefore the root $-\bar{\alpha}_k = -\alpha_k^{(1)} + i\alpha_k^{(2)}$ or $-\bar{\beta}_j = -\beta_j^{(1)} + i\beta_j^{(2)}$ will appear among the roots α_1, ..., α_m or, respectively, β_1, ..., β_n together with the root $\alpha_k = \alpha_k^{(1)} + i\alpha_k^{(2)}$ or $\beta_j = \beta_j^{(1)} + i\beta_j^{(2)}$, where $\alpha_k^{(2)} > 0$, $\beta_j^{(2)} > 0$. It readily follows that the rational spectral density of the real process $X(t)$ can always be written in the form

$$f(\omega) = C \, \frac{|(i\omega)^n + b_1(i\omega)^{n-1} + ... + b_n|^2}{|(i\omega)^m + a_1(i\omega)^{m-1} + ... + a_m|^2}$$

$$(2.130)$$

$$= C \, \frac{\omega^{2n} + B_1\omega^{2(n-1)} + ... + B_n}{\omega^{2m} + A_1\omega^{2(m-1)} + ... + A_m} \, ,$$

where all the coefficients a_1, ..., a_m, b_1, ..., b_n, and A_1, ..., A_m, B_1, ..., B_n are real.

Using formula (2.67) and the theory of residues it is easy to find the correlation function $B(\tau)$ corresponding to the spectral density (2.129). In particular, it is not hard to show that in the case of a real process $X(t)$, where $B(-\tau) = B(\tau)$, the correlation function $B(\tau)$ can be represented as

$$B(\tau) = \sum_{j=1}^{m_1} C_j(|\tau|)e^{-\alpha_j'|\tau|}$$

(2.131)
$$+ \sum_{k=1}^{m_2} \left[C_k^{(1)}(|\tau|)\cos\alpha_k^{(1)}\tau + C_k^{(2)}(|\tau|)\sin\alpha_k^{(1)}|\tau| \right]e^{-\alpha_k^{(2)}|\tau|}.$$

In (2.131) $i\alpha_1', \ldots, i\alpha_{m_1}'$ are pure imaginary roots of the denominator of $f(\omega)$ having a positive imaginary part, $\alpha_1^{(1)} + i\alpha_1^{(2)}, \ldots, \alpha_{m_2}^{(1)} + i\alpha_{m_2}^{(2)}$ are complex roots of the denominator having positive real and imaginary parts, $C_j(|\tau|)$, $C_k^{(1)}(|\tau|)$ and $C_k^{(2)}(|\tau|)$ are real polynomials in $|\tau|$, whose degrees are lower by unity than the multiplicity of the respective roots $i\alpha_j'$ or $\alpha_k^{(1)} + i\alpha_k^{(2)}$ (so that C_j, $C_k^{(1)}$, and $C_k^{(2)}$ are real constants if $A(\omega)$ has no multiple roots).[32]

Example 10. *Pólya's theorem.* The exponential correlation function (2.94) and the triangular correlation function (2.123) have the following common property: the graphs of both of these functions for $0 \leqslant \tau < \infty$ are *concave curves* tending to zero as $\tau \to \infty$. Recall that the continuous curve $y = B(\tau)$ is said to be concave if for any τ_1 and τ_2 in the range of the argument τ the inequality $B(\frac{\tau_1 + \tau_2}{2}) \leqslant$ $[B(\tau_1) + B(\tau_2)]/2$ holds. (Geometrically, this means that the midpoint of any chord of the curve $y = B(\tau)$ lies above or on the graph of $B(\tau)$. It is easy to see that all the chords of the curve will then necessarily lie as a whole above or on the graph of $B(\tau)$.) An example of a concave curve $y = B(\tau)$ is shown in Fig. 31.

It turns out that the concavity of the graph of $B(\tau)$ for $0 \leqslant \tau < \infty$ and the convergence of $B(\tau)$ to zero at infinity guarantees that the function $B(\tau)$ belongs to the class of correlation functions. Namely, *if* $B(\tau)$ *is a bounded even function of* τ *which tends to zero as* $\tau \to \infty$ *and the graph of this function in the infinite interval* $0 \leqslant \tau < \infty$ *is a concave curve, then* $B(\tau)$ *is necessarily a correlation function.* This assertion was proved by G. Pólya and is usually called *Pólya's theorem* (or *Pólya's criterion*).[33]

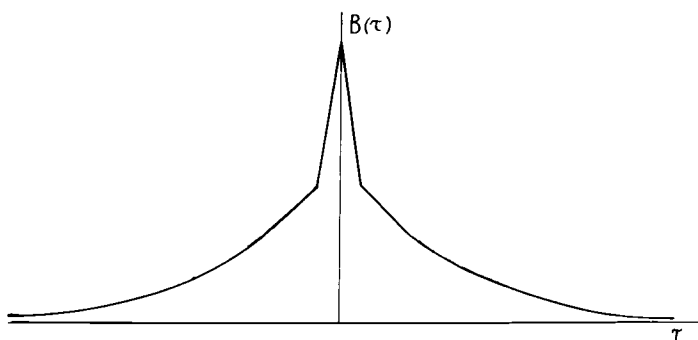For twice differentiable functions $B(\tau)$ the concavity of

Fig. 31. Example of a concave correlation function.

the graph for $\tau_1 \leqslant \tau \leqslant \tau_2$ is equivalent to the
following condition: $B''(\tau) \geqslant 0$ for $\tau_1 < \tau < \tau_2$. Thus,
any function $B(\tau)$ *which tends to zero (or*, as is easy to
see, *to some non-negative constant) as* $\tau \to \infty$ *and has a
non-negative second derivative* $B''(\tau)$ *at all points* $\tau > 0$,
is a correlation function.

Example 11. *More general exponential correlation
functions.* We have already seen that both the functions

$Ce^{-\alpha|\tau|}$ and $Ce^{-\alpha\tau^2}$, where $C > 0$ and $\alpha > 0$, are correlation
functions. It is also easy to prove that the function

(2.132) $B(\tau) = Ce^{-\alpha|\tau|^m}$, $C > 0$, $\alpha > 0$,

cannot be a correlation function if $m > 2$ (see, e.g., Notes 3
and 6 to the Introduction). On the other hand, according
to Pólya's theorem the function (2.132) will necessarily be a
correlation function if $0 < m \leqslant 1$ (because then $B''(\tau) > 0$ for
all $\tau > 0$). It is more difficult to prove that the function
(2.132) is also a correlation function for $1 < m < 2$; however,
this is true.[34] Thus, the function (2.132) is a correlation
function for all m in the interval $0 < m \leqslant 2$. The
corresponding spectral densities $f(\omega)$ for $m \neq 1$ and $m \neq 2$
cannot be expressed in a closed form in terms of elementary
functions; nevertheless, they are studied in detail in
probability theory with reference to quite a different
problem.[35]

Example 12. *Correlation functions involving Bessel*

functions. All the above expressions for the correlation function $B(\tau)$ include only elementary functions. However, there exist also many expressions for correlation functions involving some types of higher mathematical function used in applied mathematics and mathematical physics.[36] Here we consider only a few examples of this kind involving Bessel functions, which are the most widely used higher functions.

To begin with, we consider the following known integral representation of the Bessel function of order zero $J_0(t)$:

$$(2.133) \qquad J_0(t) = \frac{1}{\pi} \int_{-1}^{1} \frac{e^{itx}dx}{(1 - x^2)^{1/2}}.$$

It follows readily from (2.133) that the function

$$(2.134) \qquad B(\tau) = CJ_0(\alpha\,\tau), \quad C > 0, \ \alpha > 0,$$

is a correlation function corresponding to the spectral density

$$(2.135) \qquad f(\omega) = \begin{cases} C\pi^{-1}(\alpha^2 - \omega^2)^{-1/2} & \text{for } |\omega| < \alpha, \\ 0 & \text{for } |\omega| > \alpha \end{cases}$$

(a)   (b)
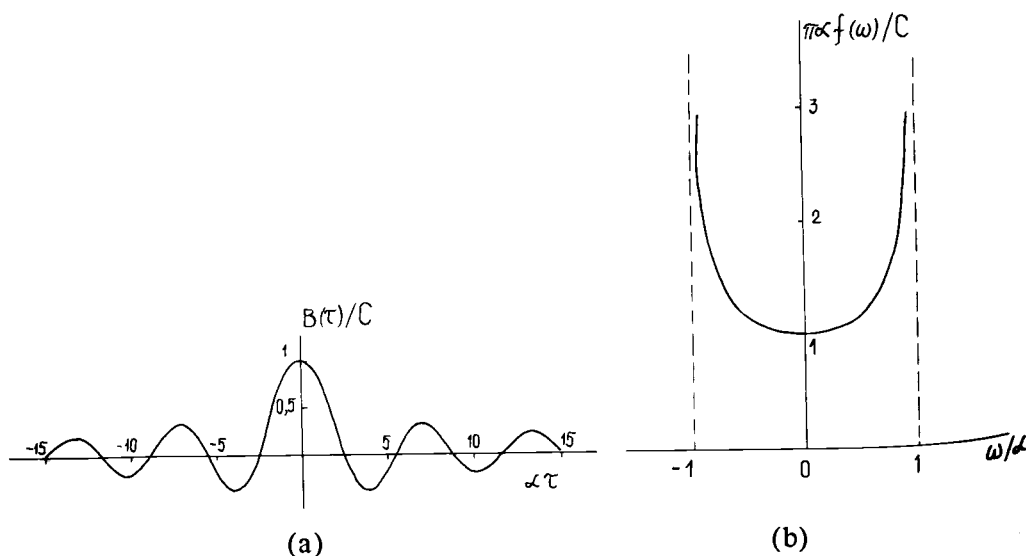
Fig. 32. (a) The correlation function (2.134); (b) the spectral density (2.135).

(see Fig. 32). The unboundedness of the density (2.135) is, of course, associated with the fact that the correlation function (2.134) does not satisfy the condition (2.66).

Equation (2.133) is a particular case of the following more general equation:

$$(2.136) \quad \begin{aligned} &\int_{-1}^{1} e^{itx}(1 - x^2)^{\nu-1/2}dx \\ &= 2^{\nu}\sqrt{\pi}\Gamma(\nu + 1/2)J_{\nu}(t)/t^{\nu}, \quad \nu > -1/2, \end{aligned}$$

where $J_{\nu}(t)$ is the Bessel function of order ν, and $\Gamma(\nu + 1/2)$ is the Γ-function.[37] Equation (2.136) implies that the function

$$(2.137) \quad B(\tau) = CJ_{\nu}(\alpha\tau)/(\alpha\tau)^{\nu}, \quad C > 0, \; \alpha > 0, \; \nu > -1/2$$

is also a correlation function which corresponds to the spectral density

$$(2.138) \quad f(\omega) = \begin{cases} \dfrac{C(\alpha^2 - \omega^2)^{\nu-1/2}}{2^{\nu}\pi^{1/2}\Gamma(\nu + 1/2)\alpha^{2\nu}} & \text{for } |\omega| < \alpha, \\ 0 & \text{for } |\omega| > \alpha. \end{cases}$$

For $\nu = 0$ and $\nu = 1/2$ (2.137) and (2.138) become (2.134), (2.135) and, respectively, (2.121) and (2.122). Note also that correlation functions of the form (2.137), where $\nu = (n-2)/2$, $n = 2,3, \ldots$, play an important part in this theory of homogeneous and isotropic random fields in multidimensional Euclidean spaces (see Chap. 4, Sec. 22).

One more interesting example of a correlation function expressible via Bessel functions is given by the formula

$$(2.139) \quad B(\tau) = C(\alpha\tau)^{\nu}K_{\nu}(\alpha\tau), \quad C > 0, \; \alpha > 0, \; \nu \geqslant 0,$$

where $K_{\nu}(x)$ is the so-called Basset's (or Macdonald's) function (i.e. a modified Bessel function of the third kind). The Fourier transform of the function (2.139) has the form[38]

$$(2.140) \quad f(\omega) = \frac{2^{\nu-1}\Gamma(\nu + 1/2)\alpha^{2\nu}C}{\pi^{1/2}(\alpha^2 + \omega^2)^{\nu+1/2}}.$$

We see that $f(\omega)$ is everywhere positive and hence (2.139) is

a correlation function. Formulae (2.139) and (2.140) have important applications in the theory of electric noise and turbulence theory;[39] in a special case, where $\nu = 1/2$, they become the familiar formulae (2.94) and (2.95).

Using the results of Sec. 4, we can now construct a large number of new examples of correlation functions proceeding from the examples given in this section. It is also clear that in cases where the correlation function $B(\tau)$ is non-negative, the corresponding spectral density $f(\omega)$ is a positive definite function of the argument ω, i.e. a correlation function. (In particular, the functions $C/(\alpha^2 + \tau^2)$, $C\sin^2(\alpha\tau)/\tau^2$ and $C/(\alpha^2 + \tau^2)^{\nu+1/2}$, where $\nu > 0$, are correlation functions according to the results of Examples 1, 7, and 12.) Moreover, many additional examples of correlation functions $B(\tau)$ can be found in the extensive mathematical literature devoted to characteristic functions of probability distributions and related topics.[40] However, we shall confine ourselves to the examples indicated above.*

11. Linear Transformations of Stationary Random Processes

It was explained in Sec. 8 that the spectral representation of a stationary random process $X(t)$ has a rather special meaning. This representation is not applicable to individual realizations $x(t)$, but necessarily requires consideration of the whole statistical ensemble of such realizations and a probabilistic interpretation of the corresponding Fourier—Stieltjes integral. Nevertheless, this representation is found to be very useful for many applications. To understand this paradox one must first of all recall the most usual method for the application of spectral decomposition of numerical (nonrandom) functions to practical problems.

Spectral (or harmonic) decomposition into sine oscillations is most useful when *linear and time-invariant transformations* of functions $x(t)$ are investigated. Hence it is reasonable to begin with the general description of such transformations. The theory of *linear transformations of functions* (or *linear filters, linear systems, linear operations on functions*, or *linear operators in function spaces* — all these terms have the same meaning) is treated in many textbooks and monographs.[41] In this section we shall only briefly elucidate some basic

relevant facts, without claiming either mathematical rigor or the coverage of even the most important applied aspects of the corresponding theory.

By *transformation* we mean the operation $\mathcal{L}$ applicable to some set of functions (constituting the *domain of applicability* of the operation $\mathcal{L}$) and transforming the function $x(t)$ from this set into another function $y(t)$:

$$(2.141) \qquad \mathcal{L}\{x(t)\} = y(t).$$

(The function $x(t)$ is then called the *input function*, and the function $y(t)$, the *output function* of the transformation $\mathcal{L}$.) The operation $\mathcal{L}$ is called a *linear transformation* if its domain of applicability contains, together with any finite set of

functions $x_1(t)$, ..., $x_n(t)$, all their linear combinations $\sum_{i=1}^{n} \alpha_i x_i(t)$

and

$$(2.142) \qquad \mathcal{L}\{\sum_{i=1}^{n} \alpha_i x_i(t)\} = \sum_{i=1}^{n} \alpha_i \mathcal{L}\{x_i(t)\}$$

for any functions $x_1(t)$, ..., $x_n(t)$, to which the operation $\mathcal{L}$ is applicable, and any numerical coefficients α_1, ..., α_n*. The number n in (2.142) is an arbitrary positive integer, but it is not always possible to put $n = \infty$, since the operation $\mathcal{L}$ may

prove inapplicable to the function $\sum_{i=1}^{\infty} \alpha_i x_i(t)$ even if it is

applicable to all functions $x_i(t)$ and $\sum_{i=1}^{\infty} \alpha_i x_i(t)$ exists. (For

instance, $\mathcal{L}$ may denote differentiation and the function $\sum_{i=1}^{\infty} \alpha_i x_i(t)$

may prove nondifferentiable, though all $x_i(t)$ are differentiable

and $\sum_{i=1}^{\infty} \alpha_i x_i(t)$ exists.) It is reasonable, however, to require in this

book that the operation $\mathcal{L}$ be such that (2.142) remains valid

---

*Instead of the general requirement (2.142) it is, of course, sufficient to require only the fulfillment of two particular conditions $\mathcal{L}\{\alpha x(t)\} = \alpha \mathcal{L}\{x(t)\}$ and $\mathcal{L}\{x_1(t) + x_2(t)\} = \mathcal{L}\{x_1(t)\} + \mathcal{L}\{x_2(t)\}$. Note also that functions $x(t)$ and coefficients α are assumed here to take complex values, as usual.

for $n = \infty$ if all the functions $x_i(t)$ and also $\sum_{i=1}^{\infty} \alpha_i x_i(t)$ belong to

the domain of applicability of $\mathcal{L}$. This requirement in fact means that the operation $\mathcal{L}$ is supposed to be *continuous*, i.e. to transform two input functions $x_1(t)$ and $x_2(t)$, which belong to its domain of applicability and are close enough to each other, into two output functions $y_1(t)$ and $y_2(t)$ which are also close to each other.

A transformation $\mathcal{L}$ is said to be *time-invariant* if, for any real τ,

$$(2.143) \quad \mathcal{L}\,\{x(t + \tau)\} = y(t + \tau),$$

where $x(t + \tau)$ and $y(t + \tau)$ are to be understood as functions of the argument t having the values at the time t equal to those of the functions $x(t)$ and $y(t)$ at the time $t + \tau$. In other words, if S_τ is the linear time-invariant transformation of a time shift by τ (i.e. $S_\tau\{x(t)\} = x(t + \tau)$), then

$$(2.143a) \quad \mathcal{L}\,\{S_\tau x(t)\} = S_\tau \mathcal{L}\,\{x(t)\}$$

for any real τ.

An important class of linear transformations is described by the integral formula

$$(2.144) \quad \mathcal{L}\,\{x(t)\} = \int_{-\infty}^{\infty} h(t;s)x(s)ds.$$

The domain of applicability of such an *integral transformation* $\mathcal{L}$ coincides with the set of functions $x(t)$ for which the integral on the right-hand side of (2.144) converges. The transformation (2.144) will be time-invariant if, for any t and τ, the value of the function $\mathcal{L}\,\{x(t + \tau)\}$ at time t is equal to the value of $\mathcal{L}\,\{x(t)\} = y(t)$ at time $t + \tau$ (so that the value of $\mathcal{L}\,\{x(t + \tau)\}$ at time $t - \tau$ is equal to $y(t)$). Therefore, if $\mathcal{L}$ is time-invariant, then

$$\int_{-\infty}^{\infty} h(t;s)x(s)ds = \int_{-\infty}^{\infty} h(t - \tau;s)x(s + \tau)$$

$$(2.145)$$

$$= \int_{-\infty}^{\infty} h(t - \tau; s - \tau)x(s)ds$$

for any τ and any $x(t)$ which makes the integral on the right-hand side of (2.144) convergent.

It is clear that (2.145) will be valid if $h(t - \tau; s - \tau) = h(t;s)$ for all τ. It is also easy to see that if the last condition fails, then a function $x(s)$ can always be found for which (2.145) also fails. Hence $h(t - \tau; s - \tau) = h(t;s)$ for any τ. In particular, putting $\tau = s$, we obtain

$$(2.146)\quad h(t;s) = h(t - s;0) = h(t - s).$$

(The function $h(t - s;0)$ depends on just one argument $t - s$, and therefore it is reasonable to denote it by $h(t - s)$.) Thus, the *integral linear transformation* (2.144) *is time-invariant only under the condition* (2.146), *when it takes the form*

$$(2.147)\quad \mathfrak{L}\{x(t)\} = \int_{-\infty}^{\infty} h(t - s)x(s)ds = \int_{-\infty}^{\infty} h(u)x(t - u)du.$$

The function $h(t)$ is called the *weighting function* or the *impulse response function*, of the integral transformation $\mathfrak{L}$. It has a simple physical meaning illuminating the term "impulse response". Consider the sequence $\delta_n(t)$, $n = 1,2, ...$, where all the functions $\delta_n(t)$ belong to the domain of applicability of $\mathfrak{L}$

and are such that $\lim_{n\to\infty} \delta_n(t) = 0$ for all $t \neq 0$ and $\int_{-\infty}^{\infty} \delta_n(t)dt = 1$

for all n. (Below we always assume that $h(t)$ is continuous; therefore, we can assume, e.g., that $\delta_n(t) = 0$ for $|t| > 1/n$ and $\delta_n(t) = n/2$ for $|t| \leqslant 1/n$.) Substituting now $\delta_n(t)$ for $x(t)$ in (2.147) and putting $n \to \infty$, we obtain

$$(2.148)\quad \lim_{n\to\infty} \mathfrak{L}\{\delta_n(t)\} = h(t).$$

But the sequence $\delta_n(t)$, $n = 1,2 ...$, is, of course, converging to the Dirac δ-function $\delta(t)$, and (2.148) is equivalent to the following well-known property of this generalized

function $\int_{-\infty}^{\infty} h(t - s)\delta(s)ds = h(t)$ (cf. Note 1 to the

Introduction). Therefore we can formally rewrite (2.148) as

$$(2.149)\quad h(t) = \mathfrak{L}\{\delta(t)\}.$$

We see that the function $h(t)$ is the output function which corresponds to the input having the form of the (generalized) function $\delta(t)$ which describes an instantaneous impulse of unit

intensity. If the term "linear filter" or "linear system" is used instead of "linear transformation", then the output function $\mathcal{L}\{x(t)\}$ is usually called the *filter* (or *system*) *response* to the input $x(t)$.   Thus, $h(t)$ is the response to the impulse at the input.

The above reasoning shows that the linear time-invariant transformation $\mathcal{L}$ can be written in the form (2.147) (i.e. is an integral transformation) only if $\lim_{n \to \infty} \mathcal{L}\{\delta_n(t)\} = h(t)$ exists. In other words, it is necessary that $\mathcal{L}$ be applicable to the Dirac δ-function.   The last condition, naturally, does not always hold: there are many simple linear transformations $\mathcal{L}$ for which the expression $\mathcal{L}\{\delta(t)\}$ is meaningless.   Even in such cases, however, (2.147) is sometimes used, but then it is assumed that h is a "generalized" function, for which "a value at the time t" $h(t)$ does not exist.   With this interpretation, generalized functions become simply synonymous to linear transformations: by specifying in some way the linear transformation $\mathcal{L}$ we also specify the generalized function h, which is its weighting function (see again Note 1 to the Introduction).   In the particular case of integral linear transformations of the form (2.147) the generalized function h becomes an ordinary function $h(t)$; hence, ordinary functions are particular cases of generalized ones.   On the other hand, the generalized function $h = \delta(x-\tau)$ corresponds to a very simple linear transformation $S_{-\tau}\{x(t)\} = x(t-\tau)$, and the generalized function $h = a_0 \delta^{(n)}(t) + a_1 \delta^{(n-1)}(t) + ... + a_n \delta(t)$ corresponds to the following transformation: $\mathcal{L}\{x(t)\} = (-1)^n a_0 d^n x(t)/dt^n + (-1)^{n-1} a_1 d^{n-1} x(t)/dt^{n-1} + ... + a_n x(t)$.

Let us now pass on to another method for describing linear time-invariant transformations $\mathcal{L}$, which actually explains the usefulness of spectral representation in the theory of such transformations.   Suppose that a transformation $\mathcal{L}$ is applicable to all the complex exponential functions ("complex sine oscillations") $x(t) = \exp(i\omega t)$. (This very general condition is always satisfied in practice and below in all the cases it will be considered true.) It turns out that then *the application of the transformation $\mathcal{L}$ to $\exp(i\omega t)$ reduces to simple multiplication of this function by a constant* (i.e. the time independent, but, generally speaking, frequency dependent) *factor* $H = H(\omega)$, so that

(2.150) $\mathcal{L}\{e^{i\omega t}\} = H(\omega)e^{i\omega t}$.

Indeed, let $\mathcal{L}$ transform $\exp(i\omega t)$ into $y_\omega(t)$, i.e. $\mathcal{L}\{\exp(i\omega t)\} =$

$y_\omega(t)$ is a function of t which is dependent on parameter ω. Then, by virtue of the linearity condition (2.142), where we put $n = 1$, and $\alpha_1 = \exp(i\omega\tau)$, we have

$$\mathfrak{L}\{e^{i\omega(t+\tau)}\} = \mathfrak{L}\{e^{i\omega\tau}e^{i\omega t}\} = e^{i\omega\tau}\mathfrak{L}\{e^{i\omega t}\} = e^{i\omega\tau}y_\omega(t).$$

On the other hand, by virtue of the time invariance of $\mathfrak{L}$ (property (2.143)) $\mathfrak{L}\{\exp[i\omega(t + \tau)]\} = y_\omega(t + \tau)$ and hence

$$y_\omega(t + \tau) = e^{i\omega\tau}y_\omega(t).$$

Now, setting $t = 0$, we get

$$y_\omega(\tau) = e^{i\omega\tau}y_\omega(0).$$

The last relation is clearly equivalent to (2.150); they differ only in that the quantity $y_\omega(0)$ in (2.150) is denoted by $H(\omega)$, and the variable τ is replaced by t.

Thus, the linear time-invariant system $\mathfrak{L}$ transforms a complex sine oscillation $\alpha\exp(i\omega t)$ of angular frequency ω into a sine oscillation $\alpha H(\omega)\exp(i\omega t)$ of exactly the same frequency, but with an amplitude multiplied by a complex factor $H(\omega)$. The function $H(\omega)$ is a very important characteristic of the transformation $\mathfrak{L}$; it is called the *transfer function* (or the *transfer coefficient*) of this transformation. If the transformation $\mathfrak{L}$ is real (i.e. if it transforms real functions $x(t)$ also into real functions), the function $\mathfrak{L}\{e^{i\omega t} + e^{-i\omega t}\} = H(\omega)e^{i\omega t} + H(-\omega)e^{-i\omega t}$ must be real for all t. It readily follows that in the case of a real transformation $\mathfrak{L}$

$$(2.151) \quad H(-\omega) = \overline{H(\omega)}.$$

Instead of a single complex function $H(\omega)$ the transformation $\mathfrak{L}$ is sometimes characterized by two real functions $A(\omega) = |H(\omega)|$ and $\psi(\omega) = \arg H(\omega)$ (so that $H(\omega) = A(\omega)\exp i\psi(\omega)$, i.e. $\mathfrak{L}\{e^{i\omega t}\} = A(\omega)e^{i[\omega t+\psi(t)]}$). The functions $A(\omega)$ and $\psi(\omega)$ are then called the *gain* and the *phase shift* (or simple *phase*) of the linear transformation (or filter, or system) $\mathfrak{L}$. In the case of a real transformation $\mathfrak{L}$, it follows from (2.151) that

$$(2.151a) \quad A(-\omega) = A(\omega), \quad \psi(-\omega) = -\psi(\omega).$$

Knowing the transfer function $H(\omega)$, one can easily describe the effect of the transformation $\mathfrak{L}$ on many more general input functions $x(t)$, than the simple complex exponent $\exp(i\omega t)$. First of all, the linearity of $\mathfrak{L}$ implies that it is applicable to all linear combinations $x(t) = \sum\limits_{k=1}^{n} \alpha_k e^{i\omega_k t}$, where n is an integer, and

$$\mathfrak{L}\,\{x(t)\} = \sum_{k=1}^{n} \alpha_{kj}\mathfrak{L}\{e^{i\omega_k t}\} = \sum_{k=1}^{n} \alpha_k H(\omega_k)e^{i\omega_k t} \ .$$

Let us now consider a much broader class of functions $x(t)$ admitting spectral decomposition into sine oscillations, i.e. representable in the form of a Fourier series, or a Fourier integral, or a Fourier–Stieltjes integral. Recall that we agreed to assume the transformations $\mathfrak{L}$ to be continuous within their domain of applicability in the following sense: if $\sum \alpha_i x_i(t)$ exists

and if $\mathfrak{L}$ is applicable to it and to all the functions $x_i(t)$, $i = 1,2, ...,$ then (2.142) holds for $n = \infty$. (This assumption is, in fact, valid under very wide regularity conditions which are fulfilled for all transformations $\mathfrak{L}$ met in practice.   In particular, it is easy to verify that it is satisfied for integral transformations of the form (2.147) and all the other specific transformations considered in this book.) Since both the infinite Fourier series and Fourier (or Fourier–Stieltjes) integrals are representable as the limits of a sequence of finite sums consisting of a finite number of complex exponents, the above-stated continuity condition implies that the linear transformation symbol   can be placed under the sign of the sum (in the case of Fourier series) or of the integral (in the case of Fourier or Fourier–Stieltjes integrals).* Hence we obtain

$$(2.152) \qquad \mathfrak{L}\,\{\textstyle\sum_k \alpha_k e^{i\omega_k t}\} = \sum_k \alpha_k \mathfrak{L}\,\{e^{i\omega_k t}\} = \sum_k \alpha_k H(\omega_k)e^{i\omega_k t},$$

---

*The situation is slightly more complicated in the case of the improper Fourier (or Fourier-Stieltjes) integral. Here, in fact, we must use the following, more general, continuity condition: if $\lim\limits_{n\to\infty} x_n(t)$ <u>exists and if $\mathfrak{L}$ is applicable to it</u>

<u>and to all functions</u> $x_n(t)$, $n = 1, 2, ...,$ <u>then</u> $\mathfrak{L}\{\lim\limits_{n\to\infty} x(t)\} = \lim\limits_{n\to\infty} \mathfrak{L}\{x(t)\}$.

This condition is also valid for the overwhelming majority of linear transformations encountered in practical applications.

(2.152a) $\mathcal{L}\{\int e^{i\omega t}\varphi(\omega)d\omega\} = \int\mathcal{L}\{e^{i\omega t}\}\,\varphi(\omega)d\omega = \int e^{i\omega t}H(\omega)\varphi(\omega)d\omega,$

(2.152b) $\mathcal{L}\{\int e^{i\omega t}d\Phi(\omega)\} = \int\mathcal{L}\{e^{i\omega t}\}d\Phi(\omega) = \int e^{i\omega t}H(\omega)d\Phi(\omega).$

We see that the transfer function $H(\omega)$ permits one to describe easily the effect of the transformation $\mathcal{L}$ on any function $x(t)$ admitting spectral decomposition.   It is this fact that determines, in the first place, the practical usefulness of spectral analysis, since time-invariant linear transformations are frequently encountered in applications and play a central role in many practical problems.

    Let $\mathcal{L}$ be an integral linear transformation of the form (2.147).   Such a transformation can be characterized either by its weighting (impulse response) function $h(t)$, or by its transfer functions $H(\omega)$.   Therefore it is clear that these two functions must be interrelated.   This relationship can be found easily: by substituting the function $\exp(i\omega s)$ for $x(s)$ in the right-hand side of (2.147) we get

$$\mathcal{L}\{e^{i\omega t}\} = e^{i\omega t}\int_{-\infty}^{\infty}h(u)e^{-i\omega u}du.$$

Hence

(2.153) $H(\omega) = \int_{-\infty}^{\infty}e^{-i\omega u}h(u)du.$

    By virtue of the inversion formula for Fourier integrals it follows that

(2.154) $h(u) = \dfrac{1}{2\pi}\int_{-\infty}^{\infty}e^{i\omega u}H(\omega)d\omega.$

We see that the functions $h(u)$ and $H(\omega)$ are simply reciprocal Fourier transforms.*   Hence, in the case of the integral

---

*Formula (2.154) can also be derived in the following way.   Since $(2\pi)^{-1}$

$\int_{-\infty}^{\infty}e^{-i\omega t}\delta(t)dt = (2\pi)^{-1}$, we have $\delta(t) = (2\pi)^{-1}\int_{-\infty}^{\infty}e^{i\omega t}d\omega$. Hence $\delta(t)$ $\delta(t)$

$= \lim\limits_{n\to\infty}\delta_n(t)$, where $\delta_n(t) = (2\pi)^{-1}\int_{-n}^{n}e^{i\omega t}d\omega$.   Recalling now that $h(t) =$

$\mathcal{L}\{\delta(t)\} = \lim\limits_{n\to\infty}\mathcal{L}\{\delta_n(t)\}$ and determining $\mathcal{L}\{\delta_n(t)\}$ by (2.152a), we

immediately obtain (2.154).

transformation (2.147), the transfer function $H(\omega)$ must decrease rapidly enough at infinity, so that its Fourier transform $h(u)$ exists. Conversely, if a Fourier transform $h(u)$ of the function $H(\omega)$ exists, this means that the linear transformation $\mathfrak{L}$ can be applied to the Dirac δ-function, and $\mathfrak{L} \{\delta(t)\} = h(t)$ (cf. the footnote on p. 147). Moreover, in this case $\mathfrak{L} \{\delta(t - s)\} = h(t - s)$ (since $\mathfrak{L}$ is time-invariant). Using

the relation $\int_{-\infty}^{\infty} \delta(t - s)x(s)ds = x(t)$, we can formally write the

chain of equalities

$$\mathfrak{L} \{x(t)\} = \mathfrak{L} \{\int_{-\infty}^{\infty} \delta(t - s)x(s)ds\} = \int_{-\infty}^{\infty} \mathfrak{L}\{\delta(t - s)\}x(s)ds$$

(2.155)
$$= \int_{-\infty}^{\infty} h(t - s)x(s)ds.$$

(Recall that $\mathfrak{L}$ is the transformation of functions of t, which does not affect $x(s)$.) The formal equalities (2.155) can easily be made rigorous; hence, if the Fourier transform (2.154) exists, $\mathfrak{L}$ is necessarily the integral transformation of the form (2.147). If, however, the integral on the right-hand side of (2.154) diverges, the function $h(t)$ does not exist in the conventional sense. As we have already noted, in this case the weighting function h of the transformation $\mathfrak{L}$ must be interpreted as a generalized function. The generalized function h can also be formally specified by (2.154), i.e. it should be assumed that its Fourier transform is an ordinary (but not decreasing rapidly enough at infinity) complex function $H(\omega)$.

Suppose now that the input function $x(t)$ describes the time variations of some physical quantity (e.g., the voltage across a conductor, the velocity of a particle, or the temperature at a fixed point of a solid or fluid). Moreover, let the linear integral transformation $\mathfrak{L}$ of $x(t)$ be effected by a physical system operating in "real time". Then the system output $y(t) = \mathfrak{L}\{x(t)\}$ at time t cannot, of course, depend on the future values of the input $x(t)$. Therefore the weighting (impulse response) function $h(u)$ must in this case satisfy the condition

(2.156) $h(u) = 0$ for $u < 0$.

This condition evidently expresses the general *causality*

principle ("the effect cannot precede the cause in time"); it is usually called the *realizability condition* and a transformation (system, filter) satisfying this condition is said to be *physically realizable* (or *nonanticipative*). Since the transfer function $H(\omega)$ uniquely determines the weighting function $h(u)$, it is clear that the realizability condition can also be expressed as some property of the transfer function. However, we shall not pursue this topic here.[42]

Let us now pass on to the main topic of this section: the effect of linear time-invariant transformations $\mathfrak{L}$ on stationary random processes $X(t)$. The class of such transformations $\mathfrak{L}$ can be specified by indicating their general properties (2.142) and (2.143), where the numerical time functions $x(t)$ must only be replaced by random processes $X(t)$. It is, however, more convenient to use a more specific description of the class of transformations $\mathfrak{L}$ by writing out an explicit formula for such transformations which includes all the linear transformations of processes $X(t)$ used in practice.

We begin with the particular class of transformations of the form

$$(2.157) \qquad \mathfrak{L}\{X(t)\} = \sum_{j=1}^{N} h_j X(t - \tau_j),$$

where N is an integer, h_j, $j = 1, ..., N$ are arbitrary complex coefficients, and τ_j, $j = 1, ..., N$ are fixed real numbers (of any sign). It is clear that (2.157) describes a linear time-invariant transformation, i.e. that $\mathfrak{L}$ possess the properties (2.142) and (2.143) (where $x(t)$ is replaced by $X(t)$). However, the class (2.157) is rather narrow and does not include many useful transformations. A much broader class of linear transformations is specified by the condition that $\mathfrak{L}$ *is a (mean square) limit of a sequence of finite sums of the form (2.157)*, i.e. that

$$(2.158) \qquad \mathfrak{L}\{X(t)\} = \lim_{n\to\infty} \sum_{j=1}^{N_n} h_j^{(n)} X(t - \tau_j^{(n)}).$$

Formula (2.158) is already sufficiently broad so that in what follows we can restrict ourselves to transformations of the form (2.158).

Let us now use the spectral representation (2.61) of the process $X(t)$. Substituting this representation in (2.158), we obtain

(2.159) $\mathfrak{L}\{X(t)\} = \lim\limits_{n\to\infty} \int_{-\infty}^{\infty} e^{it\omega} \sum\limits_{j=1}^{N_n} h_j^{(n)} e^{-i\tau_j^{(n)}\omega}\, dZ(\omega).$

It is easy to see that the (mean square) limit on the right-hand side of (2.159) exists if and only if there exists a function $H(\omega)$ such that

(2.160) $\lim\limits_{n\to\infty} \int_{-\infty}^{\infty} |H_n(\omega) - H(\omega)|^2 dF(\omega) = 0,$

where $H_n(\omega) = \sum\limits_{j=1}^{N_n} h_j^{(n)} e^{-i\tau_j^{(n)}\omega}$ and $dF(\omega) = \langle |dZ(\omega)|^2 \rangle$ (i.e. $F(\omega)$ is

the spectral distribution function of $X(t)$). Under this condition the function $H(\omega)$ evidently satisfies the inequality

(2.161) $\int_{-\infty}^{\infty} |H(\omega)|^2 dF(\omega) < \infty,$

and the limit on the right-hand side of (2.159) has the following expression:

(2.162) $\mathfrak{L}\{X(t)\} = \int_{-\infty}^{\infty} e^{it\omega} H(\omega) dZ(\omega).$

On the other hand, if $H(\omega)$ is an arbitrary complex function satisfying the condition (2.161), then such a sequence of functions $H_n(\omega)$, $n = 1, 2, \ldots$, necessarily exists that $H_n(\omega)$ has the form $H_n(\omega) = \sum h_j^{(n)} e^{-i\tau_j^{(n)}\omega}$ and the relation (2.160) holds. This means that under the condition (2.161) the transformation (2.162) can always be represented as (2.158). Thus, the class of linear transformations $\mathfrak{L}$ of stationary processes $X(t)$ of the form (2.158) coincides with the class of transformations (2.162), where $H(\omega)$ satisfies (2.161). The complex function $H(\omega)$ is called the *transfer function* of the transformation (2.162), and its modulus $|H(\omega)| = A(\omega)$ and the argument $\arg H(\omega) = \psi(\omega)$ are called the transformation *gain* and *phase shift*, respectively.

Formula (2.162) shows that if the input process $X(t)$ is stationary, then the output process $\mathfrak{L}\{X(t)\} = Y(t)$ is also a stationary random process, and the function $Z_Y(\omega)$, which appears in the spectral representation of the output process $Y(t)$, is specified by the equations

(2.163) $dZ_Y(\omega) = H(\omega) dZ(\omega) = A(\omega) e^{i\psi(\omega)} dZ(\omega).$

Hence

(2.163a) $Z_Y(\omega) = \int_{-\infty}^{\omega} H(\omega')dZ(\omega')$

and the spectral distribution function $F_{YY}(\omega)$ of the output process $Y(t)$ satisfies the equations

(2.164) $dF_{YY}(\omega) = |H(\omega)|^2 dF(\omega) = \{A(\omega)\}^2 dF(\omega),$

(2.164a) $F_{YY}(\omega) = \int_{-\infty}^{\omega} |H(\omega')|^2 dF(\omega').$

Equations (2.163) show, in particular, that any linear transformation transforms the input process $X(t)$ with a discrete spectrum $\omega_1, \omega_2, \ldots$ into the output process $Y(t)$ also with a discrete spectrum. Moreover, if the input process $X(t)$ has a spectral density $f(\omega)$, then the output process $Y(t)$ also has a spectral density $f_{YY}(\omega)$, and

(2.165) $f_{YY}(\omega) = |H(\omega)|^2 f(\omega).$

In the last case the class of transfer functions $H(\omega)$ of the transformations applicable to the process $X(t)$ is specified by a very simple condition

(2.161a) $\int_{-\infty}^{\infty} |H(\omega)|^2 f(\omega)d\omega < \infty.$

Finally, if $X(t)$ is a process with a mixed spectrum consisting of a number of discrete frequencies $\omega_1, \omega_2, \ldots$ and a continuous part, then the output process $Y(t) = \mathfrak{L}\{X(t)\}$ is also a process with a mixed spectrum.

The correlation function $B_{YY}(\tau)$ of the output process $Y(t)$ is given by the equation

(2.166) $B_{YY}(\tau) = \int_{-\infty}^{\infty} e^{i\tau\omega}|H(\omega)|^2 dF(\omega).$

In particular, if $X(t)$ has a spectral density $f(\omega)$, then

(2.166a) $B_{YY}(\tau) = \int_{-\infty}^{\infty} e^{i\tau\omega}|H(\omega)|^2 f(\omega)d\omega.$

A comparison of (2.162) with (2.152) − (2.152b) shows that the effect of a linear time-invariant transformation $\mathfrak{L}$ on a stationary process $X(t)$ is similar to the effect of such a transformation on a numerical function $x(t)$ admitting

spectral decomposition.   This is precisely what we had in mind when we noted (on p. 102) that the conventional applications of the Fourier decomposition of nonrandom functions $x(t)$ are transferred, practically without any change, to the spectral representation of stationary stochastic processes.

Let us now try to link the concept of linear transformation $\mathfrak{L}$ acting on numerical time functions $x(t)$ with the concept of linear transformation $\mathfrak{L}$ acting on stationary random processes $X(t)$.

To begin with, consider the case of an integral transformation $\mathfrak{L}$ of the form (2.147), where $h(u)$ is a continuous weighting function. Let us associate with this transformation the following transformation of stationary random processes $X(t)$:

(2.147a) $\mathfrak{L}\{X(t)\} = \int_{-\infty}^{\infty} h(u)X(t - u)du,$

where the integral on the right-hand side is the mean square integral defined in Sec. 4.   The transformation (2.147a) is clearly meaningful and is included in the class of transformations (2.158), provided that the integral on the right-hand side of (2.147a) converges, i.e. if

(2.167) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(u)\overline{h(v)}B(v - u)du\,dv < \infty$

(cf. condition (1.54)). Substituting the spectral representation of $X(t)$ on the right-hand side of (2.147a), we find that the transfer function $H(\omega)$ of the transformation (2.147a) is given by the familiar Fourier formula (2.153).   Therefore, the condition (2.167) can, in this case, be rewritten in terms of the Fourier transform $H(\omega)$ of $h(u)$, and then (2.167) takes the form of (2.161).

We have already noted in Sec. 4 (p. 69) that if a stationary process $X(t)$ is mean square continuous (and no other processes are considered in this book), then its realizations $x(t)$ are integrable functions (as a rule, either continuous, or having only jump discontinuities) and the integrals $\int_{-\infty}^{\infty} h(u)x(t - u)du$

determine the sample values (i.e. realizations) of the random variable appearing on the right-hand side of (2.147a). Therefore, applying the realization $x(t)$ of the process $X(t)$ at the input of the linear system effecting the transformation

(2.147) of numerical functions, we obtain (subject to the condition (2.167)) at the output of this system the realization of the process $\mathfrak{L}\{X(t)\} = Y(t)$ specified by formula (2.147a). We know that neither the spectral representation of the input stationary process $X(t)$, nor that of the output process $Y(t)$ can be applied directly to separate realizations $x(t)$ and $y(t)$. Nevertheless, since the output function $y(t)$ coincides with the realization of the stationary process $Y(t)$, which has a spectral representation (2.162), this function behaves in many respects as if it could be expanded in terms of the complex exponential functions $\exp(i\omega t)$, whose coefficients are obtained from those of the expansion of the input function $x(t)$ by multiplying by the transfer function $H(\omega)$. (A typical example illustrating such a behavior of the output realization $y(t)$ will be given on p. 156.) Therefore it is not surprising that the assumption about the permissibility of spectral decomposition of functions $x(t)$ has often led to quite reasonable (i.e. correct) results even when one dealt with functions which certainly could be expanded not into a Fourier series or integral, nor into a Fourier–Stieltjes integral, but which are realizations of stationary random processes. It should also be remarked that in the case of an integral linear transformation of the form (2.147a) the formula (2.166) for $B_{YY}(\tau)$ can also be rewritten in the equivalent form

$$(2.168) \quad B_{YY}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(u)\overline{h(v)}B(\tau + v - u)\,du\,dv,$$

which coincides with the result following from (1.56).

So far we have assumed that the transformation $\mathfrak{L}\{X(t)\}$ has the form (2.147a). Consider now a quite different linear transformation of the form

$$(2.169) \quad \mathfrak{L}\{X(t)\} = a_0\frac{d^n X(t)}{dt^n} + a_1\frac{d^{n-1}X(t)}{dt^{n-1}} + \dots + a_n X(t),$$

where all the derivatives are understood as mean square limits. It was noted in Sec. 4 (p. 67) that in the case of a (mean square) differentiable process $X(t)$ its realizations $x(t)$ are also differentiable functions, and $x'(t)$ coincides with the realization of the random process $X'(t)$. It follows that the realizations of the process $\mathfrak{L}\{X(t)\}$, subject to the condition that all the derivatives up to the nth order of the process $X(t)$ exist, coincide with functions $\mathfrak{L}\{x(t)\} = a_0 d^n x(t)/dt^n + \dots +$

$a_n x(t)$ obtained by applying the linear transformation $\mathfrak{L}$ to separate realizations of the process $X(t)$. In general, if $\mathfrak{L}$ is the linear transformation (2.158), which is applicable to the random process $X(t)$, then for the realizations $x(t)$ of this process there will, as a rule, exist a limit (understood in the conventional sense)

$$(2.158a) \quad \mathfrak{L}\{x(t)\} = \lim_{n \to \infty} \sum_{j=1}^{N_n} h_j^{(n)} x(t - \tau_j^{(n)}),$$

which will coincide with the corresponding realization of the process $\mathfrak{L}\{X(t)\}$. It is this important circumstance that allows us to replace linear transformations (2.158) of stationary random processes $X(t)$ by linear transformation (2.158a), applied to realizations $x(t)$ of the process $X(t)$.

The linear transformation (2.158a) of the realizations $x(t)$ of a stationary random process $X(t)$ can often be performed with the aid of a special device called a *linear filter*. We have previously used the term linear filter as a synonym for a linear transformation (or linear system). Now we note that in engineering literature this term also has a more special meaning: filter is often used to mean a device which passes harmonic (i.e. sine) oscillations in a certain frequency range (the *pass band* of the filter), while suppressing (i.e. sharply weakening or eliminating altogether) all the other harmonic oscillations. Electrical filters play a very important role in all radio engineering equipment, and therefore their use is particularly widespread. Such a filter is an electrical circuit with four free terminals (a "four-pole"), of which two represent the filter "input", and the other two, its "output". If a voltage, which is a harmonic oscillation of a given angular frequency ω, is applied at the filter input, then the output voltage will either follow the input voltage (maybe with a phase shift and sometimes with an amplitude increased by a constant factor greater than one — in the last case the filter is called an *amplifier*) or will be practically zero, depending on whether or not ω belongs to the filter pass band. The basic types of filters used in practice are *low-pass filters*, which pass all oscillations with frequencies less than a certain critical frequency ω_0 and suppress all oscillations with frequencies greater than ω_0, *high-pass filters*, which pass all oscillations with frequencies greater than a certain critical frequency ω_0 and eliminate all other harmonic oscillations,

and *band-pass filters*, which pass only oscillations with frequencies ω lying in a given interval (pass band) $\omega_1 \leqslant \omega \leqslant \omega_2$.* The reader can find detailed information on these various types of filters, including the practical design methods, circuit diagrams, and so on, in many books and manuals on electrical engineering and circuit theory.[43] The practical models of mechanical and acoustic filters are also considered in several books.[44]

The possibility of realization of linear transformations $\mathfrak{L}$ of stationary random processes $X(t)$ by using specific devices — linear filters — makes it possible to give a real physical meaning to the spectral representation of $X(t)$. Suppose that $X(t)$ is a real stationary process with the spectral representation (2.61) and that $x(t)$ is its realization. To be explicit, we assume that $x(t)$ corresponds to a fluctuating electrical voltage. If we apply this voltage to the input of an electrical filter with pass band $\omega_1 \leqslant \omega \leqslant \omega_2$, then the output voltage will be a realization of the random process $X_{\omega_1 \omega_2}(t)$ with spectral representation

$$X_{\omega_1 \omega_2}(t) = \int_{-\omega_2}^{-\omega_1} e^{i\omega t}\, dZ(\omega) + \int_{\omega_1}^{\omega_2} e^{i\omega t} dZ(\omega)$$

(2.170)

$$= 2\mathrm{Re}\int_{\omega_1}^{\omega_2} e^{i\omega t} dZ(\omega),$$

where $\mathrm{Re}X$ denotes the real part of the complex number X. The process $X_{\omega_1 \omega_2}(t)$ is the component of the process $X(t)$ corresponding to the spectral interval (or frequency range) $\omega_1 \leqslant \omega \leqslant \omega_2$. (In other words, $X_{\omega_1 \omega_2}(t)$ is the contribution of the spectral interval $\omega_1 \leqslant \omega \leqslant \omega_2$ to the values of $X(t)$.) Hence, we see that separate spectral components $X_{\omega_1 \omega_2}(t)$ (more precisely, their realizations $x_{\omega_1 \omega_2}(t)$) can be effectively singled

---

*It can be shown that an ideal band-pass (or low-pass, or high-pass) filter which completely suppresses all the oscillations with frequencies outside its pass band, is physically unrealizable. This circumstance, however, is unimportant in practice, since filters can be constructed which approximate a given ideal filter to any preassigned degree of accuracy (see Note 43).

out from the process $X(t)$ (i.e. its realization $x(t)$) by using suitably chosen band-pass filters.

Suppose now that the width of the filter pass band $\Delta\omega = \omega_2 - \omega_1$ is sufficiently narrow in comparison with the "mean frequency" $\omega_0 = (\omega_1 + \omega_2)/2$. Then we can approximate the component $X_{\omega_1\omega_2}(t) = X(\omega_0,\Delta\omega;t)$ in any preassigned finite-time interval $0 \leqslant t \leqslant T$, with any preassigned degree of (mean square) accuracy, by a random harmonic oscillation $\text{Re}[2Z(\Delta\omega)e^{i\omega_0 t}]$ where $Z(\Delta\omega) = Z(\omega_2) - Z(\omega_1)$. Note that in the case where $X(t)$ has a continuous spectrum, the realization $x(\omega_0,\Delta\omega;t)$ of $X(\omega_0,\Delta\omega;t)$ will actually not be strictly periodic for any $\Delta\omega$. (Moreover, as a realization of a stationary process with a continuous spectrum, it will not even be expandable into a Fourier or Fourier–Stieltjes integral.) Nevertheless, if $\Delta\omega/\omega_0$ is small enough, the output oscillation $x(\omega_0,\Delta\omega;t)$ will be approximated with a high degree of accuracy by a sine oscillation of angular frequency ω_0 within a large number of periods, i.e. it will behave as if $x(\omega_0,\Delta\omega;t)$ were representable in the form of a Fourier integral over a narrow frequency band $\Delta\omega$. It is clear, however, that the amplitude and phase of the corresponding sinusoid will vary from one realization of $X(\omega_0,\Delta\omega;t)$ to another. Besides, if $X(t)$ has a continuous spectrum, then the amplitude and phase of the oscillation $x(\omega_0,\Delta\omega;t)$ in two time intervals which are sufficiently far apart will also in general be different.

Being able to single out the spectral components $X_{\omega_1\omega_2}(t)$ with the aid of filters, we obtain the possibility for experimental determination of the spectral distribution function $F(\omega)$ and the spectral density $f(\omega)$ of the stationary random process $X(t)$. Note that

$$(2.171) \qquad \begin{aligned} \langle [X_{\omega_1\omega_2}(t)]^2 \rangle &= [F(\omega_2) - F(\omega_1)] + [F(-\omega_1) - F(-\omega_2)] \\ &= 2[F(\omega_2) - F(\omega_1)]. \end{aligned}$$

Therefore, the difference $F(\omega_2) - F(\omega_1)$ (which coincides with $\int_{\omega_1}^{\omega_2} f(\omega)d\omega$ if a density $f(\omega)$ exists) is equal to half the mean square of the spectral component $X_{\omega_1\omega_2}(t)$ corresponding to

the frequency range $\omega_1 \leqslant \omega \leqslant \omega_2$ (i.e. to half the average energy dissipated per unit time by the aggregate of all oscillations making up $X(t)$ which have frequencies between ω_1 and ω_2). Since the increments $Z(\omega)$ are uncorrelated, the spectral components of $X(t)$ corresponding to nonintersecting frequency intervals are also uncorrelated, and the mean square of their sum equals the sum of the mean squares of the terms. Hence, the total power of $X(t)$ equals the sum of the powers of all the spectral components corresponding to the partition of the frequency axis into nonintersecting intervals. This makes it clear that the spectral distribution function $F(\omega)$ determines the *distribution of the average power of the process $X(t)$ over the frequency spectrum.* Accordingly, the spectral density $f(\omega)$ can also be called the *power spectral density* of the process $X(t)$; in the applied literature it is usually simply called the *power spectrum* of the process $X(t)$ (see footnote on p. 105).

It will be shown in Sec. 17 that under wide regularity conditions the mean square of the stationary process $X(t)$ can be determined to a high accuracy by time averaging of the square of a single realization $x(t)$. It follows that the spectral distribution function $F(\omega)$ and the spectral density $f(\omega)$ can both be determined experimentally by analyzing the output oscillations of a number of band-pass filters; the realization $x(t)$ being applied to the inputs. The study of the methods of experimental determination of the spectral distribution functions and spectral densities of stationary processes is very important for many applications, and we shall treat it in detail in Secs. 18 and 19 of this book. Here, we shall confine ourselves to some brief remarks relating to the case where the spectral density $f(\omega)$ exists. Let us apply the realization $x(t)$, regarded as a fluctuating electrical voltage or current, to the input of an electrical low-pass filter, which only passes (practically without distortion) oscillations with frequencies less than ω. Let us also attach a wattmeter, which measures the average dissipated power (i.e. the square of the current strength or the voltage), to the filter output. Then, the needle of the instrument will register just the value proportional to

$$\int_0^\omega f(\omega')d\omega' = F(\omega) - F(0).$$ (The constant of proportionality is

determined, of course, by wattmeter calibration.) The time averaging of instantaneous power required to obtain precisely

the *mean square* of the filter output is always carried out, to some degree, by the instrument itself, because of its inertia. If, however, this is insufficient, then additional time averaging of the wattmeter readings can easily be performed. By varying ω we obtain $F(\omega) - F(0)$ as a function of ω, and then by differentiating the obtained function we find $f(\omega)$. Instead of a continuous variation of the critical frequency ω of a low-pass filter we can also use a collection of band-pass filters with narrow pass bands, which permit us to determine approximately the values of $f(\omega)$ at the mean frequencies of the pass bands of all the filters. Such a collection of narrow-band filters is often assembled as a single instrument, which is called a *spectrum* (or *wave*) *analyzer;* cf. Sec. 19.

We now return to Khinchin's formula (2.67) for the correlation function $B(\tau)$ satisfying the condition (2.66) or (2.68). This formula relates a very important statistical characteristic of the process $X(t)$ — its correlation function — with a very important physical characteristic — the spectral
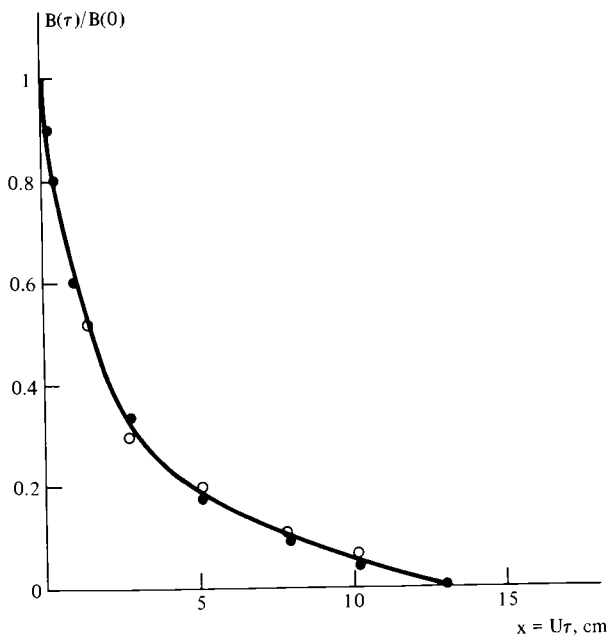
(a)

density describing the frequency distribution of the power of $X(t)$. Both characteristics can be measured directly by time averaging of the product $x(t + \tau)x(t)$ and by using a spectral analyzer (see Secs. 17-19 for more details). Therefore both Khinchin's formula (2.67) and the inverse formula (2.84) can be verified experimentally. One of the first attempts at such
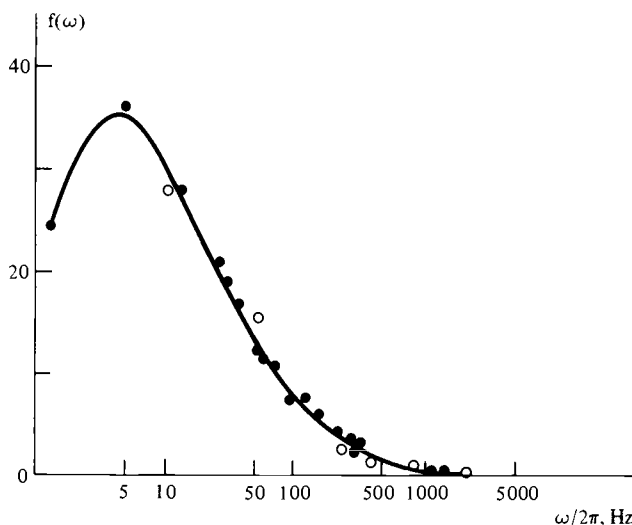
(b)

Fig. 33. (a) The normalized correlation function for fluctuation of the longitudinal velocity component of a turbulent flow in a wind tunnel behind a grid (Taylor, 1938). The abscissa is the time multiplied by the mean velocity U; the solid circles (●) correspond to the measured values of the correlation function and the light circles (o) represent the values of the correlation function calculated by using formula (2.67) and the measured values of the spectral density. (b) Similar comparison (for a turbulent flow of the same kind) of the measured values of the spectral density for longitudinal velocity fluctuations (●) with the values of $f(\omega)$ calculated by using formula (2.84) and the measured values of the correlation function (o) (Favre et al., 1954).

verification of Khinchin's formula was made in 1938 by G.I.
Taylor, an eminent English scientist, one of the founders of
modern turbulence theory.   (Khinchin's paper of 1934 was
apparently unknown to Taylor, who referred to the related
work by Wiener, 1933.) This verification applied to the case
of wind tunnel turbulence produced by a special grid. Figure
33(a) shows the directly measured values of the normalized
correlation function $R(\tau) = \langle X(t + \tau)X(t)\rangle / \langle X^2(t)\rangle$, where $X(t)$
represents the fluctuations of the longitudinal velocity (i.e.,
the velocity component along the axis of the wind tunnel),
together with the values of $R(\tau)$ calculated by using (2.67)
and the data on $f(\omega)$ obtained with the aid of a collection of
filters.[45] Figure 33(b) shows the results of a similar
verification of (2.84), which also relates to turbulent velocity
fluctuations.   As we see, the agreement between directly
measured and calculated values in both experiments presented
in Fig. 33(a),(b) is quite satisfactory.

Experiments of this type were carried out many times
with reference to random fluctuations $X(t)$ of quite a
different origin.   They always led to excellent agreement
between theory and experiment. At present, such experiments
are widely used for reciprocal control of instruments used for
measuring spectral densities $f(\omega)$ and correlation functions
$B(\tau)$. Khinchin's formula is then always considered as a
firmly established theorem relating the results of these two
types of measurement. Note also that the same formula is
widely used now in practice for determining the spectral
densities from measured values of the correlation functions
or vice versa (for more details see Secs. 18 and 19).

*The results of this section shed additional light on
one remark made during the discussion of "white noise".
We stated on p. 117 that in each specific problem only
some limited frequency band is usually of interest. The
meaning of this remark is that applied problems in the
theory of random processes are usually associated with
investigation into some linear transformations of the
processes concerned (i.e. their passage through one or
several "linear systems"). But real linear systems
practically always have a limited pass band, outside of
which their gain $A(\omega)$ is negligibly small. All the
frequencies ω contained in the pass bands of all the
linear systems transforming the given fluctuations $X(t)$

precisely constitute the "limited frequency band, which is alone of interest for the problem at hand". Hence if the spectral density $f(\omega)$ of the process $X(t)$ remains practically constant within this band, the process can readily be replaced by an idealized white-noise process.

The method described above for "experimental determination" of the correlation function by measuring the spectral density with the aid of filters can also be used as a basis for an exact proof of both Bochner's theorem on positive definite functions and the spectral representation theorem for stationary processes. We shall begin with Bochner's theorem and briefly outline how such a proof is accomplished.

For simplicity we restrict ourselves to the case of a real function $B(\tau)$ which falls off with the increase of τ so rapidly that it can be written as a Fourier integral (2.67) of a continuous function $f(\omega)$. Then, the basic assertion of Bochner's theorem is that every positive definite function $B(\tau)$ of the type described has a non-negative Fourier transform $f(\omega)$. (The converse assertion that the non-negativity of $f(\omega)$ implies the positive definiteness of $B(\tau)$ can be proved very simply by substituting eq. (2.67) onto the left-hand side of (1.59).) Thus, suppose that $B(\tau)$ is positive definite, but $f(\omega)$ is not non-negative. This means that the Fourier transform $f(\omega)$ of the positive definite function $B(\tau)$ is negative at some point $\omega = \omega_0$ (and, hence, at all the points of some sufficiently narrow band $\Delta\omega = [\omega_1, \omega_2]$, containing the point ω_0). We now use the result from Sec. 4 (see p. 58) which states that every positive definite function $B(\tau)$ is the correlation function of a stationary random process. Consider the following linear transformation of a process $X(t)$ which has $B(\tau)$ as its correlation function:

$$(2.172) \quad Y(t) = \mathfrak{L}\{X(t)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{i\omega_2 u} - e^{i\omega_1 u}}{iu} X(t - u) du.$$

It is easy to see that transformation of $X(t)$ into $Y(t)$ is equivalent to passing $X(t)$ through an ideal filter with a narrow pass band $\Delta\omega = [\omega_1, \omega_2]$; cf., e.g., eq. (2.37') in Note 43. (The fact that the filter (2.172) is physically

unrealizable and has a complex weighting function is, of course, of no importance for our arguments.) In particular, we can easily deduce from (1.55) that $\langle |Y(t)|^2 \rangle$ = $\int_{\omega_1}^{\omega_2} f(\omega)d\omega$ (cf. derivation of (2.16') in Note 17). The quantity $\langle |Y(t)|^2 \rangle$ is, however, inherently non-negative. Therefore, $\int_{\omega_1}^{\omega_2} f(\omega)d\omega$ is also non-negative and this contradicts the assumption that $f(\omega_0) < 0$. The general case of Bochner's theorem, dealing with arbitrary positive definite functions, can be proved similarly.[46]

Let us now consider the possibility of using linear filters (i.e transformations) to prove the spectral representation theorem for stationary processes. This possibility is in fact been demonstrated in Note 17. Indeed, formula (2.9') specifies a definite linear transformation of the process $X(t)$. Hence, the idea of the proof based on this formula is to select a transformation converting $X(t)$ to a function $Z(\omega)$ and then prove that this function possesses all the required properties and can be used for representing $X(t)$ as a Fourier—Stieltjes integral. Now we can make the idea of this proof still more transparent by slightly modifying the arguments. Instead of the linear transformation (2.9') we begin with a related linear transformation (2.172) describing the ideal filter with a pass band $[\omega_1,\omega_2] = \Delta\omega$. Then it is easy to prove that the processes $Y_1(t),Y_2(t), ..., Y_n(t)$, which correspond to nonintersecting pass bands $\Delta_1\omega, \Delta_2\omega, ..., \Delta_n\omega$, are pairwise uncorrelated. Moreover, if $\Delta_1\omega, \Delta_2\omega, ..., \Delta_n\omega$ form a partition of the whole spectrum of $X(t)$ (i.e. the frequency interval in which all the power of the process is stored up), then

$$X(t) = \sum_{k=1}^{n} Y_k(t).$$

Suppose first of all that all the power of the process is stored up in a finite frequency interval $-\Omega \leqslant \omega \leqslant \Omega$. Let n tend to infinity and the lengths of all the subintervals $\Delta_1\omega, ..., \Delta_n\omega$ partitioning the interval $[-\Omega,\Omega]$ tend to zero. It is clear that a spectral component $Y_k(t)$

= $X_{\Delta_k \omega}(t)$, which corresponds to a very narrow band $\Delta_k \omega$ containing the point ω'_k, can be represented as $Y_k(t) = Z_k \exp(i\omega'_k t)$ where the amplitude Z_k does not depend appreciably on t. Hence,

$$X(t) = \lim_{n \to \infty} \sum_{k=1}^{n} Y_k(t) = \lim_{n \to \infty} \sum_{k=1}^{n} e^{i\omega'_k t} Z_k.$$

In the general case of an unbounded spectrum of $X(t)$ we may also begin with the partition of the finite interval $[-\Omega, \Omega]$, and then let Ω tend to infinity to obtain the spectral representation (2.58) − (2.60).[47]*

12.   Examples of Linear Transformations of Stationary Processes

Example 1. *Differentiation.* In Sec. 4 we have defined the (mean square) derivative $X'(t) = \lim_{\Delta \to \infty} [X(t + \Delta) - X(t)]/\Delta$ of a

stationary random process $X(t)$. It is clear that differentiation is a special linear transformation of the form (2.158). If $X(t)$ has spectral representation (2.61), then

$$\frac{X(t + \Delta) - X(t)}{\Delta} = \int_{-\infty}^{\infty} \frac{e^{i\omega(t+\Delta)} - e^{i\omega t}}{\Delta} \, dZ(\omega).$$

It is a simple consequence of the definition of the integral (2.61) as the limit of the sum (2.62) that the limit $\Delta \to 0$ in the preceding equation can be taken under the integral, provided that the limit $X'(t)$ exists and that $\langle |X'(t)|^2 \rangle$ is finite. In this case, we have

$$(2.173) \quad X'(t) = \int_{-\infty}^{\infty} e^{i\omega t} i\omega \, dZ(\omega).$$

Equation (2.173) is just the spectral representation of the derivative $X'(t)$ of the stationary random process $X(t)$. At the same time this equation shows that the differentiation transformation has the transfer function $H(\omega) = i\omega$ (i.e. the gain $A(\omega) = |H(\omega)| = |\omega|$ and the phase shift $\psi(\omega) = \arg H(\omega) = \pi/2$ for $\omega \geqslant 0$ and $= -\pi/2$ for $\omega < 0$). It is obvious that differentiation is in fact a (non-ideal) high-pass filtering, since it attenuates (decreases) the power in the low frequency

range but magnifies the power in higher frequencies.

Equations (2.166) and (2.164), as applied to differentiation, take the form:

$$(2.174) \quad \langle X'(t + \tau)\overline{X'(t)} \rangle = \int_{-\infty}^{\infty} e^{i\omega\tau}\omega^2 dF(\omega),$$

$$(2.174a) \quad F_{X'X'}(\omega) = \int_{-\infty}^{\omega} \omega'^2 dF(\omega').$$

Comparing (2.174) and (2.52), we find that the right-hand side of (2.174) coincides with $- B''(\tau)$, so that this agrees with (1.50). The general condition (2.161), as applied to differentiation, shows that

$$(2.175) \quad \int_{-\infty}^{\infty} \omega^2 dF(\omega) < \infty$$

for every differentiable process $X(t)$. (This is also clear, since the spectral distribution function $F_{X'X'}(\omega)$ must be bounded.) The requirement (2.175) is clearly equivalent to the requirement for the existence of $B''(0)$, the second derivative of the correlation function $B(\tau)$ at the point $\tau = 0$ (because the left-hand side of (2.175) is equal to $-B''(0)$).

By differentiating the stationary process $X'(t)$ we obtain the second derivative $X''(t)$ of the process $X(t)$, and similarly, we can also form higher derivatives of $X(t)$. It is easy to see that the nth derivative $X^{(n)}(t)$ is represented by the formula

$$(2.176) \quad X^{(n)}(t) = \int_{-\infty}^{\infty} e^{i\omega t}(i\omega)^n dZ(\omega),$$

i.e. it is the output process of a linear transformation with a transfer function $H(\omega) = (i\omega)^n$. Such a linear transformation of $X(t)$ is possible (i.e. $X^{(n)}(t)$ exists) if, and only if, the condition

$$(2.177) \quad \int_{-\infty}^{\infty} \omega^{2n} dF(\omega) < \infty$$

is satisfied, which is equivalent to the condition that $B^{(2n)}(0)$, derivative of order $2n$ of $B(\tau)$ at the point $\tau = 0$, exists. Equation (2.176) implies that the correlation function of the process $X^{(n)}(t)$ has the representation

$$(2.178) \quad \langle X^{(n)}(t + \tau)\overline{X^{(n)}(t)} \rangle = \int_{-\infty}^{\infty} e^{i\omega\tau}\omega^{2n} dF(\omega).$$

The right-hand side of (2.178) coincides with the function

$(-1)^n B^{(2n)}(\tau).$

Example 2. *Differential operators with constant coefficients.* If a process $X(t)$ satisfies the condition (2.177) (i.e. has all derivatives up to the nth order, inclusive), then we can apply to $X(t)$ the following linear, time-invariant transformation:

$$(2.179) \quad Y(t) = \mathfrak{L}\{X(t)\} = a_0 \frac{d^n X(t)}{dt^n} + a_1 \frac{d^{n-1} X(t)}{dt^{n-1}} + \; ... \; + a_n X(t).$$

The transfer function of this transformation is clearly equal to

$$(2.180) \quad H(\omega) = a_0(i\omega)^n + a_1(i\omega)^{n-1} + \; ... \; + a_n.$$

The formulae for the correlation and spectral distribution functions of a process $Y(t)$ and the spectral representation of this process can easily be deduced from (2.166), (2.164), and (2.162).

Example 3. *Smoothing operations: uniformly and exponentially weighted averaging.* We have already seen that the transfer function $H(\omega)$ of an integral transformation (2.147) is the Fourier transform of the weighting function $h(u)$. We shall now consider two important particular cases of integral transformations.

In practice, cases are rather common where only low-frequency (i.e. long-period) variations of a timeseries $X(t)$ are of interest, while the high-frequency fluctuations of $X(t)$ are unimportant or very unreliable (e.g., greatly distorted by noise or even representing the noise alone). In such cases preliminary elimination of the unwanted high-frequency fluctuations by *smoothing* (i.e. low-pass filtering of some kind) seems appropriate. In particular, if all the oscillations with periods of the order of a given time T or less are unimportant, then *uniformly weighted averaging* of the data over a time T is often used. This averaging transforms $X(t)$ into a new process

$$(2.181) \quad Y_T(t) = \frac{1}{T} \int_{t-T}^{t} X(s)ds = \frac{1}{T} \int_0^T X(t-u)du.$$

Transformation (2.181) can also be used for an approximate description of the smoothing of a time series produced by the instrument which measures the values of $X(t)$. Such instrumental smoothing is due to inertia of the measuring instrument (usually characterized by its "time constant"). In

cases where this smoothing is substantial we can often assume
that the reading $Y_T(t)$ of an instrument with a time constant
T is related to the true value of $X(t)$ by (2.181).

$Y_T(t)$ is an integral linear transformation of the process
$X(t)$ with a weighting function

$$(2.182) \quad h(u) = \begin{cases} 1/T \text{ for } 0 \leqslant u \leqslant T, \\ 0 \quad \text{otherwise} \end{cases}$$

and a transfer function

$$(2.182a) \quad H(\omega) = \frac{1}{T} \int_0^T e^{-i\omega u} du = e^{-i\omega T/2} \frac{\sin(\omega T/2)}{\omega T/2}$$

(cf. (2.153)).   Hence in this case $A(\omega) = |H(\omega)| = \left| \dfrac{\sin(\omega T/2)}{\omega T/2} \right|$.

The gain $A(\omega)$ is close to unity for $|\omega| \leqslant 2/T$, but as $|\omega|$
increases it begins to fall off rapidly, vanishing at $\omega = \pm 2\pi/T$
(and again at $\omega = 2k\pi/T$, $k = \pm 2, \pm 3, ...$).   In the intervals
between two adjacent zeros in the positive or negative
semi-axis the function $A(\omega)$ takes small positive values (cf.
Fig. 28(b) on p. 131, where a function $f(\omega)$ proportional to
$[A(\omega)]^2$ is plotted). This behavior of $A(\omega)$ shows, of course,
that the time averaging (2.181) is a particular non-ideal
low-pass filtering.

In addition to data smoothing by the uniformly weighted
averaging over a finite time interval, *exponentially weighted
averaging* is also used in some applications. This type of
averaging is given by the formula

$$(2.183) \quad Y(t) = \frac{1}{T} \int_0^\infty e^{-u/T} X(t - u) du.$$

In   particular,   (2.183)   determines   very   accurately   the
instrumental smoothing of the measured time series $X(t)$ in those
(rather numerous) cases where the measuring instrument (with a
time constant T) is described by a first-order differential
equation (cf. p. 170–172).[48]

The weighting function $h(u)$ and the transfer function $H(\omega)$ of
the transformation (2.183) are of the form

$$(2.184) \quad h(u) = \begin{cases} T^{-1} e^{-u/T} \text{ for } t > 0, \\ 0 \quad\quad\quad \text{for } t < 0, \end{cases}$$

(2.184a) $H(\omega) = \dfrac{1}{T} \displaystyle\int_0^\infty e^{-i\omega u - u/T} du = \dfrac{1}{1 + i\omega T}$.

Hence,

(2.185) $Y(t) = \displaystyle\int_{-\infty}^\infty \dfrac{e^{i\omega t}}{1 + i\omega T} dZ(\omega), \qquad F_{YY}(\omega) = \dfrac{F(\omega)}{1 + \omega^2 T^2}$

and

$$A(\omega) = |H(\omega)| = \dfrac{1}{(1 + \omega^2 T^2)^{1/2}} .$$

The function $[A(\omega)]^2$ is proportional to the spectral density $f(\omega)$ plotted in Fig. 19(b), assuming that $\alpha = 1/T$. It is clear that transformation (2.183) also describes a particular (non-ideal) low-pass filtering.

Example 4. *Detrending operations.* Consider now the case where high-frequency (short-period) fluctuations are of primary interest, while low-frequency variations are not studied at all. Such a situation often occurs when sluggish low-frequency variations differ from high-frequency ones in their origin and/or properties (e.g., high-frequency variations are stationary, but low-frequency ones are not) or when the time of recording is insufficient for investigation of the long-period components of the given time series $X(t)$.

When only high-frequency variations are of interest, the unwanted low-frequency variations are often formally considered as *trends* (systematic variations) of $X(t)$. To eliminate the low-frequency variations (trends), high-pass filtering of the data is often used. For example, we can subtract the moving average (2.181) from the values of the process $X(t)$. Such subtraction leads to a new process:

(2.181a) $Y^{(1)}(t) = X(t) - \dfrac{1}{T} \displaystyle\int_{t-T}^t X(s) ds.$

Transformation (2.181a) is used rather often, in particular, in meteorology, to eliminate the trends ("slow variations of the mean meteorological field") in order to single out the high-frequency turbulent fluctuations which are usually studied separately from the study of the mean field.[49]

The weighting function of the transformation (2.181a) is a generalized function $h^{(1)}(u) = \delta(u) - h(u)$, where $h(u)$ is the function (2.182). Therefore (2.182a) implies that the transfer function $H^{(1)}(\omega)$ of the transformation (2.181a) is

(2.182b) $H^{(1)}(\omega) = 1 - e^{-i\omega T/2}\,\dfrac{\sin(\omega T/2)}{\omega T/2}$.

The corresponding gain $A^{(1)}(\omega) = |H^{(1)}(\omega)| = 1 - \dfrac{\sin\omega T}{\omega T/2} + \dfrac{\sin^2(\omega T/2)}{(\omega T/2)^2}$

vanishes at frequency $\omega = 0$, but increases rapidly when $|\omega|$ increases and then begins to oscillate about its limiting value $A^{(1)}(\infty) = \lim\limits_{|\omega|\to\infty} A^{(1)}(\omega) = 1$.   Such behavior of $A^{(1)}(\omega)$ shows

that the application of the transformation (2.181a) is a particular (non-ideal) high-pass filtering.

Instead of subtracting the uniformly weighted moving average (2.181) we can also subtract from $X(t)$ the exponentially weighted moving average (2.183). Then we obtain the process

(2.183a) $Y^{(2)}(t) = X(t) - \dfrac{1}{T}\int_0^T e^{-u/T}X(t-u)\,du$.

The weighting function of the transformation (2.183a) is a generalized function $h^{(2)}(t) = \delta(t) - h(t)$, where $h(t)$ is given by (2.184).   The corresponding transfer function $H^{(2)}(\omega)$ has the

form $H^{(2)}(\omega) = 1 - \dfrac{1}{1 + i\omega T} = \dfrac{i\omega T}{1 + i\omega T}$; hence $A(\omega) = \dfrac{|\omega| T}{(1 + \omega^2 T^2)^{1/2}}$. This gain also vanishes at $\omega = 0$ and tends

to 1 as $|\omega| \to \infty$.

Example 5. *Solutions of differential equations as linear transformations.* Let $Y_1(t) = \mathcal{L}_1\{X(t)\}$ be a linear transformation of a stationary process $X(t)$.   Apply now a second linear transformation $\mathcal{L}_2$ to Y_1.   Then $Y_2(t) = \mathcal{L}_2\{Y_1(t)\} = \mathcal{L}_2\{\mathcal{L}_1[X(t)]\}$ is also a linear transformation of the process $X(t)$. This new linear transformation is usually called the *product* of the transformations $\mathcal{L}_2$ and $\mathcal{L}_1$ and is denoted by $\mathcal{L} = \mathcal{L}_2\mathcal{L}_1$.   It follows readily from (2.150) that the transfer function $H(\omega)$ of the product $\mathcal{L} = \mathcal{L}_2\mathcal{L}_1$ is $H(\omega) = H_2(\omega)H_1(\omega)$, where $H_1(\omega)$ and $H_2(\omega)$ are the transfer functions of $\mathcal{L}_1$ and $\mathcal{L}_2$.

Let $\mathcal{E}$ be the *identity transformation*, defined by the condition $\mathcal{E}\{X(t)\} = X(t)$ (i.e. $\mathcal{E}$ does not affect the process $X(t)$ in any way). It is clear that the transfer function of $\mathcal{E}$ is $H_0(\omega) \equiv 1$.   The transformation $\mathcal{L}_1^-\{X(t)\} = Y(t)$ is said to be inverse to the transformation $\mathcal{L}_1$ if $\mathcal{L}_1\mathcal{L}_1^- = \mathcal{E}$, so that $\mathcal{L}_1\{Y(t)\} = X(t)$. The transfer function $H_1^-(\omega)$ of the transformation $\mathcal{L}_1^-$ is

obviously given by the equation $H_1^-(\omega) = 1/H_1(\omega)$, where $H_1(\omega)$ is the transfer function of $\mathfrak{L}_1$.

We have already considered a transformation $\mathfrak{L}$ described by the differential operators (2.179) and having a transfer function

(2.180a) $H(i\omega) = P(i\omega), \quad P(z) = a_0 z^n + a_1 z^{n-1} + ... + a_n.$

Let the transformation $\mathfrak{L}^-\{X(t)\} = Y(t)$ be the inverse to such an $\mathfrak{L}$. Then the transfer function of $\mathfrak{L}^-$ is

(2.186) $H^-(i\omega) = \dfrac{1}{P(i\omega)}$.

The condition $\mathfrak{L}\{Y(t)\} = X(t)$ takes, in this case, the form

(2.187) $a_0 \dfrac{d^n Y(t)}{dt^n} + a_1 \dfrac{d^{n-1} Y(t)}{dt^{n-1}} + ... + a_n Y(t) = X(t).$

Hence the transformed process $Y(t)$ is simply the stationary solution of the differential equation (2.187).

Let us assume that the process $X(t)$ has a continuous and everywhere positive spectral density $f(\omega)$. In this case the applicability of the transformation $\mathfrak{L}^-$ to $X(t)$ implies that

(2.188) $\displaystyle\int_{-\infty}^{\infty} \dfrac{f(\omega)}{|P((i\omega)|^2} \, d\omega < \infty.$

Hence the polynomial $P(i\omega)$ cannot have real zeros, i.e. the equation $P(\omega) = 0$ cannot have roots with a zero real part. Conversely, if the last condition is fulfilled, the condition (2.188) is certainly valid for any spectral density $f(\omega)$. Moreover the function $[P(i\omega)]^{-1}$ has in this case a Fourier transform $h(u)$. Thus the transformation $\mathfrak{L}^-$ is an integral transformation, and the output process $Y(t)$ can be represented either as

(2.189) $Y(t) = \displaystyle\int_{-\infty}^{\infty} e^{it\omega} \dfrac{dZ(\omega)}{P(i\omega)}$

or as

(2.190) $Y(t) = \displaystyle\int_{-\infty}^{\infty} h(u) X(t - u) du, \quad h(u) = \dfrac{1}{2\pi} \displaystyle\int_{-\infty}^{\infty} \dfrac{e^{i\omega u}}{P(i\omega)} d\omega.$

Equation (2.189) implies that

(2.189a) $f_{YY}(\omega) = \dfrac{f(\omega)}{|P(i\omega)|^2}$.

Equations (2.189) and (2.189a) can also be easily deduced from (2.187) if the spectral representation of the processes on both sides of the last equation is considered. Hence, subject to the condition (2.188) (i.e. in cases where equation $P(z) = 0$ has no pure imaginary roots), equation (2.187) always has precisely one stationary solution $Y(t)$ given by (2.189) or (2.190).

The representation of the random process $Y(t)$ in the form (2.190) reveals the essential difference between the case where all the roots of the equation $P(z) = 0$ have negative real parts (i.e. they lie in the left half-plane of the complex variable) and the one where this equation also has roots with a positive real part. Indeed, it is easy to show that in the former case the function $h(u)$ is equal to zero for all negative values of u, so that

(2.190a) $Y(t) = \displaystyle\int_{0}^{\infty} h(u)X(t - u)du.$

Hence, the transformation $\mathcal{L}^-$ is physically realizable in this case. Moreover, if all the roots of $P(z)$ have negative real parts, the function $h(u)$ falls off exponentially, as $u \to \infty$, and for any initial conditions $Y(t_0), Y'(t_0), ..., Y^{(n-1)}(t_0)$ at the instant t_0 the solution of (2.187) tends to the stationary solution (2.189) (i.e. (2.190a)) as $t - t_0 \to \infty$.[50] If, however, at least one of the roots of the equation $P(z) = 0$ has a positive real part, then the stationary solution $Y(t)$ also depends on the variables $X(t')$ where $t' > t$.   In this case the stationary solution is more artificial and is generally useless for applications (see, in particular, the examples given below).

Consider now two particularly simple types of differential equations of the form (2.187).

(a) First-order equations. Suppose that

(2.191) $\dfrac{dY(t)}{dt} + aY\, t) = X(t),$

where $X(t)$ is a process having the spectral density $f(\omega)$. Restricting ourselves to real processes, we must naturally suppose that the coefficient a is also real. Then the condition for the absence of roots of the equation $z + a = 0$ with a zero real part reduces to the condition $a \neq 0$. If $a > 0$, the only

root $z = -a$ has a negative real part, and if $a < 0$, a positive one.

It is easy to show that the solution of (2.191) with the initial condition $Y(t_0) = Y_0$ has the form

$$Y(t|Y_0,t_0) = Y_0 e^{-a(t-t_0)} + \int_{t_0}^{t} e^{-a(t-s)} X(s) ds$$

(2.192)

$$= Y_0 e^{-a(t-t_0)} + \int_{0}^{t-t_0} e^{-au} X(t-u) du.$$

If $a < 0$, then, as $t - t_0$ increases, this solution increases exponentially and does not tend to any limit. Moreover, in this case (2.192) is a nonstationary random process for any Y_0. The stationary solution $(2.189) - (2.190)$ of (2.191) when $a < 0$ has the form

$$(2.193) \quad Y(t) = \int_{0}^{\infty} e^{au} X(t+u) du$$

and has no relation to the solution (2.192) of the initial value problem. On the other hand, if $a > 0$, then the stationary solution $(2.189) - (2.190)$ of (2.191) has the form

$$(2.194) \quad Y(t) = \int_{0}^{\infty} e^{-au} X(t-u) du$$

and the solution (2.192) of the initial value problem tends to this process as $t - t_0 \to \infty$ for any Y_0. (The convergence to the limit of the form (2.194) as $t - t_0 \to \infty$ takes place here for both random variables $Y(t|Y_0,t_0)$ and the numerical functions $y(t|y_0,t_0)$, which are realizations of processes $Y(t|Y_0,t_0)$.) The same stationary solution can be obtained as a particular case of $Y(t|Y_0,t_0)$ if we

put $Y_0 = \int_{0}^{\infty} e^{-as} X(t_0 - s) ds$. It is easy to see that the spectral

density $f_{YY}(\omega)$ of the process (2.194) has the form $f_{YY}(\omega) = f(\omega)/(a^2 + \omega^2)$, which agrees with (2.189a).

The linear transformation (2.194) coincides, of course, with the exponentially weighted averaging (2.183), the only difference being that a was replaced in (2.183) by $1/T$, and $X(t)$ by $X(t)/T$. Such replacements convert the first-order equation (2.191) into the equation describing many measuring instruments (where $Y(t)$ is the instrument reading, $X(t)$ the measured variable, and T the time

constant of the instrument); see, e.g., the sources indicated in Note 48. Thus, the instrumental smoothing of the measured variables, produced by such instruments, is precisely described by (2.183).

When $|P(i\omega)|^{-2} = |i\omega + a|^{-2} = (\omega^2 + a^2)^{-1}$, the condition (2.188) will be fulfilled even for $f(\omega) = f_0 = \text{const}$. Hence $X(t) = cE(t)$ may be an idealized "white noise" in this case. Thus, it is also expedient to consider the equation

(2.191a) $\dfrac{dY(t)}{dt} + aY(t) = cE(t),$

which is a particular case of (2.191). The differential equation (2.191a) is a continuous time analogue of the first-order difference equation (2.10).* Therefore the stationary solution $Y(t)$ of (2.191a) is often called the *first-order autoregressive process*. This process is given by (2.194), where $X(t) = cE(t)$. Equation (2.194) in this case can also be rewritten in a more convenient form:

(2.194a) $Y(t) = \int_{-\infty}^{t} e^{-a(t-s)} dW(s), \ W(s) = \int_0^s X(s')ds' = c\int_0^s E(s')ds'.$

Substituting now the equation $\langle E(s')E(s'') \rangle = \delta(s' - s'')$ into the formula of the form (1.56) for the mean product of two integrals of the random process $X(t) = cE(t)$, we obtain

(2.195) $\langle W(t_1)W(t_2) \rangle = \begin{cases} c^2 \min (t_1, t_2) & \text{for } t_1 \geqslant 0, t_2 \geqslant 0, \\ 0 & \text{for } t_1 t_2 \leqslant 0, \\ c^2 \min(-t_1, -t_2) & \text{for } t_1 \leqslant 0, t_2 \leqslant 0. \end{cases}$

It follows readily that for any $t_1 < t_2 \leqslant t_3 < t_4$

(2.195a)
$$\langle [W(t_2) - W(t_1)]^2 \rangle = c^2(t_2 - t_1),$$
$$\langle [W(t_2) - W(t_1)][W(t_4) - W(t_3)] \rangle = 0.$$

Thus, the second moments of the process $W(t)$ contain no δ-functions, i.e. $W(t)$ is not a generalized, but an ordinary non-stationary random process, whose properties are related

---

*Note that (2.10) can also be written as $\Delta X(t) + a_1 X(t) = c_1 E(t)$ where $\Delta X(t) = X(t) - X(t - 1)$, $a_1 = (1 - a)/a$, and $c_1 = c/a$.

to those of the function $Z(\omega)$ entering the spectral representation of a stationary process $X(t)$. The integral in the first equation (2.194a) can now be defined in the usual way (cf., e.g., (2.62)). The corresponding process $Y(t)$ is then stationary and gives the stationary solution of (2.191a).

It follows from equations (2.195a) that

$$\langle [dW(s)]^2 \rangle = c^2 ds, \quad \langle dW(s)dW(s') \rangle = 0 \text{ for } s' \neq s.$$

These two relations can be formally combined and written in the form

(2.195b) $\langle dW(s)dW(s') \rangle = c^2 \delta(s - s')dsds'$

(cf. (2.81)). Substituting (2.195b) into the equation of the form (1.56) for $\langle Y(t + \tau)Y(t) \rangle$ where $\tau \geqslant 0$ and $Y(t)$ is given by the first equation (2.194a), we now obtain

(2.196) $\langle Y(t + \tau)Y(t) \rangle = c^2 \int_{-\infty}^{t} e^{-a(\tau + 2t - 2s)}ds = \dfrac{c^2}{2a} e^{-a\tau}.$

According to (2.95) the spectral density $f_{YY}(\omega) = f_0/(a^2 + \omega^2)$ corresponds to the correlation function (2.196), as it should be by virtue of (2.189a).

The representation of the first-order autoregressive process in the form (2.194a) is similar in many respects to the representation of the first-order autoregressive sequence $X(t)$ in the form of one-sided moving average sequence (2.11). In general, stationary random processes $Y(t)$ of the form $Y(t) = \int_{-\infty}^{\infty} b(s)dW(t - s)$ or $Y(t) = \int_{0}^{\infty} b(s)dW(t - s)$, where $W(s)$ is related to the white noise $E(t)$ by the second equation (2.194a), are similar to random sequences of the form (2.6) or (2.8). Such processes are sometimes called (two-sided or one-sided) *moving average processes*; see Sec. 26.2 below.

Note that the process with the correlation function (2.196) is an ordinary stationary random process. However, this process has no (mean square) derivative. This fact is not surprising, since (2.191a) shows that $dY(t)/dt = cE(t) - aY(t)$ contains a "white noise", which is a mathematical idealization, but not an ordinary random process. Therefore authors

attempting to preserve mathematical rigor sometimes prefer
to write (2.191a) in a more complicated, but more rigorous
form (e.g. as an equation including differentials or an
integral equation).[51]

*An equation of the form (2.191a) was first proposed by
the physicist Langevin in 1908 for describing the Brownian
motion of a free particle.   Therefore this equation is often
called the *Langevin equation*.   If a small particle is
immersed in a fluid, then the surrounding medium acts on
it in two ways.   Firstly, the medium offers resistance to
the motion of the particle, which takes the form of a
frictional force equal to $-wY(t)$, where $Y(t)$ is the velocity
of the particle, and the coefficient w equals $6\pi r\mu$ for a
spherical particle of radius r immersed in a fluid of
viscosity μ. Secondly, the fluctuations in the number of
collisions of molecules of the fluid with the particle appear
as an additional "purely random" force $X_1(t) = c_1E(t)$ having
a constant   spectral   density $f(\omega) = f_0 = c_1^2/2\pi$ (cf. p.
117–120). Thus, the one-dimensional equation of motion
along any fixed direction for a particle of mass m has the
form (2.191a), where $Y(t)$ is the particle velocity, $a = w/m >$
0, and $cE(t) = X_1(t)/m = (c_1/m)E(t)$ is a new white noise with
a spectral density $c^2/2\pi = f_0/m^2$.
The solution of the Langevin equation for a Brownian
particle velocity $Y(t)$ was investigated by Uhlenbeck and
Ornstein in 1930.[52] These authors showed in particular that
according to this equation the velocity $Y(t)$ represents a
stationary random process with the correlation function
$B_{YY}(\tau) = \langle Y(t + \tau)Y(t) \rangle = (\pi f_0/wm)\exp[-(w/m)|\tau|]$ and the
spectral density $f(\omega) = f_0/(m^2\omega^2 + w^2)$. These results clearly
agree with (2.196). Moreover, the general laws of statistical
physics imply that $m\langle Y^2 \rangle/2 = kT/2$, where k is the
Boltzmann constant and T is the absolute temperature.
Hence $\pi f_0/w = kT$ and $f_0 = wkT/\pi$ – this result has already
been mentioned on p. 120.
Note that according to the Ornstein–Uhlenbeck theory
of Brownian motion the velocity $Y(t)$ of the Brownian
particle   exists   (and   has   a   definite   probability
distribution), but $Y(t)$ is nondifferentiable and hence the
acceleration of the particle has no meaning. Therefore, to
study the acceleration we have to use a more exact theory.

In the original Einstein–Smoluchowski theory (see, e.g., Note 10 to the Introduction) the inertia of the particle was neglected, i.e. a limiting case of the Langevin equation, as $m \to 0$, was used. With this approximation, $wY(t) = X_1(t) = c_1E(t)$ and $Y(t) = (c_1/w)E(t)$, i.e. $Y(t)$ is not an ordinary process but a white noise. Hence the Brownian particle does not even have a velocity according to this theory. (Recall that the particle has a zero mass.) However, the integral of the velocity, i.e. the path $W(t)$ traversed by the particle, has real meaning, and is given by the second equation (2.194a), where $X(s')$ is replaced by $Y(s)$ and $c = c_1/w$. Bearing in mind that $c_1^2 = 2\pi f_0 = 2wkT$, we find, using the first equation (2.195a), that for this approximation the mean square of the path length traversed by the particle in time τ is given by the formula

$$\langle[W(t + \tau) - W(t)]^2\rangle = (2kT/w)\tau.$$

This is the famous Einstein formula of the theory of Brownian motion. In particular, this formula immediately implies that the particle cannot have a finite velocity, for otherwise the mean value of the square of the path traversed by the particle in a short time $\tau = \Delta t$ would have to be proportional to $(\Delta t)^2$ and not to Δt. (cf. the related remark concerning $Z(\omega)$ in Sec. 9). The refined form of the Einstein formula, which is valid for a Brownian particle of nonzero mass, will be given in Sec. 23.

The normality of Maxwell's distribution of molecular velocities implies that the "purely random" force $X_1(t)$ acting on a Brownian particle is a normal (i.e. Gaussian) white noise process. (This means that $X_1(t)$ must be regarded as a limit of the sequence of ordinary Gaussian processes. Another, more direct, definition of the Gaussian white noise process will be given in Sec. 24.) Thus, the process $W(t) = \int_0^t X_1(t')dt'$ is also Gaussian in this case.   In the Einstein–Smoluchowski theory of Brownian motion this process represents, to within a constant factor, a path traversed by a Brownian particle; in this relation it was studied in detail by Wiener (see again Note 10 to the Introduction). Therefore, the Gaussian random process $W(t)$, which has a mean value zero and a correlation function

of the form (2.195), is usually called the *Wiener process*, or the *Brownian motion process*. The Wiener process evidently satisfies (2.195a); on the other hand, it is easy to see that any Gaussian zero mean process $W(t)$ which satisfies conditions (2.195a) and vanishes (with probability one) at $t = 0$ is a Wiener process. Similarly, the Gaussian stationary process $X(t)$, which has a mean value zero and correlation function of the form $B(\tau) = C\exp(-a|\tau|)$, is often called the *Ornstein–Uhlenbeck process*, since in the Ornstein–Uhlenbeck theory of Brownian motion such a process represents the velocity of a Brownian particle.[53]

A differential equation of the form (2.191a) describes also the current (or voltage) fluctuations due to "thermal noises" in a circuit containing either a resistance R and a capacitance C, or a resistance R and an inductance L. Quite similar equations describe the one-dimensional Brownian motion of a harmonic oscillator when its inertia is neglected (i.e. with the Einstein–Smoluchowski approximation). In this case $Y(t)$ is the oscillator coordinate, the term dY/dt is proportional to the friction force due to the viscosity of the medium, the term $aY(t)$ represents the action of the spring keeping the oscillator, and $cE(t)$ represents the "purely random" force (cf. p. 174). There are also many other physical fluctuation problems leading to an equation of the form (2.191a). This is one of the reasons why stationary random processes with a correlation function of the form (2.94) and a spectral density of the form (2.95) very often appear in applied problems.*

(b) *Second-order equations.* Let

(2.197) $$\frac{d^2Y(t)}{dt^2} + 2a\frac{dY(t)}{dt} + bY(t) = X(t),$$

where a and b are real coefficients, and $X(t)$ is a stationary random process with a spectral density $f(\omega)$. The equation $P(z) = z^2 + 2az + b = 0$ clearly has roots with a zero real part only when $b = 0$ or when $a = 0$ and $b > 0$. Hence, in all other cases (2.197) has a single stationary solution $Y(t)$ with a spectral density

(2.198) $$f_{YY}(\omega) = \frac{f(\omega)}{|-\omega^2 + 2ia\omega + b|^2} = \frac{f(\omega)}{(\omega^2 - b)^2 + 4a^2\omega^2}.$$

Suppose, e.g., that $b > 0$, $b - a^2 = \beta^2 > 0$.   In this case the equation $P(z) = 0$ has two complex conjugate roots $z_1 = -a + i\beta$ and $z_2 = -a - i\beta$.   The solution of (2.197) with initial conditions $Y(t_0) = Y_0$, $Y'(t_0) = Y'_0$ for $t = t_0$ has the form

$$
\begin{aligned}
(2.199) \qquad T(t|Y_0,Y'_0,t_0) \ &= C_1 e^{-a(t-t_0)}\cos\beta(t - t_0) \\
&+ C_2 e^{-a(t-t_0)}\sin\beta(t - t_0) \\
&+ \beta^{-1}\int_0^{t-t_0} e^{-au}\sin\beta u X(t - u)du,
\end{aligned}
$$

where $C_1 = Y_0$, $C_2 = (Y_0 a + Y'_0)/\beta$.   If $a > 0$, this solution, as $t - t_0 \to \infty$, tends to the following unique stationary solution of (2.197), no matter what the values of Y_0 and Y'_0 (i.e. C_1 and C_2) may be:

$$
(2.200) \quad Y(t) = \beta^{-1}\int_0^\infty e^{-au}\sin\beta u \, X(t - u)du.
$$

It is easy to verify that the process (2.200) has a spectral density (2.198).   If, however, $a < 0$, the solution (2.199) does not tend to any limit, as $t - t_0 \to \infty$.   In this case the stationary solution of $Y(t)$ is expressed through future (rather than past) values of $X(s)$ with $s > t$.

The situation is quite similar when $a^2 - b = \beta_1^2 \geqslant 0$, i.e. the equation $P(z) = 0$ has real roots $z_1 = -a + \beta_1$ and $z_2 = -a - \beta_1$. If $a > 0$, $\beta_1 < a$, so that both roots z_1 and z_2 are negative, then the stationary solution $Y(t)$ of (2.197) is given by an integral of the weighted "past values" $X(t - u)$ with $u \geqslant 0$ (and this solution coincides with the limit, as $t - t_0 \to \infty$, of the solution of the initial value problem corresponding to any initial conditions at $t = t_0$). If, however, both roots z_1 and z_2 are positive, then the stationary solution equals the integral of the weighted "future values" $X(t - u)$ with $u < 0$, and if one of them is positive and the other negative, it is equal to an integral containing in the integrand the values $X(t - u)$ for all real u.

Of special interest is a particular case of (2.197) in which the right-hand side is an idealized white noise process (i.e. $X(t) = cE(t)$).   The corresponding stationary solution $Y(t)$ has a spectral density

$$
(2.201) \quad f_{YY}(\omega) = \frac{f_0}{(\omega^2 - b)^2 + 4a^2\omega^2},
$$

where $f_0 = c^2/2\pi$ is the constant spectral density of the white noise $X(t) = cE(t)$. If the equation $P(z) = z^2 + 2az + b = 0$ has no roots with a zero real part, the denominator in (2.201) will not vanish, and $Y(t)$ will be an ordinary (and even differentiable, but not twice differentiable) stationary random process. The correlation function $B_{YY}(\tau)$, of the process $Y(t)$ with the spectral density (2.201), has the form (2.111) if $b > a^2 > 0$, the form (2.114), if $b = a^2 > 0$ or $a = 0$, $b < 0$, and the form (2.115) if $b < a^2$, $a \neq 0$. When $a > 0$, $b - a^2 = \beta^2 > 0$, the stationary solution (2.200) of (2.197), where $X(t)$ is a white noise, can be written as

$$(2.202) \quad Y(t) = \beta^{-1} \int_0^\infty e^{-a(t-s)}\sin\beta(t - s)dW(s),$$

where $W(s)$ is an indefinite integral of $X(t) = cE(t)$ (see (2.194a)), which possesses the properties (2.195) and (2.195a). The process $Y(t)$ can be represented similarly when $a > 0$, $a^2 \geqslant b > 0$ (i.e. in all cases where the equation $P(z) = 0$ has only roots with negative real parts).

Equation (2.197), where $X(t) = cE(t)$, is analogous in many respects to the difference equation (2.13) with $m = 2$. For this reason a stationary random process with the spectral density of the form (2.201) is sometimes called a *second-order autoregressive process*. An *autoregressive process of order m* can be similarly defined as a stationary random process with the spectral density of the form $f(\omega) = 1/|P(i\omega)|^2$ where $P(z)$ is an mth-degree polynominal having no roots with zero real parts. Moreover, stationary processes $X(t)$ with arbitrary rational spectral densities $f(\omega)$ are natural analogues of stationary autoregressive-moving average sequences, considered in Sec. 7. However, we shall not consider this more general analogy here.[54]

*An equation of the form (2.197), where $X(t) = cE(t)$, arises naturally when considering the Brownian motion of a harmonic oscillator, i.e. of a particle of mass m kept by an elastic spring. This spring produces a force $-\lambda Y(t) = -mbY(t)$, that is proportional to the coordinate $Y(t)$ and tends to return a particle to its "equilibrium position" at $Y = 0$. In this case the term $2adY/dt$ (where $2a = w/m$) of eq. (2.197) is again due to the action of the viscous friction force, and the right-hand side of (2.197), $X(t) = X_1(t)/m$, represents a "purely random" force due to

fluctuations in the number of molecular collisions. The "oscillators" kept by a linear force of the form $- \lambda Y(t)$ play an essential part in many precision measuring instruments; the Brownian motion of such oscillators often limits the accuracy of measurements, and therefore the theory of this motion is of considerable applied interest (cf. Note 48). One more problem leading to an equation of the form (2.197) with a "white noise" on the right-hand side concerns the voltage or current fluctuations due to "thermal noises" in an electric circuit containing a resistance R, to a capacitance C, and an inductance L. There are also many other physical problems leading to the same equation, which we shall not mention here.*

13. Spectral Representation of Stationary Sequences and Their Correlation Functions

It is clear that almost all the results of Secs. 7–11 can be carried over, with a few minor changes, to the case of stationary random sequences $X(t)$, where $t = 0$, ± 1, ± 2, ... Thus, if, as in Example 1 of Sec. 7, we construct a stationary sequence $X(t)$ of the form $X\Phi(t)$, where $\Phi(t)$ is a numerical function and X is a random variable, we again find that $\langle X \rangle$ = 0 and $\Phi(t)$ = exp($i\omega t$). Therefore, we arrive at a harmonic oscillation

(2.203) $X(t) = X e^{i\omega t}$

with a random amplitude and phase. The only difference is that since now t only takes integral values, we have exp$[i(\omega + 2k\pi)t]$ = exp($i\omega t$) for all t if k is an integer, and hence ω is defined only to within an additive constant which is a multiple of 2π. This allows us to assume that in the case of stationary sequences $X(t)$ the angular frequency ω always lies between $-\pi$ and π. Therefore, if we represent an arbitrary sequence $X(t)$ as a superposition of harmonic oscillations of the form (2.203), we have to take into account only oscillations with frequencies in the interval $[-\pi, \pi]$. On the other hand, the stationarity condition for such a sum (which requires that the individual oscillations be uncorrelated) is, for the case of integer-valued t, precisely the same as for

continuously varying *t*. Thus, the general *spectral representation of stationary sequences* $X(t)$ is quite similar to the spectral representation (2.61) of stationary processes, but the improper integral over $dZ(\omega)$ extended from $-\infty$ to ∞, is replaced now by the integral from $-\pi$ to π. Therefore,

$$(2.204) \quad X(t) = \int_{-\pi}^{\pi} e^{it\omega} dZ(\omega),$$

where the properties of $Z(\omega)$ and the meaning of the integral are the same as in the case of stationary random processes.

Assuming, as in Sec. 9, that

$$(2.205) \quad \langle |dZ(\omega)|^2 \rangle = dF(\omega), \quad \langle dZ(\omega)\overline{dZ(\omega')} \rangle = 0 \text{ for } \omega' \neq \omega,$$

i.e.

$$(2.205a) \quad \langle dZ(\omega)\overline{dZ(\omega')} \rangle = \delta(\omega - \omega')dF(\omega)d\omega',$$

and then using (2.204), we obtain

$$(2.206) \quad B(\tau) = \langle X(t + \tau)\overline{X(t)} \rangle = \int_{-\pi}^{\pi} e^{i\tau\omega} dF(\omega),$$

where $F(\omega)$, $-\pi \leqslant \omega < \pi$, is a bounded nondecreasing function defined on the finite interval (cf. the similar arguments on p. 107). The function $F(\omega)$ is defined to within an arbitrary additive constant, which can be chosen so that the condition $F(-\pi) = 0$ will be fulfilled. Equation (2.206) describes the *spectral representation of the correlation function* $B(\tau)$ of a stationary random sequence $X(t)$. The function $F(\omega)$ is called the *spectral distribution function* (of the sequence $X(t)$). Moreover, if the function $F(\omega)$ is differentiable and can be represented as an indefinite integral of its derivative so that

$$(2.207) \quad F(\omega) = \int_{-\pi}^{\omega} f(\omega')d\omega', \text{ where } f(\omega) = F'(\omega) \geqslant 0,$$

then $f(\omega)$ is called the *spectral density* of $X(t)$.

The statement that it is possible to represent any correlation function $B(\tau)$ in the form of a Fourier–Stieltjes integral (2.206), where $F(\omega)$ is nondecreasing and bounded, and the converse statement that any function $B(\tau)$ of such a form is a correlation function of some stationary sequence constitute the *spectral representation theorem for*

correlation functions of stationary sequences. This theorem is due to Wold (see Note 55); it is quite similar to the Khinchin spectral representation theorem for correlation functions of stationary processes. The Wold theorem is a simple consequence of the following two facts:

(a) The class of all correlation functions of stationary random sequences coincides with the class of all positive definite sequences (see Sec. 4, p. 58 and Sec. 5, p. 71);

(b) A numerical sequence ..., $B(-2),B(-1),B(0),B(1),B(2),$ ... is positive definite if and only if it can be represented in the form (2.206) where $F(\omega)$ is a nondecreasing and bounded function.

Proposition (b) was first proved by a German mathematician, G. Herglotz, as far back as the turn of the century, and hence is known as the *Herglotz theorem* (or the *Herglotz lemma*).[55]

The fact that any function of the form (2.206), where τ = $0,\pm1,\pm2,$ ... and $F(\omega)$ is a bounded nondecreasing function, is a correlation function of a stationary sequence, can also be easily proved by considering the particular stationary sequence of the form $X(t) = Y\exp(it\Omega)$, where Y is a random variable with mean value zero and bounded variance, and Ω is a real random variable whose sample values ω can be assumed to lie between $-\pi$ and π (cf. Example 5 in Sec. 7).

By using the spectral representation theorem for correlation functions $B(\tau)$, the possibility of representing a stationary sequence $X(t)$ as an integral of the form (2.204), where $Z(\omega)$ is a complex random function with uncorrelated increments (i.e. the *spectral representation theorem for stationary sequences*), can be proved quite similarly to the corresponding proof related to the spectral representation (2.61) for stationary processes.[56]

If the sequence $X(t)$ is real, then

(2.208)
$$Z(\omega_2) - Z(\omega_1) = Z(-\omega_1) - Z(-\omega_2),$$
$$F(\omega_2) - F(\omega_1) = F(-\omega_1) - F(-\omega_2)$$

for $0 < \omega_1 < \omega_2 \leqslant \pi$. It follows that in the real case the spectral representation (2.204) of the stationary sequence $X(t)$ and the spectral representation (2.206) of its correlation function $B(\tau)$ can be written as

$$(2.209) \quad X(t) = \int_0^\pi \cos\omega t \, dZ_1(\omega) + \int_0^\pi \sin\omega t \, dZ_2(\omega),$$

and

$$(2.210) \quad B(\tau) = \int_0^\pi \cos\omega\tau \, dG(\omega),$$

where $Z_1(\omega)$ and $Z_2(\omega)$ are real random functions of ω, $0 \leqslant \omega \leqslant \pi$, with uncorrelated increments, which satisfy the relations (2.64) and (2.65), and $G(\omega)$ is a bounded nondecreasing function of ω defined for $0 \leqslant \omega \leqslant \pi$.

Equations (2.204), (2.205), and (2.206) imply that the discontinuity points ω_k of the function $F(\omega)$ (corresponding to jumps where $F(\omega)$ increases by $f_k > 0$) coincide with the discontinuity points of the random function $Z(\omega)$, which contribute the components of the form $X_k \exp(i\omega_k t)$, $\langle |X_k|^2 \rangle = f_k > 0$, to the spectral representation of the sequence $X(t)$. The set of all discontinuity points ω_k makes up a *discrete spectrum* of $X(t)$. This spectrum can be uniquely determined by the correlation function $B(\tau)$ as the set of all numbers $\omega = \omega_k$ for which

$$(2.211) \quad \lim_{T\to\infty} \frac{1}{2T+1} \sum_{\tau=-T}^{T} B(\tau)e^{-i\omega\tau} \neq 0.$$

(It is easy to see that the limit (2.211) is equal to the corresponding jump f_k at the point ω_k of the discrete spectrum.)[57] On the other hand, if, as is usually the case in real applications, the absolute value of $B(\tau)$ falls off so rapidly with $|\tau|$ that

$$(2.212) \quad \sum_{\tau=-\infty}^{\infty} |B(\tau)|^2 < \infty,$$

then the quantities $(1/2\pi)B(-\tau)$, $\tau = 0,\pm1,\pm2, \ldots$ coincide with the Fourier coefficients of the following periodic function (of a period 2π)

$$(2.213) \quad f(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} e^{-i\omega\tau} B(\tau).$$

(The infinite sum here must be defined in a special manner in some exceptional cases where (2.212) is satisfied, but the below mentioned condition (2.212a) is not valid.) If $B(\tau)$ also satisfies the following, slightly more restrictive (but almost always valid in practice) condition

(2.212a) $\sum_{T=-\infty}^{\infty} |B(T)| < \infty,$

then the function $f(\omega)$ is necessarily continuous and bounded. By virtue of the well-known formula for Fourier coefficients we have that

(2.214) $B(T) = \int_{-\pi}^{\pi} e^{i\omega T} f(\omega) d\omega,$

if condition (2.212a) (or, at least, (2.212)) is valid. Hence in this case there necessarily exists a spectral density $f(\omega)$ whose indefinite integral is the spectral distribution function $F(\omega)$. Moreover, the sequence $X(t)$ has here a *purely continuous spectrum* (cf. p. 110).

Equations (2.211) and (2.213) permit one to easily determine the spectral distribution function $F(\omega)$ by the given correlation function $B(T)$ in cases where either a spectral density exists or there is only a discrete spectrum (i.e. the spectral distribution function is either absolutely continuous or a step-function). However, in the case of an arbitrary spectral distribution function $F(\omega)$ this function can also be uniquely determined by $B(T)$ with the aid of the following general formula:

(2.215) $F(\omega_2) - F(\omega_1) = B(0)(\omega_2 - \omega_1) + \lim_{T\to\infty} \sum_{\substack{T=-T \\ (T\neq 0)}}^{T} B(T)\frac{e^{-i\omega_2 T} - e^{-i\omega_1 T}}{-iT}.$

This formula is quite similar to (2.85) (if ω_1 or ω_2 is a discontinuity point of $F(\omega)$, then here also by $F(\omega_1)$ or $F(\omega_2)$ is meant the half-sum of the values of $F(\omega)$ before and after the corresponding "jump").[58]

For a real sequence $X(t)$, (2.213), (2.214), and (2.215) can be written in the form

(2.213a) $g(\omega) = \frac{1}{\pi}[B(0) + 2\sum_{T=1}^{\infty} B(T)\cos\omega T],$

(2.214a) $B(T) = \int_0^{\pi} \cos\omega T g(\omega) d\omega,$

(2.215a) $G(\omega) = \frac{1}{\pi}[B(0)\omega + 2\sum_{T=1}^{\infty} \frac{B(T)}{T} \sin\omega T].$

However, even when $X(t)$ is real (2.213) − (2.215) are often more convenient than (2.213a) − (2.215a).

We conclude this section by formulating the discrete time analogues of (2.89), (2.91), and (2.93) that express via spectral characteristics some quantities involving the correlation $B(\tau)$. (The stationary sequence $X(t)$ and its correlation function $B(\tau)$ will now be assumed to be real.) The most natural definition of the *correlation time* T_1 of a real stationary sequence $X(t)$ with mean value zero is the following:

$$(2.216) \qquad T_1 = \frac{1}{B(0)}\left[\frac{1}{2}B(0) + \sum_{\tau=1}^{\infty} B(\tau)\right] = \frac{1}{2} + \sum_{\tau=1}^{\infty} R(\tau)$$

(cf. (2.88)). The correlation time T_1 is obviously finite in all cases where a bounded spectral density $f(\omega)$ exists according to (2.213) in such cases

$$(2.217) \qquad T_1 = \frac{\pi f(0)}{B(0)} = \frac{\pi f(0)}{2\int_0^{\pi} f(\omega)d\omega}$$

(cf. (2.89)).

The discrete time analogue of the quantity (2.90) is

$$(2.218) \qquad \beta_0^{(1)} = \lim_{T\to\infty} \frac{1}{T} \sum_{\tau=0}^{T-1} B(\tau).$$

It is clear that $\beta_0^{(1)} = 0$ if $B(\tau) \to 0$ as $\tau \to \infty$. Therefore, $\beta_0^{(1)} = 0$ in all cases where the spectral density $f(\omega)$ exists. Moreover, any periodic component of $B(\tau)$ of the form $B_k(\tau) = g_k\cos\omega_k\tau$, $\omega_k \neq 0$, contributes nothing to $\beta_0^{(1)}$ (since $\lim_{T\to\infty} T^{-1} \sum_{\tau=0}^{T-1} \cos\omega_k\tau = 0$). If, however, $\Delta F(0) = \lim_{h\to 0}[F(h) - F(0)] = \lim_{h\to 0} F(h) = g_0 > 0$, then $B(\tau)$ includes a constant term g_0 and this term gives a contribution to β_0 equal to g_0. These arguments suggest the relation

$$(2.219) \qquad \beta_0^{(1)} = \Delta F(0)$$

which is in fact true (it is a simple consequence of the relation (2.211) with $\omega_k = 0$).

Finally, the proof of the relation (2.93) must be only slightly modified to show that

$$(2.220) \qquad \beta_1^{(1)} = \lim_{T\to\infty} \frac{1}{T} \sum_{\tau=0}^{T-1} B^2(\tau)$$

satisfies the relation

(2.221) $\beta_1^{(1)} = \sum_j f_j^2 = g_0 + \sum_k g_k^2/2$,

where f_j are the jumps of the spectral distribution function
$F(\omega)$, $-\pi \leqslant \omega \leqslant \pi$, $g_0 = f_0$ is the jump at $\omega = 0$ (it can be equal to
zero, of course), and the constants $g_k = 2f_k$ are the jumps of $G(\omega)$
$= F(-\omega) + [F(\omega) - F(0)]$ at positive values of ω.

14. Discrete Samples of Random Processes and Discrete Linear Transformations

Samples of stationary random processes $X(t)$, $-\infty < t < \infty$, at
some time intervals Δ (i.e. the sequences formed by values of
$X(t)$ at a set of equally spaced time points $t_k = k\Delta$, $k =$
$0, \pm 1, \pm 2, \ldots$) constitute a very important class of stationary
random sequences.   It has already been noted in the
Introduction that such sequences are encountered in all those,
comparatively frequent, cases where observations of the
continuously varying quantity $x(t)$, which is the realization of
a stationary process $X(t)$, are repeated regularly at identical
time intervals. However, even when we have a continuous
recording of all the values of $x(t)$ within some time interval
(which are taped, for instance), these values are usually most
conveniently processed by using digital electronic computers.
Such processing evidently requires that the continuous
recording of $x(t)$ be converted into a discrete time series.
Therefore, we must sacrifice most of the existing data and
look at $x(t)$ only at certain discrete sets of time points t_k,
disregarding all the intermediate values of $x(t)$.
    The sequence

(2.222) $X(t\Delta) = X_\Delta(t)$, $t = 0, \pm 1, \pm 2, \ldots$,

where $X(t)$ is a continuous time stationary random process, is
called a *discrete sample* of the process $X(t)$ at a *sampling
interval* Δ.* Such a sample is a stationary random sequence

---

*It is, of course, possible to assume that the time origin - the point t = 0 -
does not coincide with one of the observation times t_k (i.e. to consider, in place
of X(tΔ), the more general random variables X(tΔ + t_0), where t_0 is a fixed
number). However, changes for this case are quite obvious and, therefore, we
shall not consider them here.

with the same mean value $\langle X_\Delta(t) \rangle = \langle X(t) \rangle$, as the process $X(t)$. (Below, this mean value is assumed to be zero for simplicity.) The correlation function of this discrete sample is

$$(2.223) \quad B_\Delta(\tau) = \langle X((t + \tau)\Delta)\overline{X(t\Delta)} \rangle = B(\tau\Delta), \quad \tau = 0, \pm 1, \pm 2, \dots,$$

where $B(\tau)$ is the correlation function of the process $X(t)$. From (2.223) it follows, in particular, that if $B(\tau)$ is a correlation function of some stationary random process, the sequence (2.223) is necessarily a correlation function of some stationary sequence.*

Discrete samples of the process $X(t)$ are very often used for approximate determination of the corresponding statistical characteristics. Therefore, it is very important to know the relations between the characteristics of the process $X(t)$ and of the sequence $X_\Delta(t)$. We have already noted that $X(t)$ and $X_\Delta(t)$ have the same mean value and that their correlation functions are related by the equation $B_\Delta(t) = B(\tau\Delta)$. However, it will be explained in Sec. 18 that the most important statistical characteristic of the process $X(t)$ in applications is its spectral density $f(\omega)$ or (in the general case) its spectral distribution function $F(\omega)$. Therefore the determination of these functions from the observed values of the sequence $X_\Delta(t)$ is of special interest.

The functions $f(\omega)$ and $F(\omega)$ are closely related to the spectral decomposition of the process $X(t)$ into a superposition of harmonic oscillations of the form $A(\omega)\exp(i\omega t)$. However, it is easy to see that such a decomposition cannot be uniquely restored from the data of observations at time intervals Δ. Indeed,

$$(2.224) \quad e^{i\omega t\Delta} = e^{i(\omega + k2\pi/\Delta)t\Delta} = e^{i\omega t\Delta}e^{i2\pi kt}$$

for any integers t and k. Therefore, having only the data of discrete observations at time points Δ units apart, we simply cannot distinguish an oscillation with an angular frequency ω from all the oscillations with frequencies $\omega + 2\pi k/\Delta$, where k

---

*This is, of course, quite obvious, since the validity of the positive definiteness condition (1.59) for all real $\tau_1, \dots, \tau_n$ implies its validity for $\tau_1 = k_1\Delta, \dots, \tau_n = k_n\Delta$, where $k_1, \dots, k_n$ are integers. Hence the sequence $B_\Delta(\tau)$, $\tau = 0, \pm 1, \dots$, is also positive definite in this case.

is an integer.   Moreover, in the real case, where the oscillations
with frequencies ω and −ω are also indistinguishable from each
other, the oscillation with frequency ω cannot any longer be
distinguished from all the oscillations with frequencies either
of the form ω + 2πk/Δ, or −ω + 2πk/Δ, where k is an integer (cf.
Fig. 34).   When observing oscillations of frequency ω at discrete
time points we always automatically assign to these oscillations
the least one of the frequencies ±ω + 2πk/Δ .   (This last
circumstance is used, in particular, in the stroboscope, which
permits one to "slow down" the rapid oscillations by illumination
of a vibrating or rotating body at equal time intervals.) When
applied to stationary random processes, the impossibility of
distinguishing the harmonic components with frequencies
differing by an integer multiple of 2π/Δ by observations at time
intervals Δ is called the *aliasing* effect and the frequencies ω +
2πk/Δ and −ω + 2πk/Δ , k = ±1,±2, ..., are called the *aliases* of the
frequency ω.

Using the spectral representation of the process $X(t)$, we
can write the formulae

$$(2.225) \qquad X(t) = \int_{-\infty}^{\infty} e^{it\omega} dZ(\omega), \quad X_\Delta(t) = \int_{-\infty}^{\infty} e^{it\Delta\omega} d\, Z(\omega),$$

where in the first formula, t runs through all the real
numbers, and in the second, only through the integers. By
virtue of (2.224) the values of the integrand function
$\exp(it\Delta\omega)$ in the second formula (2.225) are repeated precisely
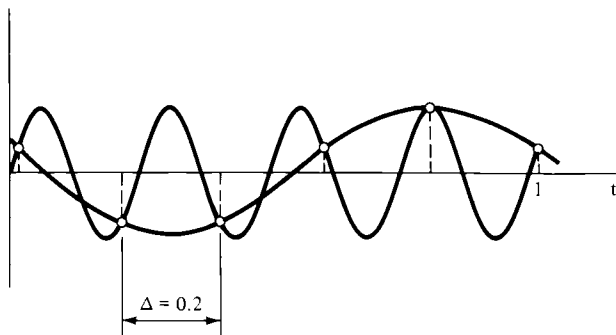
Fig. 34.   Two sinusoids of the frequencies $n = \omega/2\pi = 1$ and
$n = 4$ which are indistinguishable by sampling at
intervals of length Δ = 0.2 (Blackman and Tukey,
1959).

at intervals $2\pi/\Delta$. Thus,

$$(2.226)\quad \int_{(2k-1)\pi/\Delta}^{(2k+1)\pi/\Delta} e^{it\Delta\omega}dZ(\omega) = \int_{-\pi/\Delta}^{\pi/\Delta} e^{it\Delta\omega}dZ(\omega + 2k\pi/\Delta)$$

for any (positive or negative) integer k. Since harmonics with frequencies ω and $\omega + 2k\pi/\Delta$ are indistinguishable, it is natural to consider only frequencies which do not exceed π/Δ in absolute value, i.e. to use the identity (2.226) at all $k = \pm1,\pm2, ...$ and then rewrite the second relation (2.225) in the form

$$X_\Delta(t) = \int_{-\pi/\Delta}^{\pi/\Delta} e^{it\Delta\omega}dZ_\Delta^{(1)}(\omega),$$

(2.227)

$$Z_\Delta^{(1)}(\omega) = \sum_{k=-\infty}^{\infty} \int Z(\omega + 2k\pi/\Delta) - Z((2k - 1)\pi/\Delta)].$$

(The term $- Z((2k - 1)\pi/\Delta)$ in the second formula (2.227) does not change $dZ_\Delta^{(1)}(\omega)$ and is so selected that the condition $Z_\Delta^{(1)}(-\pi/\Delta)$ $= 0$ is satisfied.) By virtue of (2.227) the following equation for $B_\Delta(\tau) = \langle X_\Delta(t + \tau) X_\Delta(t) \rangle$ can be derived:

$$B_\Delta(\tau) = \int_{-\pi/\Delta}^{\pi/\Delta} e^{i\tau\Delta\omega}dF_\Delta^{(1)}(\omega),$$

(2.228)

$$F_\Delta^{(1)}(\omega) = \sum_{k=-\infty}^{\infty}[F(\omega + 2k\pi/\Delta) - F((2k - 1)\pi/\Delta)].$$

When the spectral density $f(\omega) = F'(\omega)$ exists, (2.228) can also be written as

$$(2.229)\quad B_\Delta(\tau) = \int_{\pi/\Delta}^{\pi/\Delta} e^{i\tau\Delta\omega}f_\Delta^{(1)}(\omega)d\omega,\quad f_\Delta^{(1)}(\omega) = \sum_{k=-\infty}^{\infty} f(\omega + \frac{2k\pi}{\Delta}).$$

The spectral representations (2.227) − (2.229) can also be written as the ordinary spectral representations of the stationary sequence $X_\Delta(t)$, $t = 0,\pm1,\pm2, ...$, and its correlation function $B_\Delta(\tau)$ (referring to the frequency band from $-\pi$ to π). To do this, we merely use the dimensionless product $\omega\Delta$ as a new variable.   Denoting this new variable again by the letter ω to retain the usual spectral representation notation, we obtain

$$(2.230)\quad X_\Delta(t) = \int_{-\pi}^{\pi} e^{it\omega}dZ_\Delta(\omega),$$

$$(2.231)\quad B_\Delta(\tau) = \int_{-\pi}^{\pi} e^{i\tau\omega}dF_\Delta(\omega) \text{ or } B_\Delta(\tau) = \int_{-\pi}^{\pi} e^{i\tau\omega}f_\Delta(\omega)d\omega,$$

where $Z_\Delta(\omega) = Z_\Delta^{(1)}(\frac{\omega}{\Delta})$, and

(2.231a)
$$F_\Delta(\omega) = \sum_{k=-\infty}^{\infty}\left[F\left(\frac{\omega + 2k\pi}{\Delta}\right) - F\left(\frac{(2k - 1)\pi}{\Delta}\right)\right],$$

$$f_\Delta(\omega) = \frac{1}{\Delta}\sum_{k=-\infty}^{\infty} f\left(\frac{\omega + 2k\pi}{\Delta}\right).$$

Equations (2.231a) give the relations between the spectral characteristics $F_\Delta(\omega)$ and $f_\Delta(\omega)$ of the sampled sequence $X_\Delta(t)$ and the corresponding characteristics of the process $X(t)$.[59]

The aliasing effect is described most clearly, however, by the initial equations (2.228) and (2.229) for $F_\Delta^{(1)}(\omega)$ and $f_\Delta^{(1)}(\omega)$. According to these equations, if observation of a continuous process $X(t)$ is carried out only at uniformly spaced time points Δ units apart, the spectrum of observations of $X_\Delta(t)$ is concentrated within the finite frequency band $-\pi/\Delta \leqslant \omega < \pi/\Delta$. (The limiting angular frequency $\omega_\Delta = \pi/\Delta$ of this band and the corresponding ordinary frequency $n_\Delta = \omega_\Delta/2\pi = 1/2\Delta$ are usually called the angular and ordinary *Nyquist frequencies* corresponding to a given sample.) Thus, the contribution to $B(0) = \langle |X(t)|^2 \rangle$ (the "power" of $X(t)$) and $B(\tau) = \langle X(t + \tau)\overline{X(t)} \rangle$ of all the frequencies lying beyond the band $-\pi/\Delta \leqslant \omega < \pi/\Delta$ is "folded" artificially into the given band: The whole frequency spectrum is partitioned into bands of length $2\pi/\Delta$ by "fold points" $(2k + 1)\pi/\Delta$, $k = 0,\pm 1,\pm 2, ...$, and the power distribution within each of the bands distinct from the *principal band* $-\pi/\Delta \leqslant \omega < \pi/\Delta$ is superimposed on the power distribution within the principal band (see Fig. 35).   Let us suppose now that frequencies exceeding the Nyquist frequency $\omega_\Delta = \pi/\Delta$ make an appreciable contribution to the initial frequency distribution of the power of the process.   It is clear that it is then possible to estimate the functions $F(\omega)$ and $f(\omega)$ from the sample at the time intervals Δ only with a considerable error because of the unavoidable aliasing. The choice of the sampling interval Δ (or, what is the same, the sampling rate $1/\Delta$ equal to the number of values sampled per unit time) is, of course, equivalent to the choice of the Nyquist frequency $\omega_\Delta = \pi/\Delta$.   If we wish that the spectral characteristics of the process $X(t)$ to be determined accurately enough from the observed sample $x_\Delta(t)$ (which is the realization of $X_\Delta(t)$), then the frequency ω_Δ must necessarily be so high that still higher frequencies ω make only a negligible contribution to the total power of the process $X(t)$. Unfortunately, we usually

do not know the frequency distribution of the power
beforehand and hence cannot be sure that the contribution of
all the frequencies above ω_Δ to the power of $X(t)$ is really
negligible. For this reason we sometimes have to use several
different sampling rates in estimating the spectral density
$f(\omega)$.   By doing this we can find experimentally for what
rate the subsequent increase of rate does not influence the
estimate of $f(\omega)$.

(a)

(b)
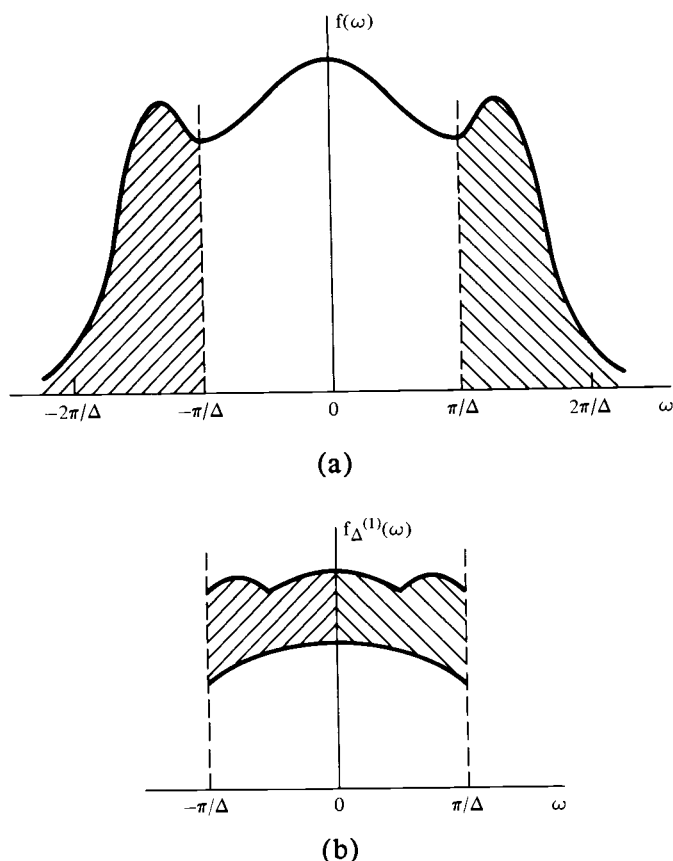
Fig. 35. (a) The spectral density $f(\omega), -\infty < \omega < \infty$, including a
         significant contribution (which is cross-hatched)
         from frequencies ω exceeding $\omega_\Delta = \pi/\Delta$. (b) A
         sketch of the spectral density $f_\Delta^{(1)}(\omega)$ of the sampled
         sequence $X(t)$; the superimposed ("aliased") power is
         cross-hatched (Koopmans, 1974).

It would be ideal, of course, if all the frequencies outside the band $-\pi\Delta \leqslant \omega < \pi/\Delta$ made no contribution at all to the process $X(t)$. Therefore, the case of the so-called band-limited random processes $X(t)$ is of special interest from the point of view of using the discrete samples $X_\Delta(t)$. Let us assume that for the

process $X(t)$ a frequency Ω exists such that $\int_{-\Omega}^{\Omega} dF(\omega) = B(0)$, i.e.

$F(\omega) = 0$ for $\omega < -\Omega$, $F(\omega) = B(0) = $ const for $\omega > \Omega$. In this case the spectral representation of $X(t)$ includes only such frequencies ω that $-\Omega \leqslant \omega \leqslant \Omega$ Then $X(t)$ is called a *band-limited stationary process* with a frequency band $|\omega| \leqslant \Omega$ (cf. the example of band-limited white noise in Sec. 10). If we now choose Δ so that the condition $\Omega \leqslant \omega_\Delta = \pi/\Delta$ is fulfilled, then $Z_\Delta^{(1)}(\omega) = Z(\omega)$ in (2.227), and hence the discrete sample $X_\Delta(t) = X(t\Delta)$, $t = 0,\pm1,\pm2,$ ..., permits us (at least in principle) to determine uniquely all the values of the spectral distribution (or density) function, i.e. the aliasing effect does not appear here.*

It is also easy to see that in the case of band-limited processes $X(t)$ with a band $|\omega| \leqslant \Omega$ the discrete sample $X_\Delta(t) = X(t\Delta)$, $t = 0,\pm1,\pm2,$ ..., where $\Delta \leqslant \pi/\Omega$ permits us to determine uniquely the variables $X(t)$ for all real values of t. In fact, it is known that the random function $Z_\Delta(\omega) = Z_\Delta^{(1)}(\omega/\Delta)$ can be uniquely determined if $X_\Delta(t)$ is given for all $t = 0,\pm1,\pm2,$ ... (see, eq. (2.41') in Note 56). However, $Z_\Delta^{(1)}(\omega) = Z(\omega)$ if $\omega_\Delta = \pi/\Delta \geqslant \Omega$.   Hence the set of the variables $X_\Delta(t)$ determines the function $Z(\omega)$, which permits one to evaluate $X(t)$ for all t with the aid of (2.61).

The explicit formula expressing $X(t)$ in terms of $X_\Delta(t) = X(t\Delta)$ can be derived as follows.   Since $\omega_\Delta \geqslant \Omega$ and the frequency band $-\Omega \leqslant \omega \leqslant \Omega$ includes the entire spectrum of $X(t)$, the spectral representation of $X(t)$ can be rewritten in the form

$$(2.232) \qquad X(t) = \int_{-\omega_\Delta}^{\omega_\Delta} e^{it\omega} dZ(\omega).$$

Let us now replace the function $\exp(it\omega), -\omega_\Delta \leqslant \omega \leqslant \omega_\Delta$, by its Fourier expansion into a series composed of the oscillations $\exp(i\omega 2\pi k/2\omega_\Delta) = \exp(ik\Delta\omega)$, $k = 0,\pm1,\pm2,$ ... .   It is easy to see

---

*Of course, in real practice even in this case the spectral characteristics of $X(t)$ will be determined from the observational data with an error, since the length of the observed realization $x(t)$ is always finite, and the observations themselves are not absolutely accurate (cf. Sec. 18 below).

that this Fourier expansion has the form

$$(2.233) \qquad e^{it\omega} = \sum_{k=-\infty}^{\infty} \frac{\sin \pi(t/\Delta - k)}{\pi(t/\Delta - k)} e^{ik\Delta\omega}, \qquad -\frac{\pi}{\Delta} \leqslant \omega \leqslant \frac{\pi}{\Delta}.$$

Taking into account (2.225), we obtain

$$(2.234) \qquad X(t) = \sum_{k=-\infty}^{\infty} \frac{\sin \pi(t/\Delta - k)}{\pi(t/\Delta - k)} X(k\Delta),$$

where the infinite sum is understood, as usual, in the mean square sense. Equation (2.234) is valid for any band-limited stationary process $X(t)$ having no power outside the band $-\pi/\Delta \leqslant \omega \leqslant \pi/\Delta$, and it permits one to express all the values of such a process in the form of linear combinations of its sampled values at time intervals Δ. This equation plays an important role in many engineering problems. In the radio-engineering literature eq. (2.234) is usually called either the *sampling theorem,* or the *Kotel'nikov theorem* (in Russian literature), or the *Shannon (sampling) theorem* (in Western literature).[60]

Equation (2.234) defines a particular *discrete linear transformation*, i.e. a linear transformation (or a linear filter, or a linear system) applied to a stationary sequence $X_\Delta(k) = X(k\Delta)$. Here we deal with a special linear transformation depending on the real parameter t and leading (for band-limited processes) to the restoration of the values of $X(t)$ from those of $X_\Delta(k)$. Let us now consider the general class of linear transformations $\mathfrak{L}$ of an arbitrary stationary sequence $X(t)$, $t = 0, \pm 1, \pm 2, \ldots$, which are representable in the form

$$(2.235) \qquad \mathfrak{L}\{X(t)\} = \lim_{n \to \infty} \sum_{j=-n}^{n} h_j^{(n)} X(t - j),$$

where $h_j^{(n)}$, $j = -n, -n + 1, \ldots, n$, are constant coefficients. By reasoning as in Sec. 11 (p. 149–150), it is easy to show that the limit (2.235) exists if and only if the sequence of the functions $H_n(\omega) = \sum_{j=-n}^{n} h_j^{(n)} e^{-ij\omega}$, $-\pi \leqslant \omega \leqslant \pi$, converges as $n \to \infty$ to some function $H(\omega)$ (in the mean square sense with respect to $dF(\omega)$, i.e. in the sense that (2.160) is valid where only the limits of integration are now replaced by $-\pi$ and π). Moreover, in this case

$$(2.236) \qquad \int_{-\pi}^{\pi} |H(\omega)|^2 dF(\omega) < \infty$$

and

$$(2.237) \quad \mathfrak{L}\{X(t)\} = \int_{-\pi}^{\pi} e^{it\omega} H(\omega) d Z(\omega).$$

From (2.237) it follows that the correlation function $B_{YY}(\tau)$ and the spectral distribution function $F_{YY}(\omega)$ of the sequence $Y(t) = \mathfrak{L}\{X(t)\}$ are equal to

$$(2.238) \quad B_{YY}(\tau) = \int_{-\pi}^{\pi} e^{i\tau\omega} |H(\omega)|^2 dF(\omega)$$

and

$$(2.239) \quad F_{YY}(\omega) = \int_{-\pi}^{\omega} |H(\omega')|^2 dF(\omega').$$

In particular, if $X(t)$ has a spectral density $f(\omega)$, then $Y(t)$ has also a spectral density $f_{YY}(\omega)$, and

$$(2.239a) \quad f_{YY}(\omega) = |H(\omega)|^2 f(\omega).$$

As in transformations of processes $X(t)$ with continuous time, the function $H(\omega)$ is called the *transfer function* of $\mathfrak{L}$, while its absolute value $A(\omega) = |H(\omega)|$ is the *gain* of $\mathfrak{L}$, and the argument $\psi(\omega) = \arg H(\omega)$ is the *phase* (or *phase-shift*) of $\mathfrak{L}$, respectively. The transfer function of the linear transformation (2.234) is obviously equal to $\exp(it\omega)$, and this fact determines the main property of the transformation.

The discrete linear transformations encountered in applications very often belong to a special subclass of *summation transformations* specified by a *weighting sequence*. Such transformations are specified by equations of the form

$$(2.240) \quad \mathfrak{L}\{X(t)\} = \sum_{j=-\infty}^{\infty} h_j X(t - j),$$

which differ from (2.235) in that the coefficients h_j are fixed (do not depend on n). The series on the right-hand side of (2.240) is (mean square) convergent if, and only if, the weighting sequence h_j, $j = 0, \pm 1, \pm 2, ...$, satisfies the condition

$$(2.241) \quad \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h_j \bar{h}_l B(j - l) < \infty,$$

where $B(k) = \langle X(t + k) \overline{X(t)} \rangle$. For summation transformations $\mathfrak{L}$ encountered in practice the sequence h_j almost always satisfies the condition

(2.242) $\sum\limits_{j=-\infty}^{\infty} |h_j| < \infty$

or at least the condition

(2.242a) $\sum\limits_{j=-\infty}^{\infty} |h_j|^2 < \infty.$

Both conditions (2.242) and (2.242a) guarantee the existence of the function $H(\omega)$, $-\pi \leqslant \omega < \pi$, with Fourier coefficients h_j:

(2.243) $H(\omega) = \sum\limits_{j=-\infty}^{\infty} h_j e^{-ij\omega}, \qquad h_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ij\omega} H(\omega)d\omega.$

The function $H(\omega)$ is then the transfer function of a linear summation transformation $\mathfrak{L}$. Conversely, if the transfer function H can be expanded into the Fourier series (2.243), where the coefficients h_j satisfy the condition (2.242) (or (2.242a)), $\mathfrak{L}$ is the summation transformation specified by the weighting sequence h_j.[61]

Realizations of the sequences $\mathfrak{L}\{X(t)\}$, where $\mathfrak{L}$ is a discrete linear transformation, can usually be represented in the form $\mathfrak{L}\{x(t)\}$, i.e. they are obtained by applying the same transformation $\mathfrak{L}$ (in the definition of which the convergence of series of the form (2.235) or (2.240) must now be understood in the usual sense of series theory) to the realization $x(t)$ of the sequence $X(t)$. The application of operations of the form (2.240) (especially in the simplest case where only a finite number of weighting coefficients h_j is nonzero) to the realizations $x(t)$ is handled quite rapidly by modern digital computers. Since such computers are widespread, investigation into wide classes of discrete linear transformations, aimed at selecting those which possess some valuable properties, has become particularly important lately. Note, also, that discrete linear transformations are now often called *digital linear transformations,* or *digital linear filters* (the term "digital" emphasizes their relationship with digital computers).[62]

The discrete linear transformation $\mathfrak{L}$, specified by the weighting sequence h_j, can, of course, be realized on a computer only when it is *physically realizable* in the sense that $h_j = 0$ for all $j < 0$. Similarly, the general linear transformation of the form (2.235) is said to be *physically realizable* if the condition $h_j^{(n)} = 0$ can be satisfied for all $j < 0$ and $n = 1, 2, \dots$. The physical realizability condition can

also be expressed in terms of the transfer function $H(\omega)$ of the transformation $\mathfrak{L}$. [63] In practice, however, this condition is usually not essential.   To see why it is not essential, we apply $\mathfrak{L}$ to the backward shifted (i.e. delayed) sequence $S_{-T}\{X(t)\} = X(t - T)$, where $X(t)$ is a stationary sequence.   We obtain then the sequence $\mathfrak{L}\{S_{-T}X(t)\}$, none of the statistical characteristics of which differ from those of the sequence $\mathfrak{L}\{X(t)\}$. Therefore, in practice one can use $\mathfrak{L}\{S_{-T}X(t)\}$ in place of $\mathfrak{L}\{X(t)\}$. When applied to $\mathfrak{L}\{S_{-T}X(t)\}$, however, the condition for physical realizability requires only that the condition $h_j^{(n)}$ $= 0$ (where $h_j^{(n)}$ are the coefficients in (2.235)) be fulfilled for all $j < -T$.   It is clear that this last condition can always be approximately satisfied, with any preassigned degree of accuracy, provided we choose T to be large enough.

15.   Examples of Discrete Linear Transformations and Correlation Functions of Stationary Sequences

In Sec. 6 we have already considered some particular examples of correlation functions of stationary random sequences and the linear transformations of such sequences. Now it is expedient to revert to such examples, giving special attention to calculating the corresponding spectral densities $f(\omega)$ and transfer functions $H(\omega)$.   We begin with examples of discrete linear transformations.

Example 1.   *Differencing.*   The discrete analog of the differentiation operation is the (*first-order*) *difference operation*

(2.244) $DX(t) = X(t) - X(t - 1)$.

This is a discrete linear transformation, which can be written as $\mathfrak{D} = \mathcal{E} - S$ where $\mathcal{E}$ is the identity transformation, and $S = S_{-1}$ is the *backward shift* (or *unit delay*) *operator*, defined by $SX(t) = S X(t - 1)$).   The transfer function of the transformation $\mathfrak{D}$ is clearly equal to

(2.245) $H(\omega) = 1 - e^{-i\omega} = 2ie^{-i\omega/2}\sin(\omega/2)$.

Hence the gain of $\mathfrak{D}$ is given by the equation

(2.245a) $A(\omega) = |H(\omega)| = 2|\sin(\omega/2)|.$

Applying n times the transformation $\mathfrak{D}$, we arrive at the *nth-order difference operation* $\mathfrak{D}^n$.   The transfer function and the gain of $\mathfrak{D}^n$ are equal to

(2.246) $H(\omega) = (2i)^n e^{-in\omega/2}[\sin(\omega/2)]^n,\ A(\omega) = 2^n|\sin\omega/2|^n.$

Equation (2.245a) shows that $\mathfrak{D}$ is a discrete (non-ideal) high-pass filter which attenuates the power in low frequencies but magnifies it in high frequencies close to the highest frequency $\omega = \pi$ (or, what is in fact the same, $\omega = -\pi$).   Such a behavior is even more pronounced for transformation $\mathfrak{D}^n$ where $n > 1$. At large values of n the transformation $\mathfrak{D}^n$ will be a filter with a very narrow pass band adjoining the frequency $\omega = \pi$ and shrinking into this point as $n \to \infty$. It follows that if the point $\omega = \pi$ belongs to the frequency spectrum of $X(t)$, and n is large enough, then any finite stretch of a fixed length T of the realization of the sequence $\mathfrak{D}^n X(t)$ will be approximated with any preassigned degree of accuracy, and with a probability arbitrarily close to unity, by the series of terms with the same absolute value but with alternating signs.[64]

Example 2. *Moving averages.* The difference operators of the form

(2.247) $\mathfrak{L} = a_n\mathfrak{D}^n + a_1\mathfrak{D}^{n-1} + \ldots + a_n\mathcal{E}$

are the natural analog of the differential operators with constant coefficients of the form (2.179). Transformations (2.247) are contained in the class of general *(one-sided) moving average transformations* of order n:

$$\mathfrak{L} = b_0\mathcal{E} + b_1 S + \ldots + b_n S^n,$$

(2.248)

$$\mathfrak{L}\{X(t)\} = b_0 X(t) + b_1 X(t-1) + \ldots + b_n X(t-n).$$

These moving averages are physically realizable linear summation transformations specified by a weighting sequence containing only a finite number, $n + 1$, of nonzero weighting coefficients. The transfer function of the transformation (2.248) is equal to

(2.249) $H(\omega) = b_0 + b_1 e^{-i\omega} + ... + b_n e^{-in\omega}.$

A widely used particular case of moving average transformation is the case of the *equally weighted average*, i.e. of the *arithmetic mean* of n successive values of a sequence:

(2.250) $\mathcal{L}\{X(t)\} = \dfrac{1}{n} \sum\limits_{k=0}^{n-1} X(t - k).$

Here, of course,

$$H(\omega) = \frac{1}{n} \sum_{k=0}^{n-1} e^{-ik\omega} = e^{-i(n-1)\omega/2}\, \frac{\sin(n\omega/2)}{n\sin(\omega/2)},$$

(2.251)

$$A(\omega) = \left| \frac{\sin(n\omega/2)}{n\sin(\omega/2)} \right|$$

(so that, e.g., $A(\omega) = |\cos(\omega/2)|$ for $n = 2$). It is clear that the equally weighted averaging (2.250) is a particular type of (non-ideal) low-pass filtering.

Note also that a sequence of moving average transformations $\mathcal{L}_n$, $n = 1, 2, ...$, can be so chosen that $\mathcal{L}_n$, at large enough values of n, will approximate arbitrarily closely an "ideal line filter" which amplifies oscillations of a given frequency ω_0 but eliminates all the other periodic oscillations. This fact explains the early result of Slutsky, who constructed long before the appearance of the spectral theory of random functions, a sequence of moving averages $\mathcal{L}_n$, $n = 1, 2, ...$, having the property that in any given finite time interval the values of any realization of $\mathcal{L}_n E(t)$, where $E(t)$ is a sequence of uncorrelated random variables and n is large enough, differs by an arbitrarily small amount, and with a probability arbitrarily close to unity, from the values of a sine wave of the form $A\sin(\omega_0 t + \varphi)$, where A, φ, and ω_0 are constants. (Slutsky considered only the case where $0 < \omega_0 < \pi$. Moreover, Example 1 above is related to the case where $\omega_0 = \pi$, and the sequence of transformations (2.250) where $n = 1, 2, ...$, corresponds to $\omega_0 = 0$.) This "sinusoidal limit theorem" of Slutsky seemed quite surprising in its day, but now it can be easily proved and extended considerably by using the spectral theory of stationary random sequences.[65]

Example 3. *Recursive filters.* Moving average

transformations of a finite order n are easily realized on digital computers, but, of course, they represent just a restricted class of rather special discrete linear transformations. In particular, all such transformations have a "short memory":   the value of $\mathfrak{L}\{X(t)\}$ depends in this case only on a finite number of past values of $X(t)$. Of course, we can also consider the moving averages of infinite order which have more "memory" than the nth-order transformation (2.248), but such moving averages of infinite order are inconvenient for practical applications. There is, however, a rather wide class of very simple physically realizable linear transformations which also includes many moving average transformations of infinite order. This is the class of so-called *recursive linear filters* $\mathfrak{L}$ of a finite order. For such filters $Y(t) = \mathfrak{L}\{X(t)\}$ (the filter output at time t) depends linearly on the values of $X(t')$ at the time $t' = t$, and at a finite number of time points $t' < t$ from the past, and also on a finite number of past values of $Y(t')$, $t' < t$. Therefore, such filters $\mathfrak{L}$ are defined by formulae

(2.252)
$$Y(t) = -a_1 Y(t-1) - \dots - a_m Y(t-m)$$
$$+ b_0 X(t) + b_1 X(t-1) + \dots + b_n X(t-n),$$

where a_j, $j = 1, \dots, m$, and b_k, $k = 0, 1, \dots, n$ are constant coefficients.

Let us denote by $H(\omega)$ the transfer function of the recursive filter specified by (2.252). Substituting into (2.252) in place of all the random variables $Y(t')$ and $X(t')$ their spectral representations, and using (2.237), we obtain

(2.253) $\int_{-\infty}^{\infty} e^{it\omega}[P(e^{-i\omega})H(\omega) - Q(e^{-i\omega})]dZ(\omega) = 0$, $t = 0,\pm 1,\pm 2,\dots$,

where

(2.254) $P(z) = \sum_{j=0}^{m} a_j z^j$, $a_0 = 1$, $Q(z) = \sum_{k=0}^{n} b_j z^j$.

Relation (2.253) must be valid for all $Z(\omega)$ for which the integral on the right-hand side of (2.253) converges, since (2.252) holds for all sequences $X(t)$ to which the filter $\mathfrak{L}$

can be applied.   Therefore it is clear that in this case*

$$(2.255) \quad H(\omega) = \frac{Q(e^{i\omega})}{P(e^{i\omega})} \ .$$

Conversely, if the linear transformation $\mathfrak{L}$ has a transfer function of the form (2.255), then $Y(t) = \mathfrak{L}\{X(t)\}$ will obviously satisfy (2.252), i.e. $\mathfrak{L}$ will be a linear recursive filter.

When all the values of $Y(t')$ with $t' < t$ are already known, the value of $Y(t)$ can be calculated very easily from the known values of $X(t')$ with the aid of (2.252).   The only problem is that it is not clear how one should behave at the beginning, when no values of $Y(t')$ are yet known. From this point of view, of greatest interest are the so-called *stable recursive filters*, which possess the following important property: If for some t_0 we choose arbitrarily the values $y(t_0 - 1)$, ..., $y(t_0 - m)$ of $Y(t_0 - 1)$, ..., $Y(t_0 - m)$ and calculate consecutively the values $y(t_0)$, $y(t_0 + 1)$, $y(t_0 + 2)$, ...   by (2.252), where the known values $x(t_0 - n)$, $x(t_0 - n + 1)$, ..., $x(t_0)$, $x(t_0 + 1)$, $x(t_0 + 2)$, ... of the stationary sequence $X(t)$ are used, then $y(t)$, as $t - t_0 \to \infty$, tends to the sampling value (i.e. realization) of $Y(t) = \mathfrak{L}\{X(t)\}$, where $\mathfrak{L}$ is a recursive filter specified by (2.252).   It is obvious that the recursive filter is stable if and only if the effect of the initial conditions on the solution of the finite-difference equation (2.252) (with respect to the unknowns $Y(t)$) gradually dies out with increasing argument t.   From the general theory of linear difference equations it follows that the last condition will hold if and only if all the roots of the equation $P(z) = 0$ exceed unity in absolute value, i.e. are located outside the unit circle of the complex plane; cf. Sec. 6 above, Examples 6 and 7 (see also Notes 3 and 6 to this chapter). The same theory implies also that the asymptotic damping rate of the effects of the initial conditions with an increase in $t - t_0$ is

the same as the damping rate of a function $|\alpha|^{-(t-t_0)}$, where α is the smallest (in absolute value) root of the equation $P(z) = 0$.   Therefore, knowing the coefficients a_j, one can easily

---

*Formula (2.255) for $H(\omega)$ can also be obtained by formal substitution of $\exp(i\omega t)$ in place of $X(t)$ and $H(\omega)\exp(i\omega t)$ in place of $Y(t) = \mathfrak{L}X(t)$ in (2.252).

estimate how many initial terms $y(t_0)$, $y(t_0+1)$, $y(t_0+2)$, ..., of the sequence $y(t)$ should be omitted to justify the assumption that the subsequent terms give the realization of the stationary sequence $Y(t)$ satisfying (2.252).   Since modern computers can calculate many values of $y(t)$, $t = t_0$, t_0+1, t_0+2, ... very rapidly, discarding even the comparatively long initial part of this series does not involve much difficulty. Hence, stable recursive filters are readily realized on computers.

We also note that a linear recursive filter $\mathfrak{L}$ specified by (2.252) can be formally expressed via the backward shift operator S with the aid of the equation

$$(2.256) \quad \mathfrak{L} = \frac{b_0 \mathcal{E} + b_1 S + ... + b_n S^n}{\mathcal{E} + a_1 S + ... + a_m S^m} = \frac{Q(S)}{P(S)}$$

(since (2.252) implies that $P(S)\mathfrak{L}\{X(t)\} = Q(S)X(t)$). In the case of a stable recursive filter $P(z) = a_m \prod_{j=1}^{m} (z - \alpha_j)$, where $|\alpha_j| > 1$ for all $j = 1, ..., m$.   Hence the function $[P(z)]^{-1}$ can be expanded into a power series which converges if $|z| \leqslant 1$.

Thus, $[P(z)]^{-1} = \sum_{j=0}^{\infty} \gamma_j z^j$, where $\gamma_0 = 1$, and (2.256) can be rewritten as

$$(2.257) \quad \mathfrak{L} = Q(S) \sum_{j=0}^{\infty} \gamma_j S^j = \sum_{j=0}^{\infty} c_j S^j, \quad \text{i.e.} \quad Y(t) = \sum_{j=0}^{\infty} c_j X(t - j).$$

(The series on the right-hand side of the formula for $Y(t)$ converges in the mean-square sense.) It is easy to show that $Y(t)$ specified by (2.257) coincides with the stationary solution of (2.252) (see, e.g., Note 6).   Hence, the stable recursive filter $\mathfrak{L}$ can also be represented in the form of a moving average transformation of infinite order. But the specification of this filter by an equation of the form (2.252) is considerably more convenient since it simplifies the application of $\mathfrak{L}$ to a given realization $x(t_0)$, $x(t_0 + 1)$, ..., of the random sequence $X(t)$.

Stable   recursive   filters   embrace   a   wide   class   of physically   realizable   discrete   linear   transformations,   and many of them possess some useful properties. In particular, a   recursive   filter   exists   which   approximates   with   good

accuracy a band-pass (or low-pass, or high-pass) filter with any given pass band. Many examples of interesting recursive filters of practical importance can be found in the literature referred to in Note 62. However, here we shall restrict ourselves to just one, very simple (and frequently encountered in applications[66]) recursive filter $\mathfrak{L}$ specified by the formula

$$(2.258) \quad Y(t) = aY(t-1) + (1-a)X(t),$$

where a is a real number and $|a| < 1$. According to (2.255), in this case

$$(2.259) \quad H(\omega) = \frac{1-a}{1-ae^{-i\omega}}, \qquad A(\omega) = \frac{1-a}{(1-2a\cos\omega + a^2)^{1/2}}$$

(see Fig. 36). For $a > 0$ the filter $\mathfrak{L}$ is a (non-ideal) low-pass filter, and for $a < 0$, a high-pass filter. Formulae (2.257) in this case clearly take the form

$$(2.260) \quad \mathfrak{L} = (1-a) \sum_{j=0}^{\infty} a^j S^j, \quad Y(t) = (1-a) \sum_{j=0}^{\infty} a^j X(t-j).$$

Hence the recursive filter (2.258) can be regarded as the discrete form of exponentially weighted averaging (2.183).*

We pass on now to examples of correlation functions and the corresponding spectral densities of stationary random sequences. (The correlation functions $B(\tau)$ which do not fall off rapidly enough, and therefore cannot be represented in terms of the spectral density, will not be considered here at all.) We begin with the simplest example, which has already been discussed in Sec. 6, Example 1.

Example 1. Let

$$(2.261) \quad B(0) = 1, \text{ and } B(\tau) = 0 \text{ for } \tau \neq 0.$$

According to formula (2.213) the corresponding spectral density is

---

*The analogy between (2.183) and (2.260) becomes especially striking if we introduce the notation $T^{-1} = \ln(a^{-1})$, which enables us to replace a^j in (2.260) by $\exp(-j/T)$.
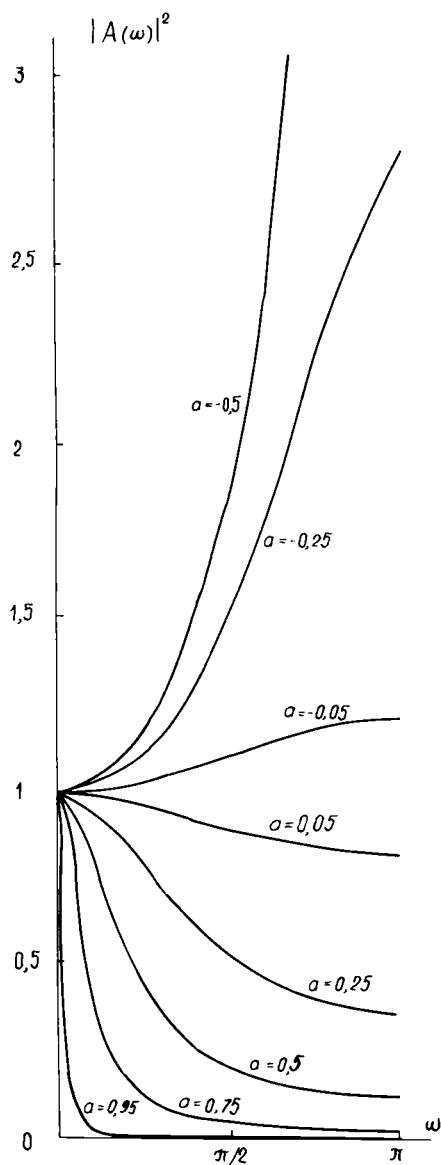
Fig. 36. The squared gain function of the recursive filter
          (2.258) for different values of the parameter a
          (Otnes and Enochson, 1972).

(2.262) $f(\omega) = 1/2\pi, \ -\pi \leqslant \omega < \pi.$

The fact that $f(\omega)$ is positive for all ω confirms that the function (2.261) is a correlation function of some stationary sequence. Recall that in Sec. 6 the same result was deduced from the consideration of a stationary sequence $E(t)$ of uncorrelated random variables normalized by the condition $\langle E^2(t) \rangle = 1$.

Similarly, if $B(\tau) = 0$ for $\tau \neq 0$, and $B(0) = C$, where $C > 0$, then $f(\omega) = C/2\pi = $ const. This correlation function (and spectral density) corresponds to the more general case of an arbitrary (non-normalized) sequence of uncorrelated random variables. Thus, stationary sequences of uncorrelated random variables are characterized by the fact that their spectral densities are constant over the whole interval $-\pi \leqslant \omega \leqslant \pi$. Therefore such a sequence is sometimes called a *white noise sequence* or a *discrete white noise*, or a *white noise with discrete parameter*.

Example 2. Let $B(\tau)$ decrease with increasing $|\tau|$ from unity to zero in arithmetic progression, and then remain equal to zero, i.e.

(2.263) $B(\tau) = \begin{cases} 1 - |\tau|/m & \text{for } |\tau| \leqslant m, \\ 0 & \text{for } |\tau| > m, \end{cases}$

where m is an integer. Then, applying (2.213), we easily obtain

(2.264) $f(\omega) = \dfrac{1}{2\pi}\Big\{1 + 2\sum_{\tau=0}^{m-1}(1 - \dfrac{\tau}{m})\cos\tau\omega\Big\} = \dfrac{\sin^2(m\omega/2)}{2\pi m \ \sin^2(\omega/2)}.$

(In deriving the last formula we used formulae 1.342.2 and 1.352.2 from Gradshteyn and Ryzhik's book, 1980.) The function (2.264) is obviously non-negative at all ω, and this confirms that the function (2.263) is a correlation function of a stationary random sequence. This result has already been proved in Sec. 6, Example 2, which shows that a sequence $X(t)$ with a correlation function of the form (2.263) can be obtained by applying the simple moving average transformation $\mathfrak{L} = (\mathcal{E} + S + ... + S^{m-1})/\sqrt{m}$ to a normalized sequence of uncorrelated random variables $E(t)$. Note also that (2.2), (2.239a), (2.251), and (2.262) allow one to obtain (2.264) very easily without any analytical transformations.

That the function (2.263) belongs to the class of correlation functions of stationary sequences also follows from the fact that this function of an integer-valued argument coincides with the values, at integer values of τ, of the continuous-time correlation function (2.123), where $T = m$, $C = m^{-1}$. In fact, it has already been explained in Sec. 14 (see p. 186) that the values of any correlation function of a stationary random process at a set of equidistant time points always form the correlation function of a stationary random sequence. (It is also easy to show that *any* correlation function of a stationary sequence can be obtained from some correlation function $B(\tau)$ of a stationary process by restricting the set of admissible values of τ to the set of integers.[67]) According to (2.124) and (2.229) the spectral density $f(\omega)$ corresponding to the correlation function (2.263) must satisfy the relation

$$f(\omega) = \sum_{k=-\infty}^{\infty} \frac{2\sin^2[\omega + 2k\pi)m/2]}{\pi m (\omega + 2k\pi)^2}$$

$$= \frac{\sin^2(\omega m/2)}{2\pi^3 m} \sum_{k=-\infty}^{\infty} \frac{1}{(\omega/2\pi + k)^2} .$$

The last equation also permits us to derive easily (2.264) by using the known identity

$$(2.265) \qquad \sum_{k=-\infty}^{\infty} \frac{1}{(x + k)^2} = \frac{\pi^2}{\sin^2 \pi x}$$

(see Gradshteyn and Ryzhik, 1980, formula 1.422.4). Conversely, assuming (2.264) to be known, we arrive at a new proof of identity (2.265).[68]

If $B(\tau)$ is given by a more general equation (2.5), then, according to (2.4), (2.239a), (2.249), and (2.262), we obtain

$$(2.266) \qquad f(\omega) = \left| \sum_{k=0}^{n} b_k e^{-ik\omega} \right|^2 / 2\pi.$$

Example 3. Let $B(\tau)$ decrease with increasing $|\tau|$ in geometric progression:

$$(2.267) \qquad B(\tau) = C a^{|\tau|}, \quad C > 0, \quad |a| < 1, \quad \tau = 0, \pm 1, \pm 2, \dots,$$

where a is a real number. Then, according to (2.213),

$$f(\omega) = \frac{C}{2\pi} \sum_{k=-\infty}^{\infty} a^{|k|} e^{-ik\omega} = \frac{C}{2\pi} \left\{ \sum_{k=1}^{\infty} (ae^{i\omega}) + \sum_{k=0}^{\infty} (ae^{-i\omega}) \right\}$$

(2.268)
$$= \frac{C}{2\pi} \left\{ \frac{ae^{i\omega}}{1 - ae^{i\omega}} + \frac{1}{1 - ae^{-i\omega}} \right\}$$

$$= \frac{C}{2\pi} \left\{ \frac{a}{e^{-i\omega} - a} + \frac{e^{i\omega}}{e^{i\omega} - a} \right\} = \frac{C}{2\pi} \frac{1 - a^2}{|e^{i\omega} - a|^2}.$$

We see that $f(\omega) \geqslant 0$ for all ω, and hence (2.267) is the correlation function of a stationary random sequence. Note that this result has already been proved, although quite differently, in Sec. 6, Example 5.   The comparison of (2.268) and (2.259) also shows that the sequence $Y(t)$ with the correlation function (2.267) can be obtained by applying the recursive filter (2.258) (or, what is the same, (2.260)) to the sequence $X(t) = cE(t)$ of uncorrelated random variables with the variance $c^2 = C(1 - a^2)/(1 - a)^2 = C(1 + a)/(1 - a)$.   (This result is also already known to us; see again Sec. 6, Example 5.) Conversely, the possibility of obtaining a sequence with the correlation function (2.267) by applying the recursive filter (2.258) to the sequence $cE(t)$, studied in Sec. 6, allows one to obtain (2.268) quite easily by using (2.239a), (2.259), and (2.262). It should finally be noted that the existence of a stationary random sequence having the function (2.267) as its correlation function follows also from the fact that the values, at integer-valued τ, of the continuous-time correlation function (2.94), where $\exp(-\alpha) = a$, coincide with (2.267). It follows, in particular, that the spectral density $f(\omega)$, corresponding to the correlation function (2.267), can also be determined from (2.95) and (2.229). Such a determination of $f(\omega)$ gives the formula

$$f(\omega) = \frac{C\alpha}{\pi} \sum_{k=-\infty}^{\infty} \frac{1}{\alpha^2 + (\omega + 2k\pi)^2}, \quad \alpha = -\ln a.$$

A comparison of this formula with (2.268) reveals a curious identity[69]

(2.269)
$$\sum_{k=-\infty}^{\infty} \frac{1}{\alpha^2 + (\omega + 2k\pi)^2} = \frac{1 - e^{-2\alpha}}{2\alpha(1 - 2e^{-\alpha}\cos\omega + e^{-2\alpha})}.$$

Example 4.   Consider now the case of an mth-order autoregressive sequence $X(t)$, which satisfies the linear difference equation (2.13). Then, according to (2.239a), (2.255), and (2.262), the spectral density is

$$
f(\omega) = \frac{c^2}{2\pi|1 + a_1 e^{-i\omega} + ..., + a_m e^{-im\omega}|^2}
$$

(2.270)
$$
= \frac{c^2}{2\pi|e^{im\omega} + a_1 e^{i(m-1)\omega} + ... + a_m|^2}
$$

$$
= \frac{c^2}{2\pi|(e^{i\omega} - \alpha_1) ... (e^{i\omega} - \alpha_m)|^2},
$$

where $\alpha_1, ..., \alpha_m$ are the roots of the equation $z^m + a_1 z^{m-1} + ... + a_m = 0$. The corresponding correlation function $B(\tau)$ can then be readily found from (2.214). Suppose, in particular, that $m = 2$. Then we can rewrite equation (2.214) in the form

$$
B(k) = \frac{c^2}{2\pi}\int_{-\pi}^{\pi} \frac{e^{ik\omega}d\omega}{(e^{i\omega} - \alpha_1)(e^{-i\omega} - \bar{\alpha}_1) \ (e^{i\omega} - \alpha_2)(e^{-i\omega} - \bar{\alpha}_2)}
$$

(2.271)
$$
= \frac{c^2}{2\pi i}\oint_{|z|=1} \frac{z^{k-1}dz}{(z - \alpha_1)(z - \alpha_2)(1/z - \bar{\alpha}_1)(1/z - \bar{\alpha}_2)},
$$

where $z = e^{i\omega}$. Now, applying the residue theorem, we can easily show that $B(k)$ coincides with the coefficient of $1/z^k$ in the expansion of the rational function

$$
f^*(z) = \frac{c^2}{(z - \alpha_1)(1/z - \bar{\alpha}_1)(z - \alpha_2)(1/z - \bar{\alpha}_2)}
$$

(2.272)
$$
= \frac{c^2 z^2}{(z - \alpha_1)(1 - \bar{\alpha}_1 z)(z - \alpha_2)(1 - \bar{\alpha}_2 z)}
$$

in a Laurent series. If the roots α_1 and α_2 are distinct, the expansion of the function $f^*(z)$ in partial fractions takes the form

$$f^*(z) = \frac{c^2}{(\alpha_1 - \alpha_2)(1 - \alpha_1\alpha_2)} \left[\frac{\alpha_1}{1 - \alpha_1^2} \left(\frac{\alpha_1}{z - \alpha_1} + \frac{1}{1 - \alpha_1 z} \right) \right.$$

$$\left. - \frac{\alpha_2}{1 - \alpha_2^2} \left(\frac{\alpha_2}{z - \alpha_2} + \frac{1}{1 - \alpha_2 z} \right) \right].$$

Let us assume now that both α_1 and α_2 are less then unity in absolute value. Then the power series expansions of the functions $1/(1 - \alpha_1 z)$ and $1/(1 - \alpha_2 z)$ clearly contain only non-negative powers of z, while expansions of $\alpha_1/(z - \alpha_1)$ and $\alpha_2/(z - \alpha_2)$ contain only positive powers of z^{-1}. It is easy to see that, in the case where α_1 and α_2 are real, the coefficient of $z^{-\tau}$, $\tau \geqslant 0$ in the Laurent expansion of $f^*(z)$ is equal to the right-hand side of eq. (2.15). Thus, we have proved the validity of this equation. Assuming that $\alpha_1 = \alpha_2 = \alpha$ or that $\alpha_1 = \alpha \exp(i\theta)$ and $\alpha_2 = \alpha \exp(-i\theta)$ are complex conjugate, we can quite similarly obtain (2.15a) and (2.15b). It is also clear that the same method can be applied to autoregressive sequences of a higher-than-second order.

Example 5. A generalization of the autoregressive sequence is the autoregressive-moving average sequence $X(t)$, which satisfies the difference equation (2.17). Such a sequence can be obtained by applying a recursive filter, specified by (2.252) with $b_0 = 1$, to the white noise sequence $X(t) = cE(t)$. According to (2.239a), (2.255), and (2.262), the spectral density $f(\omega)$ in this case has the form

$$f(\omega) = \frac{|c|^2|e^{in\omega} + b_1 e^{i(n-1)\omega} + \dots + b_n|^2}{2\pi|e^{im\omega} + a_1 e^{i(m-1)\omega} + \dots + a_m|^2}$$

(2.273)

$$= \frac{C}{2\pi} \frac{|(e^{i\omega} - \beta_1) \dots (e^{i\omega} - \beta_n)|^2}{|(e^{i\omega} - \alpha_1) \dots (e^{i\omega} - \alpha_m)|^2},$$

where $C = |c|^2 > 0$, $b_n \neq 0$, $a_m \neq 0$, $\alpha_1 \dots, \alpha_m$ are the roots of the equation $z^m + a_1 z^{m-1} + \dots + a_m = 0$, and $\beta_1, \dots, \beta_n$ are the roots of the equation $z^n + b_1 z^{n-1} + \dots + b_n = 0$. It is easy to show that there will be no loss of generality if we assume that all the roots $\alpha_1, \dots, \alpha_m$ have absolute values less than 1, while all the roots $\beta_1, \dots, \beta_n$ have absolute values which do not exceed 1. Under these conditions (2.273) represents, nevertheless, a general spectral density which is rational in $\exp(i\omega)$.[70]

As in the previous example we can use the residue theorem to prove that the correlation function $B(\tau)$ which corresponds to the spectral density (2.273) is, for $\tau \geqslant 0$, equal to the coefficient

of z^{-T} in the Laurent expansion of the rational function

$$f^*(z) = C\frac{(z - \beta_1)...(z - \beta_n)(1/z - \overline{\beta}_1)...(1/z - \overline{\beta}_n)}{(z - \alpha_1)...(z - \alpha_n)(1/z - \overline{\alpha}_1)...(1/z - \overline{\alpha}_m)} .$$

Therefore $B(\tau)$ can be calculated in this case by expanding the function $f^*(z)$ in partial fractions.

Suppose, for instance, that $n = 1$, $m = 1$, and both the roots $\beta_1 = \beta$ and $\alpha_1 = \alpha$ are real and have absolute values less than 1. In this case

$$(2.274) \quad f(\omega) = \frac{C}{2\pi}\frac{|e^{i\omega} - \beta|^2}{|e^{i\omega} - \alpha|^2} = \frac{C}{2\pi}\frac{(e^{i\omega} - \beta)(e^{-i\omega} - \beta)}{(e^{i\omega} - \alpha)(e^{-i\omega} - \alpha)} ,$$

and the expansion of $f^*(z)$ in partial fractions has the form

$$f^*(z) = \frac{C(\alpha - \beta)(1 - \alpha\beta)}{\alpha(1 - \alpha^2)}\left[\frac{\alpha}{z - \alpha} + \frac{1}{1 - \alpha z}\right] + \frac{C\beta}{\alpha}.$$

The function $1/(1 - \alpha z)$ is equal to 1 at $z = 0$ and is regular in the unit circle. Hence, its series expansion contains only non-negative powers of z and begins with a constant term, which equals 1. Moreover, $\alpha/(z - \alpha) = \sum\limits_{k=1}^{\infty} \alpha^k/z^k$.   Therefore, the coefficient of z^{-T} in the expansion of $f^*(z)$ is equal to $C(\alpha - \beta)(1 - \alpha\beta)\alpha^{T-1}/(1 - \alpha^2)$ for $\tau > 0$ and to $C(\alpha - \beta)(1 - \alpha\beta)/\alpha(1 - \alpha^2) + C\beta/\alpha = C(1 - 2\alpha\beta + \beta^2)/(1 - \alpha^2)$ for $\tau = 0$. Thus,

$$(2.275) \quad B(\tau) = \begin{cases} \dfrac{C(\alpha - \beta)(1 - \alpha\beta)}{1 - \alpha^2}\alpha^{|\tau|-1} & \text{for } \tau \neq 0, \\[4mm] \dfrac{C(1 - 2\alpha\beta + \beta^2)}{1 - \alpha^2} & \text{for } \tau = 0. \end{cases}$$

A great many other examples of correlation functions of stationary sequences can be found rather easily. In particular, a number of new examples can be obtained by examining the examples of correlation functions of stationary processes collected in Sec. 10 and restricting the values of their argument τ to the set of integers. However, we shall not pursue these examples here. Note only, to conclude this section, that if $B(\tau)$, $-\infty < \tau < \infty$, is a correlation function of a stationary process corresponding to rational spectral density in ω, then the correlation function $B(\tau)$, $\tau = 0, \pm1, \pm2, ...$, of a stationary sequence will necessarily correspond to the spectral density which is a rational function of $\exp(i\omega)$.[71]

Chapter 3
DETERMINATION OF THE STATISTICAL CHARACTERISTICS OF A STATIONARY RANDOM FUNCTION FROM EXPERIMENTAL DATA

In Chapter 1 we defined stationary random functions $X(t)$, introduced their main statistical characteristics — the mean values m and correlation functions $B(\tau)$ — and discussed some basic properties of the functions $X(t)$. Then in Chap. 2 we considered a number of specific models of stationary functions and studied the spectral representations of both the functions $X(t)$ themselves and the corresponding correlation functions $B(\tau)$. These spectral representations naturally led us to new important statistical characteristics — the spectral distribution functions $F(\omega)$ and the spectral densities $f(\omega)$. We showed that the exact knowledge of the correlation function $B(\tau)$ of a stationary function $X(t)$ is fully equivalent to the knowledge of the corresponding spectral distribution function $F(\omega)$ or (in cases where the spectral density exists) of the spectral density $f(\omega) = F'(\omega)$. However, in practical problems we seldom know beforehand the precise shape of either the correlation function or the spectral distribution function (or the spectral density); also, the mean value m of a stationary function $X(t)$ is quite often unknown to us. Therefore, it is very important to see whether it is possible to determine, if only approximately, the values of the main statistical characteristics of the function $X(t)$ (i.e. to *estimate* the values of these characteristics) from experimental data

related to the function $X(t)$.

The problem of estimating unknown quantities related to probability distributions of any kind belongs to the field of *statistics* (or, more precisely, of *statistical inference*). This chapter is devoted to statistical problems of estimating the mean value, correlation function, spectral density, and spectral distribution function of a stationary random function $X(t)$ by using the observed values of this function. The function $X(t)$ will always be considered real in this chapter; this restriction simplifies some of the arguments but is of no practical importance since the overwhelming majority of stationary functions encountered in applications are in fact real.

16.   Determination of the Mean Value of a Stationary Function $X(t)$

In this section we consider the problem of estimating the mean value $m = \langle X(t) \rangle$ of a stationary random function $X(t)$ from experimental data.   It is, however, worth recalling at first some general statistical concepts and then applying them to the simpler problem of determining the mean value of a random variable X.

In applications of probability theory, one ordinarily deals with observations of phenomena which repeat themselves many times. If a random variable X characterizes the actual observed phenomenon, then every observation permits us to determine one *sample value* (or *realization*, or *observed value*) x of X.   Hence, N independent observations supply us with observational data consisting of N independent sample values $x_1, ..., x_N$ of the variable X.   To estimate the unknown parameter c of the probability distribution of the variable X, we construct an *estimate* $c_N^* = c_N^*(x_1, ..., x_N)$ of c, which is a suitable function of all the observations, and set parameter c equal to the value of c_N^*.

Of course, the function $c_N^*(x_1 ..., x_N)$ will be useful as an estimate of c only if it possesses some "good properties". To formulate these properties we must note that c_N^* is a sample value of the random variable $c_N^*(X_1 ..., X_N)$, where $X_1 ..., X_N$ are independent random variables, all of which have the same probability distribution as X. The random variable

$c_N^*(X_1, ..., X_N) = C_N^*$ is called an *estimator* of c corresponding to the *estimate* $c_N^* = c_N^*(x_1, ..., x_N)$. (In most statistical texts the estimate and the estimator are denoted by the same symbol and the reader must decide from the context whether an estimate or an estimator is meant. Such usage is convenient in some respects and sometimes we also adopt it in this book.) The probability distribution of the estimator C_N^* describes the variability of the estimate c_N^* over distinct samples of N independent observations $x_1, ..., x_N$; it is called the *sampling distribution* of the estimate c_N^*. For a "good estimate" c_N^* the sampling distribution must, of course, be concentrated in the vicinity of the true value of the estimated parameter c.

Let us now formulate the stated requirement in terms of the first and second moments of the estimator C_N^*. Its mean value $\langle C_N^* \rangle$ describes the average value of the estimate c_N^* over all the possible samples of N observations. Therefore, the difference $\delta(C_N^*) = \langle C_N^* \rangle - c$ (which is called the *bias* of the estimate c_n^* or of the estimator C_N^*) characterizes the *systematic error* of the estimate c_N^*. If the bias is equal to zero, i.e. $\langle C_N^* \rangle = c$, then c_N^* is called an *unbiased estimate* (and C_N^*, an *unbiased estimator*). .

The unbiasedness of c_N^* is, of course, a desirable property but, in fact, it is not necessary and not sufficient for practical usefulness of the estimate. In practice, the case of a large sample size N is of the greatest importance, since if N is small, then any estimate of an unknown parameter usually has a low accuracy and hence is useless. Thus, in many practical situations it is sufficient to require only that $\delta(C_N^*)$ be small for large values of N. This requirement will clearly be satisfied if $\delta(C_N^*) \to 0$ as $N \to \infty$, i.e. if the estimate c_N^* is *asymptotically unbiased,* but it does not necessarily have to be strictly unbiased at any finite value of N.

It is also clear that unbiasedness alone does not guarantee at all that the estimate is useful in applications. A much more important requirement is that the estimate should get "closer and closer" to the true value of the estimated parameter as the number N of the used observations (i.e. the sample size) increases. The closeness of the estimate c_N^* to the true value of c is measured most conveniently by the *mean square error* $\Delta^2(C_N^*) = \langle (C_N^* - c)^2 \rangle$ *of the estimate* c_N^* (i.e. of the corresponding estimator C_N^*). If $\Delta^2(C_N^*) \to 0$ as $N \to \infty$, i.e. if $C_N^* \to c$ in the mean square sense as $N \to \infty$, then the

estimate c_N^* (and the estimator C_N^*) are said to be *consistent.** The consistency of the estimate c_N^* signifies that by a sufficient increase in the number of observations N one can achieve any desirable degree of accuracy (no matter how high). This property is, of course, particularly important from the standpoint of the applications.[1]

Note that according to (0.11)

$$\Delta^2(C_N^*) = \langle(C_N^* - \langle C_N^* \rangle)^2\rangle + (\langle C_N^* \rangle - c)^2 = \sigma^2(C_N^*) + \delta^2(C_N^*).$$

Hence, the estimate c_N^* (in other words, the estimator C_N^*) is consistent if and only if both its variance and bias tend to zero as $N \to \infty$. We see that *any consistent estimator is necessarily asymptotically unbiased.*

Consider now the estimation problem for the unknown mean value $m = \langle X \rangle$ of a random variable X. If the observational data consist of n sample values $x_1, ..., x_N$ of X, then it is natural to take the *sample mean* (i.e. the arithmetic mean of all the observed values) $m_N^* = (1/N) \sum_{i=1}^{N} x_i$ as the estimate of $m = \langle X \rangle$. The sample mean, as an estimate of m, has an obvious intuitive appeal and also has several attractive properties. In particular, m_N^* is an *unbiased estimate* of m, since the corresponding estimator

$$(3.1) \qquad M_N^* = \frac{X_1 + X_2 + ... + X_N}{N}$$

obviously satisfies the condition

$$(3.2) \qquad \langle M_N^* \rangle = m$$

---

*In statistical tests the estimate c_n^* is most often said to be consistent if C_N^* tends to c in probability as $N \to \infty$, i.e. if for any $\epsilon > 0$

$$\lim_{N \to \infty} P\{|C_N^* - c| > \epsilon\} = 0.$$

With this definition the convergence of $\Delta^2(C_N^*)$ to zero as $N \to \infty$ becomes only a sufficient (but not a necessary) condition for consistency (cf. the deduction of (1.40) from (1.37)). In this book, however, it is more convenient to take the mean square convergence of C_N^* to c as the definition of consistency.

for any N. Moreover, the estimate m_N^* is also consistent, i.e.

(3.3) $\langle (M_n^* - m) \rangle \to 0$, as $N \to \infty$.

In fact, it is easy to show that

(3.4) $\Delta^2 (M_N^*) = \langle (M_N^* - m)^2 \rangle = \sigma^2(M_N^*) = \dfrac{\sigma_x^2}{N}$

where σ_x^2 is the variance of X, which, as usual, is assumed to be finite. (The fact that $\sigma^2(M_N^*) \to 0$, i.e. $M_N^* \to m$, as $N \to \infty$, is called the *law of large numbers* in probability theory.) We see that the *root-mean-square error* $\Delta(M_N^*)$ of the estimator M_N^* falls off rather rapidly (like $N^{-1/2}$) as $N \to \infty$. This circumstance makes the sample mean m_N^* a valuable estimate of the mean value m.[2]

Let us now pass on to the problem of estimating the mean value $m = \langle X(t) \rangle$ of a stationary random function $X(t)$. The above considerations applied to this problem show that if we can make observations of N independent realizations of $X(t)$, written as $x_1(t), x_2(t), ..., x_N(t)$, then the arithmetic mean $m_N^*(t) = $

$(1/N) \sum\limits_{i=1}^{N} x_i(t)$ of $x_i(t)$ for every fixed value of t will be an

unbiased and consistent estimate of m. It is, however, clear that in order to obtain a sufficiently accurate estimate the number N must be chosen rather large, i.e. a large number of independent realizations must be used. At the same time, in practice, repeated observations of a random function (i.e. multiple recordings of the values $x(t)$ at fixed conditions associated with a random function $X(t)$) usually turn out to be quite complicated and expensive, and sometimes even impossible, for the necessary experiment cannot be repeated at all. Therefore it is highly desirable to be able to get along with as few realizations as possible. Indeed, the special practical value of the correlation theory of *stationary* random functions is due, to a considerable extent, to the fact that if $X(t)$ is stationary, its mean value m (and also its correlation function $B(\tau)$) can usually be easily calculated from just one realization of $X(t)$ (of the kind illustrated by Figs. 5 and 6).

The possibility of these calculations stems from the fact that for stationary random functions $X(t)$ all the time instants t are practically equivalent. Therefore, it seems expedient

here to try to replace the averaging over a set of different realizations by time averaging of the data relating to a single realization $x(t)$. With reference to the problem of estimating the mean value m of a stationary function $X(t)$ from its observed values $x(t)$ at $t = 1, 2, ..., T$ (in the case of discrete time t) or at $0 \leqslant t \leqslant T$ (in the case of continuous t), such an approach means that the following *statistics* (i.e. function of observations) is taken as an estimate of m:

$$(3.5) \qquad m_T^* = \frac{1}{T} \sum_{t=1}^{T} x(t)$$

or, accordingly,

$$(3.5a) \qquad m_T^* = \frac{1}{T} \int_0^T x(t)dt.$$

It is clear that the estimate m_T^* is unbiased. In fact, since $X(t)$ is stationary, the corresponding estimator

$$(3.6) \qquad M_T^* = \frac{1}{T} \sum_{t=1}^{T} X(t) \quad \text{or} \quad M_T^* = \frac{1}{T} \int_0^T X(t)dt$$

satisfies the condition

$$(3.7) \qquad \langle M_T^* \rangle = \langle X(t) \rangle = m \quad \text{for all } T$$

whether the time t is discrete or continuous. However, to be suitable for practical use the estimate m_T^* must also be consistent, i.e. its accuracy must increase unboundedly with an unbounded increase in the length T of the observation interval (this length now plays the role of the "sample size"). Thus, in practice, *the mean value* m *of the stationary random function* $X(t)$ *can be determined from a single realization if and only if the following condition holds*:

$$(3.8) \qquad \langle (M_T^* - m)^2 \rangle \to 0 \text{ as } T \to \infty,$$

i.e.

$$(3.9) \qquad \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} X(t) = m \quad \text{or} \quad \lim_{T \to \infty} \frac{1}{T} \int_0^T X(t)dt = m,$$

where the limit is understood, as usual, in the mean square sense. Note that under the condition (3.8) (or, what is the same, (3.9)) the value of m can be determined from a single realization with any desirable degree of accuracy, provided the observation time

T is chosen sufficiently long. Therefore, by using just one realization of $X(t)$ one can satisfy any accuracy requirements related to the determination of the mean value m.[*]

The problem of determining the conditions under which the time averages of the physical quantities approach, with increasing averaging time, the corresponding "ensemble averages" (i.e. probabilistic mean values, obtained by averaging the corresponding quantities over the whole space of experimental outcomes), first arose in statistical mechanics, where systems possessing the indicated property are traditionally called *ergodic systems* (see, e.g., Khinchin, 1949). Therefore, the mathematical theorems which state that the time averages over an infinite time interval coincide, for some classes of random functions, with the probabilistic mean values, are often called *ergodic theorems*.[**] Accordingly, the random functions $X(t)$, for which such theorems hold, are sometimes called *ergodic random functions*. (In order to avoid confusion we note that the terms *ergodic theorems* and *ergodic functions* sometimes refer also to more specific results and types of random functions; cf., e.g., p. 234 and Notes 10 and 20 to this chapter.)

The simplest ergodic theorem gives conditions under which (3.8) holds for a stationary random function $X(t)$. It is easy to see that this equation cannot be valid for any stationary random function $X(t)$. Indeed, suppose, e.g., that $X(t) = X(0)$ with probability one for any t, but $X(0)$ is not a constant. (All the realizations $x(t)$ are represented here by straight lines parallel to the t-axis.) The function $X(t)$ is evidently stationary, but the specification of its realization $x(t)$ is equivalent to the specification of a single sample value $x = x(0)$ of the random variable $X(0)$. In this case $m_T^* = x(0)$

---

[*]This does not mean, of course, that observation of several realizations is altogether useless for the determination of m. If the values of several realizations are available, then one can supplement the time averaging for each of the realizations by the subsequent averaging of the obtained time averages over the realization set. Such a procedure allows one to increase considerably the accuracy of the obtained estimate of m (or to achieve the same accuracy within a much shorter averaging time).

[**]Another widespread term for theorems of the same type (which is borrowed from probability theory and not from physics) is the laws of large numbers.

for any T, so that to determine $m = \langle X(t) \rangle = \langle X(0) \rangle$ with an acceptable accuracy we must have at our disposal a large number of different realizations $x(t)$ (i.e. values $x = x(0)$).

The given example is, of course, rather artificial: in any applied problem such a function $X(t)$ will always be regarded as a random variable $X = X(0)$, and nobody will apply the theory of random function to it. A more interesting (and also more general) example is as follows. Let $X(t) = X(t,A)$ be some set of distinct stationary random functions numbered by the values of some parameter A (which may have either a finite or denumerable, or even a continuous set of values). Suppose now that at the beginning the value of A is chosen at random in accordance with some probability distribution (so that A itself is a random variable), and then one caries out an experiment generating the random function $X(t,a)$, where a is the sample value of A. It is easy to see that the obtained composite random function $X(t) = X(t,A)$, which unites all the functions $X(t,a)$, is also stationary. However, if even for random functions $X(t,a)$ at any a the condition (3.8) holds, then for $X(t)$ it will nevertheless be erroneous. Indeed, according to (3.8) the variable $M_T^* = M_T^*(a)$ corresponding to a fixed value a of A tends to $m(a) = \langle X(t,a) \rangle$ as $T \to \infty$. However, the particular mean value $m(a)$ at a given a is, as a rule, different from the general mean value $m = \langle X(t) \rangle = \langle X(t,A) \rangle$ obtained by additional averaging of $m(a)$ over all the values of a. The situation described is the most common source of stationary, but non-ergodic random functions arising in practical problems. In any application non-ergodicity usually just means that the random function concerned is, in fact, an artificial union of a number of distinct ergodic stationary functions. A typical example of this kind is offered by air temperature fluctuations shown in Fig. 5(b). We have already noted on p. 24 that the curve in Fig. 5(b) can be regarded as a realization of a random function describing the results of all the measurements performed at a fixed point of the atmosphere, at a fixed time of the year and the day, and at fixed weather conditions, but it may, if so desired, also be related to the ensemble of the observations at any points and time instants and under quite arbitrary meteorological conditions (in particular, at an arbitrary mean temperature). It is clear that with this last condition the relevant function $X(t)$ will definitely be non-ergodic: the time averaging of one of its realizations will characterize only the mean temperature

at which the specific measurement was made, but not the "general mean temperature of the Earth's atmosphere," which is described by the value $\langle X(t) \rangle$.

The above reasoning shows that a general stationary random function may well be a mixture of different ergodic functions, for which the value m_T^* characterizes only one of the mixture components, but not the entire mixture as a whole. (See, in this respect, the concluding part of Note 10 to this chapter in Volume II, where the mathematical results are formulated which show that a non-ergodic stationary function can, in practice, always be regarded as a mixture of ergodic components.) If, however, the stationary function $X(t)$ is not an obvious "mixture", i.e. an artificial union of some collection of essentially different functions (i.e. if the "statistical ensemble" of realizations $x(t)$ is not to be chosen excessively wide), then ergodicity will, as a rule, occur. The point is that for relations (3.8) and (3.9) to hold, it is only necessary that certain very general conditions (very often met in practice) are fulfilled. Indeed, let us consider for definiteness the case of continuous time t and divide the segment $0 \leqslant t \leqslant T$ into n equal parts of length $T_1 = T/n$. Then the time average M_T^* of formula (3.6) can be represented as the arithmetic mean of n random variables

$$M_{T_1}^{(i)} = \frac{1}{T_1} \int_{(i-1)T_1}^{iT_1} X(t)dt, \quad i = 1, 2, ..., n.$$

If, as is usually the case, $b(\tau) = \langle (X(t + \tau) - m \rangle \langle X(t) - m) \rangle \to 0$ as $\tau \to \infty$, then at a sufficiently large $T_1 = T/n$ the variables $M_{T_1}^{(i)}$ and $M_{T_1}^{(j)}$ with $|j - i| > 1$ will be almost completely

uncorrelated (because they include only the values of $X(t)$ separated by an interval of length equal at least to T_1).

Moreover, even neighboring variables $M_{T_1}^{(i)}$ and $M_{T_1}^{(i+1)}$ will

be correlated rather weakly, since the vast majority of the values of $X(t)$ contained in $M_{T_1}^{(i)}$ are far removed from all the

values of $X(t)$ contained in $M_{T_1}^{(i + 1)}$. Therefore it is natural

to expect that if $b(\tau) \to 0$, as $\tau \to \infty$, one can apply to the variables $M_{T_1}^{(i)}$ (where $T_1 = T/n$ is large enough) the same

law of large numbers which holds for uncorrelated random variables. However, this means that in the case at hand the arithmetic mean of the n variables $M_{T_1}^{(i)}$ (which is equal to

$M_{nT_1}^* = M_T^*$) must tend to $\langle M_T^{(1)} \rangle = \langle X(t) \rangle = m$, as $n \to \infty$ (i.e. T

$\to \infty$), i.e. the relation (3.9) must hold.   Moreover, it can even be expected that the law of large numbers will be applicable to the variables $M_{T_1}^{(i)}$ (and, hence, the relation (3.9) will be

valid) under some more general conditions, when $b(\tau)$ does not necessarily tend to zero, as $\tau \to \infty$, but nevertheless the correlation between the variables $M_{T_1}^{(i)}$ and $M_{T_1}^{(j)}$ with $i \neq j$ for a large T_1 is rather weak.

The above-described expectations turn out to be true. It can be shown that *for a stationary random function $X(t)$ the relations (3.8) and (3.9) hold, if and only if*

$$(3.10) \qquad \lim_{T \to \infty} \frac{1}{T} \sum_{\tau=0}^{T-1} b(\tau) = 0$$

(in the case of discrete time) or, accordingly,

$$(3.10a) \qquad \lim_{T \to \infty} \frac{1}{T} \int_0^T b(\tau) d\tau = 0$$

(in the case of continuous time). The italicized statement was first proved by Slutsky and is called *Slutsky's ergodic theorem.* It is easy to see the *Slutsky's conditions* (3.10) and (3.10a) are certainly valid, if $\lim_{\tau \to \infty} b(\tau) = 0$, but they may hold even when

$b(\tau)$ does not tend to zero, as $\tau \to \infty$. [3]

To prove Slutsky's theorem we must express the variance $\sigma^2(M_T^*) = \langle (M_T^* - m)^2 \rangle$ in terms of the centered correlation function $b(\tau)$ of the process $X(t)$. By using (1.55) or its analog for the case of discrete time it is easy to show that

$$\sigma^2(M_T^*) = \frac{2}{T^2} \sum_{\tau_1=0}^{T-1} \sum_{\tau=0}^{\tau_1} b(\tau) - \frac{1}{T} b(0) =$$

$$(3.11)$$

$$= \frac{2}{T^2} \sum_{\tau=0}^{T-1} (T - \tau) b(\tau) - \frac{1}{T} b(0)$$

in the case of discrete time t and, accordingly,

$$(3.11a) \qquad \sigma^2(M_T^*) = \frac{2}{T^2} \int_0^T \int_0^{T_1} b(\tau) d\tau d\tau_1 = \frac{2}{T^2} \int_0^T (T - \tau) b(\tau) dt$$

in the case of continuous time t.[4] It only remains to prove that condition (3.10) is necessary and sufficient for the right-hand side of (3.11) to tend to zero, as $T \to \infty$, and the condition (3.10a), for the right-hand side of (3.11a) to tend to zero. This proof is also very simple (see Note 5).

We have already noted in Sec. 4 (see p. 61) that for stationary random functions $X(t)$ encountered in practice the condition $\lim_{\tau \to \infty} b(\tau) = 0$ is usually valid. In this case the

condition (3.10) or (3.10a) is clearly satisfied as well and, hence, Slutsky's theorem is applicable to $X(t)$. Then, by virtue of (3.8) and of Chebyshev's inequality (1.39), where $X_n - X$ is now replaced by $M_T^* - m$, one can choose, for any $\epsilon > 0$ and $\eta > 0$, a $T_0 = T_0(\epsilon, \eta)$ such that

$$(3.12) \qquad P\{\left| \frac{1}{T} \sum_{t=1}^T X(t) - m \right| < \epsilon\} > 1 - \eta$$

or, accordingly,

$$(3.12a) \qquad P\{\left| \frac{1}{T} \int_0^T X(t) dt - m \right| < \epsilon\} > 1 - \eta$$

for $T > T_0$. The last relations give us sufficient grounds to believe that for a large enough T,

$$(3.13) \qquad \frac{1}{T} \sum_{t=1}^T x(t) \approx m$$

or, accordingly,

$$(3.13a) \qquad \frac{1}{T} \int_0^T x(t) dt \approx m.$$

In the case of continuous time t and a function $x(t)$ realized in the form of a fluctuating electric current, the integral on the right-hand side of (3.13a) can often be easily calculated with the aid of an analog integrator performing the integration of the input current. One can also use a digital computer to perform numerical integration. In this case one must replace the right-hand side of (3.13a) with an integral sum, i.e. put

$$(3.13b) \qquad \frac{1}{N + 1} \sum_{k=0}^N x\left[\frac{kT}{N} \right] \approx m,$$

where T and N are large enough. To estimate the accuracy of the approximations (3.13) and (3.13a), one should use (3.11) and (3.11a), which explicitly express the mean square error $\sigma^2(M_T^*)$ via the correlation function $b(\tau)$. A similar formula for the mean square error of the approximation (3.13b) can easily be obtained from (3.11) by replacing $b(\tau)$ in the right-hand side of (3.11) with $b_{T/N}(\tau) = b(\tau T/N)$, $\tau = 0, 1, ..., N$ (cf. Sec. 14, eq. (2.223)).[6]

The outlined proof of Slutsky's theorem (due to Slutsky himself) does not use at all the spectral theory of stationary random functions. We now note that, since $b(\tau)$ is a correlation function of the stationary function $\overset{\circ}{X}(t) = X(t) - \langle X(t) \rangle = X(t) - m$, the function $b(\tau)$ can be represented in the form

$$(3.14) \qquad b(\tau) = \int_{-\Lambda}^{\Lambda} e^{i\tau\omega}\, d\overset{\circ}{F}(\omega),$$

where $\Lambda = \pi$ in the case of discrete time t, $\Lambda = \infty$ in the case of continuous time t, and $\overset{\circ}{F}(\omega)$ is the spectral distribution function of the stationary function $\overset{\circ}{X}(t)$. (The relation $B(\tau) = \langle X(t + \tau)X(t) \rangle = b(\tau) + m^2$ readily implies that $\overset{\circ}{F}(\omega) = F(\omega) - m^2 D(\omega)$, where $F(\omega)$ is the spectral distribution function of $X(t)$ and $D(\omega)$ is a "unit jump function", i.e. $D(\omega) = 0$ for $\omega \leqslant 0$ and $D(\omega) = 1$ for $\omega > 0$.) Moreover, according to (2.218) − (2.219) and (2.91),

$$(3.15) \qquad \lim_{T \to \infty} \frac{1}{T} \sum_{\tau=0}^{T-1} b(\tau) = \Delta\overset{\circ}{F}(0)$$

and also

$$(3.15a) \qquad \lim_{T \to \infty} \frac{1}{T} \int_0^T b(\tau)d\tau = \Delta\overset{\circ}{F}(0),$$

where $\Delta\overset{\circ}{F}(0) = \overset{\circ}{F}(+0) - \overset{\circ}{F}(-0)$ is a jump of the function $\overset{\circ}{F}(\omega)$ at the point $\omega = 0$. We see that Slutsky's theorem has also a rather simple spectral formulation: *the relation (3.9) holds for a stationary random function* X(t) *if and only if* $\Delta\overset{\circ}{F}(0) = 0$, *i.e. the spectral distribution function of* $\overset{\circ}{X}(t) =$ X(t) − $\langle$X(t)$\rangle$ *is continuous at the point* $\omega = 0$. This spectral formulation throws a new light upon the meaning of Slutsky's condition (3.10) or (3.10a). If $\Delta\overset{\circ}{F}(0) = f_0 > 0$, then the spectral representation (2.83) of $\overset{\circ}{X}(t)$ includes a constant random component X_0 satisfying the conditions $\langle X_0 \rangle = 0$, $\langle |X_0|^2 \rangle = f_0$, and the spectral representation of $X(t) = \overset{\circ}{X}(t) + m$ includes a constant term $m + X_0$. It is clear

that in this case the mean value $\langle X(t) \rangle = m$ cannot be uniquely determined by time averaging a single realization $x(t)$ of $X(t)$ depends on a sample value x_0 of the random variable X_0 and since $x(t)$ fluctuates about the constant level $m + x_0$ and not about m.   (Cf. the similar reasoning on pp. $215 - 216$, where the case of a function $X(t)$ of the form $X(t) = X_0 = \text{const}$ is considered.) Moreover, it is easy to show that $M_T^* \to m + X_0$ as $T \to \infty$ *in the general case of an arbitrary stationary function* $X(t)$, where $m = \langle X(t) \rangle$ and $m + X_0$ is the jump of the random function $Z(\omega)$ on the right-hand side of (2.61) at the point $\omega = 0$ (i.e. X_0 is the jump at $\omega = 0$ of the function $\overset{\circ}{Z}(\omega)$ appearing in the spectral representation of $\overset{\circ}{X}(t) = X(t) - \langle X(t) \rangle$); see the final part of Note 22 to Chap.   2 in   Vol. II.   This last result is a simple generalization of Slutsky's theorem.

Let us now return to the original (non-spectral) formulation of Slutsky's theorem. For the time being we regard the time t as continuous and assume that the function $b(\tau)$ not only approaches zero as $\tau \to \infty$, but also falls off so rapidly that

$$(3.16) \qquad \int_0^\infty b(\tau)d\tau = C < \infty.$$

(This last condition is, in fact, valid for almost all the processes $X(t)$ encountered in practice.) If a bounded spectral density $f(\omega)$ of the fluctuation process $\overset{\circ}{X}(t) = X(t) - \langle X(t) \rangle$ does exist, then, by virtue of (2.84), $C = \pi f(0)$; hence, C cannot be negative and is equal to zero only if $f(0) = 0$.   Let us first assume that $C > 0$ (the exceptional case where $C = 0$ will be briefly discussed a little later jointly with the case where $C = \infty$).   Then the correlation time (2.88) of the process $\overset{\circ}{X}(t)$ is finite and positive:

$$(3.17) \qquad T_1 = \frac{1}{b(0)} \int_0^\infty b(\tau)d\tau = \int_0^\infty R(\tau)d\tau = \frac{C}{b(0)}.$$

If the observation time T does not exceed very much the correlation time T_1, the estimate m_T^* will usually have a low accuracy and therefore will be of little use.   If, however, $T \gg T_1$ (this case is the most important for the applications), then, for all such values of τ that $b(\tau)$ is not very small, the term $Tb(\tau)$ in the integrand of the right-hand side of (3.11a) will greatly exceed the term $\tau b(\tau)$.   Moreover, if τ/T is not small and $T \gg T_1$, then $b(\tau)$ will be very small; therefore, the contribution from all such values of τ to the integral in the right-hand side of (3.11a) is negligible.   Thus the term $\tau b(\tau)$

of the integrand in the right-hand side of (3.11a) can be entirely neglected if $T \gg T_1$. Using now the relation $\int_0^T b(\tau)d\tau \approx \int_0^\infty b(\tau)d\tau = b(0)T_1$ which has a high accuracy for $T \gg T_1$, we get the following simple estimate of $\sigma^2(M_T^*) = \langle (M_T^* - m)^2 \rangle$, applicable when $T \gg T_1$:

$$(3.18)^* \qquad \sigma^2(M_T^*) \approx \frac{2b(0)T_1}{T} = \frac{2T_1}{T}\sigma_X^2.$$

Formula (3.18) shows that $\sigma(M_T^*)$ falls off approximately as $T^{-1/2}$ with increasing T; it is similar to the expression (3.4) for the variance of the arithmetic mean of N independent observations, but now the number $N_1 = T/2T_1$ appears in place of N. The number N_1 can be called the *effective number of independent observations*, because a comparison of (3.18) with (3.4) shows that at large values of T the determination of the mean value m from a realization of the length T is close in its accuracy to the determination of m from N_1 independent observations. (Thus, an increase in observation time by $2T_1$ is practically equivalent to the addition of one more independent measurement.)

For discrete time t, formulae (3.16) - (3.18) take the form:

$$\sum_{\tau=0}^\infty b(\tau) = C < \infty, \qquad T_1 = \frac{1}{b(0)}\left[\frac{1}{2}b(0) + \sum_{\tau=0}^\infty b(\tau)\right],$$

$$\sigma^2(M_T^*) \approx \frac{2b(0)T_1}{T} = \frac{2T_1}{T}\sigma_X^2.$$

Here, the formula for T_1 differs from (2.216) only by replacement of $B(\tau)$ with $b(\tau)$, while the formula for $\sigma^2(M_T^*)$ coincides exactly with (3.18) and can be derived quite similarly to the derivation of this result in the case of continuous t. Note that equation (3.18) can be used to estimate an adequate averaging time T in cases where an acceptable level of error is picked for dimensionless ratio

---

*Note that T need not be excessively large for formula (3.18) to be decently accurate. Thus, e.g., if $b(\tau) = A \exp(-\alpha|\tau|)$, then when $T \geqslant 4T_1 = 4/\alpha$ this formula allows one to determine $\sigma^2(M_T^*)$ with an accuracy of not less than 13%; if $b(\tau) = A \sin\alpha\tau/\alpha\tau$, then the error of the estimate (3.18) of $\sigma^2(M_T^*)$ will not exceed 5% for $T \geqslant 4T_1 = 2\pi/\alpha$ and 2.5% for $T \geqslant 8T_1 = 4\pi/\alpha$.

$\sigma(M_T^*)/m$, say $\sigma(M_T^*)/m = \epsilon$. In fact, according to (3.18), if ϵ is small enough (i.e. the necessary averaging time T is large enough), $\sigma(M_T^*)/m \leqslant \epsilon$ if and only if

$$(3.18a) \quad T \geqslant T_\epsilon = \frac{\sigma_X^2}{m^2} \frac{2T_1}{\epsilon^2}.$$

Formulae (3.18) and (3.18a) are simple enough, but their practical application is often greatly complicated by the absence of reliable data on the variance $B(0) = \sigma_X^2$ and on the time scale T_1. Nevertheless, a crude estimate of the typical "fluctuation level" $\sigma_X = [b(0)]^{1/2}$ and the corresponding "characteristic time" T_1 can be obtained in many cases from visual observation of the realization $x(t)$ or from physical considerations relating to the nature of the random function $X(t)$. Such an estimate is already sufficient for determining, with the aid of (3.18) and (3.18a), the order of magnitude of the mean square error in the estimate M_T^* at a given value of T or of the value of T which guarantees the given error level ϵ.[7]

Clearly, formulae (3.18) and (3.18a) are inapplicable to stationary random functions $X(t)$ having a zero or infinite correlation time T_1. In practical applications, processes $X(t)$ with a zero correlation time most often appear as derivatives $X(t) = Y'(t)$ of some stationary process $Y(t)$ with a twice differentiable correlation function $B_{YY}(\tau)$. In such cases, however, it is usually known beforehand that $\langle X(t)\rangle = \langle Y'(t)\rangle = 0$, so that the problem of estimating the value of $\langle X(t)\rangle$ is of no importance. Nevertheless, it is worth noting that if $T_1 = 0$, then, by virtue of (3.11) and (3.11a), the value of $\sigma(M_T^*)$ decreases with increasing T not as $T^{-1/2}$, but more rapidly, and hence the acceptable averaging times are generally shorter here then for cases where $T_1 > 0$. Similarly, if $T_1 = \infty$, then formulae (3.11) and (3.11a) imply that $\sigma(M_T^*)$ falls off more slowly than $T^{-1/2}$ as $T \to \infty$, i.e. the acceptable averaging times are longer than for cases where $T_1 < \infty$. (In particular, if the spectral density $f(\omega)$ exists, then $T_1 = 0$ when $f(0) = 0$ and $T_1 = \infty$ when $f(0) = \infty$. In these cases the rate of decrease of $\sigma(M_T^*)$, as $T \to \infty$, is simply related to the rate of decrease or increase of $f(\omega)$, as $\omega \to 0$.[8])

If $T \gg T_1$, then, under rather wide regularity conditions, the probability distribution of the estimator M_T^* of the mean value $m = \langle X(t)\rangle$ (and also of the related

estimator $M_{TN}^* = \dfrac{1}{N + 1} \sum\limits_{k=0}^{N} X\left(\dfrac{kT}{N}\right)$ where N is large enough) can be closely approximated by a normal (Gaussian) probability distribution.   The validity of this statement follows from the general *central limit theorem* for stationary random functions. Using a normal approximation for the probability distribution, one can easily construct the *confidence interval* for m, i.e.   the interval $[M_T - \Delta , M_T + \Delta]$ containing the true value of m with a probability p that is not less than a fixed (and picked beforehand) number which is sufficiently close to unity (for instance, equal to 0.95 or 0.99, or 0.995).[9]

   Relation (3.8) evidently guarantees that for any fixed sufficiently large T, the sample value m_T^* of the random variable M_T^* will be close to m with a probability that is only slightly less than unity.   However, it does not yet imply, of course, that the difference $|m_T^* - m|$ will necessarily be small for all sufficiently large values of T simultaneously (cf. p. 65, where a similar situation is considered in connection with the discussion of the continuity of the realization $x(t)$).   In applications, the values m_T^* are never considered simultaneously for all large values of T, but usually one restricts oneself to a fixed single value of T.   Besides, it is always assumed in applied problems that an event having a sufficiently low probability is practically impossible.   Therefore, from the practical point of view, the relation (3.8) (or even any of the weaker relations (3.12) and (3.12a)) fully justifies the use of estimates of the form (3.13) − (3.13b).   However, the problem on the conditions which guarantees that m_T^* will be close to m for all sufficiently large T simultaneously (i.e. that the numerical function m_T^* of the argument T will tend to a constant m, as $T \to \infty$) is, of course, very interesting from the standpoint of the mathematical theory of probability. Hence, it is not surprising that the problem is widely discussed in the mathematical literature.   The principal conclusion from this discussion is that, although theoretically the mean square convergency of M_T^* to m, as $T \to \infty$, does not necessarily imply the convergence of the numerical function m_T^* of T to m for any (or almost any, i.e. with probability one) realization $x(t)$, in actual applications convergence of the numerical function m_T^* to m takes place practically always with probability one.[10]

17. Determination of the Mean Square and Correlation Function of $X(t)$

So far we have considered only the simplest problem of estimating the mean value $m = \langle X(t) \rangle$ of a stationary function $X(t)$ from a single realization $x(t)$. However, some other statistical characteristics of the random function $X(t)$ are also often needed in applications. In this section we focus our attention on the problem of estimating the correlation function $B(\tau)$, but at the beginning we consider the case where the mean value $m^{(g)} = \langle g(X(t)) \rangle$ of some function g of the argument $X(t)$ is of interest.*

We already know that if the random function $X(t)$ is strictly stationary, then for any $g(x)$ the function $Y(t) = g(X(t))$ is also stationary (see Sec. 3, p. 51). Therefore, $m^{(g)}$ can be calculated approximately by time averaging of $g(x(t))$ if the function

$$b_g(\tau) = \langle [g(X(t + \tau)) - \langle g(X(t + \tau)) \rangle][g(X(t)) - \langle g(X(t)) \rangle] \rangle$$

satisfies Slutsky's condition (3.10) or (3.10a). The mean square error (i.e. the variance) of the estimator $M_T^{(g)} = (1/T)\sum_{t=1}^{T} g(X(t))$ or $M_T^{(g)} = (1/T)\int_0^T g(X(t))dt$ is expressed via the correlation function $b_g(\tau)$ with the aid of the usual formulae (3.11) or (3.11a) which were given in the previous section with reference to the special case where $Y(t) = g(X(t))$ coincides with $X(t)$ (i.e. $g(x) = x$). Moreover, if T is large enough, the probability distribution of the random variable $M_T^{(g)}$ can usually be assumed, with sufficient accuracy, to be normal (for the same reason as in the case where $g(x) = x$).

It will be shown later in this section that the function $b_g(\tau)$ can be approximately calculated from a realization $x(t)$ which allows one to determine also the realization $y(t) = g(x(t))$ of $Y(t) = g(X(t))$. However, the exact determination of $b_g(\tau)$ requires knowledge of the two-dimensional probability

---

*In the particular case where $g(x) = 1$ for $x < x_0$ and $g(x) = 0$ otherwise, the mean value $m^{(g)} = \langle g(X(t)) \rangle$ evidently coincides with the one-dimensional probability distribution function $F(x_0) = P\{X(t) < x_0\}$ of the variable $X(t)$.

probability distribution of random variables $\{X(t), X(t + \tau)\}$, and only in some special cases can it produce an explicit formula. Therefore we shall not seek generality but will concentrate on a special (but very important) case of the Gaussian function $X(t)$.* Here, the situation is noticeably simplified, because all the finite-dimensional probability distributions of $X(t)$ (and hence the functions $b_g(\tau)$ for all $g(x)$) are determined uniquely by the mean value m and the correlation function $b(\tau)$ of the random function $X(t)$.

For simplicity we assume, for the time being, that $m = 0$ and first consider the relatively simple question of determining, by time averaging, the mean square (i.e. the variance) $\mu^2 = b(0) = \langle X^2(t) \rangle$ of $X(t)$. It is easy to verify that if $X(t)$ is a Gaussian stationary function, then[11]

$$(3.19) \qquad b_{x^2}(\tau) = \langle X^2(t + \tau) X^2(t) \rangle - \langle X^2(t) \rangle^2 = 2b^2(\tau).$$

Thus, the estimator**

$$b_T^*(0) = \frac{1}{T} \sum_{t=1}^{T} X^2(t)$$

or

$$b_T^*(0) = \frac{1}{T} \int_0^T X^2(t) dt$$

of the variance $b(0)$ will be consistent if and only if the the following condition is fulfilled:

$$(3.20) \qquad \lim_{T \to \infty} \frac{1}{T} \sum_{\tau=0}^{T-1} b^2(\tau) = 0 \quad \text{or} \quad \lim_{T \to \infty} \frac{1}{T} \int_0^T b^2(\tau) d\tau = 0.$$

It is clear that if $b(\tau) \to 0$, as $\tau \to \infty$, condition (3.20) will necessarily be fulfilled (as well as Slutsky's condition (3.10) — (3.10a). However, (3.20) is generally more restrictive than (3.10)

---

*Some information related to several specific examples of non-Gaussian functions $X(t)$ can be found in Note 17.

**Since both the letters b and B are utilized in this book, later in this section we shall, as is often done in statistical books, make no distinction between the estimators (depending on the random function $X(t)$) and the corresponding estimates (the sampling values of the estimators depending on the realizations $x(t)$). Thus, we shall denote the estimator and the estimate by the same letter.

$- $ (3.10a): It follows from (3.20) that condition (3.10) or (3.10a) is also valid, but it is quite possible that Slutsky's condition (3.10) or (3.10a) holds for a given correlation function $b(\tau)$, but the corresponding condition (3.20) does not. This follows at once from the fact that the validity of Slutsky's condition is equivalent to the continuity at the point $\omega = 0$ of the spectral distribution function $\overset{\circ}{F}(\omega)$ of the stationary function $\overset{\circ}{X}(t) = X(t) - m$, while the validity of condition (3.20) is equivalent to the absence of a discrete spectrum of $\overset{\circ}{X}(t)$, i.e. to the continuity of $\overset{\circ}{F}(\omega)$ at all points ω; cf. p. 220 and eqs. (2.91), (2.218) $-$ (2.219), (2.93), and (2.220) $-$ (2.221). (The possibility of deriving Slutsky's condition from (3.20) can also be proved easily without any reference to the fact that $b(\tau)$ is a correlation function admitting a spectral representation.[12])

Formula (3.19) shows that the variance of the simplest estimate $b_T^*(0)$ of the mean square of the Gaussian stationary random function $X(t)$ with $\langle X(t) \rangle = 0$ can be obtained by simple replacement of the function $b(\tau)$ by $2b^2(\tau)$ in (3.11) or (3.11a). However, if the mean value m of the function $X(t)$ is also unknown, the situation is somewhat more complicated.* If one wishes to estimate the variance $\sigma_X^2 = b(0) = \langle X(t) - m)^2 \rangle$ of $X(t)$, it is natural to resort to time averaging of the square of the difference $x(t) - m_T^*$, i.e. in the case of continuous t, to use the estimate

$$b_T^*(0) = \frac{1}{T}\int_0^T \left[x(t) - \frac{1}{T}\int_0^T x(s)ds \right]^2 dt = \frac{1}{T}\int_0^T x^2(t)dt - \left[\frac{1}{T}\int_0^T x(t)dt \right]^2.$$

The mean value of the estimate $b_T^*(0)$ (i.e., of the corresponding estimator, where $x(t)$ is replaced by $X(t)$) can be expressed rather simply in terms of the centered correlation function $b(\tau) = \langle X(t + \tau)X(t) \rangle - \langle X(t) \rangle^2$ of the random function $X(t)$; in the Gaussian case, the variance of $b_T^*(0)$ can also be expressed via $b(\tau)$. However, in this case, all the formulae are considerably more cumbersome than with a known value of m. In particular, it can readily be verified that $\langle b_T^*(0) \rangle \neq b(0)$ if $m \neq 0$, so that the estimate $b_T^*(0)$ is biased. Therefore the mean square error $\Delta_T^2(0) = \langle [b_T^*(0) - b(0)]^2 \rangle$ of the estimate $b_T^*(0)$ does not coincide with its

---

*If $m = \langle X(t) \rangle$ is known but not equal to zero, it can, of course, be subtracted from all the observed values of x(t). Therefore, this case actually does not differ from the one where it is known that $\langle X(t) \rangle = 0$.

variance, but is equal to the sum of the variance $\sigma^2(b_T^*(0))$ and the square of the bias $\mathfrak{b}(b_T^*(0)) = \langle b_T^*(0) \rangle - b(0)$ (by virtue of (0.11)).   It can, however, be shown that if Slutsky's condition is valid, the square of the bias of $b_T^*(0)$ falls off rapidly, as $T \to \infty$, and the main contribution to $\Delta_T^2(0)$ at large values of T gives the term which coincides with the expression for the variance of $b_T^*(0)$ in the case where $m = \langle X(t) \rangle$ is known.[13] It follows that for an unknown value of $\langle X(t) \rangle$ the condition (3.20) also guarantees the consistency of the estimate $b_T^*(0)$ of the variance $b(0)$ of the Gaussian stationary random function $X(t)$.   Moreover, in the case, say, of a Gaussian stationary process $X(t)$ having a correlation

function $b(\tau)$ such that $b(\tau) \to 0$, as $\tau \to \infty$, and $\int_0^\infty b^2(\tau)d\tau = C_2$

$< \infty$, (3.18) can be applied to the process $X^2(t)$ whether $\langle X(t) \rangle$ is known or not.   Since $b_{x^2}(0) = 2b^2(0)$ by virtue of (3.19), the

mean square error of the estimate $b_T^*(0)$ (equal to $\sigma^2[b_T^*(0)]$ if $\langle X(t) \rangle$ is known) at sufficiently large values of T satisfies the relation

$$(3.21) \qquad \frac{\sigma^2[b_T^*(0)]}{[b(0)]^2} \approx \frac{4T_2}{T}, \quad T_2 = \frac{1}{b^2(0)} \int_0^\infty b^2(\tau)d\tau$$

(when $\langle X(t) \rangle$ is unknown, $\sigma^2[b_T^*(0)]$ must be replaced by $\Delta_T^2(0)$).

  *Let us now say a few more words about determining the higher moments $\mu^{(n)} = \langle X(t)^n \rangle$, $n > 2$, of a stationary random function $X(t)$ from a single realization $x(t)$.   We confine ourselves to the case where t is continuous (the case of discrete t is analyzed quite similarly) and first assume, for simplicity, that $X(t)$ is a Gaussian random process with $\langle X(t) \rangle = 0$.   It is clear that then all the moments $\mu^{(n)}$ of odd orders n vanish, so that only the determination of even moments is of interest. Using the results of Note 11 to this chapter in Vol. II, we can easily show that the centered correlation function $b_{x^{2k}}(\tau)$

$= \langle X^{2k}(t + \tau)X^{2k}(t) \rangle - \langle X^{2k}(t) \rangle^2$ of the process $X^{2k}(t)$, where k is a positive integer, is equal to a special linear combination of the functions $b^{2k}(t)$, $b^{2k-2}(\tau)$, ..., $b^2(\tau)$.   It can also be easily shown that the validity of condition

(3.20) implies that a similar condition holds for all the functions $b^n(\tau)$, where $n > 2$, and hence also for all the functions $b_{x^{2k}}(\tau)$, $k = 2, 3, \ldots$.[14] Thus, condition (3.20)

is sufficient for the possibility to accurately determining all the higher moments $\mu^{(2k)} = \langle X^{2k}(t) \rangle$ of a Guassian stationary process $X(t)$ by time averaging of the observed functions $x^{2k}(t)$ (i.e. for consistency of the estimators

$$M_T^{(2k)} = T^{-1} \int_0^T X^{2k}(t) dt \text{ of } \mu^{(2k)}).$$

Note now that if the integral of $b^2(\tau)$ from 0 to ∞ converges, then the integrals of $b^4(\tau)$, ..., $b^{2k}(\tau)$ from 0 to ∞ converge all the more. Hence, (3.18) is applicable in this case to all the estimators $M_T^{(2k)}$. Let us express, with the aid of the Isserlis formula given in Note 11, the variance $b_{x^{2k}}(0) = \langle X^{4k}(t) \rangle - \langle X^{2k}(t) \rangle^2$ of $X^{(2k)}(t)$ via the

moment $\langle X^{2k}(t) \rangle^2 = (\mu^{(2k)})^2$. Then it is easy to show that, e.g., asymptotic formulae for mean square errors of the estimators $M_T^{(4)}$ and $M_T^{(6)}$ of the moments $\mu^{(4)}$ and $\mu^{(6)}$ can be written in the form

$$(3.22) \quad \frac{\sigma^2(M_T^{(4)})}{(\mu^{(4)})^2} \approx \frac{64 T_4}{3T}$$

and

$$(3.23) \quad \frac{\sigma^2(M_T^{(6)})}{(\mu^{(6)})^2} \approx \frac{452 T_6}{5T},$$

where T_4 and T_6 are the correlation times of the processes $X^4(t)$ and $X^6(t)$.[15] It is natural to expect that the time scale T_4 is, as a rule, less than T_2, and T_6 less than T_4 (because $b^4(\tau)$ falls off, as $\tau \to \infty$, more rapidly than $b^2(\tau)$, and $b^6(\tau)$ more rapidly than $b^4(\tau)$). However, the differences between these time scales can hardly be so significant in most cases as to offset the very fast growth of the numerical coefficients in (3.21), (3.22), and (3.23) (at first 4, then 64/3 $\approx$ 20, and then 452/5 $\approx$ 90). Indeed, let us assume, e.g., that $B(\tau) = b(\tau) = C \exp(-\alpha |\tau|)$. Then the correlation time of $X(t)$ is $T_1 = 1/\alpha$, and all the time scales T_2, T_4, and T_6 can be expressed explicitly via T_1. In this case (3.21), (3.22),

and (3.23) take the form[16]

$$\frac{\sigma^2(b_T^*(0))}{[b(0)]^2} = \frac{\sigma^2(M_T^{(2)})}{(\mu^{(2)})^2} \approx \frac{2T_1}{T}, \quad \frac{\sigma^2(M_T^{(4)})}{(\mu^{(4)})^2} \approx \frac{28T_1}{3T},$$

(3.24)

$$\frac{\sigma^2(M_T^{(6)})}{(\mu^{(6)})^2} \approx \frac{466T_1}{15T}.$$

Thus, let $X(t)$ be a Gaussian random process with an exponential correlation function of the form (2.94), and one wishes to determine, by time averaging, the second, the fourth, and the sixth moments of $X(t)$ with the same relative error $\sigma(M_T^{(2k)})/\mu^{(2k)}$, $k = 1, 2, 3$. Then in the case of the fourth moment one must use averaging over an approximately 5-fold time interval, and in the case of the sixth moment, over an approximately 15-fold time interval as compared with the second moment. Similar results are valid for other, not too exotic, forms of the correlation function $B(\tau)$ of the process $X(t)$.

The approximate formulae $(3.21) - (3.24)$ are sometimes also used when the random function $X(t)$ is not exactly Gaussian but when there are grounds to expect that its finite-dimensional probability distributions do not differ significantly from the normal distributions. In fact, however, in such cases one must be cautious and take into account that even a small deviation from normality may alter considerably the results referring to moments of a relatively high order. Besides, in the case of non-Gaussian random functions $X(t)$, there arises an additional problem of estimating central moments $\langle [X(t) - \langle X(t) \rangle]^n \rangle$ of odd orders n (equal to zero in the Gaussian case). These odd-order moments are often estimated by time averaging the odd power of a realization $x(t)$ less accurately than the moments of the neighboring even orders. Hence, to obtain sufficiently reliable estimates of odd moments, one must choose a longer averaging interval than in the case of neighboring even order. Note also that if there are no grounds to assume that the probability distributions of $X(t)$ are rather close to normal distributions, formulae of the form $(3.21) - (3.24)$ cannot be used at all, since in such cases they can sometimes lead to absolutely wrong results.[17] *.

We now pass on to the problem of estimating the values of the correlation function $B(\tau) = \langle X(t + \tau)X(t)\rangle$ (or $b(\tau) = B(\tau) - m^2$, or $R(\tau) = b(\tau)/b(0)$), which, along with the value of m, are always assumed known in the correlation theory. (Recall that earlier in this section the function $b(\tau)$ was also assumed known). To be able to determine the value of the function $B(\tau) = \langle X(t + \tau)X(t)\rangle$, where τ is a fixed nonnegative number, with any desirable degree of accuracy by time averaging the data relating to a single realization $x(t)$, the following relation must hold:

$$(3.25) \qquad \lim_{T \to \infty} \frac{1}{T - \tau} \sum_{t=1}^{T-\tau} X(t + \tau)X(t) = B(\tau)$$

or, accordingly,

$$(3.25a) \qquad \lim_{T \to \infty} \frac{1}{T - \tau} \int_0^{T-\tau} X(t + \tau)X(t)dt = B(\tau).$$

If the time t is discrete and (3.25) holds, then

$$(3.26) \qquad B_T^*(\tau) = \frac{1}{T - \tau} \sum_{t=1}^{T-\tau} x(t + \tau)x(t)$$

is, for $\tau \geqslant 0$, an unbiased and consistent estimate of $B(\tau)$. Similarly, if t is continuous and (3.25a) holds, then both the estimates

$$(3.26a) \qquad B_T^*(\tau) = \frac{1}{T - \tau} \int_0^{T-\tau} x(t + \tau)x(t)dt$$

and

$$(3.26b) \qquad B_{TN}^*(\tau) = \frac{1}{N + 1} \sum_{k=0}^{N} x\left[\frac{k(T - \tau)}{N} + \tau\right] x\left[\frac{k(T - \tau)}{N}\right]$$

are unbiased and consistent for any $\tau \geqslant 0$. (The consistency means here, as usual, that the accuracy of the estimate can be made arbitrarily high, provided T is chosen, or T and N are chosen, sufficiently large.) If $\tau < 0$, we need only replace τ by $|\tau|$ in the right-hand sides of (3.26) − (3.26b).

The necessary and sufficient condition for the validity of (3.25) and (3.25a) can easily be obtained from Slutsky's theorem if we apply it to the τ-dependent random function (of argument t) $Y_\tau(t) = X(t + \tau)X(t)$. (Note that if the realization of $X(t)$ is known for $t = 1, 2, ..., T$ or $0 \leqslant t \leqslant T$, then the realization of $Y_\tau(t)$ will be known only for $t = 1, 2, ..., T - \tau$ or, accordingly, $0 \leqslant t \leqslant T - \tau$.) The function $Y_\tau(t)$

clearly is wide-sense stationary if and only if the function $\langle X(t + \tau)X(t)\rangle$ is independent of t and

$$(3.27) \quad \langle X(t + s + \tau)X(t + s)X(t + \tau)X(t)\rangle = B^{(4)}(\tau,s)$$

is also independent of t (but can depend on τ and s). Thus, we must now require that the original function $X(t)$ be not only wide-sense stationary, but also stationary up to at least the fourth order (i.e. that its fourth moments remain unaltered on any simultaneous time shift of all the four respective time points). Let us assume that the time t is continuous; the situation with random functions $X(t)$ of the discrete argument t is quite similar and we need not dwell on it. By virtue of Slutsky's theorem *the estimate* (3.26a) *of the correlation function* B(τ) *is consistent* (i.e. the relation (3.25a) holds) *if and only if*

$$(3.28) \quad \lim_{T \to \infty} \frac{1}{T} \int_0^T \{B^{(4)}(\tau,s) - [B(\tau)]^2\}ds = 0.$$

Based on the estimate $B_T^*(\tau)$ of $B(\tau)$ we can form related estimates $b_T^*(\tau) = B_T^*(\tau) - m_T^{*2}$ (or $\breve{b}_T^*(\tau) = (T - \tau)^{-1}\int_0^{T-\tau}(x(t + \tau) - m_T^*)(x(t) - m_T^*)dt$)* of the function $b(\tau) = B(\tau) - m^2$ and $R_T^*(\tau) = b_T^*(\tau)/b_T^*(0)$ (or $\breve{R}_T^*(\tau) = \breve{b}_T^*(\tau)/\breve{b}_T^*(0)$) of the function $R(\tau) = b(\tau)/b(0)$. The consistency of the estimate $B_T^*(\tau)$ implies that all four estimates $b_T^*(\tau)$, $\breve{b}_T^*(\tau)$, $R_T^*(\tau)$, and $\breve{R}_T^*(\tau)$ are also consistent. However, the estimates $b_T^*(\tau)$, $\breve{b}_T^*(\tau)$, $R_T^*(\tau)$, and $\breve{R}_T^*(\tau)$, unlike the estimate $B_T^*(\tau)$, are not unbiased but only asymptotically unbiased (see the above discussion of this topic for the case where $\tau = 0$).

In most applications one can, proceeding from physical

---

*The estimates $b_T^*(T)$ and $\breve{b}_T^*(T)$ coincide for $T = 0$, but for $T \neq 0$ they are not identical. However, their difference is slight for large values of T, and they are equivalent asymptotically.

intuition, be confident that the correlation coefficient between the two random variables

$$Y_\tau(t + s) = X(t + s + \tau)X(t + s) \text{ and } Y_\tau(t) = X(t + \tau)X(t)$$

tends to zero, as $s \to \infty$. Then condition (3.28) is evidently valid. The general formulae for the mean square errors (i.e. the variances) of the estimates $B_\tau^*(\tau)$ of the correlation function $B(\tau)$ both for discrete and for continuous time, can be easily deduced from (3.11) and (3.11a) by replacing the function $b(\tau)$ of τ with the function $B^{(4)}(\tau,s) - [B(\tau)]^2$ of s. The formula for the mean square error of the estimate $B_{TN}^*(\tau)$ can be deduced quite similarly from the expression for the mean square error of the estimate (3.13b) of $m = \langle X(t) \rangle$. However, the utilization of these formulae for the mean square errors of $B_\tau^*(\tau)$ and $B_{TN}^*(\tau)$ is considerably impeded by the fact that they include the fourth moment $B^{(4)}(\tau,s)$ which is usually unknown and cannot be easily estimated (see Note 19).

The estimates $B_\tau^*(\tau)$ and $B_{TN}^*(\tau)$ are, of course, realizations of random variables which are similar to the variables M_τ^* and M_{TN}^* discussed in the previous section, but which correspond to the random function $Y_\tau(t)$, and not to $X(t)$. Therefore, it is clear that under some, rather general, conditions the probability distributions of the estimators $B_\tau^*(\tau)$ and $B_{TN}^*(\tau)$ for sufficiently large T must be close to the normal probability distributions. For the same reasons it can also be asserted that, as a rule, the variable $B_\tau^*(\tau)$ tends to the true value of $B(\tau)$, as $T \to \infty$, not only in the mean square, but also with probability one (i.e. $|B_\tau^*(\tau) - B(\tau)|$ is small for all sufficiently large values of T simultaneously; cf. above p. 224). We shall not discuss here, however, the specific conditions ensuring the asymptotic (for $T \to \infty$) normality of sampling distributions of the estimators $B_\tau^*(\tau)$ and $B_{TN}^*(\tau)$ and their almost sure convergence to $B(\tau)$, as $T \to \infty$ (see, however, Note 18).

Let us now discuss again in some more detail the important particular case of Gaussian stationary functions $X(t)$, which are often encountered in various applications. In this particular case the fourth-order moments $B^4(\tau,s)$ (as well as all the other statistical characteristics of the function $X(t)$) are determined uniquely by the constant m and the function $B(\tau)$ (or $b(\tau)$). By using the general formula for the moments

of multidimensional normal distributions (see Note 11), it is not hard to show that *if* X(t) *is Gaussian, then*

$$(3.29)\quad B^4(\tau, s) = B^2(s) + B^2(\tau) + B(s + \tau)B(s - \tau) - 2m^4$$

(in the case where $s = 0$ and $m = 0$ this formula coincides with (3.19)). By inserting (3.29) into (3.28) (or into a similar formula referring to the case of discrete t) it can further be proved that for Gaussian stationary random functions the estimates (3.26) − (3.26b) are consistent for any value of τ if and only if the condition (3.20) holds (i.e., if the corresponding estimate is consistent for $\tau = 0$). The mean square error of the estimate $B_T^*(\tau)$ (or $B_{TN}^*(\tau)$) is expressed, in the Gaussian case, via the constant m and the function $B(\tau)$; the values of m and $B(\tau)$ also determine, in the case at hand, the errors of the estimates $b_T^*(\tau)$, and $\hat{b}_T^*(\tau)$ of the function

$b(\tau)$ and $R_T^*(\tau)$ and $\widehat{R}_T^*(\tau)$ of the function $R(\tau)$.[19] The corresponding formulae for the mean square errors can be used in practice by replacing in them the unknown values of the correlation function by the same estimates whose accuracy we seek to evaluate. However, it is clear that the accuracy of the results thus obtained will not be very high.

It should also be noted that the general formula for moments of the multidimensional Gaussian distribution can be used to prove the following statement: *if* X(t) *is a Gaussian stationary function, then the validity of condition* (3.20) *implies that the time average over a time interval of length* T *of a product* $x(t)x(t + \tau_1) \dots x(t + \tau_{n-1})$ *tends, as* $T \to \infty$, *to the corresponding* n-*th order moment* $\langle X(t)X(t + \tau_1) \dots X(t + \tau_{n-1})\rangle$ *for any n and* $\tau_1, \dots, \tau_{n-1}$. Moreover, it is also possible to prove that if $X(t)$ is a Gaussian stationary function and $\Phi(x_0, x_1, \dots, x_{n-1})$ is an arbitrary function of n variables satisfying the condition that $\langle\Phi(X(t), X(t + \tau), \dots, X(t + \tau_{n-1}))\rangle = m_\Phi(\tau_1, \dots, \tau_{n-1})$ is finite, then *the validity of condition* (3.20) *also guarantees that the time average of* $\Phi(x(t), x(t + \tau_1), \dots, x(t + \tau_{n-1}))$ *over a time interval of length* T *tends, as* $T \to \infty$, *to the corresponding probabilistic mean value* $m_\Phi(\tau_1, \dots, \tau_{n-1})$. This last fact constitutes the content of the *general ergodic theorem for Gaussian stationary random functions*.[20]

*Since all the probability distributions of a Gaussian stationary random function $X(t)$ with mean value zero

are determined uniquely by the values of the correlation function $B(\tau) = b(\tau)$, it is clear that all the statistical characteristics of $X(t)$ can be expressed in terms of $b(\tau)$. This circumstance can be used as a basis for a number of specific methods of approximate determination of the functions $b(\tau)$ and $R(\tau) = b(\tau)/b(0)$ from a single realization $x(t)$ of the random function $X(t)$. Thus, it is easy to verify that for a Gaussian stationary function $X(t)$ having mean value zero and an arbitrary function $g(x)$, the following relation holds:

$$(3.30) \qquad \langle g(X(t + \tau)X(t) \rangle = cR(\tau), \; c = \langle g(X(t))X(t) \rangle,$$

if $\langle g(X(t))X(t) \rangle < \infty$.[21] Therefore the estimation of the value of $\langle g(X(t + \tau))X(t) \rangle$ by time averaging the function $g(x(t+\tau))x(t)$ of t can be used for approximate determination of the values of $R(\tau)$. In the particular case where $g(x) = \mathrm{sgn}\,x$ is a sign function (i.e. $g(x) = 1$ for $x > 0$, $g(x) = 0$ for $x = 0$, and $g(x) = -1$ for $x < 0$), the time averaging of $g(x(t + \tau))x(t)$ can be computed very easily since here $\sum_{k} \mathrm{sgn}(x(t_k + \tau))x(t_k)$ involves only

additions but no multiplications. Moreover, in this case $c = \langle |X(t)| \rangle = (2/\pi)^{1/2}\sigma_X = (2b(0)/\pi)^{1/2}$ and hence

$$(3.31) \qquad \langle \mathrm{sgn}(X(t + \tau))X(t) \rangle = (2b(0)/\pi)^{1/2}R(\tau).$$

According to (3.31) the approximate determination of $b^{(r)}(\tau) = \langle \mathrm{sgn}(X(t + \tau))X(t) \rangle$ and $b^{(r)}(0) = \langle \mathrm{sgn}(X(t))X(t) \rangle$ by time averaging of $\mathrm{sgn}(x(t + \tau))x(t)$ and $\mathrm{sgn}(x(t))x(t)$ allows one, in the case of a Gaussian function $X(t)$ with $\langle X(t) \rangle = 0$, to determine approximately also the value of $R(\tau) = b^{(r)}(\tau)/b^{(r)}(0)$ and $b(\tau) = \pi b^{(r)}(0)b^{(r)}(\tau)/2$. In the engineering literature the function $b^{(r)}(\tau)$ is sometimes called the *relay correlation function* (or the *hybrid-sign correlation function*) of the random function $X(t)$, while the methods for estimating the correlation functions $R(\tau)$ and $b(\tau)$, which use the function $b^{(r)}(\tau)$, are accordingly called *relay methods* (or *hybrid-sign methods*) of correlation analysis.[22]

It is also clear that if $X(t)$ is a Gaussian random

function and $\langle X(t) \rangle = 0$, then the mean value $\langle g_1(X(t + \tau))g_2(X(t)) \rangle = b_{g_1 g_2}(\tau)$ is determined uniquely by the

values of $b(\tau)$ and $b(0)$ no matter what the functions $g_1(x)$ and $g_2(x)$ may be.   (This follows from the fact that the two-dimensional probability distribution of the vector $(X(t+\tau), X(t))$ depends only on $b(0)$ and $b(\tau)$.) Moreover, in many cases the value of $b(\tau)$ (or, at least, of $R(\tau) = b(\tau)/b(0)$) can also be expressed in terms of $b_{g_1 g_2}(\tau)$.[23] Let us consider, for example, a simple case

where $g_1(x) = g_2(x) = \text{sgn } x$.   It is easy to show that if $X(t)$ is Gaussian and $\langle X(t) \rangle = 0$, then

$$(3.32) \qquad \langle \text{sgn}(X(t + \tau))\text{sgn}(X(t)) \rangle = \frac{2}{\pi} \sin^{-1}R(\tau). \quad [24]$$

Consequently, the estimate of the correlation function $b^{(p)}(\tau)$ of the auxiliary random function $Y(t) = \text{sgn}(X(t))$ can also be used for approximate determination of the values of the normalized correlation function $R(\tau) = \sin[\pi b^{(p)}(\tau)/2]$ of the original function $X(t)$.

The function $\langle \text{sgn}(X(t+\tau))\text{sgn}(X(t)) \rangle = b^{(p)}(\tau)$ is sometimes called a *polarity correlation function* (or a *sign correlation function*) of the random functions $X(t)$, and the methods for estimating the normalized correlation function $R(\tau)$ of the Gaussian random function $X(t)$, which are based on the relation (3.32) or on some equivalent relation, are called *polarity coincidence methods* (or *sign methods*).[25]

In the case of a general Gaussian stationary function $X(t)$ with an unknown mean value $\langle X(t) \rangle = m$, the relations (3.30), (3.31), and (3.32) can be applied to the function $X^{(1)}(t) = X(t) - M_T^*$, where M_T^* is the simplest estimator (3.6) of the mean value m. However, all these relations are no longer exact in this case, and they hold only approximately at large values of T. As regards the non-Gaussian random functions $X(t)$, the possibility of applying to them similar relations or some of their modifications naturally requires special investigation; see Note 26 where some information on this topic can be found.*

Let us now revert again to general random functions $X(t)$ and to the simplest estimates (3.26), (3.26a), and (3.26b) of the

corresponding correlation functions $B(\tau)$.   According to the foregoing, these estimates are consistent under some quite general conditions which can almost always be considered fulfilled in practice.   Hence, by using such an estimate, one can determine the values of $B(\tau)$ from a single realization $x(t)$ to any degree of accuracy, provided that the observation interval T is long enough.   Calculations by formulae (3.26) or (3.26b), which formerly seemed very tedious, can now be performed quite rapidly with the aid of digital computers. Moreover,   a   great   number   of   quite   different   analog computing   machines   — correlators —   have   been   specially designed   to   perform   calculations   of   integrals   of   the   form (3.26a).[27] As an example, Fig. 37 shows the estimates of the normalized correlation function $R(\tau) = B(\tau)/B(0)$ for fading of radio signals, calculated from the curve given in Fig. 5(a), by using formula (3.26b) with $T = 20$ sec.   and $T/N = 0.016$ sec., and for the Beveridge wheat price series of Fig.   6(e), calculated with the aid of formula (3.26). We see that the empirical normalized correlation function $R_T^*(\tau) = B_T^*(\tau)/B_T^*(0)$ behaves quite similarly in both cases: near the point $\tau = 0$ it falls off rapidly with the increase of τ, then passes zero and begins to oscillate smoothly, with a small amplitude, about zero.   But are these oscillations real?   This is an important question, which requires special investigation.

In the foregoing we only showed that for any fixed value of τ all three estimates (3.26), (3.26a), and (3.26b) converge to the true value of $B(\tau)$, as $T \to \infty$ (and at a fixed or a decreasing T/N), under very general conditions practically always valid in application. But in application it is usually the entire function $B(\tau)$, rather than a single value of $B(\tau)$, that is of interest. From the convergence of the estimate $B_T^*(\tau)$ or $B_{TN}^*(\tau)$ to $B(\tau)$ for each fixed τ, it does not yet follow that the convergence will take place simultaneously for all τ, i.e. that the function $B_T^*(\tau)$ (or $B_{TN}^*(\tau)$) will tend to the function $B(\tau)$. Besides, it is clear that all the indicated estimates become extremely unreliable when τ approaches T, because here the real averaging interval $T - \tau$ is relatively small. Therefore, in practice one usually confines oneself to the values of τ constituting only a small fraction of T (no more than 10-20%, as a rule). However, even in this case, estimates at comparatively large values of τ (at which the true value of $B(\tau)$ is very small) are usually highly unreliable, since their root-mean-square error appreciably
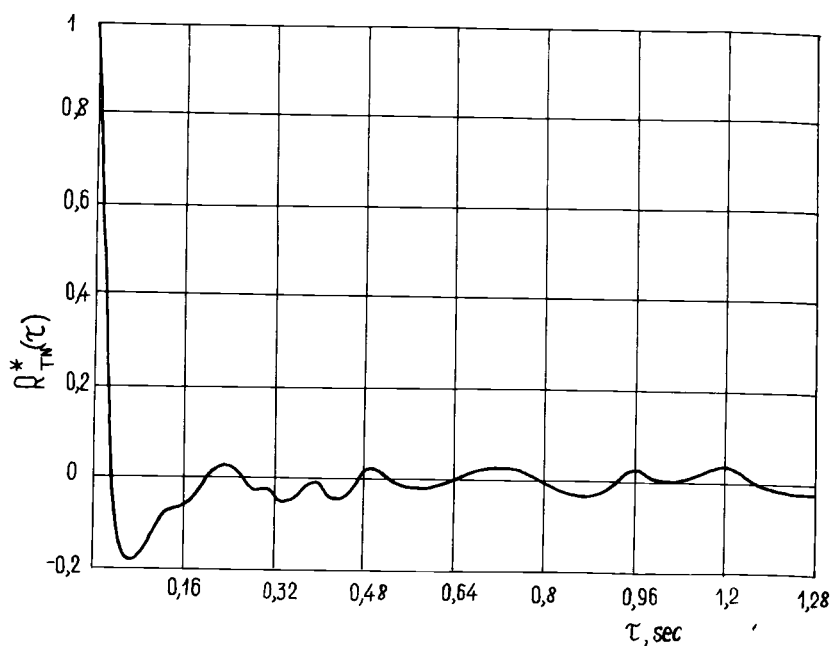
(a)
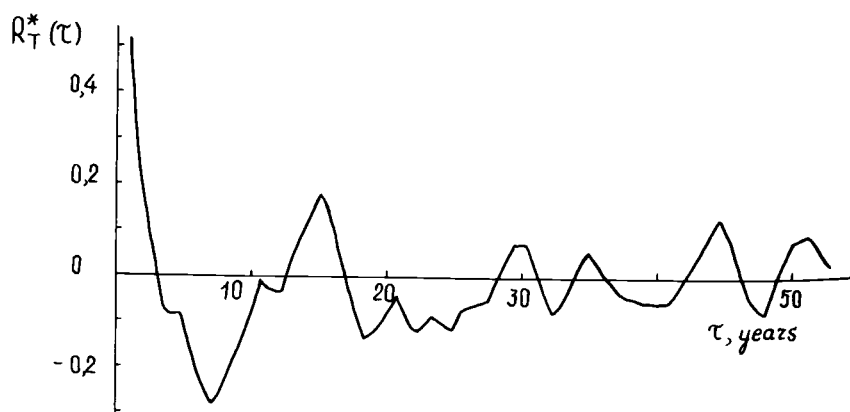
(b)

Fig. 37. The calculated normalized correlation function
(a) for fading of radio signals (James et al., 1947);
(b) for Beveridge wheat price series (Kendall, 1946).

exceeds the value of $B(\tau)$ being estimated.   Moreover, the values of $B_T^*(\tau)$ (or $B_{TN}^*(\tau)$) at close values of τ are usually rather strongly correlated (see Note 19).   This makes the functions $B_T^*(\tau)$ and $B_{TN}^*(\tau)$ vary smoothly with τ, often creating the false impression that the variation of the function $B(\tau)$ terminates in a number of regular "waves".

As an illustration to the above remarks, Fig. 38(a) shows the function $R_T^*(\tau) = B_T^*/B_T^*(0)$, where $T = 480$, calculated for the artificial series $x(t)$, $t = 1, 2, ..., 480$, simulated with the aid of "random numbers" in such a way that the correlation function $R(\tau) = B(\tau)/B(0)$ of the relevant stationary sequence $X(t)$ could be determined theoretically.   (For the description of a method of such simulation see Sec. 6, Example 6; in fact, $x(t)$ used in Fig. 38(a) is a realization of the second-order autoregressive sequence corresponding to $a_1 = -1.1$, $a_2 = 0.5$.) Another similar example is given in Fig. 38(b), where two empirical correlation functions $R_T^*(\tau)$, $T = 200$, are presented, calculated from two independent realizations $x^{(1)}(t)$ and $x^{(2)}(t)$, $t = 1, 2, ..., 200$, together with the corresponding "theoretical" function $R(\tau)$. (The realizations $x^{(1)}(t)$ and $x^{(2)}(t)$ both correspond here to the second-order autoregressive sequence, where $a_1 = -0.7$, $a_2 = 0.49$.) We can see that in both Figs. 38(a) and 38(b) the theoretical correlation functions $R(\tau)$ decay to zero rapidly, while the empirical correlation functions $R_T^*(\tau)$ do not decay but oscillate smoothly about zero.   (Moreover, in Fig. 38(b) the oscillations of the two functions $R_T^*(\tau)$, calculated from two realizations, clearly diverge.) It would be natural to believe that the disordered oscillations of the functions $R_{TN}^*(\tau)$ and $R_T^*(\tau)$ about zero in Figs. 37(a) and (b) have the same origin as the oscillations of the empirical correlation functions about zero on Figs. 38(a) and (b), i.e. that they are also unrelated in any way with the true shape of the function $R(\tau)$ (which is unknown to us in the cases shown in Figs. 37(a) and (b).

Of course, the amplitude of the spurious "waves" on the graph of the empirical correlation function depends on the sample size (i.e. on the "averaging interval" T and, when (3.26b) is applied, on the "discretization period" T/N), and this amplitude can be appreciably decreased by increasing the number of the values $x(t)$ used in computation. This is seen, in particular, from Fig. 38(c), where another example of the empirical normalized correlation function $R_T^*(\tau)$ is depicted, which refers to the same stationary sequence $X(t)$ as in the

case of Fig. 38(a), but which is based on a much greater sample size, namely $T = 15{,}000$.   We see that the values of $R^*_{15,000}(\tau)$ are much closer to the values of the theoretical correlation function $R(\tau)$ than the values of $R^*_{480}(\tau)$. Nevertheless, the graph of $R^*_{15,000}(\tau)$ also terminates in "waves" (this time, of a small amplitude) which are absent in the graph of $R(\tau)$.
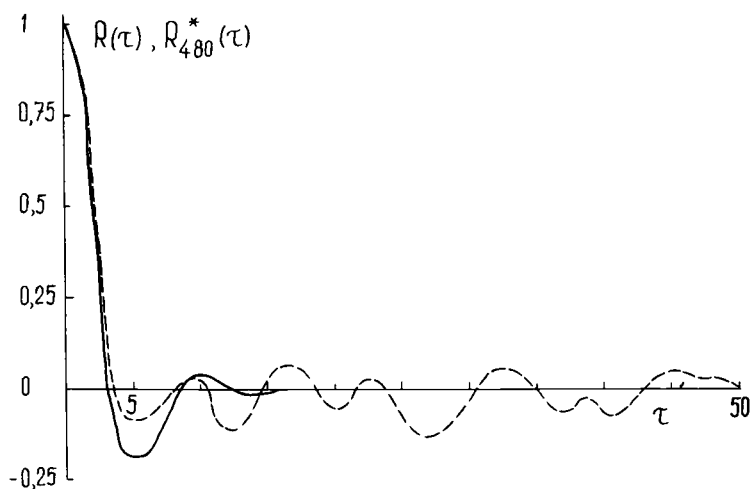
(a)
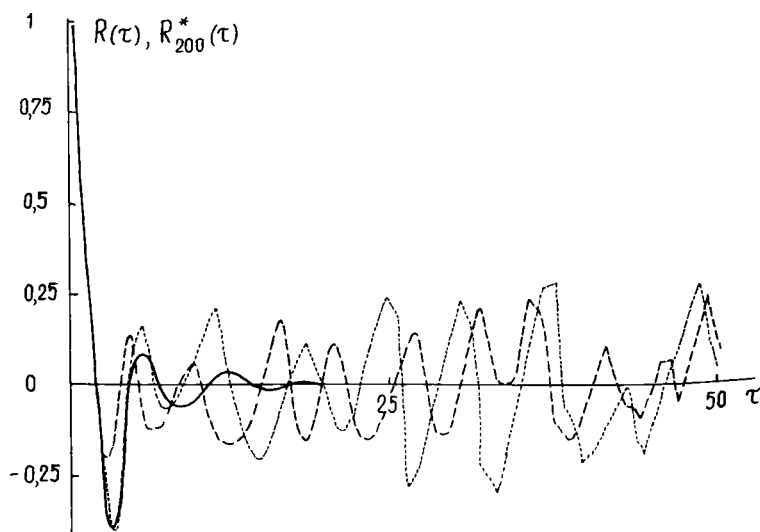
(b)
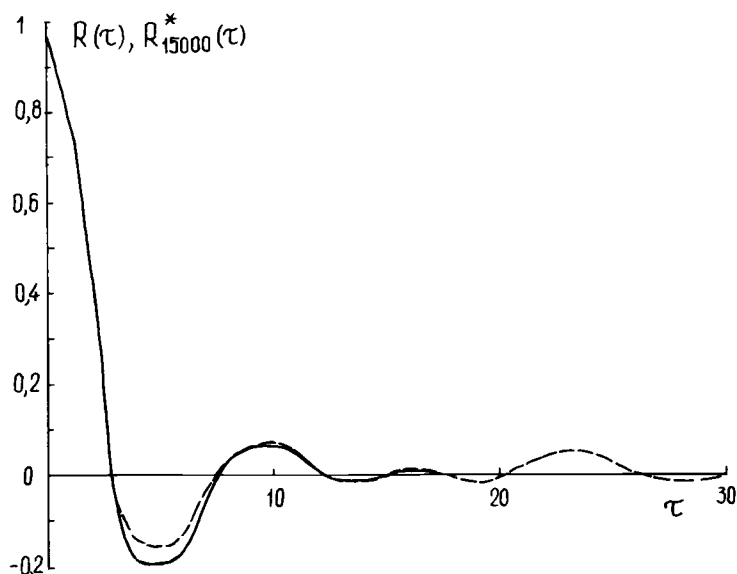
(c)

Fig. 38. (a) Theoretical ("true") normalized correlation function $R(\tau)$ of artificial stationary sequence (the solid line) compared to the estimated correlation function calculated from 480 observations (the dashed line) (Hannan, 1960). (b) Theoretical correlation function $R(\tau)$ of another stationary sequence (the solid line) compared with two empirical correlation functions calculated from two independent realizations of length $T = 200$ (Wold, 1965). (c) The same theoretical correlation function as in (a) (the solid line) compared with the empirical correlation function calculated from 15,000 observations (Brown and Hardin, 1973).

Since all the estimates of the correlation function are unreliable at large values of τ, the question also arises whether it is possible to improve, e.g., the estimate $B_T^*(\tau)$ by artificially reducing its values for all not-too-small τ. The simplest method for such reduction is to include into the right-hand side of (3.26) and (3.26a) an additional factor $(T - \tau)/T = 1 - \tau/T$, which equals unity at $\tau = 0$ and falls off smoothly to zero at the point $\tau = T$, where the estimates (3.26)

and (3.26a) are completely unreliable. Then we arrive at the estimates

$$(3.33) \qquad B_T^{**}(\tau) = \frac{1}{T} \sum_{t=1}^{T-\tau} x(t + \tau)x(t)$$

or

$$(3.33a) \qquad B_T^{**}(\tau) = \frac{1}{T} \int_0^{T-\tau} x(t + \tau)x(t)dt,$$

which are, unlike the estimates $B_T^*(\tau)$, biased, since $\langle B_T^{**}(\tau) \rangle = (1 - \tau/T)B(\tau)$. (The same symbol $B_T^{**}(\tau)$ in the last formula denotes the estimator obtained when the nonrandom realization $x(t)$ is replaced by the random function $X(t)$ in the right-hand sides of (3.33) and (3.33a).) When applied to the estimate (3.26b), the role of $B_T^{**}(\tau)$ is played by the biased estimate

$$(3.33b) \qquad B_{TN}^{**}(\tau) = \frac{1}{N + 1} \sum_{k=0}^{[(T-\tau)N/T]} x\left[\frac{kT}{N} + \tau\right] x\left(\frac{kT}{N}\right),$$

where $[(T - \tau)N/T]$ is the largest integer not exceeding the number $(T - \tau)N/T = (1 - \tau/T)N$ (i.e. the largest integer k satisfying the inequality $\tau + kT/N \leqslant T$). However, all the properties of the estimate $B_{TN}^{**}(\tau)$ are quite similar to those of $B_T^{**}(\tau)$, and therefore later we shall speak, for simplicity, only of the estimates $B_T^{**}(\tau)$.

It is important that estimates $B_T^{**}(\tau)$ are always asymptotically unbiased, i.e. their bias $\delta_T(\tau) = \langle B_T^{**}(\tau) \rangle - B(\tau)$ tends to zero, as $T \to \infty$. It is still more important that estimates $B_T^{**}(\tau)$ are practically always consistent (because at $\tau \ll T$ they hardly differ from $B_T^*(\tau)$).[28] From now on we shall characterize the accuracy of estimates $B_T^{**}(\tau)$ and $B_T^*(\tau)$ by their mean square errors $\Delta_{1T}^2(\tau) = \langle [B_T^{**}(\tau) - B(\tau)]^2 \rangle$ and $\Delta_T^2(\tau) = \langle [B_T^*(\tau) - B(\tau)]^2 \rangle$. Then, apparently, the inequality $\Delta_{1T}^2(\tau) \leqslant \Delta_T^2(\tau)$ holds for all τ and T (i.e. the estimate $B_T^{**}(\tau)$ is in all cases not less accurate than $B_T^*(\tau)$) at least for all the Gaussian stationary functions $X(t)$ with $\langle X(t) \rangle = 0$ for which both $\Delta_{1T}^2(\tau)$ and $\Delta_T^2(\tau)$ can be expressed in terms of $B(\tau) = b(\tau)$. True, this statement has not been proved rigorously so far, but it is confirmed by all the specific examples for which the values $\Delta_{1T}^2(\tau)$ and $\Delta_T^2(\tau)$ were calculated.[29] The last circumstance alone makes estimates $B_T^{**}(\tau)$ superior to estimates $B_T^*(\tau)$. However, apart from this the estimate $B_T^{**}(\tau)$ has an additional important advantage over

$B_T^*(\tau)$. Indeed, it can easily be shown that the function $B_T^{**}(\tau)$ (extended to negative and larger-than-T values of τ with the aid of natural relations $B_T^{**}(-\tau) = B_T^{**}(\tau)$ and $B_T^{**}(\tau) = 0$ for $|\tau| > T$) is always positive definite, i.e. it can be a correlation function of some random stationary function $X(t)$, while the function $B_T^*(\tau)$ does not possess this property.[30]

The indicated advantages of the estimate $B_T^{**}(\tau)$ over $B_T^*(\tau)$ forced most of the serious investigators to prefer the former of those two estimates, and nowadays precisely the estimate $B_T^{**}(\tau)$ is used almost universally in practical situations. Nevertheless, this estimate also has certain shortcomings. The positive definiteness of the function $B_T^{**}(\tau)$ does not imply, of course, that this function is close at all τ to the correlation function of precisely that random function $X(t)$ whose realization is the function $x(t)$ appearing in (3.33) or (3.33a). Note in this connection that the function $B_T^{**}(\tau)$ is positive definite only if formula (3.33) or (3.33a) is used at all $|\tau| < T$, whereas it was noted above that it is practically expedient to consider the estimate of $B(\tau)$ only at values of τ that do not exceed some fixed small fraction of T. [31] It is also clear that the spurious "waves" which are observed on the graphs of the functions $R_T^*(\tau)$ in Fig. 38 will also be present on the graphs of the functions $B_T^{**}(\tau)$ and $R_T^{**}(\tau) = B_T^{**}(\tau)/B_T^{**}(0)$ (although, generally speaking, these waves will be somewhat smaller and will decay with increasing τ). Therefore, it is often reasonable to use, instead of $B_T^{**}(\tau)$, a still more reduced estimate of the form

$$(3.34) \quad B_T^{(a)}(\tau) = a_T(\tau)B_T^{**}(\tau),$$

where $a_T(0) = 1$, but $a_T(\tau)$ falls off with increasing $|\tau|$. It is reasonable to choose the function $a_T(\tau)$ in such a way that $a_T(\tau) = 0$ for $|\tau| \geqslant k_T$, where k_T is such a number that $B_T^{**}(\tau)$ only oscillates disorderly about zero at $\tau > k_T$ and k_T/T is small enough (say, of the order of 0.1). One can, e.g., assume that $a_T(\tau) = a(\tau/k_T)$, where $a(0) = 1$ and $a(x) = 0$ at $x > 1$. If we put $a(x) = 1$ for $|x| \leqslant 1$ (i.e., $a_T(\tau) = 1$ for $|\tau| < k_T$) and $a(x) = 0$ for $|x| > 1$ (i.e. $a_T(\tau) = 0$ for $|\tau| > k_T$), then formula (3.33) or (3.33a) will be used only for $\tau \leqslant k_T$, while all the values of $B_T^{**}(\tau)$ for $\tau > k_T$ will be fully neglected (i.e. replaced by zero). However, this choice of $a(\tau)$ (shown in Fig. 39(a)) is defective in the following respect: the corresponding function $B_T^{(a)}(\tau)$ not only lacks positive definiteness but is even discontinuous. If, however, the function

$a_T(\tau)$ is chosen to be continuous and positive definite, then the product $a_T(\tau)B_T^{**}(\tau)$ will also be a continuous positive definite function by virtue of the statement on p. 60. For instance, it can be assumed that $a_T(\tau) = a(\tau/k_T)$, where $a(x) = 1 - |x|$ for $0 \leqslant |x| \leqslant 1$ and $a(x) = 0$ for $|x| > 1$ (see Fig. 39(b)).

The function $a_T(\tau)$ appearing in the right-hand side of (3.34) is usually called the *lag window*; if $a_T(\tau) = a(\tau/k_T)$, then $a(x)$ is called the *lag window generator*. It is often expedient to include a reasonably chosen lag window into the formula for estimating $B(\tau)$. In particular, the above-mentioned "triangular" lag window $a(\tau/k_T)$, where $a(x) = \max(1 - |x|, 0)$, is called the *Bartlett window*, and the corresponding estimate $B_T^{(a)}(\tau)$, the *Bartlett estimate*. Many other forms of lag

(a)

(b)

(c)
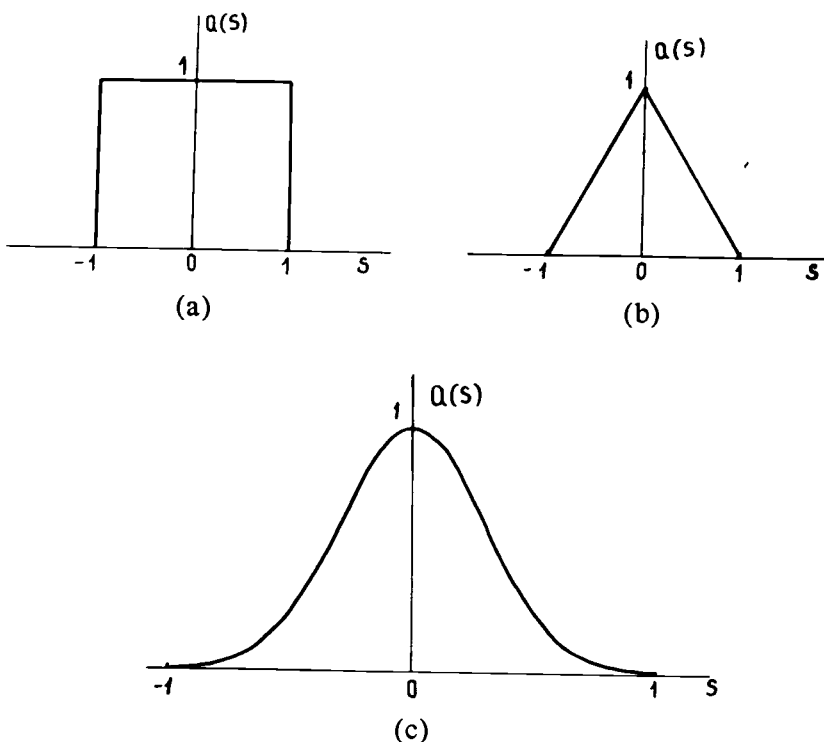
Fig. 39. Three examples of lag window generators $a(s)$.
(a)   The rectangular window generator; (b) the Bartlett window generator; (c) the Parzen window generator.

windows are also known, e.g., the *Parzen window* shown in Fig. 39(c), which falls off much less rapidly in the vicinity of zero.[32] Estimates of the form (3.34) will be encountered in the next two sections, where further examples of lag windows will be given and quite a different motivation of the usefulness of such estimates will be presented. It will be clear from the contents of Secs. 18 and 19 that at present it is most convenient in many respects to use the lag windows $a_T(\tau)$ that do not vanish identically at $|\tau| \geqslant k_T$, but only fall off rapidly enough with $|\tau|$. Such behavior of $a_T(\tau)$ is, of course, sufficient to eliminate the spurious waves in the graphs of the correlation functions, which are clearly seen in Fig. 38.

18. Statistical Spectral Analysis. Determination of the Spectral Density Function

In the two previous sections of this chapter we considered the methods for approximate determination of the mean value m and the correlation function $B(\tau)$ of a stationary random function $X(t)$ from its single realization $x(t)$ observed over a finite time interval. However, it is not the mean value and the correlation function but the spectral density function $f(\omega)$ and the spectral distribution function $F\omega)$ that play the main role in most of the practical applications of the theory of stationary random functions. (The practical importance of the spectral properties of stationary random functions is clearly demonstrated by the appearance of quite a number of books devoted to various special applications of spectral methods.[33]) It is obvious that the spectral theory of stationary functions can be useful only when we know the spectral distribution function $F(\omega)$ or the spectral density function $f(\omega)$ of the random function of interest $X(t)$. Since the values of the functions $f(\omega)$ and $F(\omega)$ are unknown in almost all the practical problems, these functions must be estimated somehow from observational data prior to any application of spectral theory. Note also that the shape of a curve representing the spectral density function $f(\omega)$ varies much more than the shape of a curve representing $F(\omega)$. Hence, wherever the spectral density $f(\omega)$ exists, it describes the spectral properties of $X(t)$ much more clearly than does its indefinite integral − the monotone nondecreasing function

$F(\omega)$. Therefore the methods for spectral density estimation are especially important. (All these methods taken together are often called *statistical spectral analysis.**) A few remarks concerning the experimental determination of the spectral density have been made earlier (see pp. 157 − 158 and Figs. 23 and 33(b)), but the importance of this problem requires its more detailed consideration. Such a consideration will be the main subject of this and the following sections of the book.

The literature on statistical spectral analysis (i.e. on methods of spectral density estimation) is enormous − it numbers several hundreds of special monographs, chapters in books of more general contents and in expository articles, and several thousands of research papers, reports, and purely applied works.[34] Of course, only a small portion of the material pertaining to this topic will be covered in the present book. Nevertheless we shall try to outline all the general principles of statistical spectral analysis and to discuss the main practical methods for spectral density estimation. However, most attention will be given to one very important numerical method of spectral analysis; therefore, we shall revert to the entirely different methods mentioned in Secs. 10 and 11 only in Sec. 19.

We shall always assume in this and the following sections (except in the final part of Sec. 19) that $\langle X(t) \rangle = 0$ and that the correlation function $\langle X(t + \tau)X(t) \rangle = B(\tau) = b(\tau)$ decreases so rapidly, as $|\tau| \to \infty$, that there exists a spectral density $f(\omega)$ which is a continuous function of the frequency ω. Moreover, we shall also assume, as in Secs. 16 and 17, that only one realization $x(t)$, where $t = 1, 2, ..., T$ or $0 \leqslant t \leqslant T$, is observed and can be used for the estimation of $f(\omega)$.**

The problem of estimating the spectral density $f(\omega)$ is clearly related to the one of estimating the correlation function $B(\tau)$. In particular, it is easy to see that a consistent

---

*Similarly the whole complex of statistical techniques used to determine <u>any</u> statistical characteristics of a stationary time function from observational data is known as <u>time series analysis</u>.

**Of course, if several realizations are known, we can estimate the spectral density from each of them and then improve the results by calculating the average of all the estimates obtained (cf. the first footnote on p. 215).

estimate $f_T^*(\omega)$ of $f(\omega)$ (i.e. such an estimate that the corresponding estimator tends to the true value of $f(\omega)$ as $T \to \infty$) can exist only under the "condition of second-order ergodicity", which ensure, the existence of a consistent estimate of the correlation function $B(\tau)$. Therefore, it will always be assumed below that the stationary function $X(t)$ under consideration is second-order ergodic.

In the preceding section we considered several reasonable estimates of the function $B(\tau)$, but most attention was given to the estimate $B_T^*(\tau)$ defined by equations (3.26) $-$ (3.26a) on p. 231 and to its modification $B_T^{**}(\tau)$ defined on p. 242.* We recall that the estimate $B_T^*(\tau)$, as distinct from $B_T^{**}(\tau)$, is strictly unbiased. However, the estimate $B_T^{**}(\tau)$ is usually more accurate (i.e. its mean square error is smaller) and, besides, it is a positive definite function of τ, whereas $B_T^*(\tau)$ does not possess this property. The last circumstance is particularly important when the estimation of the spectral density $f(\omega)$ is the main task: it implies that in using the Fourier transform of $B_T^{**}(\tau)$ as the estimate of $f(\omega)$ we can be confident, at least, that whatever ω be, we shall not obtain a meaningless negative value. Therefore it seems expedient to attempt to estimate $f(\omega)$ by replacing the values of the unknown function $B(\tau)$ by its estimate $B_T^{**}(\tau)$ in equations (2.84) and (2.213), which express $f(\omega)$ in terms of $B(\tau)$.

Since $B_T^{**}(\tau) = 0$ for $|\tau| \geqslant T$, the corresponding estimate of $f(\omega)$ can be written in the form

$$(3.35) \qquad i_T(\omega) = \frac{1}{2\pi} \int_{-T}^{T} e^{-i\omega\tau} B_T^{**}(\tau) d\tau$$

or

$$(3.35a) \qquad i_T(\omega) = \frac{1}{2\pi} \sum_{\tau=-T+1}^{T-1} e^{-i\omega\tau} B_T^{**}(\tau).$$

With the aid of simple transformations it is easy to prove[35] that (3.35) and (3.35a) can also be rewritten as

---

*The estimates $B_{TN}^*(T)$ and $B_{TN}^{**}(T)$, which were also mentioned in Sec. 17, coincide with the estimates $B_T^*(T)$ and $B_T^{**}(T)$ referring to the stationary sequence $X_{1/N}(t) = X(t/N)$, $t = 0,\pm1,\pm2, ...,$ which is a discrete sample of the process $X(t)$ at a sampling interval $\Delta = 1/N$. Therefore the estimates $B_{TN}^*(T)$ and $B_{TN}^{**}(T)$ need not be considered separately.

$$(3.36) \qquad i_T(\omega) = \frac{1}{2\pi T} \left| \int_0^T e^{-i\omega t} x(t)\, dt \right|^2$$

or

$$(3.36a) \qquad i_T(\omega) = \frac{1}{2\pi T} \left| \sum_{t=1}^T e^{-i\omega t} x(t) \right|^2.$$

The last equations confirm once more that $i_T(\omega) \geqslant 0$ for all ω (i.e. that $B_T^{**}(\tau)$ is a positive definite function of τ). The estimate $i_T(\omega)$ is clearly a sample value (i.e. a realization) of the estimator

$$(3.37) \qquad I_T(\omega) = \frac{1}{2\pi T} \left| \int_0^T e^{-i\omega t} X(t)\, dt \right|^2$$

or

$$(3.37a) \qquad I_T(\omega) = \frac{1}{2\pi T} \left| \sum_{i=1}^T e^{-i\omega t} X(t) \right|^2.$$

Both the nonrandom function $i_T(\omega)$ and the random function $I_T(\omega)$ of the argument ω are called the *periodogram* of a random function $X(t)$. The periodogram was first introduced at the end of the last century by the British physicist Schuster in connection with another (though related) problem,[36] and now it plays a very important role in statistical spectral analysis. It is clear that for a real random function $X(t)$ (recall that we agreed to consider only such functions in Chap. 3 of the book) the periodogram $i_T(\omega)$ in the case of continuous time can be rewritten as

$$(3.38) \qquad i_T(\omega) = \frac{1}{2\pi T} \left\{ \left[\int_0^T x(t)\cos\omega t\, dt \right]^2 + \left[\int_0^T x(t)\sin\omega t\, dt \right]^2 \right\}$$

and that a similar expression is also possible for $I_T(\omega)$ in the case of continuous time and for both functions $i_T(\omega)$ and $I_T(\omega)$ in the case of discrete time.

By using the relation $\langle B_T^{**}(\tau) \rangle = (1 - |\tau|/T)B(\tau)$ and then expressing $B(\tau)$ in terms of $f(\omega)$, it is easy to show that in the case of continuous time

$$(3.39) \qquad \langle I_T(\omega) \rangle = \int_{-\infty}^{\infty} \frac{2}{\pi T} \frac{\sin^2[T(\omega' - \omega)/2]}{(\omega' - \omega)^2} f(\omega')\, d\omega'.$$

(More detailed derivations of this and some subsequent equations are given, together with some relevant additional comments, in Note 37.) In discrete time, (3.39) is replaced by

$$(3.39a) \quad \langle I_T(\omega) \rangle = \int_{-\pi}^{\pi} \frac{1}{2\pi T} \frac{\sin^2[T(\omega' - \omega)/2]}{\sin^2[(\omega' - \omega)/2]} f(\omega')d\omega'.$$

Equations (3.39) and (3.39a) show that, in general, $\langle I_T(\omega) \rangle \neq f(\omega)$, i.e. that $I_T(\omega)$ is a biased estimator of the spectral density $f(\omega)$. However, the last circumstance is of no importance. Indeed, at small T a more or less satisfactory estimate of $f(\omega)$ is impossible, and at large T the difference between $\langle I_T(\omega) \rangle$ and $f(\omega)$ is very small, because both the integrand functions of $\omega' - \omega$ in the right-hand sides of (3.39) and (3.39a) tend to the δ-function $\delta(\omega' - \omega)$ as $T \to \infty$. The last comment implies that the estimator $I_T(\omega)$ of the continuous spectral density $f(\omega)$ (as well as the estimator $B_T^{**}(\tau)$ of the function $B(\tau)$) is *asymptotically unbiased*, i.e. its bias tends to zero as $T \to \infty$. This would naturally be expected because the functions $B_T^{**}(\tau)$ and $I_T(\omega)$ are Fourier transforms of one another.

Calculation of the variance $\sigma^2[I_T(\omega)] = \langle (I_T(\omega) - \langle I_T(\omega) \rangle)^2 \rangle$ and of the centered correlation function $b_I(\omega_1,\omega_2) = \langle (I_T(\omega_1) - \langle I_T(\omega_1) \rangle)(I_T(\omega_2) - \langle I_T(\omega_2) \rangle) \rangle$ of the non-stationary random function $I_T(\omega)$ of a variable ω is, of course, a more complicated task. Nevertheless, under some additional assumptions explicit expressions can also be obtained for these characteristics. We shall restrict ourselves to the rather simple (but practically very important) case where the function $X(t)$ is Gaussian, although in fact close results can also be derived under much wider assumptions (see Note 37). In the Gaussian case it is easy to prove that, for instance, in continuous time

$$b_I(\omega_1,\omega_2) = \left[\int_{-\infty}^{\infty} \frac{2}{\pi T} \frac{\sin[T(\omega' - \omega^1)/2]}{\omega' - \omega_1} \frac{\sin[T(\omega' - \omega_2)/2]}{\omega' - \omega_2} f(\omega')d\omega'\right]^2$$

$$(3.40)$$

$$+ \left[\int_{-\infty}^{\infty} \frac{2}{\pi T} \frac{\sin[T(\omega' - \omega_1)2]}{\omega' - \omega_1} \frac{\sin[T(\omega' + \omega_2)/2]}{\omega' + \omega_2} f(\omega')d\omega'\right]^2$$

The corresponding equation for $b_I(\omega_1,\omega_2)$ in the case of discrete time differs from (3.40) only by replacement of the limits of integration on the right-hand side by $-\pi$ and π, and of the functions $g_T(x) = (2/\pi T)^{1/2}[\sin(Tx/2)/x]$ (where $x = \omega' - \omega_1$, or $\omega' - \omega_2$, or $\omega' + \omega_2$) by

$$g_T^{(1)}(x) = \frac{1}{(2\pi T)^{1/2}} \frac{\sin(Tx/2)}{\sin(x/2)} \,.$$

Putting $\omega_1 = \omega_2 = \omega$ in (3.40), we obtain an explicit expression for the variance $\sigma^2[I_T(\omega)]$. It is not hard to show that for $T \to \infty$ this expression implies the result

(3.41) $\lim\limits_{T\to\infty} \sigma^2[I_T(\omega)] = \begin{cases} 2f^2(\omega) \text{ for } \omega = 0, \\ f^2(\omega) \text{ for } \omega \neq 0. \end{cases}$

A similar result is valid for discrete time, but here $\lim\limits_{T\to\infty} \sigma^2[I_T(\omega)] = 2f^2(\omega)$ both for $\omega = 0$ and for $\omega = \pm\pi$ (although again $\lim\limits_{T\to\infty} \sigma^2[I_T(\omega)] = f^2(\omega)$ for all other ω). Moreover, if $\omega_2 \neq \pm\omega_1$ (recall that by definition $I_T(-\omega) = I_T(\omega)$), then (3.40) leads to the result

(3.42) $\lim\limits_{t\to\infty} b_I(\omega_1,\omega_2) = 0 \text{ for } \omega_2 \neq \pm\omega_1.$

It should also be noted that in fact, as $T \to \infty$, the periodogram ordinates $I_T(\omega)$ at points ω_1 and $\omega_2 \neq \pm\omega_1$ tend to be not only uncorrelated, but even mutually independent random variables (see Note 37).

It follows from (3.41) that the periodogram variance $\sigma^2[I_T(\omega)]$ does not tend to zero, as $T \to \infty$, but remains finite. Hence, only the bias (i.e. the systematic error) of the periodogram estimate $i_T(\omega)$ for the spectral density $f(\omega)$ tends to zero as $T \to \infty$ (because $\langle I_T(\omega)\rangle \to f(\omega)$), but the root-mean-square random error of this estimate $\sigma(I_T) = \langle (I_T(\omega) - \langle I_T(\omega)\rangle)^2\rangle^{1/2}$ is finite at all T (and at large values of T it is of the same order of magnitude as $f(\omega)$.) We see that the periodogram $i_T(\omega)$ is not a consistent estimate of the spectral density $f(\omega)$ despite the fact that its Fourier transform — the function $B_T^{**}(\tau)$ — is a consistent estimate of the Fourier transform $B(\tau)$ of the function $f(\omega)$. (The explanation of this discrepancy in the properties of the estimates $B_T^{**}(\tau)$ and $i_T(\omega)$ will be discussed later.) Moreover, (3.42) shows that when T is large the random variables $I_T(\omega)$ at different frequencies ω are (at least in the case of a Gaussian $X(t)$, but in fact also under much wider conditions) mutually uncorrelated (and even independent). The last fact implies that at high values of T the non-random periodogram $i_T(\omega)$ (which is the realization of $I_T(\omega)$) must be an extremely irregular function of the frequency ω: its values must change abruptly with insignificant changes in argument, fluctuating disorderly about the mean value $\langle I_T(\omega)\rangle$ (which differs little

from the spectral density $f(\omega)$).   Indeed, whenever the values of $i_T(\omega)$ were actually determined from a sufficiently long observed stationary time series $x(t)$, the frequency variation of $i_T(\omega)$ was found to be quite erratic. This fact will now be illustrated by several examples.

The first example which we shall consider has played an important role in the development of the whole field of time series analysis — it is the Beveridge periodogram of the wheat price index for years 1500-1869 shown in Fig. 6(e).   This periodogram is presented in Fig. 40(a).* We can see that the function $i_T(T_0)$, $T_0 = 2\pi/\omega$, is in this case characterized by a large number of maxima ("peaks") separated by deep troughs. Beveridge was inclined to think that at least 18 of these maxima are significant, i.e. they reflect some real periodic changes in economic conditions.   However, the previous considerations (in particular, equations (3.41) and (3.42)) cast considerable doubt on this conclusion; it would rather be expected that either all or almost all the "peaks" in Fig. 40(a) were actually produced by very strong random fluctuations in values $i_T(\omega)$.**

The second, more modern example of the same kind is given in Fig. 40(b), which shows the periodogram of the series consisting of 200 successive observations of mean yearly air temperature in Basel.[39] Here also the periodogram is very jagged and irregular; therefore, it is difficult to believe that its peaks are unrelated to random fluctuations of

---

*As was customary in the past, the x-axis in Fig.   40(a) denotes not the frequency ω (or $n = \omega/2\pi$), but the period $T_0 = 1/n = 2\pi/\omega$. This choice of the independent variable affects the overall shape of the graph, but not the number of "peaks" and "troughs" in it.[38]

**Apparently, graphs of the type depicted in Fig. 40 led the prominent British applied mathematician and geophysicist, Harold Jeffreys, to the pessimistic conclusion, related to periodograms, that usually "periodicities found by harmonic analysis and not predicted by previous theoretical considerations should be mistrusted, as many complications are capable of giving spurious periods; not more than a tenth of those that have been asserted will bear a proper statistical examination" (see Jeffreys and Jeffreys, 1950, p. 452).
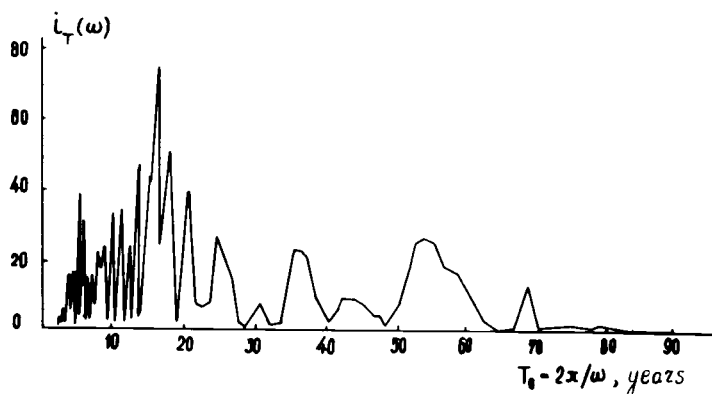
(a)
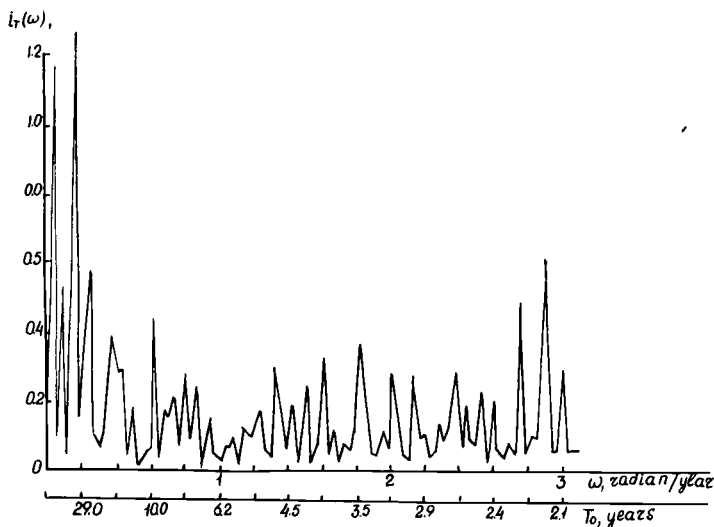
(b)

Fig. 40.   (a) The Beveridge periodogram of the wheat price
index (Yule and Kendall, 1950). (b) The periodogram
of the mean yearly temperatures in Basel calculated
from 200 observations (Polyak, 1975).

periodogram values, but reflect some true properties of the
series considered.

In both examples shown in Figs. 40(a) and 40(b) we naturally
do not know the true form of the spectral density $f(\omega)$ and can

only   suspect   that   this   density   is   smoother   than   the corresponding periodogram.   If, however, we wish to compare the values of $i_T(\omega)$ with the true values of $f(\omega)$, we must consider the periodograms of some artificial time series $x(t)$ simulating   the   realizations   of   the   stationary   sequences $X(t)$ with   known   correlation   functions $B(\tau)$ and   spectral   densities $f(\omega)$ (cf. Sec. 6, p. 79).   The first examples of periodograms of such   artificial   time   series   were   calculated   by   Kendall   (1946). One of his examples is reproduced in Fig. 41(a), where the solid line shows the true spectral density $f(\omega)$, and the dashed line the corresponding periodogram $i_T(\omega)$ for $T = 480$.   We see that   the   values   of $i_T(\omega)$ fluctuate   disorderly   about   the   values of $f(\omega)$, forming a great number of "peaks" and "troughs", which have nothing to do with the true form of the spectral density.   Another   similar   example   (due   to   Wold,   1965)   is depicted in Fig. 41(b), where two periodograms $i_T(\omega)$, $T = 200$, are   presented   together   with   the   graph   of   the   true   spectral density $f(\omega)$.   The periodograms in Fig. 41(b) are calculated from two different realizations $x_1(t)$ and $x_2(t)$ of length $T =$ 200 of the same stationary sequence $X(t)$ having the spectral density $f(\omega)$.   Note that both periodograms are quite erratic and that their deviations from the smooth graph of $f(\omega)$ bear no resemblance to one another.   Finally, in Fig.   41(c) a periodogram is shown which is calculated from a realization of   the   "discrete   white   noise"   (i.e.   of   the   sequence   of independent random variables with a spectral density $f(\omega) =$ constant).   All   these   figures   show   quite   clearly   that   a periodogram is in fact a very poor estimate of the spectral density. Moreover, the mathematical results presented above (see,   in   particular   (3.42)   and   Note   37)   imply   that   with   an increase in realization length T the behavior of $i_T(\omega)$ must become   increasingly   erratic.   It   is   therefore   clear   that   the assertion that $f(\omega) = \lim_{T \to \infty} I_T(\omega)$ (which is rather often made in

the engineering literature and sometimes is even used as the definition   of   the   spectral   density $f(\omega)$)   is   erroneous   and devoid of any basis.
     The   above-mentioned   mathematical   results   not   only explain the extreme irregularity of the periodogram graphs in   Figs.   40   and   41,   but   also   suggest   that   acceptable (consistent)   estimates   of   the   spectral   density $f(\omega)$ can   be
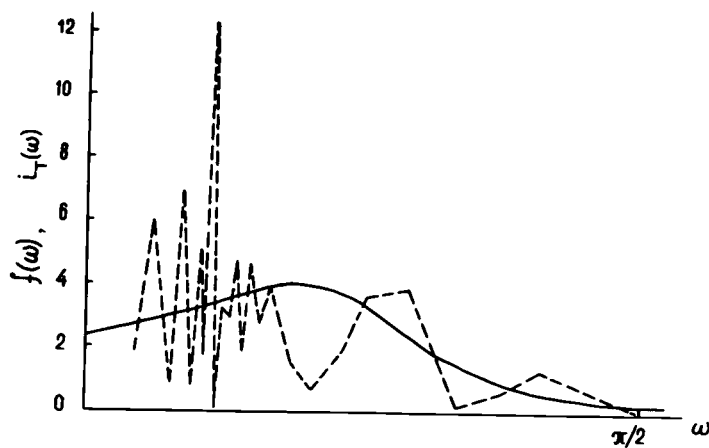
(a)
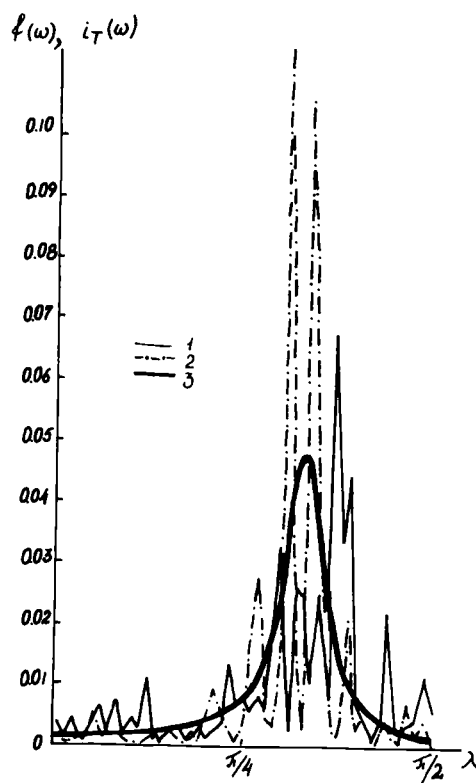
(b)
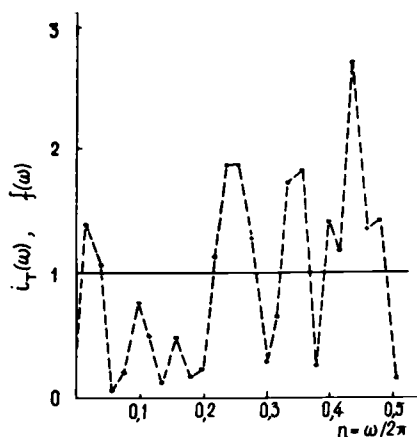
(c)

Fig. 41. (a) The periodogram of an artificial sequence of
length $T = 480$ (the dashed line) compared with the
true spectral density of a stationary sequence (the
solid line) (Kendall, 1946). (b) The true spectral
density of a stationary sequence (the solid line 3) and
two periodograms calculated from two independent
realizations of length $T = 200$ (the lines 1 and 2)
(Wold, 1965). (c) The periodogram $i_{200}(\omega)$ (the dashed
line) and the true spectral density for the sequence of
independent random variables (Jenkins and Watts,
1968).

obtained with the aid of appropriate averaging (or
smoothing) of the periodogram ordinates. The necessary
averaging can be realized in different ways, and below we
consider two constructions of consistent estimates of $f(\omega)$
which involve periodograms. The first construction is based
on the fact (clearly demonstrated by Fig. 41(b)) that if
$i_T^{(1)}(\omega)$ and $i_T^{(2)}(\omega)$ are periodograms calculated from two
independent (or only weakly dependent) realizations $x^{(1)}(t)$
and $x^{(2)}(t)$ of length T of the same random function $X(t)$,
then the two differences $i_T^{(1)}(\omega) - \langle I_T(\omega) \rangle \approx i_T^{(1)}(\omega) - f(\omega)$
and $i_T^{(2)}(\omega) - \langle I_T(\omega) \rangle \approx i_T^{(2)}(\omega) - f(\omega)$ at any fixed ω will be
independent (or, at least, almost independent). Let us now
divide a realization $x(t)$ of sufficiently great length T into
a large number n of subsets of length $T_1 = T/n$ and

compute the periodograms $i_{T_1}^{(k)}(\omega)$, $k = 1, ..., n$, for each of these subsets. Then the arithmetic mean $\varphi_T^{(n)}(\omega) = [\sum_{k=1}^{n} i_{T_1}^{(k)}(\omega)]/n$ will, as an estimate of $f(\omega)$, have a bias $\langle I_{T_1}(\omega) \rangle - f(\omega)$ (clearly tending to zero as $T_1 = T/n \to \infty$) and a variance close to $\sigma^2[I_T(\omega)]/n$ (tending to zero as $n \to \infty$).[*] Hence, if $n = n_T$ is chosen in such a way that $n \to \infty$ and $T/n \to \infty$ as $T \to \infty$, then the estimate $\varphi_T^{(n)}(\omega)$ of $f(\omega)$ will be consistent.[40]

We shall now pass to the second construction of a consistent estimate of $f(\omega)$ that is also based on periodogram computation, but does not require partition of the available lengthy realization and determination of the periodograms for each of the parts. Instead of it, the new method uses averaging over some frequency band (i.e. frequency smoothing) of the periodogram ordinates $i_T(\omega)$ computed from the whole observed realization $x(t)$, $t = 1, 2, ..., T$ or $0 \leqslant t \leqslant T$. Since the values of $I_T(\omega)$ at different (but close) frequencies ω are asymptotically uncorrelated (and even asymptotically independent) random variables with mean values $\langle I_T(\omega) \rangle \approx f(\omega)$, the averaging of $I_T(\omega)$ over a narrow frequency band considerably reduces the variance. Moreover, such an averaging affects the mean value only slightly if the variation of the spectral density $f(\omega)$ within the band used is immaterial. It can be deduced from (3.40) (or the similar equation for discrete time) that at a large, but finite value of T the correlation between random variables $I_T(\omega_1)$ and $I_T(\omega_2)$ practically disappears when $|\omega_2 - \omega_1|$ is of the order of $1/T$ (see Note 37). Therefore, if we average the values of $i_T(\omega)$

---

[*]If T_1 is great enough (much greater than the correlation time of the random function $X(t)$), even the adjacent realizations of length T_1 will be very weakly interdependent. Hence all the corresponding periodograms $I_T^{(k)}(\omega)$, $k=1, ..., n$, will be practically independent. Thus, the variance of their sum will be close to $n\sigma^2[I_{T_1}(\omega)]$ and the variance of the arithmetic mean $\Phi_T^{(n)}(\omega)$ will be close to $\sigma^2[I_{T_1}(\omega)]/n$.

over a frequency band of length Δ_T, where $\Delta_T \to 0$ as $T \to \infty$, but $\Delta_T \gg 1/T$, i.e. $T\Delta_T \to \infty$ (e.g., if $\Delta_T \sim T^{-\alpha}$, where $0 < \alpha < 1$), we arrive at a consistent estimate

$$f_T^*(\omega) = \frac{1}{\Delta_T} \int_{\omega-\Delta_T/2}^{\omega+\Delta_T/2} i_T(\omega)d\omega.$$

Precisely these considerations were adduced in the late 1940s, when the possibility of consistent estimation of any continuous spectral density $f(\omega)$ was first proven.[41] (Note 41 contains also the reference to the amazing early note by Einstein anticipating in 1914 this approach to the spectral density estimation.)

Consider now at greater length the averaging procedures used for periodogram smoothing. The simplest uniformly weighted averaging over a frequency band of length $\Delta = \Delta_T$ is, of course, only one of many possibilities. A much more general averaging operation is described by a convolution integral involving some weighting function $A_T(\omega)$:

$$(3.43)^* \qquad \varphi_T^{(A)}(\omega) = \int A_T(\omega - \omega')i_T(\omega')d\omega'.$$

In (3.43), as everywhere below in this and the next sections, an integral without any limits of integration is assumed to be taken over the entire axis $-\infty < \omega' < \infty$ in the case of continuous time and over the interval $-\pi \leqslant \omega' < \pi$ in the case of discrete time. (In the last case the weighting function $A_T(\omega)$ should be considered periodic with period 2π.) The *smoothed periodogram estimate* (3.43) is a sample value (a realization) of the corresponding *smoothed periodogram estimator*

$$(3.43a) \qquad \Phi_T^{(A)}(\omega) = \int A_T(\omega - \omega')I_T(\omega')d\omega',$$

which is a random function of ω.

The right-hand sides of (3.43) and (3.43a) have the form of convolution of two functions to which, as is well known, the Fourier transform can conveniently be applied. Indeed, let $a_T(\tau)$ be a Fourier transform of the weighting function $A_T(\omega)$, i.e.

---

*It is sometimes convenient to replace $A_T(\omega - \omega')$ in the right-hand side of (3.43) by $[A_T(\omega - \omega') + A_T(\omega + \omega')]/2$. Since $I_T(-\omega) = I_T(\omega)$, such a replacement clearly does not change the value of the integral.

(3.44) $a_T(\tau) = \int e^{i\omega\tau} A_T(\omega)d\omega$, $-\infty < \tau < \infty$ or $\tau = 0,\pm 1,\pm 2,\dots$.

Then, taking into account the relation $B_T^{**}(\tau) = \int e^{i\omega\tau} i_T(\omega)d\omega$,

we obtain from (3.43), by virtue of (3.35) or (3.35a),

(3.45) $\int e^{i\omega\tau} \varphi_T^{(A)}(\omega)d\omega = a_T(\omega)B_T^{**}(\tau)$.

Equation (3.45) expresses the well-known "convolution theorem", which states that the Fourier transform of a convolution of two functions is equal to the product of their Fourier transforms. This equation also implies the relation

(3.46) $\varphi_T^{(A)}(\omega) = \dfrac{1}{2\pi} \displaystyle\int_{-T}^{T} e^{-i\omega\tau} a_T(\tau)B_T^{**}(\tau)d\tau$

or

(3.46a) $\varphi_T^{(A)}(\omega) = \dfrac{1}{2\pi} \displaystyle\sum_{\tau=-T+1}^{T-1} e^{i\omega\tau} a_T(\tau)B_T^{**}(\tau)$.

The estimator (3.43a) can, of course, also be represented in the form (3.46) or (3.46a), but in this case by $B_T^{**}(\tau)$ we must mean the corresponding estimator (and not the estimate) of the function $B(\tau)$ (i.e. the time-averaging of $X(t + \tau)X(t)$, rather than $x(t + \tau)x(t)$).

   *We note in passing that the estimate $B_T^{**}(\tau)$ of $B(\tau)$ in the right-hand side of (3.46) and (3.46a) can be replaced by any other consistent estimate of $B(\tau)$ and then we again obtain a consistent estimate of $f(\omega)$. We shall mention one special case which is of some applied interest. Consider a Gaussian random function $X(t)$ which has mean value zero and let $\hat{R}_T^{(p)}(\tau) = \sin[\pi\hat{B}_T^{(p)}(\tau)/2]$, where

$$\hat{B}_T^{(p)}(\tau) = (1/T)\int_0^{T-|\tau|} \mathrm{sgn}\,x(t + |\tau|)\mathrm{sgn}\,x(t)dt$$

if t is continuous and

$$\hat{B}_T^{(p)}(\tau) = (1/T)\sum_{t=1}^{T-|\tau|} \mathrm{sgn}\,x(t + |\tau|)\mathrm{sgn}\,x(t)$$

if t is discrete, be the polarity coincidence estimate of $R(\tau) = B(\tau)/B(0)$ (see p. 236 and Note 25 to this chapter in Vol. II). If we now replace $B_T^{**}(\tau)$ by $\hat{R}_T^{(p)}(\tau)$ in (3.46)

and (3.46a), we obtain a consistent estimate of the normalized spectral density $h(\omega) = f(\omega)/B(0) = f(\omega)/\int f(\omega')d\omega'$ (coinciding with $f(\omega)$ in cases where $B(0) = 1$). This new estimate is very easy to compute, since it depends only on a very simple binary function $y(t) = \mathrm{sgn}\,x(t)$ (i.e., if t is continuous, only on the position of zeros of a realization $x(t)$).[42] *

Of course, the smoothed periodogram $\varphi_T^{(A)}(\omega)$ is a reasonable estimate of the spectral density $f(\omega)$ only if the integration in the right-hand side of (3.43) describes the averaging of $i_T(\omega)$ over a narrow frequency band enclosing the given point ω. To ensure this we must choose a weighting function $A_T(\omega)$ having a peak (maximal value) at the point $\omega = 0$, and decreasing with increasing $|\omega|$; at relatively large $|\omega|$ this function must vanish or, at least, take very small values. It also appears expedient to assume that the function $A_T(\omega)$ is even (i.e. $A_T(-\omega) = A_T(\omega)$), and that it is normalized by the condition

(3.47) $\int A_T(\omega)d\omega = 1.$

Moreover, we must also require that, as $T \to \infty$, $A_T(\omega)$ becomes more and more "narrow" (i.e. that it concentrates more and more closely about zero). Then the function $a_T(\tau)$ will also have a maximum at the point $\tau = 0$ (where $a_T(0) = 1$ by virtue of (3.47)) and will decrease with increasing $|\tau|$, taking small values at large $|\tau|$. Note that the evenness of $A_T(\omega)$ implies that $a_T(\tau)$ is real and even, so that $a_T(-\tau) = a_T(\tau)$. According to the "uncertainty principle" of function theory (see Sec. 10, p. 116, and also Note 24 to Chap. 2) the "effective width" of the function $a_T(\tau)$ is inversely proportional to that of $A_T(\omega)$. Therefore an unbounded narrowing of the function $A_T(\omega)$, as $T \to \infty$, corresponds to an unbounded broadening of the function $a_T(\tau)$. In other words, at large values of T the function $a_T(\omega)$ must be chosen so that it remains close to its maximum value $a_T(0) = 1$ in an appreciable range of not-too-large values of $|\tau|$ and takes small values only at relatively large τ (say, those exceeding a specified fraction of T). The practical methods for a rational choice of functions $a_T(\tau)$ and $A_T(\omega)$ and examples of some widely used forms of these functions will be considered below. At present we note only that the function $A_T(\omega)$ is usually called the *spectral window* of the estimate $\varphi_T^{(A)}(\omega)$ and the function $a_T(\tau)$, its *lag window*.[43]

The representation (3.46) − (3.46a) of the estimate $\varphi_T^{(A)}(\omega)$ sheds additional light on the role of "frequency smoothing" in (3.43). We have already noted above that $B_T^{**}(\tau)$ for any fixed τ is a consistent estimate of $B(\tau)$, but nevertheless the periodogram $i_T(\omega)$ is an inconsistent estimate of $f(\omega)$. It is easy to see the cause of this. Indeed, according to (3.35) − (3.35a), $i_T(\omega)$ depends on *all* the values of $B_T^{**}(\tau)$, while at not-too-small values of $|\tau|/T$ these values are, for any T, rather poor estimates of $B(\tau)$. Recall that the graph of $B_T^{**}(\tau)$ in the range of large τ contains "spurious waves" (or "ripples"), which are unrelated to the true form of $B(\tau)$; see Sec. 17, pp. 237 - 243, and Figs. 37 and 38. As T increases, the amplitude of these "ripples" decreases but the length of the time interval in which they appear increases. Therefore, it is no wonder that the properties of the estimate $i_T(\omega)$ do not improve with increasing T. We have already noted on p. 243 that the spurious waves can be eliminated if we multiply the estimate $B_T^{**}(\tau)$ by a rationally chosen weighting function (called a "lag window") $a_T(\tau)$, which imparts more weight to the most reliable values of $B_T^{**}(\tau)$ at $|\tau| \ll T$ and allows one either to neglect altogether the highly doubtful values of $B_T^{**}(\tau)$ at large values of $|\tau|$ or to take them into account, but with a very small weight. Now we see that multiplication of the estimate $B_T^{**}(\tau)$ by a reasonable lag window is strictly equivalent to the smoothing of the periodogram in accordance with equation (3.43).

Let us now proceed to the mathematical investigation into the properties of estimates (3.43) and estimators (3.43a). At very large values of T, evidently

(3.48)
$$\langle \Phi_T^{(A)}(\omega) \rangle = \int A_T(\omega - \omega')\langle I_T(\omega')\rangle d\omega'$$
$$\approx \int A_T(\omega - \omega')f(\omega')d\omega',$$

while at smaller T we can use exact equations (3.39) and (3.39a). Therefore calculation of the bias $\delta\{\Phi_T^{(A)}(\omega)\} = \langle \Phi_T^{(A)}(\omega) \rangle - f(\omega)$ of the estimator $\Phi_T^{(A)}(\omega)$ involves no difficulty. However, the variance $\sigma^2\{\Phi_T^{(A)}(\omega)\}$ of $\Phi_T^{(A)}(\omega)$ is considerably harder to calculate. For this we must use, first of all, equation (1.55) on p. 68, which permits one to express $\sigma^2\{\Phi_T^{(A)}(\omega)\} = \langle[\Phi_T^{(A)}(\omega) - \langle\Phi_T^{(A)}(\omega)\rangle]^2\rangle$ in terms of the correlation function $b_I(\omega_1, \omega_2)$ of $I(\omega)$. The function $b_I(\omega_1, \omega_2)$ depends on fourth moments of $X(t)$ and hence the variance $\sigma^2\{\Phi_T^{(A)}(\omega)\}$ is specified by the function $B(\tau)$ (or $f(\omega)$) alone

only in some special cases.   Let us restrict ourselves, for
simplicity, to the important special case where the function
$X(t)$ is Gaussian. Then the function $b_I(\omega_1, \omega_2)$ is given by (3.40)
(or by a similar equation referring to discrete time).   Using
this equation it is possible to obtain, after a number of
analytical manipulations, the following important asymptotic
formula for the variance of $\Phi_T^{(A)}(\omega)$ at large values of T:

$$
(3.49) \qquad \sigma^2\{\Phi_T^{(A)}(\omega)\} = \frac{2\pi}{T} \int [A_T^2(\omega - \omega')
$$
$$
+ A_T(\omega - \omega')A_T(\omega + \omega')] f^2(\omega') d\omega'.
$$

Similar relations can in fact also be derived for some
classes of stationary random functions $X(t)$ much wider
than the class of Gaussian functions.[44]
We are, of course, particularly interested in the case where
the spectral window $A_T(\omega)$ has a sharp peak at $\omega = 0$ and falls off
rapidly on both sides of it.   If $A_T(\omega)$ is negligibly small
everywhere except in a small vicinity of zero, where the
variation of the spectral density $f(\omega)$ can be neglected to a first
approximation, then (3.48) and (3.49) can be rewritten as

$$
(3.48a) \qquad \langle \Phi_T^{(A)}(\omega) \rangle \approx f(\omega) \int A_T(\omega') d\omega',
$$

$$
(3.49a) \qquad \sigma^2\{\Phi_T^{(A)}(\omega)\} \approx \frac{2\pi}{T} f^2(\omega) \int A_T^2(\omega') d\omega' \quad \text{for } \omega \neq 0.
$$

For $\omega = 0$ equations (3.49) and $A_T(-\omega) = A_T(\omega)$ imply that the
factor $2\pi/T$ in the right-hand side of (3.49a) must be replaced
by $4\pi/T$.   Moreover, in the case of discrete time, when $A_T(\omega)$
is an even periodic function of period 2π, such a replacement
of the factor $2\pi/T$ by $4\pi/T$ must also be made for $\omega = \pm\pi$.
If (3.48a) is valid, then the condition (3.47) evidently
guarantees asymptotic unbiasedness of the estimator $\Phi_T^{(A)}(\omega)$
(and estimate $\varphi_T^{(A)}(\omega)$) of the spectral density $f(\omega)$.   Moreover,
equation (3.49a) shows that the variance of $\Phi_T^{(A)}(\omega)$ tends to
zero, as $T \to \infty$, only when the following condition is fulfilled
as well:

$$
(3.50) \qquad T^{-1} \int A_T^2(\omega') d\omega' \to 0 \text{ for } T \to \infty.
$$

Hence, *two conditions*, (3.47) *and* (3.50), *are sufficient* (at least
in the case of a Gaussian function $X(t)$, but also in fact

much wider conditions) *for the consistency of the estimate* $\varphi_T^{(A)}(\omega)$ *of* $f(\omega)$.

Consider at first the case of continuous time, i.e. suppose that $X(t)$ is a stationary process. Then the fulfillment of conditions (3.47) and (3.50) can most simply be achieved by choosing a T-dependent frequency function $A_T(\omega)$ in the form

(3.51) $A_T(\omega) = k_T A(k_T \omega),$

where $k_T \to \infty$, as $T \to \infty$, and the frequency function $A(\omega)$ is even and satisfies the conditions

(3.52) $\int A(\omega)d\omega = 1, \quad |A(\omega)| \leqslant A(0), \quad A(\omega) \to 0 \text{ for } |\omega| \to \infty.$

(Note that we did not include the condition $A(\omega) \geqslant 0$, which ensures the nonnegativity of the estimate $\varphi_T^{(A)}(\omega)$ for all T and ω, in (3.52). In fact, this last condition is not strictly necessary and sometimes is even detrimental; this circumstance will be discussed in more detail in the next section.) The spectral window of the form (3.51) is sometimes called a *scale parameter spectral window* (or a *spectral window of scale parameter form*), corresponding to a *scale parameter* k_T and a *spectral window generator* $A(\omega)$. It follows from (3.51) and (3.44) that the lag window $a_T(\tau)$, which corresponds to a scale parameter spectral window, has the form

(3.51a) $a_T(\tau) = a(\tau/k_T).$

Such a lag window is called a *scale parameter lag window* corresponding to a *lag window generator*[45] $a(\tau) = \int \exp(i\omega\tau) A(\omega)d\omega$. The lag window generator $a(\tau)$ is an even function of τ having the properties

(3.52a) $a(0) = 1, \quad |a(\tau)| \leqslant a(0), \quad a(\tau) \to 0 \text{ for } |\tau| \to \infty.$

The specific examples of lag window generators and spectral window generators will be given at the beginning of the next section.

If a spectral window is of scale parameter form, then condition (3.47) is evidently a simple consequence of (3.52), i.e. is always valid. As for condition (3.50), it amounts here to the requirement that the relation $k_T/T \to 0$ as $T \to \infty$ should

hold. Thus, *if* $k_T \to \infty$, *but* $k_T/T \to 0$ *as* $T \to \infty$, *then the estimate* $\varphi_T^{(A)}(\omega)$, *where* $A_T(\omega) = k_T A(k_T\omega)$ *and conditions* (3.52) *hold, is a consistent estimate of* $f(\omega)$.

For discrete time the situation is slightly more complicated since here $A_T(\omega)$ is defined only for $-\pi \leqslant \omega \leqslant \pi$ (and continues periodically with period 2π outside this interval). Therefore, if $A(\omega)$ does not vanish outside a finite segment (say, for $|\omega| > 1$), neither (3.47) nor (3.52a) follows from the validity of (3.51) for $-\pi \leqslant \omega \leqslant \pi$. If, however, $A(\omega) \equiv 0$ for $|\omega| > 1$ and $k_T > 1/\pi$, all the above-mentioned statements concerning estimates $\varphi_T^{(A)}(\omega)$ with $A_T(\omega) = k_T A(k_T\omega)$ will be valid without any change in the case of discrete time as well. Note also that when spectral and lag windows are chosen in the case of discrete time, one often proceeds not from equations (3.51) and (3.52) related to $A_T(\omega)$, but from equations (3.51a) and (3.52a) for $a_T(\tau)$. Then the Fourier transform $A_T(\omega)$ of the function $a_T(\tau) = a(\tau/k_T)$ of discrete argument τ will satisfy (3.51) only approximately.* Note, however, that most interesting for applications is the case where T and k_T are large, so that the function $A_T(\omega)$ has a sharp peak at zero frequency and falls off rapidly to very low values on both sides of it. In this case the behavior of the function $A_T(\omega)$, corresponding to $a_T(\tau) = a(\tau/k_T)$, at relatively small values of $|\omega|$ is well described by (3.51) also when time is discrete. Therefore, we shall, as a rule, use simple equations (3.51) and (3.52) everywhere without mentioning inessential modifications which are, strictly speaking, necessary if $a_T(\tau) = a(\tau/k_T)$, but the time t is discrete (this situation is considered comprehensively in the literature referred to in Note 44). We emphasize, in particular, that if $a_T(\tau) = a(\tau/k_T)$, *where the lag window generator* $a(\tau)$ *satisfies* (3.52a) *and* $k_T \to \infty$, $k_T/T \to 0$, *as* $T \to \infty$, *then the estimate* (3.46) *of the spectral density of a random stationary sequence* $X(t)$ *is consistent* (at least for a Gaussian sequence $X(t)$, but also in fact under much wider conditions).

Suppose again that the function $A_T(\omega)$ is of the form

---

*Using Poisson's summation formula (2.42^1) (see Vol. II, Note 59 to Chap. 2), it is easy to show that if time t is discrete and $a_T(\tau) = a(\tau/k_T)$ where $a(\tau) = \int_{-\infty}^{\infty} \exp(i\omega\tau)A(\omega)d\omega$, the exact formula for $A_T(\tau)$ has the form $A_T(\tau) = k_T \sum_{j=-\infty}^{\infty} A(k_T(\omega + 2\pi j))$.

(3.51), where the spectral window generator $A(\omega)$ satisfies (3.52). According to (3.49a), in this case $\sigma^2\{\Phi_T^{(A)}(\omega)\}$ is proportional to k_T/T (see (3.53) below), i.e. the variance of $\Phi_T^{(A)}(\omega)$ decreases with a decrease in the scale parameter k_T. Hence, taking into account only the estimate variance, it would be expedient to choose k_T to be as small as possible (i.e. to increase as slowly as possible with increasing T). It is easy to see, however, that a slow increase in k_T is disadvantageous from the viewpoint of the estimate bias. Indeed, (3.51) shows that $1/k_T$ characterizes the "width" of a spectral window (the so-called *spectral bandwidth*), so that windows $A_T(\omega)$ of large bandwidth correspond to small values of k_T. It is clear that if the spectral bandwidth is large, the variance of $\Phi_T^{(A)}(\omega)$ (i.e. the mean square of the random error of an estimate) will be small, but simultaneously the bias $\mathfrak{b}\{\Phi_T^{(A)}(\omega)\} = \langle\Phi_T^{(A)}(\omega)\rangle - f(\omega)$ of $\Phi_T^{(A)}(\omega)$ (the systematic error of an estimate) will be large. Also the resolvability of the estimator $\Phi_T^{(A)}(\omega)$ (i.e. its possibility of distinguishing the two spectral peaks at the close frequencies ω_1 and ω_2) will be poor if k_T is small (i.e. the spectral bandwidth is large).* Therefore, in practice one has to find some reasonable compromise between estimates with low values of k_T having a small variance, but a rather large bias (and a poor resolvability), and estimates with high values of k_T having a large variance, but a small bias.[46]

---

*Equations (3.49a) and (3.51) clearly imply that in cases where the spectral window is of scale parameter form

$$(3.53) \qquad \sigma^2\{\Phi_T^{(A)}(\omega)\} \approx 2\pi f^2(\omega) \frac{k_T}{T} \int A^2(\omega')d\omega'$$

at large values of T. However, the determination of the dependence of the bias $\mathfrak{b}\{\Phi_T^{(A)}(\omega)\} = \langle\Phi_T^{(A)}(\omega)\rangle - f(\omega)$ on the scale parameter k_T is a more difficult task. In this case one cannot restrict oneself to the approximate equation (3.48a), which was used above to prove the asymptotic unbiasedness of $\Phi_T^{(A)}(\omega)$. Instead of this, one should revert to the more precise formula (3.48) for

---

*Note that when a high resolvability of an estimate is especially important, it is usually expedient to apply quite different methods for spectral estimation instead of periodogram smoothing (see Sec. 19 below).

$\langle \Phi_T^{(A)}(\omega) \rangle$. Let us expand the function $f(\omega')$ on the right-hand side of (3.48) into the Taylor series at the point $\omega' = \omega$ and then use (3.51). Assuming now that $f(\omega)$ is a twice differentiable function and $\int \omega^2 A(\omega)d\omega = 2L$, where $0 < L < \infty$ (it is easy to see that the latter condition is equivalent to the validity of the relation $a(\tau) \approx 1 - L\tau^2$ for small values of τ), we easily find that

(3.54) $\delta\{\Phi_T^{(A)}(\omega)\} \approx L f''(\omega) k_T^{-2}$

at large values of T and k_T. If, however, $\int \omega^2 A(\omega)d\omega = \infty$ or $\int \omega^2 A(\omega)d\omega = 0$ (the latter relation can, of course, be valid only when the spectral window generator $A(\omega)$ alternates between positive and negative values), then (3.54) is inapplicable. In a more general case, where $a(\tau) \approx 1 - L|\tau|^q$, $0 < L < \infty$, at small $|\tau|$ and the function $f(\omega)$ is smooth enough, the asymptotic equation for $\delta\{\Phi_T^{(A)}(\omega)\}$ takes the form

(3.54a) $\delta\{\Phi_T^{(A)}(\omega)\} \approx L M_q(\omega) k_T^{-q}$,

where $M_q(\omega)$ is determined by the spectral density $f(\omega)$ (in particular, $M_q(\omega) = f^{(q)}(\omega)$ when q is an even integer).[47] Both equations (3.54) and (3.54a) confirm that with a decrease in k_T the bias $\delta\{\Phi_T^{(A)}(\omega)\}$ increases.

Systematic and random errors of a given spectral estimator sometimes play different roles and must be considered separately. Moreover, quite often the overall form of the spectral density function is of greatest interest, while the individual spectral ordinates $f(\omega)$ and the mean square errors of their estimators are of no special importance. Nevertheless, the determination of the "optimal" value of the scale parameter k_T from the minimization condition applied to the mean square error

$$\Delta^2\{\Phi_T^{(A)}(\omega)\} = \langle [\Phi_T^{(A)}(\omega) - f(\omega)]^2 \rangle$$

$$= \delta^2\{\Phi_T^{(A)}(\omega)\} + \sigma^2\{\Phi_T^{(A)}(\omega)\}$$

seems reasonable in many cases. If (3.54) is valid, then the condition $\Delta^2\{\Phi_T^{(A)}(\omega)\}$ = minimum implies that $k_T \sim T^{1/5}$,

$\Delta^2\{\Phi_T^{(A)}(\omega)\} \sim T^{-4/5}$ (and also $\delta^2\{\Phi_T^{(A)}\} \sim T^{-4/5}$, $\sigma^2\{\Phi_T^{(A)}\} \sim T^{-4/5}$) at large values of T, where $\sim$ means "is proportional to". If, however, (3.54a) is valid, then, according to the same minimization condition, $k_T \sim T^{1/(2q+1)}$, $\Delta^2\{\Phi_T^{(A)}\} \sim T^{-2q/(2q+1)}$

(and also $\delta^2\{\Phi_T^{(A)}\} \sim T^{-2q/(2q+1)}$, $\sigma^2\{\Phi_T^{(A)}\} \sim T^{-2q/(2q+1)}$) at large T.[48] Note, however, that the coefficients in the asymptotic relations for "optimal" values of k_T and $\Delta^2\{\Phi_T^{(A)}\}$ depend on an unknown spectral density $f(\omega)$ and its derivatives (the explicit form of these coefficients can be obtained easily from the equations given in Note 48). This dependence limits severely the practical usefulness of the relations. Therefore, other considerations must be used in practice for choosing the value of k_T; some of them will be discussed at the beginning of the next section.*

19. Some Practical Aspects and Additional Methods of Statistical Spectral Analysis. Determination of the Spectral Distribution Function

In Sec. 18 we considered the consistent estimates of the spectral density function $f(\omega)$ obtained by smoothing the periodogram $i_T(\omega)$. However, the computation of such estimates requires a number of practical decisions to be made, related in particular to the selection of a shape for the spectral window generator $A(\omega)$ (or the lag window generator $a(\tau)$) and a value of the scale parameter k_T. (For simplicity, only windows of scale parameter form will be considered below.) Moreover, the smoothed periodogram estimates of $f(\omega)$ form only one type from among several types of spectral estimates used at present in applied works. In this section some practical aspects of spectral estimation by the smoothed-periodogram method will be briefly discussed and a few other methods of such estimation will be outlined. (A more detailed presentation of all these topics can be found in sources referred to in Note 34.)

Let us begin with short remarks concerning the choice of the scale parameter value k_T. If the estimate $B_T^{**}(\tau)$ of the correlation function $B(\tau)$ is calculated, then k_T can be selected from the requirement that multiplication by lag window $a_T(\tau) = a(\tau/k_T)$ would truncate all the spurious "waves" in the graph

of $B_T^{**}(\tau)$, but change only slightly the smooth initial part of this graph. This method has been applied with reasonable success in some practical situations, but, of course, the necessity of plotting the correlation curve often makes such an approach to the choice of k_T rather inconvenient.

Many sources recommend choosing such a value of k_T that an appropriately defined *overall width* of the spectral window $A_T(\omega)$ (i.e. *spectral bandwidth;* cf. Note 46) will be a specified fraction (e.g., one-half) of the width of the narrowest essential detail of the spectral density $f(\omega)$. However, this recommendation also does not seem to be sufficiently widely applicable. Indeed, it is often impossible to estimate beforehand, even crudely, the width of the narrowest essential detail of $f(\omega)$. Moreover, the notion of "essential detail" is rather vague: sometimes it is reasonable to apply this notion to isolated peaks and troughs in $f(\omega)$, but it is also occasionally reasonable to take the length of the smallest interval between adjacent peaks as the "essential width" of $f(\omega)$.

In general, there is no consensus of opinion as to the choice of k_T in practical calculations of the spectral density. This is, of course, only natural, since a single "best" recommendation applicable in all cases does not exist at all. Let us mention, nevertheless, some practical suggestions given by various authors. Blackman and Tukey recommend a choice of k_T such that the lag window $a(\tau/k_T)$ proves to be practically equal to zero for all τ exceeding some "truncation point" T_0 lying within 5 to 10% of the length T of the realization used. (Later some other researchers increased the recommended value of T_0/T up to 20% or even 30%.) Parzen recommends calculating $\varphi_T^{(A)}(\omega)$ for three different values of k_T corresponding to three different values $T_0^{(1)}$, $T_0^{(2)}$, and $T_0^{(3)}$, of the "truncation point" T_0 of the relevant lag window $a_T(\tau)$ such that $0.05 \leqslant T_0^{(1)}/T_1 \leqslant 0.1$, $0.1 \leqslant T_0^{(2)}/T \leqslant 0.25$, and $T_0^{(3)}/T \approx 0.5$, and then to choose "by eye", from among the three graphs of the spectral density obtained, the one that seems most likely. Jenkins and Watts recommend calculations at several values of k_T, starting with a "small" value of k_T (i.e. a large spectral bandwidth) which gives a very smooth function $\varphi_T^{(A)}(\omega)$ and then progressively increasing k_T (i.e. decreasing the bandwidth). Then the graphs of the estimates $\varphi_T^{(A)}(\omega)$ will become more and more complicated and we must choose such a value of k_T that any further increase in it does not

change the main features of the spectral shape and only adds some unlikely "erratic details" to the graph of $\varphi_T^{(A)}(\omega)$ (cf. Fig. 44).[49] In fact, however, none of the possible procedures is universal, and only practical experience can teach one to act rationally in this matter.   Let us just emphasize that any prior information (even very rough) on the spectral properties of the random function considered can be of great help in choosing an appropriate value of k_T, and that trial calculations at several different values of k_T often prove practically expedient.

Now we shall consider some special examples of spectral windows $A_T(\omega)$ and lag windows $a_T(\tau)$. At the end of Sec. 17 (pp. 243 − 245) we indicated three examples of scale parameter lag windows $a_T(\tau) = a(\tau/k_T)$:   the *rectangular window* depicted in Fig. 39(a), the triangular *Bartlett window* (Fig. 39(b)), and the so-called *Parzen window* (Fig. 39(c); the explicit equation for this $a(\tau)$ is given in Vol. II in Note 32 to this chapter). It is easy to check that if the time t is continuous (i.e. a random process $X(t)$ is studied), the corresponding spectral windows are also of scale parameter form $A_T(\omega) = k_T A(k_T\omega)$, where $A(\omega) = \sin\omega/\pi\omega$ for the rectangular lag window, $A(\omega) = 2\sin^2(\omega/2)/\pi\omega^2$ for the Bartlett window, and $A(\omega) = (3/8\pi)(\sin[\omega/4]/[\omega/4])^4$ in the case of the Parzen window. (Slightly more complicated equations for $A_T(\omega)$ are obtained if $X(t)$ is a stationary sequence, i.e. the time is discrete.)[50] Note that all three above-mentioned lag windows vanish at $|\tau| \geqslant k_T$. This circumstance, as will be explained below, had long been regarded as considerably simplifying the calculations, but at present it has essentially lost its former importance. Spectral windows $A_T(\omega)$ corresponding to the Bartlett and Parzen lag windows turned out to be strictly nonnegative, whereas in the case of the rectangular window $a(\tau)$ the function $A(\omega)$ takes both positive and negative values. This circumstance confirms the statement made in Sec. 17 that the Bartlett and Parzen lag windows (but not the rectangular lag window) are positive definite functions of τ.

Two more lag windows of scale parameter form $a_T(\tau) = a(\tau/k_T)$ were proposed by Tukey. These windows are specified by the lag window generator

$$(3.55) \qquad a(\tau) = \begin{cases} (1 - 2a) + 2a\cos\pi\tau & \text{for } |\tau| \leqslant 1, \\ \\ 0 & \text{for } |\tau| > 1, \end{cases}$$

where either $a = 0.23$ (the *"Hamming"* or *"Tukey–Hamming"* window), or $a = 0.25$ (the *"Hanning"* or *"Tukey–Hanning"* window, which is also often referred to simply as the *"Tukey window"*). Both the commonly used lag window generators of the form (3.55) are shown in Fig. 42; neither of them is strictly positive definite, but their Fourier transforms $A(\omega)$ only slightly invade the region of negative values.[51] Many other examples of lag and spectral windows can be found in the available vast literature on spectral analysis of stationary random functions, but the overwhelming majority of them will not even be mentioned in this book.[52]

Recall, however, the simplest (and earliest proposed) method, referred to on p. 257, for obtaining a consistent estimate of the spectral density $f(\omega)$ by averaging the periodogram over an interval of length $\Delta_T = 2/k_T$ (where $\Delta_T \to 0$ and $T\Delta_T \to \infty$, i.e. $k_T \to \infty$ and $k_T/T \to 0$, as $T \to \infty$) centered on the frequency ω. This method gives an estimate which is evidently a special case of (3.43) corresponding to the *rectangular spectral window* of scale parameter form $A_T(\omega) = k_T A(k_T\omega)$, where $A(\omega)$ is depicted in Fig. 43(a). The rectangular spectral window is often also called the *Daniell spectral window* (but it may also be called the *Einstein–Daniell window;* cf. Note 41). The lag window corresponding to this spectral window has the form $a_T(\tau) = a(\tau/k_T)$, where $a(\tau) = \sin\tau/\tau$ (both for continuous and discrete t). Two other reasonable spectral window generators $A(\omega)$, vanishing outside a finite frequency interval are the *triangular spectral window generator* $A(\omega) = \max\{1-|\omega|, 0\}$ (Fig. 43(b)) and the *parabolic spectral window generator* $A(\omega) = \max\{3(1-\omega^2)/4, 0\}$
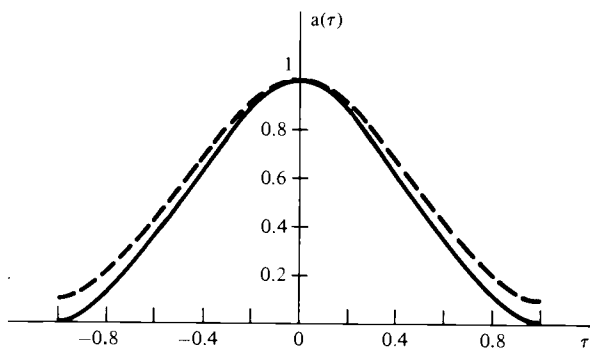
Fig. 42. The Tukey lag window generators: Hanning window generator (the solid line) and Hamming window generator (the dashed line).
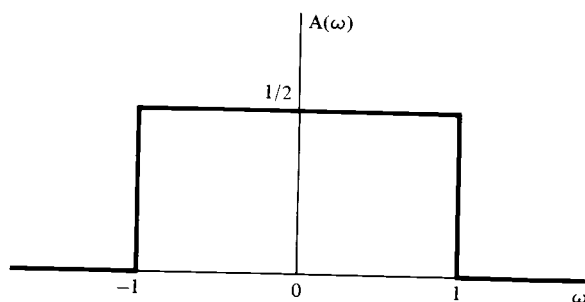
(a)
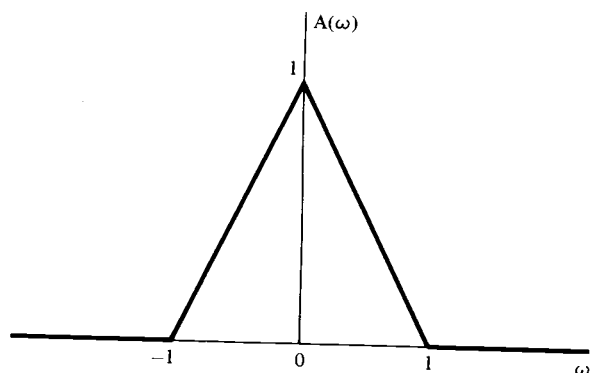
(b)
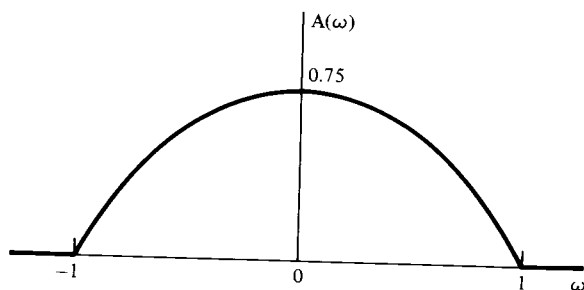
(c)

Fig. 43. Three simple spectral window generators: (a) The rectangular (or Daniell's) window generator; (b) the triangular window generator; (c) the parabolic window generator.

(Fig. 43(c)). All three spectral windows depicted in Fig. 43 are strictly non-negative; therefore, the corresponding lag windows $a_T(\tau) = a(\tau/k_T)$ and lag window generators $a(\tau)$ are positive definite functions of τ. Note that all three lag windows do not vanish after a certain value $|\tau|$, i.e. they have no "truncation point". However, the simplicity of the corresponding spectral windows makes them very convenient for direct computation of the estimates $\varphi_T^{(A)}(\omega)$ with the aid of the so-called fast Fourier transform (FFT) algorithm, which will be discussed later in this section.   Therefore these three windows have attracted much attention recently.[53]

*In   recent   years   spectral   windows $A_T(\omega)$ with alternating signs have also attracted much attention. For a long time the condition $A_T(\omega) \geqslant 0$ seemed to be quite necessary, since it guarantees non-negativity of the estimate $\varphi_T^{(A)}(\omega)$ of the non-negative function $f(\omega)$.   It is clear, however, that if T is large enough, then any reasonable   spectral   window $A_T(\omega)$ must   be   very "narrow", i.e. all its values outside a narrow frequency band centered at the zero frequency must be very close to zero.   Therefore the probability of obtaining a negative value of $\varphi_T^{(A)}(\omega)$ is very low at large T even if $A_T(\omega)$ takes both positive and negative values. (Negative values of $\varphi_T^{(A)}(\omega)$ can appear only if the true spectral density $f(\omega)$ changes its value quite abruptly at some frequencies.   Of course, if $\varphi_T^{(A)}(\omega) < 0$ for some ω, then it is reasonable to replace this value of $\varphi_T^{(A)}(\omega)$ by zero.) On the other hand, it is easy to check that if an alternating spectral window generator $A(\omega)$ satisfies the

conditions $\int_{-\infty}^{\infty} \omega^{2k} A(\omega)d\omega = 0$, for $k = 1, 2, ..., m - 1$, then q

$= 2m$ in (3.54a). Hence, in this case the estimate bias decreases with an increase in k_T much more rapidly than in the case of a strictly nonnegative spectral window with $q = 2$.   This implies that if T is sufficiently large, it is possible to reduce considerably the bias $\mathfrak{b}\{\Phi_T^{(A)}(\omega)\}$ of the estimator $\Phi_T^{(A)}(\omega)$ and also its mean square error

$$\Delta^2\{\Phi_T^{(A)}(\omega)\} = \mathfrak{b}^2\{\Phi_T^{(A)}(\omega)\} + \sigma^2\{\Phi_T^{(A)}(\omega)\} \quad \text{(in comparison}$$

with the values of $\mathfrak{b}\{\Phi_T^{(A)}(\omega)\}$ and $\Delta^2\{\Phi_T^{(A)}(\omega)\}$ accessible

with strictly non-negative spectral windows) by using spectral windows taking both positive and negative values. This conclusion is confirmed by the results of some recent numerical experiments involving simulated time series having known spectral densities. The available literature also contains a number of examples of spectral window generators $A(\omega)$ satisfying the conditions $\int_{-\infty}^{\infty} \omega^{2k} A(\omega) d\omega = 0$ for several first integral values of k. However, so far such spectral windows have been only rarely used in practice.[54]

When we speak of reducing the bias of spectral estimates, it is worth mentioning one more method which is rather widely utilized in practice at present. This method is based not on selection of some special "bias reducing" spectral windows, but on modification of the periodogram $i_T(\omega)$ by so-called *tapering* − a special preliminary transformation of the data. This transformation consists of multiplication of all the data $x(t)$ by some numerical function $h_T(t)$ (depending on the sample length T), i.e. in replacing the values $x(t)$ by $x^{(h)}(t) = h_T(t)x(t)$, $t = 1, 2, ..., T$ or $0 \leqslant t \leqslant T$. The function $h_T(t)$, which is normalized by the condition

$$(3.56) \qquad \sum_{t=1}^{N} h_T^2(t) = T \quad \text{or} \quad \int_0^T h_T^2(t) dt = T,$$

is called a *data window* (or *taper*, or *fader*). The utilization of non-transformed data is clearly equivalent to the choice of the *rectangular data window* $h_T(t) = 1$, for $t = 1, ..., T$ or $0 \leqslant t \leqslant T$. However, data windows gradually decreasing from the middle of the observed realization (the point $t = T/2$) to its edges are most useful for bias reduction (precisely these windows are usually called "tapers" or "faders").

Let $I_T^{(h)}(\omega)$ be the so-called *modified periodogram*, i.e. the periodogram of the tapered random function $X^{(h)}(t) = h_T(t)X(t)$. Similar to the derivation of (3.39) and (3.39a), the following equation for $\langle I_T^{(h)}(\omega) \rangle$ can easily be obtained:

(3.57) $\langle I_T^{(h)}(\omega)\rangle = \int K_T^{(h)}(\omega - \omega')f(\omega')d\omega'$,

where

(3.58) $K_T^{(h)}(\omega) = \dfrac{1}{2\pi T}\left|\displaystyle\sum_{t=1}^{T} e^{-i\omega t}h_T(t)\right|^2$

if the time t is discrete, and

(3.58a) $K_T^{(h)}(\omega) = \dfrac{1}{2\pi T}\left|\displaystyle\int_0^T e^{-i\omega t}h_T(t)dt\right|^2$

if the time is continuous.[55]

The data windows $h_T(t)$ that are most widely used in practice are of the form

(3.59) $h_T(t) = h(t/T)$,

where $h(\tau)$, $0 \leqslant \tau \leqslant 1$, is a bounded continuous function which satisfies the relation*

(3.60) $\displaystyle\int_0^1 h^2(\tau)d\tau = 1$,

has its maximum at $\tau = 1/2$, and decreases to zero at $\tau = 0$ and $\tau = 1$. It is easy to check that in this case $K_T^{(h)}(\omega)$ tends to the δ-function $\delta(\omega)$ as $T \to \infty$. The same is also true if $h_T(t)$ is not of the form (3.59) but the following weaker condition holds: the reasonably defined "effective width" of the window $h_T(t)$ tends to infinity as $T \to \infty$ (see again Note 55). Thus, in all these cases the modified periodogram $I_T^{(h)}(\omega)$ is an asymptotically unbiased estimator of the spectral density $f(\omega)$. Note also that, as is explained in Note 55, at rather large but finite values of T the weighting function $K_T^{(h)}(\omega)$, which corresponds to the continuous function $h(\tau)$ gradually decreasing from the point $\tau = 1/2$ to $\tau = 0$ and $\tau = 1$, is considerably more closely concentrated in the vicinity of the zero frequency than the weighting function in (3.39) or (3.39a). Therefore, at large but finite values of T the modified periodogram bias $\delta\{I_T^{(h)}(\omega)\}$ $= \langle I_T^{(h)}(\omega)\rangle - f(\omega)$ will be considerably smaller than

---

*According to (3.56), relation (3.60) will be only approximately valid in the case of discrete time. However, even in this case the inaccuracy of (3.60) will be negligibly small at large values of T.

than $\langle I_T(\omega) \rangle - f(\omega)$.

Of course, the asymptotic unbiasedness alone does not make the modified periodogram $I_T^{(h)}(\omega)$ an acceptable estimator of $f(\omega)$. It is not hard to show that $I_T^{(h)}(\omega)$ is an inconsistent estimator of the spectral density and that the variance of $I_T^{(h)}(\omega)$ (and also the covariance between the values of $I_T^{(h)}(\omega)$ at two distinct frequencies) tends, as $T \to \infty$, to the same limit as in the case of not tapering (i.e. when the conventional periodogram $I_T(\omega)$ is considered).[56] A consistent estimator of $f(\omega)$ can be obtained from $I_T^{(h)}(\omega)$ by averaging it over a narrow frequency band, i.e. by substituting $I_T^{(h)}(\omega)$ for $I_T(\omega)$ in the right-hand side of (3.43a), where $A_T(\omega) \to \delta(\omega)$ as $T \to \infty$ and condition (3.50) is also fulfilled. However, by so acting we lose, to a certain extent, the main advantage of the modified periodogram since averaging over a frequency band produces a considerable additional bias. (Recall that when estimating the bias of the smoothed periodogram estimator $\Phi_T^{(A)}(\omega)$ on pp. 264-265 we replaced $\langle \Phi_T^{(A)}(\omega) \rangle$ by the right-hand side of (3.48), i.e. we fully neglected the periodogram bias in comparison with the bias due to frequency averaging.) Moreover, data tapering leads to an increase in the variance of the smoothed periodogram estimator, and in this respect such transformation of the data is even harmful. However, the increase in the variance due to tapering is often rather slight (see Note 56). Therefore, it can be hoped that, at least on some occasions, the variance increase would be more than offset by the corresponding reduction in the bias of the estimator.

The bias of a spectral estimate may play an especially inportant role if the true spectral density function $f(\omega)$ has a complicated form and at some frequencies the value of $f(\omega)$ changes very abruptly. Therefore, if there are reasons to believe that $f(\omega)$ shows such a behavior, it is advisable to taper the data before computing a spectral estimate. Note also that a more accurate equation for the mean value $\langle \Phi_T^{(A)}(\omega) \rangle$ of a smoothed periodogram estimator than that given by the right-hand side of (3.48) can be written as

$$(3.61) \qquad \langle \Phi_T^{(A)}(\omega) \rangle = \int A_T^*(\omega - \omega') f(\omega')\, d\omega',$$

where $A_T^*(\omega)$ is the convolution of the spectral window $A_T(\omega)$ with the kernel $K_T(\omega)$ appearing in the integrand on the right-hand side of (3.39) or (3.39a).   Most often the width of the spectral window $A_T(\omega)$ is of the same order as k_T^{-1}, i.e. it is much greater than the width of the kernel $K_T(\omega)$ appearing in the formulae for $\langle I_T(\omega)\rangle$ (this latter width is of the same order as T^{-1}). Therefore, if $A_T(\omega)$ is continuous, we can simply consider $K_T(\omega)$ as a δ-function, i.e. use the simplified equation (3.48).   Of course, little will be gained then by data tapering.   If, however, $A_T(\omega)$ has sharp discontinuities, then the difference between a discontinuous kernel $A_T(\omega)$ and a continuous kernel $A_T^*(\omega) = A_T(\omega)*K_T(\omega)$, where the star denotes convolution, can sometimes have a noticeable effect on the bias. Thus, in situations involving a spectral window with sharp discontinuities (e.g., the Daniell window), tapering of the data, which transforms $K_T(\omega)$ into a narrower kernel $K_T^{(h)}(\omega)$, may sometimes be useful.

We have already seen that if the periodogram is smoothed by applying (3.43), then data tapering often gives little gain.   However, we know that a consistent estimate of the spectral density $f(\omega)$ can also be obtained without the utilization of any periodogram averaging over a frequency band. To this end the observed realization of length T is divided into a great number n of non-overlapping or, what is often preferable, partially overlapping subsets of fixed length T_1 (where $T_1 \ll T$). Then the periodograms $i_{T_1}^{(k)}(\omega)$,

$k = 1, ..., n$, are for each of these subsets and the arithmetical mean $\varphi_T^{(n)}(\omega)$ of n calculated periodograms is used as the estimate of $f(\omega)$.   In this case the bias of the obtained estimate (which coincides, by definition, with the bias of the corresponding estimator) is clearly equal to $\langle I_{T_1}(\omega)\rangle - f(\omega)$, while its variance tends to zero as $T \to \infty$ and $n \to \infty$. Note also that the bias $\delta\{I_{T_1}(\omega)\} = \langle I_{T_1}(\omega)\rangle - f(\omega)$ is considerably greater than $\langle I_T(\omega)\rangle - f(\omega)$ since $T_1 \ll T$. Therefore, it is important to make the bias of $I_{T_1}(\omega)$ as small as possible, and this can be achieved by skillful selection of the data window $h_{T_1}(t)$ applied to each subset of

the length T_1 used for periodogram computation.[57] A slightly more complicated combined method for spectral estimation, which uses the arithmetic mean $\varphi_T^{(n)}(\omega)$ as a first-stage estimate and supplements it by computation of a more precise second-stage estimate, will also be described below in this section.*

Now we shall digress from discussion of data tapering and consider, for simplicity, only the conventional periodograms $i_T(\omega)$ (or $I_T(\omega)$) and the corresponding smoothed periodogram estimates $\varphi_T^{(A)}(\omega)$ (or estimators $\Phi_T^{(A)}(\omega)$). A number of spectral analysis papers discuss the choice of the "optimal shape" of the spectral window $A_T(\omega)$ (or the lag window $a_T(\tau)$).[58] Since different authors have used quite different "optimality criteria" in these discussions, and, moreover, there is no single "optimal window" which is "most effective" simultaneously for all possible forms of the true spectral density $f(\omega)$, it is no wonder that the obtained conclusions also turn out to be contradictory in some respects. It is significant, however, that both theoretical arguments and practical calculations show that the estimates $\varphi_T^{(A)}(\omega)$ corresponding to different, but reasonable, window shapes (e.g., to the Parzen window, two Tukey windows, and the three spectral windows depicted in Fig. 43) are close enough when the window bandwidths do not differ too much (see, e.g., Fig. 45). Therefore, in actual calculations of spectral densities for time series encountered in practice, the window shape is usually of no great concern and can be selected primarily from subjective preferences and reasons related to the convenience of calculation (e.g., the availability of a computer program).

Much space in the literature on spectral analysis is also given to the probability distributions of estimators $\Phi_T^{(A)}(\omega)$ (i.e. sampling distributions for the estimates $\varphi_T^{(A)}(\omega)$). These distributions are necessary for constructing the confidence intervals for the spectral estimates which characterize the estimate reliability (cf. above, p. 224). However, we shall not consider this problem here.[59]

We now proceed to a brief survey of the various practical methods for finding spectral densities. The principal

techniques used for this purpose at present can be conventionally classified into four broad categories.

1. *Methods of spectral density estimation by spectrum, or wave, analyzers.* We have spoken on pp. 157-158 of these methods for approximate determination of the spectral density $f(\omega)$ of a stationary process $X(t)$ from its single realization $x(t)$. Here, to find the function $f(\omega)$, the realization $x(t)$ (recorded in the form of an electric signal) is passed first through a narrow band-pass linear filter. Then the filter output $y(t)$ is squared and this square is integrated over a time interval of reasonable length T and is divided by T (i.e. $y^2(t)$ is averaged over a time interval of length T). The obtained time average of $y^2(t)$ is taken as the estimate of the value $f(\omega)$ of the spectral density at the point ω coinciding with the center of the pass band of the filter. The squaring of the filter output $y(t)$ and the time-averaging of $y^2(t)$ is usually carried out automatically by analog computers (a "square law detector" and an "integrator") included in the spectrum analyzer. A typical analog spectrum analyzer includes a number m (e.g., a dozen or two) of narrow band linear filters (analyzer "cells") with different pass bands, which give a number of values $f(\omega_i)$, $i = 1, ..., m$. Such a set of values $f(\omega_i)$ gives an idea of the entire continuous frequency function $f(\omega)$. There is also another type of spectrum analyzer which uses only one fixed narrow band filter, but includes also a generator of harmonic oscillations of various frequencies and "tunes" the given filter to the required frequency by multiplying $x(t)$ by the function of the form $\exp(i\omega t)$. (The radio engineering technique for such multiplication is known as "heterodyning"; we shall not discuss it here.) Spectrum analyzers are widely used in many applied fields; the various engineering realizations of such analyzers are described in numerous special works.[60]

The spectral density estimate given by an analog spectrum analyzer is not identical to any smoothed periodogram estimate $\varphi_T^{(A)}(\omega)$, but these estimates are of the same basic form in the sense that they are both rather similar quadratic functions of the realization $x(t)$, $0 \leqslant t \leqslant T$. Therefore, the bias and variance (i.e. the systematic and random errors) of a reading of the spectrum analyzer can be investigated quite similarly to the investigation of these errors for a smoothed periodogram estimator $\Phi_T^{(A)}(\omega)$. Such an investigation shows that the statistical properties of spectrum analyzer readings

differ only slightly from those of the estimators $\phi_T^{(A)}(\omega)$ discussed above.[61]

The extensive development in recent years of practical methods for digital signal processing has led to frequent replacement of analog filtering of signals by their digital filtering performed by a computer. For this the continuous signal $x(t)$ has to be converted first of all into a digital record of the discrete sample $x_\Delta(t) = x(t\Delta)$, $t = 1, 2, ...,$ where Δ is an appropriately chosen small number; then the sample $x(t\Delta)$ can easily be filtered digitally (see Note 62 to Chap. 2 for the literature on digital filters). To obtain a spectral density estimate we must now only square the filtered signal $y(t)$ and average $y^2(t)$ over time; both these operations can, of course, also be carried out on a computer. Such a method of spectral density estimation is theoretically quite similar to the estimation of $f(\omega)$ by an analog spectrum analyzer. Therefore, both methods must be included in the same category, though the practical realization of the estimation looks quite different in the two cases.

We also note that quite a different type of desk *digital spectrum analyzer* is rather widely used at the present time. The appearance of these spectrum analyzers is due to the recent considerable improvement and miniaturization of digital computers and the introduction of new, much faster methods for Fourier transform calculations, which will be discussed below. The new analyzers include, first, a special convertor which converts the continuous electric signal $x(t)$ into a magnetic tape record of the numerical sequence $x_\Delta(t) = x(t\Delta)$, $t = 1, 2, ...,$ where Δ is small enough and, second, a small computer with a fixed program for calculating the smoothed periodogram $\varphi_T^{(A)}(\omega)$ of the sequence $x_\Delta(t)$. (In this calculation the form of the spectral window generator $A(\omega)$ is usually fixed, but the value of the scale parameter k_T, which determines the window bandwidth, can be varied.) Digital spectrum analyzers may strongly resemble in their outward appearance the old desk analog spectrum analyzers which use a set of electrical filters, but in fact these new analyzers represent a special engineering realization of the numerical methods forming the third category of practical methods for spectrum density estimation (see below).

2. *Numerical methods for determining the smoothed periodogram* $\varphi_T^{(A)}(\omega)$ *from the correlation function estimate (initial approach of Blackman and Tukey).* The mushrooming of digital computers since the late 1940s made the numerical methods for computerized determination of the spectral density very popular in all the applied fields.   By now, these numerical methods have irrevocably displaced all the other methods for finding the spectral density in most of the practical situations (though occasionally the analog spectrum analyzers nevertheless prove irreplaceable).

We have already indicated one of the numerical methods for estimating the values of $f(\omega)$, namely the method based on digital filtering of the sequence $x_\Delta(t) = x(t\Delta)$ (or $x(t)$, $t = 0, 1, ...,$ if the time is discrete from the beginning), squaring the filtered signal $y(t)$ and averaging of $y^2(t)$. This method, however, is none other than the computerized realization of the operations carried out by analog spectrum analyzers. At the same time it would seem more natural, once computers are used, not to imitate analog methods, but to develop different techniques specially adapted to digital computers.   Among such special digital methods, the one which was considered principal and was used most widely in practice for a long time was the method based on calculating the smoothed periodogram ordinates $\varphi_T^{(A)}(\omega)$

with the aid of equation (3.46a), which represents $\varphi_T^{(A)}(\omega)$ as a

discrete Fourier transform of the estimate $B_T^{(a)}(\tau) = a_T(\tau)B_T^{**}(\tau)$

of the correlation function $B(\tau)$. The application of (3.46a), rather than (3.43), which represents $\varphi_T^{(A)}(\omega)$ in the form of a periodogram $i_T(\omega')$ smoothed by convolution with a spectral window $A_T(\omega)$, was dictated by purely practical, but very important, considerations.

In fact, if T is large enough, the conventional methods for calculating the Fourier transform of a series of length T are extremely laborious even for modern computers.   The number of required elementary arithmetical operations (and, hence, the computing time) increased approximately as T^2 with increasing T.   On the other hand, calculation of the estimate $B_T^{**}(\tau)$ of the correlation function is relatively simple and does not require much time.   Moreover, if $a_T(\tau)$ on the right-hand side of (3.46a) is a function such that $a_T(\tau) = 0$ for $|\tau| \geq k_T$, then the values of $B_T^{**}(\tau)$ had to be determined only

for $\tau = 0, 1, ..., k_T - 1$, where $k_T \ll T$ (e.g., $k_T = 0.1T$ or $k_T = 0.05T$). When the values $B_T^{**}(\tau)$, $\tau = 0, 1 ..., k_T - 1$, are computed, the computation of $\varphi_T^{(A)}(\omega)$ requires only the finding of the discrete Fourier transform of the relatively short series $B_T^{(a)}(0), B_T^{(a)}(1), ..., B_T^{(a)}(k_T-1)$, consisting of k_T numbers. Thereby for reasonable values of the ratio k_T/T the total computing time for the smoothed periodogram $\varphi_T^{(A)}(\omega)$ was reduced (as compared with the procedure based on the use of (3.43)), by a factor of a hundred, or even hundreds, while the application of (3.43) often required hundreds of hours for estimate computation (when the observation series was rather long).[62] We emphasize that the condition $a_T(\tau) = 0$ for $|\tau| > k_T$ and the smallness of k_T in comparison with T are essential for substantial reduction of the computation time. Therefore, it is not surprising that even comparatively recently the spectral windows $A_T(\omega)$, whose Fourier transforms did not vanish outside a limited band (for instance, the Daniell window of Fig. 43(a)) were supposed to be exclusively of theoretical interest, but useless practically.

(a)

(b)
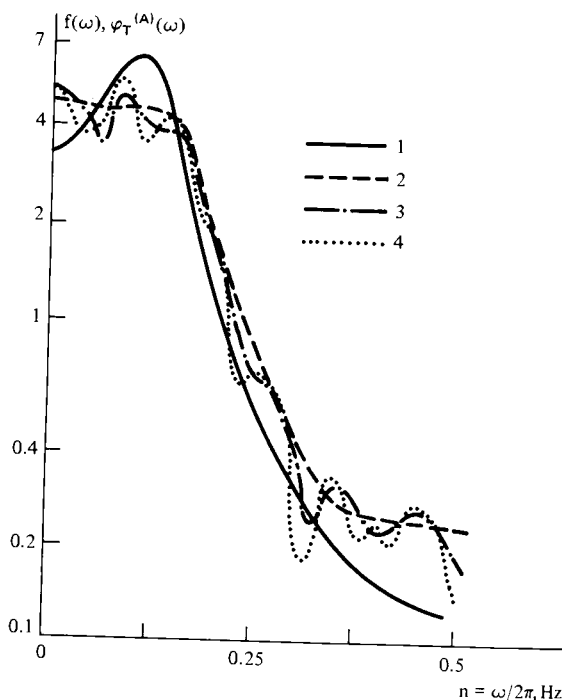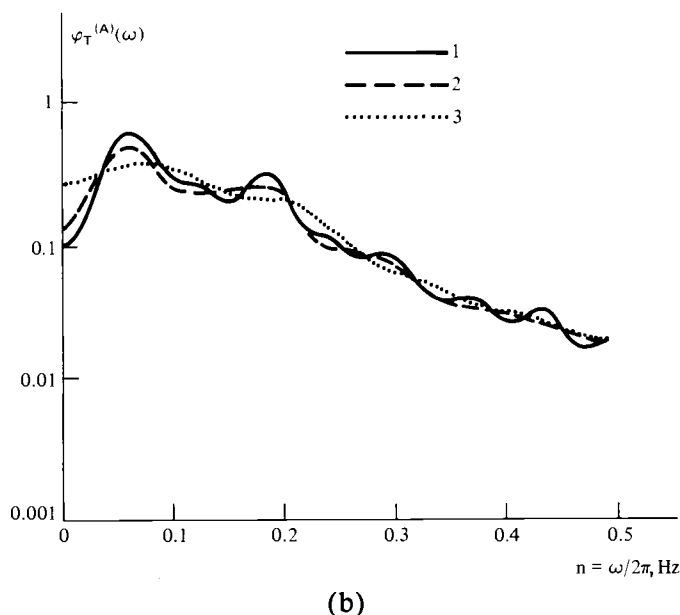
Fig. 44. (a) The true spectral density $f(\omega)$ of a simulated stationary sequence (line 1) and three estimates $\varphi_T^{(A)}(\omega)$ calculated by using the formula (3.46a) where $a_T(\tau) = a(\tau/k_T)$ is the Parzen lag window and $k_T = 16$ (line 2), or $k_T = 32$ (line 3), or $k_T = 48$ (line 4) (Jenkins and Watts, 1968). (b) The smoothed periodograms $\varphi_T^{(A)}(\omega)$ of the Beveridge series calculated by using the "Hanning" lag window at $k_T = 30$ (line 1), $k_T = 20$ (line 2), and $k_T = 10$ (line 3) (Anderson, 1970).

During the 1950s and the early 1960s the above-described method of spectral density estimation (which is often called the Blackman–Tukey or Tukey method[63]) was generally regarded as the best numerical method. One of its important advantages over the method based on the use of analog spectrum analyzers is the fact that a computer can easily perform computations at several different values of k_T, whereas the pass bands of the individual filters ("cells") of an analog analyzer (which play here the role of the parameter k_T) are always fixed beforehand (e.g., all of them are equal to 1/2 or 1/3 octave). Two specific examples of the evaluation of $\varphi_T^{(A)}(\omega)$ by (3.46a) are given in Fig. 44. Figure 44(a) refers to an artificial time series $x(t)$, $t = 1, 2, ..., 400$,

representing the realization of a stationary sequence $X(t)$ with a known spectral density $f(\omega)$ (also shown in the figure). We can see that with an increase in k_T (i.e. with a decrease in the width of the frequency band over which the periodogram is averaged) the estimate $\varphi_T^{(A)}(\omega)$ becomes an increasingly erratic function of frequency, but its overall agreement with the theoretical curve $f(\omega)$ improves. In particular, the graphs of $\varphi_T^{(A)}(\omega)$ at $k_T = 32$ and $k_T = 48$ (but not at $k_T = 16$) both reveal a spectral peak at the frequency $n \gtrsim 0.125$ Hz; however, the width and height of this peak are greatly underestimated in both cases.* Figure 44(b) exhibits the values of $\varphi_T^{(A)}(\omega)$ for the "Beveridge wheat price series" (see Fig. 6(e) on p. 29). In this case the spectral estimates $\varphi_T^{(A)}(\omega)$ have a rather smooth form (in contrast to the behavior of the periodogram $i_T(\omega)$ in Fig. 40(a), and $\varphi_T^{(A)}(\omega)$ changes only slightly on passing from $k_T = 20$ to $k_T = 30$.

The method for evaluating $\varphi_T^{(A)}(\omega)$ by (3.46a) has retained its great practical importance up to now, but (at least for large values of T) in quite a different realization, which will be explained below.

3. *Application of the fast Fourier transform algorithm: direct smoothing of the periodogram.* Cardinal changes in evaluating the practical importance of different methods for approximate determination of the spectral density took place in the late 1960s, after the so-called fast Fourier transform (FFT) algorithm had gained wide popularity. The FFT algorithm (which has several forms, i.e. is in fact a group of related algorithms) provides a very efficient numerical technique for fast and easy calculation of a Fourier transform of a finite sequence. (For brevity we shall speak here only of one, the simplest form of the algorithm.) If a sequence consists of T numbers (either complex or real), then the FFT algorithm permits one to compute the corresponding discrete Fourier transform with a number of arithmetical operations close to $2T\log_2 T$ (instead of about T^2 operations needed for conventional computation of a Fourier transform). It is easy to see that at $T =$

---

*The last circumstance can be attributed to the insufficiency of the sampling length $T = 400$, but may also be due to the poor choice of the "pseudorandom numbers" used for numerical simulation of the realization $x(t)$ (cf. Note 4 to Chap. 2 in Vol. II).

1000 the FFT method reduces the time of computation of a Fourier transform to about 1/50, and with a further growth of T this gain rapidly increases. As a result, the computation of Fourier transforms, which had been considered very cumbersome and time consuming, suddenly became very simple and fast.

The FFT method was introduced by J. Cooley and J. Tukey in 1965 in a short note which then seemed entirely new and original to almost everybody. However, later it was discovered that a similar method for Fourier transform computation had been proposed (but did not attract any attention and was forgotten almost immediately) as early as 1942, and then it was based on some results going back to the very beginning of this century. Still later it was also found that the FFT algorithm was in fact described by C.F. Gauss in a manuscript written apparently in 1805 (i.e. before the publication of the famous book by Fourier) and first published in 1866 (almost 100 years before the discovery of this method by Cooley and Tukey). Moreover, several related methods of numerical Fourier analysis were also used during the nineteenth century but were later forgotten. At present, a comprehensive description of various forms of FFT algorithm (and of some other simplified algorithms for discrete Fourier transform computations) can be found in many sources.[64]

The FFT algorithm is easily adaptable to modern computers, and its reinvention in 1965 has drastically changed the whole field of statistical spectral analysis. Using an FFT computer program the periodogram ordinates $i_T(\omega)$ can be computed directly at a set of equidistant discrete frequencies ω (e.g., at all the frequencies of the form $\omega_k = 2\pi k/T$ or $\pi k/T$, where k is an integer) even if the time series $x(t)$ is rather long (i.e. for large T). The needed computation time is in this case appreciably less than the time needed for determination of the estimate $B_T^{**}(\tau)$ at $\tau = 0, 1, ..., k_T - 1$. When the periodogram is computed we can easily evaluate $\varphi_T^{(A)}(\omega)$ by replacing the integral in the right-hand side of (3.43) with a sum over a discrete set of frequencies ω_k with weights given by an appropriate spectral window $A_T(\omega - \omega')$. In particular, a computationally very simple spectral estimate corresponds to the use of the "rectangular spectral window" proposed by Daniell in 1946 (and anticipated by Einstein in 1914): in this case $\varphi_T^{(A)}(\omega)$ is equal to the arithmetic mean of several neighboring periodogram ordinates. Thus we see that the FFT algorithm made the Daniell

window very convenient for practical applications.

If the length T of the series $x(t)$ is too large, it is reasonable to divide it into n nonoverlapping or partially overlapping pieces of smaller length $T_1 < T$ (cf. pp. 255 − 256 and Notes 40 and 57). Then we can compute the periodograms (or, what is sometimes preferable, the modified periodograms of the data multiplied by a data window; see pp. 272 − 276) for each of these pieces and use the arithmetic mean of all the periodograms at the frequency ω as the estimate of $f(\omega)$. It is also possible to evaluate the smoothed periodograms (3.43) for each of the pieces of length $T_1 < T$ and then form the arithmetic mean of these smoothed periodograms. Both methods require computing many Fourier transforms and therefore the FFT algorithm is, in these cases, of great help. By combining all the described methods it is now possible to find fully satisfactory spectral density estimates over a very wide frequency range, using up to several millions of numbers $x(t)$ if necessary.[65]

Figure 45 presents, as an example, several spectral estimates $\varphi_T^{(A)}(\omega)$ for a numerically simulated series $x(t)$, $t = 1, ..., 1024$, with a known spectral density $f(\omega)$. Some of these estimates are of the form (3.46a) and are computed via the estimates $B_T^{**}(\omega)$ without any use of the FFT algorithm, while the others are of the form (3.43) and are evaluated via the periodogram $i_T(\omega)$, which is computed with the aid of the FFT algorithm. In all cases the values of k_T were selected to obtain approximately identical overall bandwidth of the corresponding spectral window $A_T(\omega)$. We can see that the accuracy of estimates of the two types is quite comparable, but on the average the estimates obtained by direct periodogram smoothing seem to be slightly preferable to the estimates computed via $B_T^{**}(\tau)$. Note also that the times of computation of the estimates (3.43) were considerably shorter in this case than for Blackman–Tukey estimates (3.46a). Therefore, even for moderately long series $x(t)$ the utilization of (3.46a) for spectral estimate computation can now be considered justified only when the calculations are carried out with a wide application of the FFT algorithm (see below).

3a. *Computation of estimates of correlation functions by using the FFT algorithm. Modern approach to computation of* $\varphi_T^{(A)}(\omega)$

via $B_T^{**}(\omega)$. The development of the FFT algorithm greatly influenced also the computational aspects of *statistical correlation analysis* dealing with the estimation of the correlation functions $B(\tau)$. As noted above, the computation of the periodogram ordinates $i_T(\omega)$ is now often a much simpler task than direct computation of the values of $B_T^{**}(\omega)$ by equation (3.33) or (3.33a). Therefore, even when the main interest lies in estimating the correlation function $B(\tau)$, it is
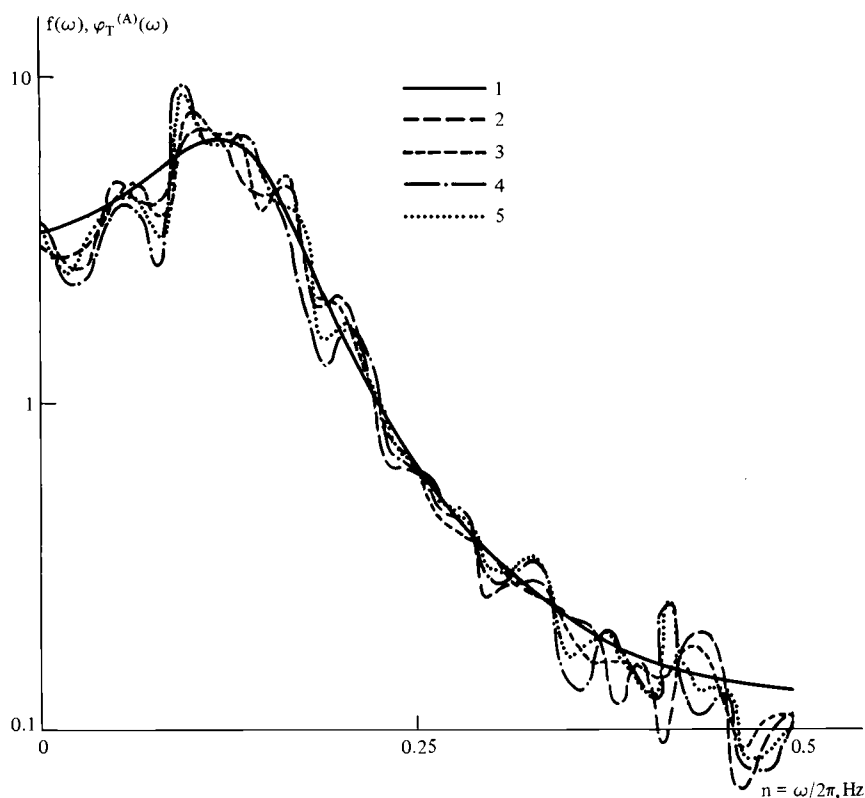
Fig. 45. The true spectral density $f(\omega)$ of a numerically simulated stationary sequence (line 1) and the estimates $\varphi_T^{(A)}(\omega)$, $T = 1024$, of this density calculated by I. A. Kozhevnikova (Moscow State University) by using FFT algorithm and parabolic (line 2) or rectangular (line 3) spectral window and also by using the formula (3.46a) with the "Hanning" lag window (line 4) or the Parzen lag window (line 5).

often expedient to calculate at first the periodogram $i_T(\omega)$ by using the fast Fourier transform and then evaluate $B_T^{**}(\tau)$ as an inverse Fourier transform of the periodogram (also with the aid of the FFT algorithm).[66]

According to the "convolution theorem" of the Fourier transform theory, the Fourier transform of a convolution of any two numerical functions is equal to the product of their Fourier transforms, i.e. it has a very simple form. (It makes no difference whether these functions are of a continuous or a discrete argument.) Therefore the computational simplicity of discrete Fourier transforms makes it expedient to resort to Fourier transforms each time one has to compute the convolution of two sufficiently long numerical series (see again Note 66). Since the right-hand side of (3.43), which determines the smoothed periodogram $\varphi_T^{(A)}(\omega)$, is a convolution of the functions $i_T(\omega)$ and $A_T(\omega)$, the last comment can be applied, in particular, to (3.43). The result is somewhat unexpected: applying at first the Fourier transform to both sides of (3.43) and then applying the inverse Fourier transform to both sides of the resulting equation (which coincides with (3.45)), we again arrive at the familiar formula (3.46a) conveying the estimate $\varphi_T^{(A)}(\omega)$ in the form of the Fourier transform of the series $a_T(\tau)B_T^{**}(\tau)$, $\tau = 0, \pm1, ..., \pm(T-1)$. Recall that, as we have said before, (3.46a) had played the main role in statistical spectral analysis before the introduction of the FFT. The reason was that direct computation of the periodogram $i_T(\omega)$ for large values of T was considered impractical at the time. Now we see that evaluation of $\varphi_T^{(A)}(\omega)$ via (3.46), i.e. as an inverse Fourier transform of the sequence $a_T(\tau)B_{TT}^*(\tau)$, $\tau = 0, \pm1, ..., \pm(T-1)$, may often be practically quite expedient even now, in the age of FFT, but only on condition that the FFT algorithm is used three times in the course of calculations: the first time to find the Fourier transform of the series $x(t)$, $t = 1, ..., T$, whose square determines the periodogram $i_T(\omega)$, the second time to evaluate the estimate $B_T^{**}(\tau)$ via the periodogram $i_T(\omega)$, and the third time to compute $\varphi_T^{(A)}(\omega)$ as an inverse Fourier transform of the series $B_T(\tau)$, $\tau = 0, \pm1, ..., \pm(T-1)$, tapered with the aid of multiplication by a lag window $a_T(\tau)$.

*The last method for spectral estimation allows also a further generalization. Instead of applying the Fourier transformation to the whole available series $x(t)$, $t = 1, ..., T$, we can first partition this series into P pieces of length T_1,

where $P \geqslant 1$ is an integer, $T_1 \leqslant T$, the pieces may overlap, and the time T is assumed to be discrete for definiteness. Then each piece of data is multiplied by the same symmetric data window $h_{T_1}(t)$, $t = 1, ..., T_1$, (where $h_{T_1}(t) = h_{T_1}(T_1 - t)$). After this the FFT is computed for all the obtained modified pieces of data so that the squared absolute values of these Fourier transforms determine P modified periodograms $i_{T_1}^{(h,k)}(\omega)$, $k = 1, ..., P$. The arithmetic mean of these periodograms

$$(3.62) \qquad f_T^{(P,T_1)}(\omega) = \frac{1}{P} \sum_{k=1}^{P} i_{T_1}^{(h,k)}(\omega)$$

is the first-stage spectral estimate (cf. Note 57).

Computing now the inverse Fourier transform of $f_T^{(P,T_1)}(\omega)$ by the FFT method, we obtain the first-stage correlation function estimate $B_T^{(P,T_1)}(\tau)$. The second-stage correlation function estimate $\tilde{B}_T(\tau)$ is then formed by multiplying $B_T^{(P,T_1)}(\tau)$ by some suitable symmetric lag window $a_{T_1}(\tau)$ (where $a_{T_1}(-\tau) = a_{T_1}(\tau)$):

$$(3.63) \qquad \tilde{B}_T(\tau) = B_T^{(P,T_1)}(\tau) a_{T_1}(\tau).$$

Finally, the second-stage spectral estimate $\tilde{f}_T(\omega)$ is computed as the FFT of $\tilde{B}_T(\tau)$. An advantage of this method is that when the estimate $B_T^{(P,T_1)}(\tau)$ has already been computed, the computation of second-stage estimates $\tilde{B}_T(\tau)$ and $\tilde{f}_T(\omega)$ could be repeated many times with different lag windows $a_{T_1}(\tau)$ (cf. Fig. 44).

Of course, when $P = 1$, $T_1 = T$, and the rectangular data window is used (i.e. $h_T(t) = 1$ for $t = 1, ..., T$), the method becomes the usual Blackman–Tukey method in its modern version based on the FFT. If the rectangular lag window is selected in (3.63) (i.e. $a_{T_1}(t) = 1$ for $t = $

$0, \pm 1, \ldots, \pm (T_1 - 1)$, the computations of $\hat{B}_T(\tau)$ and $\hat{f}_T(\tau)$ are unnecessary and the method becomes the Welch method (see pp. 275-276 and Note 57). In the general case it is easy to show that the mean value of the final estimate $\hat{f}_T(\omega)$ is very close to the convolution of a true spectral density $f(\omega)$ with a kernel ("the effective spectral window") $A_e(\omega)$ which is itself the convolution of a Fourier transform $A_{T_1}(\omega)$ of the lag window $a_{T_1}(\tau)$ and

a kernel $K_{T_1}^{(h)}(\omega)$ (see (3.58)) which corresponds to the

data window $h_{T_1}(t)$ (cf. (3.61); see also Note 67).

Therefore, we can select one of the two windows $a_{T_1}(t)$

and $h_{T_1}(t)$ quite arbitrarily and still obtain the desired

form of the effective spectral window $A_e(\omega)$ by appropriate selection of the second window. In particular, the computation time can be significantly reduced if we utilize the rectangular data window $h_{T_1}(t) = 1$ for $t = 1, \ldots, T_1$, since

then we do not have to multiply each piece of data by a special numerical function. Note, however, that if such a selection of $h_{T_1}(t)$ is combined with a reasonable selection of

the effective spectral window $A_e(\omega)$ having a Fourier transform $a_e(\tau)$, which is a symmetric and monotonically decreasing (for $\tau \geqslant 0$) function of τ, then this combination may necessitate the utilization of an unusual lag window $a_{T_1}(\tau)$ which is not a monotonically decreasing function of

the lag τ for $\tau \geqslant 0$.[67] *

4. *Parametric methods of spectral density estimation.* Recall the example of spectral density determination for fading of radio signals, whose results are presented in Fig. 23, p. 124. In this example, dating back to the 1940s, none of the above-described methods for estimating the values of $f(\omega)$ was used. Instead, use was made of the fact that the corresponding correlation function estimate $B_{TN}^*(\tau)$ resembled, at relatively small values of τ, a damped cosinusoid, i.e., a function $C \exp(- \alpha |\tau|) \cos \omega_0 \tau$. The investigators simply began with selecting the values of the parameters α and ω_0 at which

the function $\exp(-\alpha|\tau|)\cos\omega_0\tau$ fits fairly well, at small $|\tau|$, the empirical normalized correlation function $R^*_{TN}(\tau)$ = $B^*_{TN}(\tau)/B^*_{TN}(0)$; see Fig. 22.   After this the Fourier transform of the selected function $\exp(-\alpha|\tau|)\cos\omega_0\tau$ was taken for the unknown normalized spectral density $f(\omega)/B(0)$.

A similar example is depicted in Fig. 26, which shows the approximation of the estimated normalized correlation function $R^*_{TN}(\tau)$ for the averaged atmospheric pressure by a function of the form (2.116). This approximation indicates that in this case the spectral density $f(\omega)$ is apparently close to a function of the form (2.118) with parameters determined by (2.119).

The approximate determination of the spectral density $f(\omega)$ by fitting the estimated correlation function to some simple analytical formula, whose Fourier transform can be explicitly evaluated, was utilized rather often in the early practical applications of the theory of stationary random functions. The shortcomings of this method strike the eye at once: it seems highly unwieldy, rather subjective, and hardly realizable on a digital computer.   On the other hand, the spectral densities thus obtained are usually reasonably smooth functions, whose graphs do not contain numerous unnatural alternating "peaks" and "troughs" (see again Fig. 23).   It is perhaps still more important that the spectral densities obtained by this method are specified by simple analytical equations (usually representing $f(\omega)$ as a rational function of ω or $\exp(i\omega)$). The last circumstance is often very advantageous since it greatly facilitates the utilization of the determined spectral density in those, rather frequent, cases where it is necessary to solve some linear approximation problem relating to the random function $X(t)$ (e.g., the extrapolation problem on the best linear prediction of the future of $X(t)$ from its past or the filtering problem on such "optimum smoothing" of the observed values of $X(t)$ which eliminates the influence of the noise distorting the observations; see Notes 31 and 70 to Chap. 2 in Vol. II).

There exist some analog instruments of a special kind (so-called "parametric correlometers"; see Note 27 to this chapter) which can be used for finding an approximation of the estimated values of a correlation function by a simple analytical equation. Of much greater importance, however, is the development of numerical methods for the realization of such an approximation (for the case of discrete time) on digital computers. Numerous investigations have been carried out along these lines in recent years. They showed that, contrary to the

first impression, parametric estimation of the spectral density by a numerical process is, in fact, feasible and even has some important advantages over the more conventional non-parametric spectral estimation described above. At present the parametric methods of spectral analysis compete quite successfully with non-parametric periodogram smoothing in many fields dealing with fluctuating series of observations. Therefore it is not surprising that the literature devoted to the applications of parametric spectral estimation has been expanding rapidly in recent years.[68]

Of course, the accuracy of a parametric spectral estimate (just as that of any other spectral estimate) depends greatly on the form of the true spectral density $f(\omega)$. Moreover, a considerable role in parametric estimation is also played by the proper choice of the class of function $f(\omega_1; c_1, ..., c_M)$ (depending on M unknown parameters $c_1, ..., c_M$) from among which one then chooses the function approximating $f(\omega)$. Subject to this choice the whole set of parametric spectral estimates can be divided into several distinct subsets.

4a. *Autoregressive, or maximum entropy, spectral estimation.* The simplest (and most widely used in practice) class of functions $f(\omega; c_1, ..., c_M)$ selected to approximate unknown spectral densities is the class of strictly positive functions of the form

$$f(\omega) = \frac{c^2}{2\pi\left|1 + a_1 e^{-i\omega} + ... + a_m e^{-im\omega}\right|^2}$$

(3.64)

$$= \frac{1}{\left|c_0 + c_1 e^{-i\omega} + ... + c_m e^{-im\omega}\right|^2}$$

where $a_1, ..., a_m, c^2$ (or, what is the same, $c_0, c_1, ..., c_m$) are unknown parameters and hence $M = m + 1$. The functions (3.64) coincide with the spectral densities of the autoregressive stationary sequences of finite orders (see p. 206). Therefore the estimates of the spectral density $f(\omega)$ obtained by approximating it with a function of the form (3.64) are often called *autoregressive spectral estimates* (or *AR spectral estimates*).[69]

There is also another name often applied to the same estimates, which reflects quite a different approach to parametric spectral estimates of the form (3.64). Suppose that

we know the first $m + 1$ values of the correlation function of a stationary sequence, $B(0), B(1), ..., B(m)$ or, more generally, that we have some good (i.e. reliable enough) estimates of them, $B^*(0), B^*(1), ..., B^*(m)$. Then the corresponding spectral density $f(\omega)$ cannot be quite arbitrary, but must satisfy equations

$$(3.65) \qquad \int_{-\pi}^{\pi} e^{i\omega\tau} f(\omega)d\omega = B^*(\tau), \qquad \tau = 0, 1, ..., m,$$

either exactly (if $B^*(\tau)$ are the true values of the correlation function) or, at least, with a high accuracy. Seeking a good estimate of $f(\omega)$, it is reasonable to restrict oneself to the class of functions $f^*(\omega)$ for which

$$(3.65a) \qquad \int_{-\pi}^{\pi} e^{i\omega\tau} f^*(\omega)d\omega = B^*(\tau), \qquad \tau = 0, 1, ..., m.$$

Now the question arises how to choose a spectral density estimate $f^*(\omega)$ if nothing is known about the true spectral density except the constraints (3.65).

It was suggested in the late 1960s that in such a situation a proper spectral density estimate would be given by a function $f^*(\omega)$ that maximizes the so-called "entropy"

$$(3.66) \qquad H = \int_{-\pi}^{\pi} \ln f^*(\omega)d\omega$$

subject to conditions (3.65a).* It can be shown that in the case of Gaussian sequences $X(t)$ the maximization of the entropy means that the spectral density $f^*(\omega)$ corresponds to the most random (i.e. most unpredictable) sequence the correlation function $B(\tau)$ of which is consistent with a set of $m + 1$ known values.[70] However, conditions (3.65a) and $H =$ maximum make sense for any stationary sequence $X(t)$, and it is natural to assume that they always specify a reasonable spectral density estimate.

Maximization of the integral (3.66) subject to conditions (3.65a) is a rather simple variational problem, and it is easy to show that its solution is of the form (3.64), where the parameters $c_0, c_1, ..., c_m$ (or $c^2, a_1, ..., a_m$) must be determined

---

*As usual, ln in the right-hand side of (3.66) denotes a natural (base e) logarithm. It is clear, however, that in fact any base of logarithms can be used here, since a base change only implies multiplication of the right-hand side of (3.66) by a constant factor.

from the conditions (3.65a).[71] Hence we again arrive at autoregressive spectral estimates which are therefore often called *maximum entropy spectral estimates*.[72] In practice it is reasonable to take the conventional estimates $B_T^{**}(\tau)$ for $B_T^{*}(\tau)$ in (3.65a). Thus m can be determined (at least theoretically) from the condition that all the estimates $B_T^{**}(0)$, $B_T^{**}(1)$, ..., $B_T^{**}(m)$ are reliable enough (discussion of the practical methods for determining m will be given below).

Let us suppose, at first, that the order of the autoregression m has already been chosen and that one only has to determine the values of the parameters a_1, ..., a_m, c^2 (or c_0, c_1, ..., c_m) which agree best with the given observations $x(t)$, $t = 1$, ..., T.   Then, to find the estimates for a_1, ..., a_m, we can use, e.g., the fact that for an mth-order autoregression the correlation sequence $B(\tau)$, $\tau = 0, \pm 1, \pm 2$, ..., satisfies the Yule–Walker equations (2.16) - (2.16a). On the basis of this we can obtain appropriate values of a_1, ..., a_m by solving a linear system of m first equations (2.16) with $k = 1, 2, ..., m$, where the unknown coefficients $B(k)$, ..., $B(k - m)$ in the left-hand sides are replaced by the corresponding estimates $B_T^{**}(k)$, ..., $B_T^{**}(k - m)$.   (Recall that the estimates $B_T^{**}(\tau)$ at large values of T are best computed by applying the FFT algorithm, i.e. by first computing the periodogram $i_T(\omega)$.). The solutions $a_1 = a_1^{*}$, ..., $a_m = a_m^{*}$ of the system of m first equations (2.16) depend on $B_T^{**}(0)$, $B_T^{**}(1)$, ..., $B_T^{**}(m)$ (i.e. on the observations $x(1)$, ..., $x(T)$).   They determine reasonable estimates of a_1, ..., a_m, the so-called *Yule–Walker estimates*, which are consistent and in general have, at large values of T, good statistical properties. Substituting now the estimates a_j^{*}, $j = 1$, ..., m, for a_j in (2.16a), where $B(0)$, ..., $B(m)$ are again replaced by $B_T^{**}(0)$, ..., $B_T^{**}(m)$, we find the consistent estimate $(c^2)^{*}$ of the parameter c^2 in (3.64).[73]

The conventional method of solving the linear system of first m equations (2.16), where the coefficients $B(\tau), \tau = 0, 1, ..., m - 1$, are replaced by $B_T^{**}(\tau)$, requires inversion of the $m \times m$ matrix $\mathbf{B}_T^{**} = \|B_T^{**}(k - j)\|$, $k = 1, ..., m$, $j = 1, ..., m$, which is, at large values of m, a rather tedious and time-consuming operation. Therefore a special numerical technique, the so-called *Levinson*, or *Levinson–Durbin algorithm*, was developed to speed up and simplify computations.[74] This new technique is recursive: it permits one to determine successively the parameters $a_1^{(1)}$, $c_{(1)}^2$ of

a first-order autoregressive model fitting the given data $x(1)$, ..., $x(T)$, the parameters $a_1^{(2)}$, $a_2^{(2)}$, $c_{(2)}^2$ of an appropriate second-order autoregressive model, the parameters $a_1^{(3)}$, $a_2^{(3)}$, $a_3^{(3)}$, $c_{(3)}^2$ of a third-order model, and so on till the parameters

$a_1^{(m)} = a_1^*$, ..., $a_m^{(m)} = a_m^*$, $c_{(m)}^2 = (c^2)^*$ are determined which coincide with the sought-for estimates of a_1, ..., a_m, c^2. In the extensive literature on autoregressive (maximum entropy) spectral estimation one can also find several other effective numerical methods for quickly computing appropriate estimates of a_1, ..., a_m, c^2 from the observations $x(1)$, ..., $x(T)$. An important example is *Burg's method*, which is also recursive (in the same sense as above) and, in addition, does not require preliminary computation of the correlation function estimates $B_T^{**}(\tau)$ for any τ.[75]

Replacing, in the right-hand side of (3.64), the unknown parameters a_1, ..., a_m, c^2 (or c_0, c_1, ..., c_m) by some of their estimates, we obtain the *autoregressive* (or *maximum entropy*) *estimate* of $f(\omega)$, which from now on will be denoted by $\varphi_T^{AR}(\omega)$. As for the order of autoregression m (on which value the estimate $\varphi_T^{AR}(\omega)$ essentially depends), it plays here a role somewhat similar to that of the scale parameter k_T in the case of smoothed periodogram estimates $\varphi_T^{(A)}(\omega)$ corresponding to the windows of scale parameter form (3.51) and (3.52). Based on this similarity we may try to choose the value of m by the method recommended above for choosing the value of k_T: We select m to be very small at first (say, equal to 1) and then increase its values successively until we convince ourselves that any further increase in m changes the overall form of $\varphi_T^{AR}(\omega)$ only slightly. However, there are also several other, more objective, methods for choosing the autoregression order m. One can, in particular, plot the estimated values $c_{(j)}^2$ of the parameter c^2 for a number of successively increasing values of order j. (Recall that the values $c_{(j)}^2$ for $j = 1, 2, ..., m$ are computed anyway whenever a recursive algorithm is used for estimating autoregressive parameters.) As a rule, the values of $c_{(j)}^2$ will decrease with j at first but will "level off" beginning with some j. Then a reasonable estimate of m is the first value of j for which the "levelling off" occurs and $c_j^{(2)}$ shows no signs of further substantial decrease.[76]

*There are also some "automatic" criteria for discerning

the value of m, which lead to a unique selection of the autoregression order and hence to a unique form of the autoregressive spectral estimate $\varphi_T^{AR}(\omega)$ implied by the data. Most of these criteria are also based on examination of the sequence $c_{(1)}^2$, $c_{(2)}^2$, ..., $c_{(m)}^2$, ... of the estimates for the parameter c^2 in (3.64) corresponding to a number of successively increasing values of the autoregression order. In particular, Akaike's *final prediction error criterion* is defined by

$(3.67)^*$ $FPE(m) = \dfrac{T + m}{T - m}\, c_{(m)}^2.$

$FPE(m)$ always has a minimum at some definite value $m = m_{FPE}$, which is the optimal value for the order of autoregression according to the final prediction error criterion. Thus, to find m_{FPE} one must only compute $FPE(m)$ for a number of successive values of m beginning with $m = 1$.[77]

Another criterion, also suggested first by Akaike, is usually called *Akaike's information criterion* and is defined by

(3.68) $AIC(m) = \ln(c_{(m)}^2) + \dfrac{2m}{T}.$

$AIC(m)$ is a function of m that also takes a minimal value at some point $m = m_{AIC}$; the value m_{AIC} is the optimal selection for m according to Akaike's information criterion. At large values of T the two criteria (3.67) and (3.68) are practically equivalent since it is easy to see that asymptotically (as $T \to \infty$)

(3.69) $\ln[FPE(m)] = AIC(m).$ [78]

---

*Whenever the mean value of the sequence $X(t)$ is unknown and the sample mean m_T^* is subtracted from all the observations $x(t)$ to obtain a realization of a sequence $X(t)$ with mean zero, the definition of FPE(m) must be slightly changed: in this case

$(3.67a)$ $FPE(m) = \dfrac{T + m + 1}{T - m - 1}\, c_{(m)}^2.$

A similar modification must also be applied to some other criteria in the case where the mean value of $X(t)$ is unknown (e.g., to CAT; see (3.70)).

The more complicated *Parzen criterion for autoregressive transfer function* (CAT) in its simplest form is defined by

$$(3.70) \quad CAT(m) = \frac{1}{T} \sum_{j=1}^{m} \frac{T-j}{Tc_{(j)}^2} - \frac{T-m}{Tc_{(m)}^2}.$$

The optimal value $m = m_{CAT}$ of the order according to the stated criterion is that m for which $CAT(m)$ attains its minimum value.[79]

Some other criteria for selecting the most suitable autoregression order m and many examples of the applications of various criteria to some observations (both real and simulated, i.e. arising from a random sequence with a known spectral density) can be found in the vast literature on the subject. In particular, the study of the applications of the order-selecting criteria to simulated time series generated by genuine autoregressive sequences of the fixed order m_0 shows that all three criteria very often result in correct determination of the order, but nevertheless there is also a finite probability of error which does not tend to zero as the realization length T tends to infinity. In other words, the order estimates m^* determined by all three criteria are not consistent estimates of m_0, i.e. they do not tend to m_0 as $T \to \infty$ (they tend to overestimate m_0, as a rule). This fact is distressing from the mathematician's point of view, and therefore several (successful) attempts were made to modify the criteria in such a way that the new order estimates would be consistent. However, the inconsistency of the order estimates is, in fact, not such an important defect for applications usually dealing with the series of observations which are not generated by a genuine autoregressive sequence. Most of the results obtained through the application of order-selection criteria to various specific problems of autoregressive spectral analysis show that the orders determined by different criteria do not, as a rule, differ significantly* and the obtained spectral estimates $\varphi_T^{AR}(\omega)$ are usually reasonable and quite useful. Only rather short observation series (i.e. small T) are exceptional

---

*At least when the length T of series x(t), t = 1, ..., T, is long enough. For short series most of the authors recommend the use of the AIC.

in this respect: here, none of the criteria yields satisfactory results, and therefore it was even suggested that some number between $T/2$ and $T/3$ be taken as the autoregression order m in such instances.[80*]

It was proved in Sec. 18 that a very important property of smoothed periodogram spectral estimates $\varphi_T^{(A)}(\omega)$ is that such an estimate can always be made consistent by an appropriate selection of the corresponding spectral window $A_T(\omega)$ (or, what is the same, the lag window $a_T(\tau)$), where T is an unboundedly increasing variable. The same is true, under very wide regularity condition, for autoregressive spectral estimates $\varphi_T^{AR}(\omega)$. Of course, for an estimate $\varphi_T^{AR}(\omega)$ the form of the dependence of the order $m = m_T$ on T must be strictly restricted to make $\varphi_T^{AR}(\omega)$ consistent, while for an estimate $\varphi_T^{(A)}(\omega)$ one must select the form of the dependence on T of the spectral window $A_T(\omega)$ (or of the scale parameter k_T).

Let us assume that $X(t)$ is a Gaussian stationary sequence having an everywhere positive and continuous spectral density $f(\omega)$. (The assumption that $X(t)$ is Gaussian can, in fact, be also considerably relaxed; see Note 81.) It is easy to prove then that the spectral density $f(\omega)$ can always be represented as the limit of a sequence of functions of the form (3.64). Let $\varphi_{T,\,m_T}^{AR}(\omega)$ be an autoregressive spectral estimate of the spectral density $f(\omega)$ computed from observational data $x(t)$, $t = 0, ..., T$ and having the order $m = m_T$, and let $\Phi_{T,\,m_T}^{AR}(\omega)$ be the corresponding estimator (which differs from $\varphi_{T,\,m_T}^{AR}(\omega)$ in the replacement of the non-random observations $x(t)$ by the random variables $X(t)$). Then it can be proved that *if $m_T \to \infty$ but $m_T/T \to 0$ as $T \to \infty$* (and if also some additional mild restrictions are valid which are imposed on the rate of increase of m_T with T), *the following relation holds*:

$$(3.71) \quad \lim_{T\to\infty} \frac{T}{m_T} \left\langle [\Phi_{T,m_T}^{AR}(\omega) - f(\omega)]^2 \right\rangle = \begin{cases} 2f^2(\omega) & \text{for } \omega \neq 0,\ \pm\pi\,, \\ 4f^2(\omega) & \text{for } \omega = 0 \text{ or } \omega = \pm\pi. \end{cases}$$

We see that at large values of T the mean square error of the estimator $\Phi_{T,\,m_T}^{AR}(\omega)$ is of the order of m_T/T. Therefore, since

$m_T/T \to 0$ as $T \to \infty$, the estimator $\Phi_{T,\,m_T}^{AR}(\omega)$ (and also the estimate

$\varphi_{T, m_T}^{AR}(\omega))$ is consistent.[81] Note that according to (3.53) the mean square error of the smoothed periodogram estimator $\Phi_T^{(A)}(\omega)$ is also proportional to $f^2(\omega)$ and is of the order of k_T/T if the spectral window is of scale parameter form, while the estimator bias (its "systematic error") $\delta\{\Phi_T^{(A)}(\omega)\}$ is smaller than, or of the same order as, the standard deviation of the estimator (its "random error") $\sigma\{\Phi_T^{(A)}(\omega)\}$. (The approximate equality of $\delta\{\Phi_T^{(A)}(\omega)\}$ and $\sigma\{\Phi_T^{(A)}(\omega)\}$ holds, in particular, when k_T is selected to minimize the mean square error of $\Phi_T^{(A)}(\omega)$; cf. pp. 265-266.) Thus, the asymptotic expressions for mean square errors of the estimators $\Phi_T^{(A)}(\omega)$ and $\Phi_T^{AR}(\omega)$ are often quite similar, but with an autoregressive estimator the autoregression order $m = m_T$ plays the role of the scale parameter k_T.

The theoretical results on the mean square errors for two types of spectral estimators are asymptotic, i.e. are valid only for long enough series $x(t)$, $t = 1, ..., T$ (i.e. for large T). Moreover, the theory gives no indications as to how long the series must be to justify the utilization of the asymptotic equations. Therefore a number of authors compared directly the traditional (smoothed periodogram) and the autoregressive spectral estimates computed for the same series $x(t)$, which can be either natural or artificially simulated to have a given spectral density. (Note, in addition, that autoregressive estimates differing in the order m or in the method used for coefficient estimation are also compared in some works.)[82] These comparisons and, more generally, all the experience in applying autoregressive spectral estimates to various series $x(t)$, $t = 1, 2, ...,$ show that, as a rule, the accuracies (the biases and standard deviations) of an intelligently selected smoothed periodogram $\varphi_T^{(A)}(\omega)$ and autoregressive estimate $\varphi_T^{AR}(\omega)$ are, for long and moderately long series $x(t)$, of the same order. One typical example of comparison of the estimates $\varphi_T^{AR}(\omega)$ and $\varphi_T^{(A)}(\omega)$, where $T = 648$, for a natural observation series is shown in Fig. 46. The nonparametric estimate $\varphi_T^{(A)}(\omega)$ in Fig. 46 was computed using Parzen's window with a reasonably selected value of the scale parameter k_T. In this example (and in some others, too) the curve depicting $\varphi_T^{AR}(\omega)$ is appreciably smoother than the other curve. Based on graphs of this type, some authors came to the conclusion that the autoregressive estimates represent the overall form of the true spectral density better, while the smoothed periodograms are usually more erratic and contain more spurious details. However, in fact, results of the
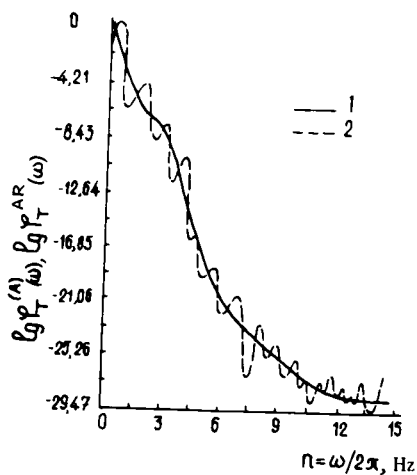
Fig. 46. Estimates of the spectral density for the electroencephalogram from a new-born baby (Jones, 1974). Line 1 shows the autoregressive estimate $\varphi_T^{AR}(\omega)$, $T = 648$, where the autoregression order $m = 6$ was determined by Akaike's information criterion. Line 2 represents the Blackman–Tukey estimate $\varphi_T^{(A)}(\omega)$ for $T = 648$ calculated by using Parzen's lag window.

comparison between estimates $\varphi_T^{AR}(\omega)$ and $\varphi_T^{(A)}(\omega)$ depend strongly on the form of the true spectral density $f(\omega)$. It can be shown, e.g., that for the moving average stationary sequence of the form $X(t) = E(t) + 0.95E(t - 1)$ (cf. (2.4)) the spectral density can be estimated quite satisfactorily by a smoothed periodogram $\varphi_T^{(A)}(\omega)$, but the autoregressive estimate $\varphi_T^{AR}(\omega)$ is here, for large values of T, rather erratic and does not represent accurately enough the true form of the spectral density.[83] It is clear, therefore, that any completely general statements on the comparative merits and demerits of autoregressive and smoothed periodogram spectral estimates must be treated with great caution.

Note now that from the point of view of computational efficiency the autoregressive estimates $\varphi_T^{AR}(\omega)$ are, at any rate, not inferior to the non-parametric estimates $\varphi_T^{(A)}(\omega)$. (In fact, the estimates $\varphi_T^{AR}(\omega)$ can often be computed even slightly faster than the traditional estimates $\varphi_T^{(A)}(\omega)$.) There are also some specific aspects of spectral estimation which demonstrate the advantages of autoregressive spectral analysis in comparison

with the more traditional technique based on periodogram smoothing. First of all, the autoregressive estimates $\varphi_T^{AR}(\omega)$ are usually rather smooth and have quite a reasonable form (though, of course, they are inaccurate and unreliable) even in rather short observation series $x(t)$, while the nonparametric estimates are, in fact, very erratic and hence fully inapplicable at low values of series length T. Therefore it is not surprising that the autoregressive spectral estimates are widely used at present in many applied fields where only short records or observation series are available.[84] Besides, the autoregressive spectral estimates $\varphi_T^{AR}(\omega)$ often have greater resolvability than the conventional smoothed periodograms $\varphi_T^{(A)}(\omega)$ with the same value of T, i.e. the estimates $\varphi_T^{AR}(\omega)$ distinguish more effectively between two peaks at close frequencies.[85] These properties of autoregressive estimates naturally explain the great popularity of the estimates $\varphi_T^{AR}(\omega)$ among the workers in applied fields.

4b. *Autoregressive-moving average (ARMA) spectral estimation.* Of course, the class of functions of the form (3.64) does not exhaust all the classes of nonnegative functions $f(\omega, c_1, ..., c_m)$ theoretically suitable for approximating the spectral densities $f(\omega)$. However, only a few such classes are mentioned in the entire vast literature on parametric methods of spectral analysis,[86] and only one class of functions not of the form (3.64) is really used rather widely in spectral estimation. This other class is a more general class of functions rational in $\exp(i\omega)$, i.e. having the form

$$\text{(3.72)} \quad \begin{aligned} f(\omega) &= \frac{c^2|1 + d_1 e^{-i\omega} + ... + d_n e^{-in\omega}|^2}{2\pi|1 + a_1 e^{-i\omega} + ... + a_m e^{-im\omega}|^2} \\[2mm] &= \frac{|b_0 + b_1 e^{i\omega} + ... + b_n e^{-in\omega}|^2}{2\pi|1 + a_1 e^{-i\omega} + ... + a_m e^{-im\omega}|^2} \end{aligned}$$

where $c^2, d_1, ..., d_n, a_1, ..., a_m$ (or, what is the same, $b_0, b_1, ..., b_n, a_1, ..., a_m$) are the unknown parameters (and hence $M = n + m + 1$). The functions (3.72) clearly do not differ from the spectral densities (2.273) of general autoregressive-moving average sequences $X(t)$. Therefore, the parametric estimates of spectral densities obtained by using the approximation of an unknown function $f(\omega)$ by a function of the form (3.72) are called *autoregressive-moving average spectral estimates* (or *ARMA spectral estimates*).[87]

The spectral densities (2.266) of moving average sequences make up a special subclass of functions (3.72) characterized by condition $a_1, = \ldots = a_m = 0$ (i.e. $m = 0$). This subclass is similar in some respects to the subclass of autoregressive spectral densities (3.64), and the functions of the form (2.266) can also be used to approximate the unknown spectral density $f(\omega)$. However, the estimation of the moving average parameters $b_0, b_1, \ldots, b_n$ from the data $x(1), \ldots, x(T)$ is a much more difficult task than the estimation of the autoregressive parameters $a_1, \ldots, a_m$ (cf. p. 301 below). For this reason, the *moving average spectral estimates* (corresponding to $m = 0$ in (3.72)) are rarely used within the framework of parametric spectral analysis.[88]

The class of all rational functions of the form (3.72) is, of course, much more extensive than the class of functions of the form (3.64). Therefore, if we are to limit the number of adjustable parameters by a given, not too large, value M, we have reason to expect that a better approximation to the true spectrum will be obtained when general models of the form (3.72), and not more specialized autoregressive models, are used.

Let us suppose at first that the autoregression and moving average orders m and n have already been chosen and one must only estimate the parameters $a_1, \ldots, a_m, b_0, \ldots, b_n$ from the observations $x(1), \ldots, x(T)$. This estimation problem is covered by a vast literature,[89] only a small part of which will be used in the exposition below. To determine the approximate values of the autoregression parameters $a_1, \ldots, a_m$ one can use the fact that the generalized Yule-Walker equations (2.18) must be valid if the spectral density $f(\omega)$ is of the form (3.72). Consider first m of these equations (with $k = n + 1, n + 2, \ldots, n + m$) and replace in them the unknown coefficients $B(k), \ldots, B(k - m)$ by their estimates $B_T^{**}(k), \ldots, B_T^{**}(k - m)$. Then we obtain a system of m linear equations with m unknowns $a_1, \ldots, a_m$, and it is easy to show that its solutions $a_1 = a_1^*, \ldots, a_m = a_m^*$ are consistent estimates of the parameters $a_1, \ldots, a_m$.[90] However, the estimates of moving average parameters $b_0, \ldots, b_n$ cannot be determined so easily. One possible method for finding the estimates of these parameters is based on consideration of the so-called *residual sequence*

$$(3.73) \qquad Y(t) = \sum_{j=0}^{m} a_j X(t - j), \qquad a_0 = 1.$$

It is easy to verify that if the spectral density of $X(t)$ is

given by (3.72) then the residual sequence $Y(t)$ will have a spectral density of the form (2.266). It follows that the correlation function $B_{YY}(\tau)$ of the sequence $Y(t)$ must have the form (2.5). On the other hand, (3.73) implies that the correlation function $B_{YY}(\tau)$ can also be written as

$$(3.74) \qquad B_{YY}(\tau) = \sum_{j=0}^{m} \sum_{k=0}^{m} a_j a_k B(\tau - j + k).$$

Therefore, if consistent estimates a_j^* and $B^{**}(\tau)$ for the coefficients a_j and the correlation function $B(\tau)$ of $X(t)$ have been already determined, the consistent estimates $b_0^*, ..., b_n^*$ of the parameters $b_0, ..., b_n$ can be found as solutions of the following system on $n + 1$ nonlinear equations

$$(3.75) \qquad \sum_{k=0}^{n-\tau} b_{k+\tau}^* b_k^* = \sum_{j=0}^{m} \sum_{k=0}^{m} a_j^* a_k^* B_T^{**}(\tau - j + k), \quad \tau = 0,1,...,n.$$

(In the particular case of a moving average sequence $X(t)$, where $m = 0$, these equations take the form: $\sum_{k=0}^{n-\tau} b_{k+\tau}^* b_k^* = B_T^{**}(\tau)$, $\tau = 0, ..., n$. However, even in this case we arrive at a rather complicated nonlinear system.)

A special numerical technique was developed for computer solution of the system (3.75).[91] Note, however, that if only the estimate of the spectral density $f(\omega)$ is of interest, it is unnecessary to compute the estimates of the parameters $b_1, ..., b_n$. In fact, we have already noted that the spectral density of the residual sequence (3.73) coincides with the numerator of the right-hand side of (3.72) divided by 2π. Therefore, equation (3.72) can also be rewritten as

$$(3.72a) \qquad f(\omega) = \frac{B_{YY}(0) + 2 \sum_{\tau=1}^{n} B_{YY}(\tau)\cos\omega\tau}{|1 + a_1 e^{-i\omega} + ... + a_m e^{-im\omega}|^2},$$

where $B_{YY}(\tau) = \langle Y(t + \tau)Y(\tau) \rangle$. This implies that if $a_1^*, ..., a_m^*$ are consistent estimates of $a_1, ..., a_m$ and

$$(3.76) \qquad B_{YY}^*(\tau) = \sum_{j=0}^{m} \sum_{k=0}^{m} a_j^* a_k^* B_T^{**}(\tau - j + k),$$

then

$$(3.77) \qquad \varphi_T^{ARMA} (\omega) = \frac{B_{YY}^*(0) + 2 \sum_{\tau=1}^{n} B_{YY}^*(\tau)\cos\omega\tau}{\left| 1 + a_1^* e^{-i\omega} + ... + a_m^* e^{-im\omega} \right|^2}$$

is a natural estimate of $f(\omega)$.[92]

Replacing the parameters a_1, ..., a_m, b_0, b_1, ..., b_n in the right-hand side of (3.72) or the parameters a_1, ..., a_m, $B_{YY}(0)$, $B_{YY}(1)$, ..., $B_{YY}(n)$ in the right-hand side of (3.72a) by some proper estimates of these parameters, we obtain an autoregressive-moving average estimate (or ARMA estimate) $\varphi_T^{ARMA}(\omega)$ of the unknown spectral density $f(\omega)$. The estimate (3.77) gives one particular example of such an estimate. However, if a high accuracy is needed, the simplest consistent estimates a_1^*, ..., a_m^*, $B_{YY}^*(0)$, ..., $B_{YY}^*(n)$ (or a_1^*, ..., a_m^*, b_0^*, b_1^*, ..., b_n^*) must be used only as preliminary (first-stage) estimates of the unknown parameters which are then utilized to construct more accurate second-stage parameter estimates.[93]

It was assumed above that the autoregressive and moving average orders m and n have been given beforehand. However, in practice the values of these two orders must always be determined somehow before we proceed to the parameter estimation and computation of the estimate $\varphi_T^{ARMA}(\omega)$. Almost the entire discussion on pp. 293-296 of the problem on choosing the autoregression order m can be repeated, with proper modifications, in connection with the choice of the two orders m and n for ARMA spectral estimates. Here also a subjective selection of the "best looking" estimate $\varphi_{T, m, n}^{ARMA}(\omega)$ (of the orders m and n) is advisable when such estimates are computed for a number of pairs (m,n); moreover, some more objective procedures of order selection are also possible.

*In particular, some authors recommend computing the estimates $c_{(m,n)}^2$ of the parameter c^2 in the right-hand side of (3.72) for a number of ARMA models of various orders m and n and then selecting the pair (m,n) corresponding to the point of the plane for which the "levelling off" of $c_{(m,n)}^2$ occurs, i.e. $c_{(m,n)}^2$ cease to decrease substantially with a further increase of the numbers m and n.[94] Also, *Akaike's information criterion* (3.68) can be generalized to ARMA spectral estimates; then it can be defined as

$$(3.78) \quad AIC(m,n) = \ln(c^2_{(m,n)}) + \frac{2(m + n)}{T} .$$

To apply this criterion we must compute the values $AIC(m,n)$ for a number of pairs (m,n) and then choose the orders m and n that correspond to the minimal value of $AIC(m,n)$. There are also some other reasonable criteria for choosing the orders of an ARMA model.[95*]

The class of ARMA spectral estimates $\varphi^{ARMA}_{T, m, n}(\omega)$ is more general, of course, than the class of autoregressive (AR) spectral estimates $\varphi^{AR}_{T, m}(\omega)$. Therefore, it is clear that the estimates of $\varphi^{ARMA}_{T, m, n}(\omega)$ are also consistent under very wide regularity conditions in the sense that the following relation holds at least for some special selections of the functions $m(T) = m_T$ and $n(T) = n_T$:

$$(3.79) \quad \lim_{T\to\infty} \Phi^{ARMA}_{T,m_T,n_T}(\omega) = f(\omega),$$

where lim denotes the mean square limit and $\Phi^{ARMA}_{T, m_T, n_T}(\omega)$ is the estimator corresponding to the estimate $\varphi^{ARMA}_{T, m_T, n_T}(\omega)$. In fact, we can always take that $n_T = 0$ for all T, and m_T satisfies the conditions necessary for the validity of (3.71). It is natural to expect that the relation (3.79) holds also in all the cases where $m_T \to \infty$, $n_T \to \infty$, $m_T/T \to 0$, $n_T/T \to 0$ as $T \to \infty$, and some mild restrictions are imposed on the rates of increase of m_T and n_T with T. However, no results of such a type were strictly proved until now.

Many specific examples of ARMA spectral estimates can be found in the sources indicated in Note 87. One additional example of this kind is presented in Fig. 47, where a smoothed periodogram $\varphi^{(A)}_{86}(\omega)$ of a geophysical series $x(t)$, $t = 1, ..., 86$, computed by using the Blackman–Tukey method and the Parzen lag window with a reasonable value of the scale parameter k_T, is compared with the estimates $\varphi^{ARMA}_{86, 1,1}(\omega)$

(corresponding to the minimal value of $AIC(m,n)$) and $\varphi^{AR}_{86,3}(\omega)$ (corresponding to the minimal value of $AIC(m)$). We see that the estimates $\varphi^{ARMA}_{86}(\omega)$ and $\varphi^{AR}_{86}(\omega)$ are comparatively close to each other, while the estimate $\varphi^{(A)}_{86}(\omega)$ is clearly more erratic.

We also note in conclusion that in applications to stationary processes with continuous time the ARMA
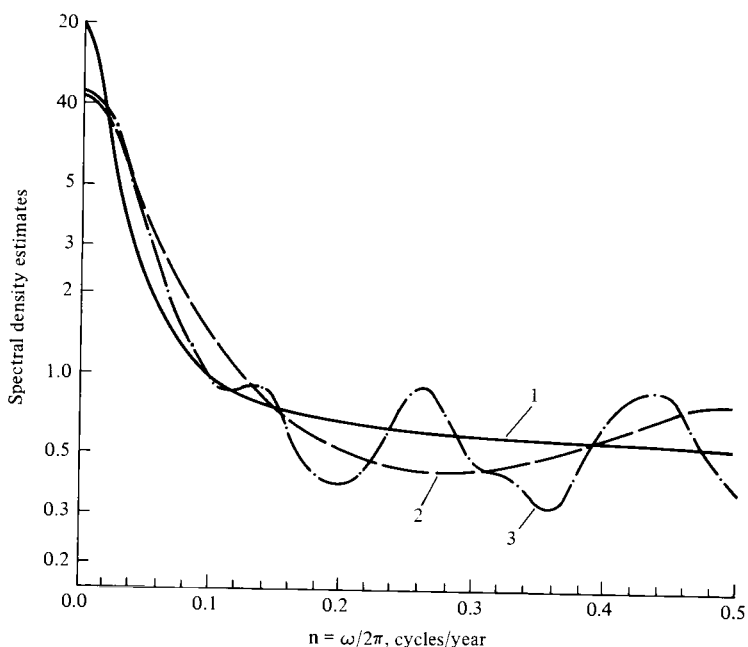
Fig. 47. Three estimates of the spectral density (measured in
$(°C)^2$ year) for the series of mean yearly
temperatures of Earth's belt from 75°N to 80°N. The
estimates were calculated by V.E. Privalsky from
data for 86 years. Line 1 represents the
ARMA-estimate $\varphi^{ARMA}_{86,1,1}(\omega)$. Line 2 shows the
AR-estimate $\varphi^{AR}_{86,3}(\omega)$. Line 3 corresponds to
smoothed periodogram estimate $\varphi^{(A)}_{86}(\omega)$ calculated by
using Parzen's lag window.

estimates include, in particular, the examples depicted in Figs.
22, 23, and 26, where the analytical functions approximating the
empirical correlation functions were chosen in both cases from
among the functions having rational Fourier transforms.

So far, in this section and the previous ones, only the
estimation of the spectral density function $f(\omega)$ has been
considered, but the estimation of the spectral distribution
function $F(\omega)$ has been completely ignored. This arrangement
of the material is justified, first of all, by the fact that in the
overwhelming majority of applied investigations the spectral

distribution function is not even mentioned and all the attention is focused on the study of the spectral density $f(\omega)$. (As we know, this density is quite often just called the *spectrum*, or the *power spectrum*, of the stationary process, or sequence, $X(t)$; see the footnote on p. 105.) Thus, the problem of the spectral density estimation is much more important practically than a similar problem related to the spectral distribution function and, therefore, it is only natural that the former problem occupied the central place in the entire enormous literature (intended primarily for workers in applied fields) which is devoted to spectral estimation.

Note now that the problem of determination of the spectral density $f(\omega)$ is not only more important practically but is also much more difficult theoretically than the problem of determination of the indefinite integral $F(\omega) = \int^{\omega} f(\omega')d\omega'$. The comparative simplicity of determining the spectral distribution function $F(\omega)$ is a simple consequence of the results from Sec. 18 concerning the properties of the periodogram $I_T(\omega)$. Indeed, we have already seen that under some definite, rather general conditions the periodogram ordinates $I_T(\omega)$ are random variables with a mean value close to $f(\omega)$ (provided the spectral density exists) and a bounded variance; moreover, as $T \to \infty$, the variables $I_T(\omega_1)$ and $I_T(\omega_2)$ tend to become uncorrelated (and even independent), provided $\omega_2 \neq \omega_1$. It was actually emphasized above (although not explicitly formulated) that the following important result holds: *if Δ is an arbitrary interval of the frequency axis ω, then $\int_\Delta i_T(\omega')d\omega'$ is a consistent estimate of* $\int_\Delta f(\omega')d\omega'$ (cf. pp. 256-257). Let us choose, as Δ, the interval $[-\pi,\omega]$ (when we deal with sequences $X(t)$ depending on the discrete argument t) or a half-line $(-\infty,\omega]$ (when we deal with stationary processes, i.e. when t is continuous). Then we evidently get the following result:

$$(3.80) \qquad F_T^*(\omega) = \int_{\omega_0}^{\omega} i_T(\omega')d\omega',$$

where $\omega_0 = -\pi$ for discrete time and $\omega_0 = -\infty$ for continuous time, is a consistent estimate of the spectral distribution function $F(\omega)$. (Recall again that consistency means that the estimate is asymptotically unbiased and its variance tends to zero as

$T \to \infty$, i.e. that the corresponding estimator tends to the estimated quantity.)

The above reasoning directly refers only to the case where a continuous spectral density $f(\omega)$ exists and the function $X(t)$ is Gaussian. In fact, such assumptions were made in Sec. 18 in studying the properties of a periodogram (though it was also remarked there that the assumption on the normality of $X(t)$ can be considerably relaxed). However, the statement on the consistency of $F_T^*(\omega)$ holds under much wider conditions having no relation to the assumptions used in Sec. 18 to study the periodogram $I_T(\omega)$. It is easy to show that proving this statement requires only second-order ergodicity of a stationary function $X(t)$, i.e. convergence of the estimator $B_T^{**}(\tau)$, as $T \to \infty$, to the true correlation function $B(\tau)$.[96] Therefore, by integrating the periodogram we can obtain, under very wide conditions, a fully satisfactory estimate of the spectral distribution function $F(\omega)$.

Of course, it does not yet follow from the consistency of the estimate $F_T^*(\omega)$ of the spectral distribution function $F(\omega)$ that, when $F(\omega)$ is an indefinite integral of the spectral density $f(\omega)$, the derivative of $F_T^*(\omega)$ with respect to ω is a consistent estimate of $f(\omega)$. In fact, differentiating $F_T^*(\omega)$, we again arrive at the periodogram $i_T(\omega)$, which is, as we know, a very poor estimate of $f(\omega)$. On the other hand, estimation of the function $F(\omega)$ alone is not very useful from the practical point of view. The thing is that all the monotonically nondecreasing functions of the frequency ω often seem rather alike (as opposed to the spectral densities $f(\omega)$, which differ sharply in form). Therefore, the spectral distribution functions $F(\omega)$ do not characterize the spectral properties of $X(t)$ vividly enough, though in some cases these functions are, nevertheless, undoubtedly useful.[97]

The proof of the consistency of the estimate (3.80) by no means exhausts the statistical theory of estimating the spectral distribution function $F(\omega)$ from the observation data. The available statistical literature contains many additional interesting and important results relating, e.g., to the rate of convergence to zero, as $T \to \infty$, of the mean square error $\Delta^2\{F_T^*(\omega)\} = \langle [F_T^*(\omega) - F(\omega)]^2 \rangle$, the asymptotic (at large T) distributions of the random function $F_T^*(\omega)$ or of $\max_{\omega}|F_T^*(\omega) -$

$F(\omega)|$, etc.[98] In particular, much attention in such investigations was given to a special case of stationary

functions $X(t)$ having a *mixed spectrum,* i.e. containing both a discrete spectrum and a continuous one. Let us assume, for simplicity, that a discrete spectrum is finite, i.e. consists only of a finite number K (which usually is unknown but in some exceptional cases can also be known beforehand) of frequencies. Then the stationary function $X(t)$ can be represented as

$$(3.81) \qquad X(t) = \sum_{k=1}^{K} X_k \cos(\omega_k t + \Theta_k) + N(t),$$

where X_k and Θ_k, $k = 1, ..., K$, are random variables and $N(t)$ is a stationary random function having a continuous spectral density $f(\omega)$ (cf. (2.83)). The determination of the spectral distribution function $F(\omega)$ is equivalent here to the determination of the set of frequencies (the discrete spectrum) $\omega_1, ..., \omega_K$, the jumps of $F(\omega)$ at these frequencies $f_k = \langle |X_k|^2 \rangle$, and the spectral density $f(\omega)$ of the continuous component $N(t)$. A number of results related to this problem can be found in the available literature.[99]

The determination of the purely periodic components of a stationary time series $X(t)$, i.e. components of the form $X_k \cos(\omega_k t + \Theta_k)$, is a classical statistical problem (having many important practical applications) of the search of *hidden periodicities* in a given series of observations. Many works are devoted to this problem; in particular, precisely in this respect the concept of a periodogram was first introduced by Schuster in his investigations of several geophysical and astrophysical time series. A classical scheme of hidden periodicities regards an observational time series $x(t)$ as composed of an unknown (but finite) number of purely sinusoidal oscillations and a purely random ("white") noise $n(t)$. Much later some authors suggested that in many practical situations better agreement with the data can be achieved if we assume that the noise is colored, i.e. its spectrum may be non-uniform. At present the search of hidden periodicities must be regarded as one special problem of spectral estimation theory related to the theory of estimating the spectral distribution function in the case of a mixed spectrum. However, we cannot pursue this subject at greater length in this book.[100]

Chapter 4
SOME GENERALIZATIONS OF THE CONCEPTS OF A STATIONARY RANDOM FUNCTION AND OF A SPECTRAL REPRESENTATION

In this final chapter we shall briefly consider several practically important classes of random functions more general than the class of ordinary (one-dimensional) random sequences and processes. We shall devote most of our attention to the results which generalize the theorems from Secs. 8, 9, and 13 on the spectral representation of stationary random functions and of their correlation functions. In discussing this subject we shall make wide use of the theorem on generalized spectral representation of random functions formulated in Note 17 to Chap. 2 (see Vol. II and also pp. 447–448 below). Note also that for the sake of brevity a number of results in this chapter will be given without any proof or with only a short sketch of a proof.

20. Multidimensional Stationary Random Functions

So far we have been speaking exclusively about a single random function $X(t)$. In applications, however, one often has to consider simultaneously several random functions $X_1(t)$, $X_2(t)$, ..., $X_n(t)$, statistically related to each other. Thus, when the Brownian motion of a particle is observed, one may wish to study simultaneously the time variations $X(t)$, $Y(t)$, and

$Z(t)$ of all three coordinates of the particle; the turbulence measurements are in many cases aimed at determining the simultaneous fluctuations of, say, the three velocity components $U(t)$, $V(t)$, and $W(t)$ and the temperature $\Theta(t)$; in investigating a given linear (or nonlinear) system $\mathfrak{L}$ one often has to consider simultaneously both the input function $X(t)$ and the output function $Y(t)$, etc.   It is, of course, easy to give many more practical examples of this type. Therefore simultaneous theoretical study of several random functions is undoubtedly a very important task.

By analogy with the case of a multidimensional random variable $\mathbf{X} = (X_1, ..., X_n)$, a set of several (real or, more generally, complex) random functions $X_1(t)$, ..., $X_n(t)$ can also be considered a single *multidimensional* (or *vector-valued*) *random function* $\mathbf{X}(t) = \{X_1(t), X_2(t), ..., X_n(t)\}$ defined on a set T of elements t.[1] The functions $X_j(t)$, $j = 1, ..., n$, are called the *components* of $\mathbf{X}(t)$.   For a full specification of a random function $\mathbf{X}(t)$ we have to give a set of all the multidimensional probability distributions of the random variables $\mathbf{X}(t_1^{(1)}, ..., t_{m_1}^{(1)}; t_1^{(2)}, ..., t_{m_2}^{(2)}; ...; t_1^{(n)}, ..., t_{m_n}^{(n)}) = \{X_1(t_1^{(1)}),$

$..., X_1(t_{m_1}^{(1)}); X_2(t_1^{(2)}), ..., X_2(t_{m_2}^{(2)}), ..., X_n(t_1^{(n)}), ..., X(t_{m_n}^{(n)})\}$, where

$(t_1^{(1)}, ..., t_{m_1}^{(1)}), (t_1^{(2)}, ..., t_{m_2}^{(2)}), ..., (t_1^{(n)}, ..., t_{m_n}^{(n)})$ are n arbitrary

finite sets of the elements in T. [2] However, we shall not consider these rather complicated multidimensional probability distributions but shall restrict ourselves to the correlation theory of multidimensional random functions $\mathbf{X}(t)$. This means that we shall consider only the moments of the function $\mathbf{X}(t)$ of the first two orders — the *vector of the mean values* $\mathbf{m}(t) = \{m_1(t),$ ..., $m_n(t)\} = \{\langle X_1(t)\rangle, ..., \langle X_n(t)\rangle\}$ and the *matrix of the correlation functions* (*correlation matrix*, for short)

$$\mathcal{B}(t,s) = \|B_{jk}(t,s)\| = \|\langle X_j(t)\overline{X_k(s)}\rangle\|, \quad j,k = 1,...,n.$$

It is clear that the components $m_j(t)$ of the vector $\mathbf{m}(t)$ coincide with the mean values of the separate components $X_j(t)$ of the multidimensional function $\mathbf{X}(t)$; similarly, the diagonal elements $B_{jj}(t,s)$ of the correlation matrix $\mathcal{B}(t,s)$ are the ordinary correlation functions of the random functions $X_j(t)$. Moreover, the correlation matrix $\mathcal{B}(t,s)$ includes also the

so-called *cross-correlation functions* $B_{jk}(t,s)$ of two random functions $X_j(t)$ and $X_k(t)$ defined by the equation

(4.1) $B_{jk}(t,s) = \langle X_j(t)\overline{X_k(s)} \rangle$

(as in Chap. 2, we consider here right from the start, the general case of complex functions $\mathbf{X}(t)$). It follows from (4.1) that the cross-correlation functions always satisfy the condition

(4.2) $B_{jk}(t,s) = \overline{B_{kj}(s,t)}$.

Below in this section we shall always assume that the set T of elements t coincides either with the set of all the real numbers $-\infty < t < \infty$ or with the set of all the integers $t = 0, \pm 1, \pm 2 \ldots$. Just as in the one-dimensional case, we shall call the functions $\mathbf{X}(t)$ of a real argument t *multidimensional random processes* and the functions $\mathbf{X}(t)$ of an integral argument *multidimensional random sequences*; when speaking simultaneously about multidimensional processes and sequences we shall use the term *multidimensional random functions*. Within the framework of correlation theory, a multidimensional random function $\mathbf{X}(t)$ is said to be *stationary* if its mean-value vector $\langle \mathbf{X}(t) \rangle = \mathbf{m}$ is a constant vector and the correlation matrix $\mathcal{B}(t,s) = \|B_{jk}(t,s)\| = \|B_{jk}(t - s)\| = \mathcal{B}(t - s)$ depends only on the difference $t - s$. In this case all the components $X_j(t)$, $j = 1, \ldots, n$, of the function $\mathbf{X}(t)$ are evidently stationary functions, and any two of them possess the property that their cross-correlation function depends only on $t - s$. Two random functions possessing this last-mentioned property are said to be *stationarily correlated (with each other)*; thus, *the multidimensional random function $\mathbf{X}(t)$ is stationary if, and only if, all its components $X_1(t), \ldots, X_n(t)$ are stationary and are also stationarily correlated in pairs.* This definition is obviously analogous to the definition of wide-sense stationarity given in Sec. 3 for one-dimensional random functions. It is, of course, also not hard to give a definition of stationarity for the multidimensional random function $\mathbf{X}(t)$ which resembles the strict-sense definition given in Sec. 3.[3] However, in this section we shall consider no strict-sense concepts at all, and therefore only the formulated definition of stationarity will be of interest to us.

The correlation theory of n-dimensional stationary random functions $\mathbf{X}(t) = \{X_1(t), \ldots, X_n(t)\}$ deals only with a constant

n-dimensional vector $\underline{m} = \langle X(t) \rangle$ and an $(n \times n)$-matrix $B(\tau) = \|B_{jk}(\tau)\| = \|\langle X_j(t + \tau)\overline{X_k(t)} \rangle\|$, all the elements of which depend on a single variable τ (taking integral values for stationary sequences $X(t)$ and arbitrary real values for stationary processes $X(t)$). Just as in the one-dimensional case we can always consider, instead of the given multidimensional stationary function $X(t)$, its "fluctuation" (or centered random function) $X(t) - m$. Therefore, the mean value $m = \langle X(t) \rangle$ can be assumed equal to zero (i.e. to the zero vector $0 = (0, ..., 0)$) without loss of generality; this is what we shall do from now on. As for the elements of the correlation matrix $B(\tau)$, by virtue of (4.2), they must satisfy the condition

$$(4.3) \qquad B_{jk}(\tau) = \overline{B_{kj}(-\tau)}.$$

In the case of a real random function $X(t)$ all the correlation functions $B_{jk}(\tau)$ are, of course, also real. According to (4.3) all the diagonal elements $B_{jj}(\tau)$, $j = 1, ..., n$, of the correlation matrix $B(\tau)$ are in this case even functions of τ, as the correlation functions of real random function $X_j(t)$ actually should be. We note, however, that the cross-correlation functions $B_{jk}(\tau)$, where $j \neq k$, need not be even in the case of a real function $X(t)$. It is also clear that the general Cauchy–Buniakovsky–Schwarz inequality (0.23) implies the following inequality related to the function $B_{jk}(\tau)$:

$$(4.4) \qquad |B_{jk}(\tau)| \leqslant [B_{jj}(0)B_{kk}(0)]^{1/2}.$$

The inequality $|B_{jk}(\tau)| \leqslant |B_{jk}(0)\|$, however, may not be fulfilled, i.e. the cross-correlation function $B_{jk}(\tau)$ can readily take its maximum value at $\tau \neq 0$.

The general spectral representation theorem for the correlation functions of stationary random functions (i.e (2.52) and (2.206)) implies that the diagonal elements of the correlation matrix $B(\tau)$ can be represented in the form

$$(4.5) \qquad B_{jj}(\tau) = \int e^{i\omega\tau}dF_{jj}(\omega), \ j = 1, ..., n,$$

where, as in all the subsequent equations, the limits of integration are $-\infty, +\infty$ in the case of multidimensional stationary processes, and $-\pi, +\pi$ in the case of stationary sequences, and $F_{jj}(\omega)$ are real nondecreasing functions of ω. Proceeding from the same spectral representation theorem it is

also easy to prove that all the nondiagonal elements $B_{jk}(\tau)$, $j \neq k$, of the matrix $\mathcal{B}(\tau)$ (i.e. the cross-correlation functions of different components of $X(t)$) can also be represented in the form of Fourier–Stieltjes integrals

$$(4.6) \qquad B_{jk}(\tau) = \int e^{i\omega t} dF_{jk}(\omega), \quad j,k = 1, ..., n,$$

where, however, the functions $F_{jk}(\omega)$ are now, in general, complex functions of bounded variation (i.e. such that both their real and imaginary part is a difference of two monotonically nondecreasing functions).[4] Equation (4.6) describes the *spectral representation of the cross-correlation function* $B_{jk}(\omega)$, and the complex function $F_{jk}(\omega)$ appearing in it is called the *cross-spectral distribution function* of the stationary (and stationarily correlated) random functions $X_j(t)$ and $X_k(t)$. Consider now the most important special case where the correlation functions $B_{jk}(\tau)$, j, $k=1$, ..., n, fall off sufficiently rapidly at infinity, so that all of them satisfy condition (2.66) or (2.68) if $X(t)$ is a stationary process, or, respectively, condition (2.212) or (2.212a) if $X(t)$ is a stationary sequence. In this case all the functions $B_{jk}(\tau)$, j, $k=1$, ..., n, can be represented as ordinary Fourier integrals, i.e. the following simpler relation can be used instead of (4.5) and (4.6)

$$(4.7) \qquad B_{jk}(\tau) = \int e^{i\omega\tau} f_{jk}(\omega) d\omega, \quad j,k = 1, ..., n,$$

where $f_{jk}(\omega) = dF_{jk}(\omega)/d\omega$. Formula (4.7) shows that the cross-spectral distribution functions $F_{jk}(\omega)$ can be written as indefinite integrals of their derivatives:

$$(4.8) \qquad F_{jk}(\omega) = \int_{\omega_0}^{\omega} f_{jk}(\omega') d\omega',$$

where $\omega_0 = -\infty$ for continuous time and $\omega_0 = -\pi$ for discrete time. The functions $f_{jj}(\omega)$, $j = 1$, ..., n, are, of course, the ordinary spectral densities of the stationary functions $X_j(t)$, $j = 1$, ..., n. Therefore $f_{jj}(\omega)$ must always be real and everywhere nonnegative, while the *cross-spectral densities* $f_{jk}(\omega)$ (of the two stationary functions $X_j(t)$ and $X_k(t)$) are generally complex. According to the formula for the inversion of a Fourier integral, if the cross-correlation function $B_{jk}(\tau)$ is known, the cross-spectral density $f_{jk}(\omega)$ can be found from the relation

$$(4.9) \; f_{jk}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\omega\tau}B_{jk}(\tau)d\tau \; \text{ or } \; f_{jk}(\omega) = \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty} e^{-i\omega\tau}B_{jk}(\tau).$$

Using (4.9) and (4.8) it is now easy to find the values of the cross-spectral distribution functions $F_{jk}(\omega)$. The expression for the cross-spectral distribution function $F_{jk}(\omega)$ via the cross-correlation function $B_{jk}(\tau)$ obtained by integrating both sides of (4.9) is evidently of the form (2.85) or (2.215). It is not hard to prove that this expression is also valid (with the same reservations as in the one-dimensional case) in the general case where the cross-spectral density $f_{jk}(\omega)$ does not necessarily exist.[5]

The representability of the cross-correlation function $B_{jk}(\tau)$ as the Fourier-Stieltjes integral (4.6) does not essentially restrict the set of such functions because the class of functions of the form (4.6) is extremely broad. (Recall that the class of the ordinary Fourier integrals (4.7), which is considerably narrower than the class of Fourier–Stieltjes integrals, includes all the functions which fall off sufficiently rapidly at infinity.) If, however, instead of considering a single function $B_{jk}(\tau)$ we pass on to consideration of the whole correlation matrix $B(\tau) = \|B_{jk}(\tau)\|$, then it is easy to obtain additional conditions, which essentially restrict both the class of the correlation matrices $B(\tau) = \|B_{jk}(\tau)\|$ and the class of the *spectral distribution matrices* $F(\omega) = \|F_{jk}(\omega)\|$ (whose elements are all the spectral and cross-spectral distribution functions $F_{jk}(\omega)$, $j,k = 1, ..., n$ of a multidimensional stationary random function $X(t)$). First of all, the correlation matrix $B(\tau)$ must satisfy condition (4.3) and this condition implies that

$$(4.10) \qquad F_{jk}(\omega + \Delta\omega) - F_{jk}(\omega) = \overline{F_{kj}(\omega + \Delta\omega)} - \overline{F_{kj}(\omega)}$$

for any ω and $\Delta\omega > 0$. Hence *the matrix* $F(\Delta\omega) = F(\omega + \Delta\omega) - F(\omega)$ *is always Hermitian.*[*] Let now $c_1, ..., c_n$ be n arbitrary complex numbers which we use to form a new one-dimensional complex stationary random function $X^{(c)}(t) =$

---

*Recall that a complex matrix $A = \|a_{jk}\|$ is said to be Hermitian if $a_{jk} = \bar{a}_{kj}$ for all j and k. The Hermiticity of the matrix $F(\Delta\omega)$ obviously guarantees the validity of (4.3) by virtue of (4.6). On the other hand, using the above-mentioned relation expressing $F_{jk}(\omega)$ in terms of $B_{jk}(\tau)$ we can easily show that the Hermiticity of $F(\Delta\omega)$ is an immediate consequence of (4.3).[6]

$\sum_{j=1}^{n} c_j X_j(t)$.   It is easy to see that the correlation function of $X^{(c)}(t)$ is equal to

$$(4.11) \qquad B^{(c)}(\tau) = \sum_{j,k=1}^{n} B_{jk}(\tau) \; c_j \overline{c}_k.$$

Hence, the correlation matrix $\mathcal{B}(\tau) = \| B_{jk}(\tau) \|$ must have the property that the function (4.11) is a positive definite function of τ for any $c_1, ..., c_n$. By virtue of (4.5) and (4.6) the function $B^{(c)}(\tau)$ can also be represented as the Fourier-Stieltjes integral, and the role of the relevant spectral distribution function $F^{(c)}(\omega)$ is now played by the sum

$$(4.12) \qquad F^{(c)}(\omega) = \sum_{j,k=1}^{n} F_{jk}(\omega) \; c_j \overline{c}_k.$$

Since the spectral distribution function of the stationary function $X^{(c)}(t)$ (as well as of any other stationary random function) must be real and nondecreasing, *for any ω and $\Delta\omega >$ 0, the matrix*

$$(4.13) \qquad \mathcal{F}(\Delta\omega) = \mathcal{F}(\omega + \Delta\omega) \; - \mathcal{F}(\omega) = \| F_{jk}(\omega + \Delta\omega) - F_{jk}(\omega) \|$$

must be not only Hermitian but also positive definite, i.e. be such that

$$(4.14) \qquad \sum_{j,k=1}^{n} [F_{jk}(\omega + \Delta\omega) - F_{jk}(\omega)] \; c_j \overline{c}_k \geqslant 0$$

for any complex numbers $c_1, ..., c_n$.[*] The property (4.14) of the spectral distribution matrix $\mathcal{F}(\omega) = \| F_{jk}(\omega) \|$ is a generalization to the multidimensional case of the main property of the spectral distribution function $F(\omega)$ of a one-dimensional stationary function $X(t)$ according to which $F(\omega)$ is always a real nondecreasing function. It is also easy to prove the converse assertion that for any matrix $\mathcal{F}(\omega) = \| F_{jk}(\omega) \|$ satisfying condition (4.14) one can find a multidimensional stationary random function $X(t)$ which has $\mathcal{F}(\omega)$ as its spectral distribution matrix (i.e. which has a correlation matrix $\mathcal{B}(\tau) = \int \exp(i\omega\tau)d\mathcal{F}(\omega)$).[7]

The assertion that *the class of correlation matrices $\mathcal{B}(\tau) =$*

---

[*]In fact (4.14) implies that $\mathcal{F}(\Delta\omega)$ is Hermitian and the same is also true for (4.15) and the matrix $\| f_{jk}(\omega) \|$.

‖$B_{jk}(\tau)$‖ *of multidimensional stationary random functions* $X(t)$ *coincides with the class of matrices representable as the Fourier-Stieltjes integral (4.6), where all the increments* $\mathfrak{F}(\Delta\omega)$ = $\mathfrak{F}(\omega + \Delta\omega) - \mathfrak{F}(\omega)$ *of a matrix-function* $\mathfrak{F}(\omega)$ = ‖$F_{jk}(\omega)$‖ *are Hermitian positive definite matrices,* was first proved by Cramér.[8] This assertion evidently generalizes the theorem by Khinchin (see p. 93) related to the one-dimensional case. In the important case where the spectral and cross-spectral densities $f_{jk}(\omega)$ exist (i.e. all the functions $B_{jk}(\tau)$, j, k = 1, ..., n, fall off rapidly enough at infinity) the condition (4.14) can, of course, be replaced by the condition

$$(4.15) \qquad \sum_{j,k=1}^{n} f_{jk}(\omega) c_j \bar{c}_k \geqslant 0.$$

Thus, *the spectral density matrix* ‖$f_{jk}(\omega)$‖ *must be Hermitian and positive definite at all values of* ω.

Conditions (4.14) and (4.15) can be most easily verified in the particular case where n = 2. It is easy to see that in this case condition (4.14) is equivalent to the validity of the three inequalities

$$(4.14a) \qquad \Delta F_{11}(\omega) \geqslant 0, \quad \Delta F_{22}(\omega) \geqslant 0, \quad |\Delta F_{12}(\omega)|^2 \leqslant \Delta F_{11}(\omega)\Delta F_{22}(\omega)$$

and, similarly, condition (4.15) can be written as

$$(4.15a) \qquad f_{11}(\omega) \geqslant 0, \quad f_{22}(\omega) \geqslant 0, \quad |f_{12}(\omega)|^2 \leqslant f_{11}(\omega) f_{22}(\omega).$$

*Example. Let $X(t)$ = $\{X_1(t), X_2(t)\}$ be a two-dimensional stationary random process such that both the components $X_1(t)$, $X_2(t)$ have an exponential correlation function: $B_{11}(\tau)$ = $C_1 \exp(-\alpha_1|\tau|)$, $B_{22}(\tau)$ = $C_2 \exp(-\alpha_2|\tau|)$. In this case

$$f_{11}(\omega) = \frac{A_1}{\omega^2 + \alpha_1^2}, \quad f_{22}(\omega) = \frac{A_2}{\omega^2 + \alpha_2^2},$$

where $A_1 = C_1\alpha_1/\pi$, $A_2 = C_2\alpha_2/\pi$. Hence the last inequality (4.15a) implies that $f_{12}(\omega)$ falls off when $|\omega| \to \infty$ at least as rapidly as ω^{-2}. In particular, we can assume that

$$f_{12}(\omega) = \frac{A_{12} e^{is\omega}}{(i\omega)^2 + 2a_1(i\omega) + a_2},$$

where A_{12}, s, a_1, and a_2 are real parameters.   The cross-correlation function $B_{12}(\tau)$, which corresponds to the cross-spectral density $f_{12}(\omega)$, can easily be evaluated by using the theory of residues. The form of the resulting expression clearly depends on the location of the roots μ_1, μ_2 of the equation $\mu^2 + 2a_1\mu + a_2 = 0$.   If $a_1^2 - a_2 = d^2 > 0$ (i.e. the roots μ_1 and μ_2 are real) and $\mu_2 < \mu_1 < 0$, then

$$B_{12}(\tau) = \begin{cases} (\pi A_{12}/d)(e^{\mu_1(\tau-s)} - e^{\mu_2(\tau-s)}) & \text{for } \tau > s, \\[2mm] 0 & \text{for } \tau \leqslant s, \end{cases}$$

while if $\mu_1 > \mu_2 > 0$, then $B_{12}(\tau) = 0$ for $\tau \geqslant s$ but $B_{12}(\tau)$ differs from zero for $\tau < s$.   Moreover, if $\mu_1 < 0 < \mu_2$, then

$$B_{12}(\tau) = \begin{cases} (\pi A_{12}/d)e^{\mu_1(\tau-s)} & \text{for } \tau \geqslant s, \\[2mm] (\pi A_{12}/d)e^{\mu_2(\tau-s)} & \text{for } \tau < s. \end{cases}$$

Similarly, if $a_1^2 - a_2 = -d_1^2 < 0$ (i.e. $\mu_1 = -a_1 + id_1$ and $\mu_2 = -a_1 - id_1$ are complex conjugate) and $a_1 > 0$, then

$$B_{12}(\tau) = \begin{cases} (2\pi A_{12}/d_1)e^{-a_1(\tau-s)}\sin d_1(\tau - s) & \text{for } \tau > s, \\[2mm] 0 & \text{for } \tau \leqslant s, \end{cases}$$

while if $a_1^2 - a_2 < 0$ and $a_1 < 0$, then $B_{12}(\tau) = 0$ for $\tau \geqslant s$ and $\neq 0$ for $\tau < s$.

Parameter s in the expression for $f_{12}(\tau)$ can be selected quite arbitrarily, but to satisfy (4.15a) parameters A_{12}, a_1, and a_2 must be selected in such a way that

$$A_{12}^2 \leqslant A_1 A_2, \quad A_{12}^2 \alpha_1 \alpha_2 \leqslant A_1 A_2 a_2^2$$

and either

$$A_{12}^2(\alpha_1^2 + \alpha_2^2) \leqslant A_1 A_2 (4a_1^2 - 2a_2)$$

or, if the last inequality does not hold,

$$\{A_{12}^2(\alpha_1^2 + \alpha_2^2) - A_1A_2(4a_1^2 - 2a_2)\}^2$$

$$\leqslant 4(A_1A_2 - A_{12}^2)(A_1A_2a_2^2 - A_1^2\alpha_1^2\alpha_2^2). *$$

The results presented above are closely connected with the existence of the spectral representations of the multidimensional stationary random functions $X(t) = \{X_1(t), ..., X_n(t)\}$ themselves. According to (2.61) and (2.204), each component $X_j(t)$ of a multidimensional stationary random function $X(t)$ can be represented in the form

$$(4.16) \qquad X_j(t) = \int e^{i\omega\tau}dZ_j(\omega), \quad j = 1,...,n,$$

where $Z_j(\omega)$, $j = 1, ..., n$ are random functions with uncorrelated increments (and zero mean values, if $\langle X(t)\rangle = 0$). The last equation can also be written in vector notation as

$$(4.17) \qquad X(t) = \int e^{i\omega t}dZ(\omega),$$

where $Z(\omega) = \{Z_1(\omega),...,Z_n(\omega)\}$ is a vector random function. It is not hard to show that the fact that the components of $X(t)$ are not only stationary but also stationarily correlated in pairs implies that the random functions $Z_j(\omega)$ and $Z_k(\omega)$ for any i and k must satisfy the following relations:

$$(4.18) \quad \langle[Z_j(\omega_1 + \Delta\omega_2) - Z_j(\omega_1)] \overline{[Z_k(\omega_2 + \Delta\omega_2) - Z_k(\omega_2)]}\rangle = 0$$

if the intervals $[\omega_1, \omega_1 + \Delta\omega_1]$ and $[\omega_2, \omega_2 + \Delta\omega_2]$ are disjoint (i.e. nonintersecting), and

$$(4.19) \quad \langle[Z_j(\omega+\Delta\omega) - Z_j(\omega)] \overline{[Z_k(\omega + \Delta\omega) - Z_k(\omega)]}\rangle$$

$$= F_{jk}(\omega + \Delta\omega) - F_{jk}(\omega),$$

where $F_{jk}(\omega)$ is the cross-spectral distribution function of the random functions $X_j(t)$ and $X_k(t)$.[9] It is easy to see that the relation (4.6) also follows immediately from (4.16) and (4.18), if (4.19) is used as the definition of $F_{jk}(\omega)$. Moreover, (4.19) also implies the result

$$(4.20) \quad \sum_{j,k=1}^n [F_{jk}(\omega + \Delta\omega) - F_{jk}(\omega)]c_j\bar{c}_k = \langle\left|\sum_{j=1}^n [Z_j(\omega + \Delta\omega) - Z_j(\omega)]c_j\right|^2\rangle.$$

Thus, the left-hand side of (4.20) coincides with the mean value of a nonnegative random variable, and this explains the meaning of the inequality (4.14).

Suppose now that among the numbers $c_1, ..., c_n$ in the left-hand side of (4.14) only c_j and c_k differ from zero. Then we obtain

$$F_{jj}(\Delta\omega) |c_j|^2 + F_{kk}(\Delta\omega) |c_k|^2 + F_{jk}(\Delta\omega) c_j\bar{c}_k + F_{jk}(\Delta\omega) c_k\bar{c}_j \geqslant 0,$$

where, as usual, $F_{pq}(\Delta\omega)$ denotes $F_{pq}(\omega + \Delta\omega) - F_{pq}(\omega)$. It follows from this that

(4.21) $$|F_{jk}(\Delta\omega)|^2 \leqslant F_{jj}(\Delta\omega)F_{kk}(\Delta\omega).$$

Inequality (4.21) implies, in particular, that if the spectral densities $f_{jj}(\omega)$ and $f_{kk}(\omega)$ exist (i.e. $F_{jj}(\Delta\omega) \approx f_{jj}(\omega)\Delta\omega$, $F_{kk}(\Delta\omega) \approx f_{kk}(\omega)\Delta\omega$ for sufficiently small $\Delta\omega$), then there necessarily exists the cross-spectral density $f_{jk}(\omega)$, and

(4.22) $$|f_{jk}(\omega)|^2 \leqslant f_{jj}(\omega)f_{kk}(\omega).$$

Another conclusion from (4.21) is that only the discontinuity points of both the functions $F_{jj}(\omega)$ and $F_{kk}(\omega)$ can be the discontinuity points of the function $F_{jk}(\omega)$.

Even for a real multidimensional stationary function $X(t) = \{X_1(t), ..., X_n(t)\}$ the cross-spectral distribution functions $F_{jk}(\omega)$, where $j \neq k$, are, in general, complex. Since, however, in this case $B_{jk}(\tau)$ is a real function, (4.6) implies that $F_{jk}(\Delta\omega) = F_{jk}(-\Delta\omega)$, where $-\Delta\omega$ is an interval symmetric to the interval $\Delta\omega$ with respect to the point $\omega = 0$. In particular, the cross-spectral densities $f_{jk}(\omega)$ of a real stationary function

$X(t)$ satisfy the condition $f_{jk}(\omega) = f_{jk}(-\omega)$, and the jump $f_{jk}^{(0)}$ of the function $F_{jk}(\omega)$ at the point $\omega = 0$ is always real.   Let $F_{jk}(\Delta\omega) = C_{jk}(\Delta\omega) - iQ_{jk}(\omega)$, where $C_{jk}(\Delta\omega)$ and $Q_{jk}(\Delta\omega)$ are real interval functions; then evidently $C_{jk}(-\Delta\omega) = C_{jk}(\Delta\omega)$ and $Q_{jk}(-\Delta\omega) = -Q_{jk}(\Delta\omega)$. Replacing now the complex function $\exp(i\omega\tau)$ in the right-hand side of (4.6) by $\cos\omega\tau + i\sin\omega\tau$ and $dF_{jk}(\omega) = F_{jk}(d\omega)$ by $C_{jk}(d\omega) - iQ_{jk}(d\omega) = dC_{jk}(\omega) - idQ_{jk}(\omega)$, and discarding the odd-terms of the integrand which make no contribution to the integral, we arrive at the following real

form of the spectral representation of the cross-correlation function $B_{jk}(\tau)$ of real stationary functions $X_j(t)$ and $X_k(t)$:

(4.23) $B_{jk}(\tau) = \int\cos\omega\tau\, dC_{jk}(\omega) + \int\sin\omega\tau\, dQ_{jk}(\omega).$

In the most important case of the existence of a cross-spectral density $f_{jk}(\omega)$ we similarly put $f_{jk}(\omega) = c_{jk}(\omega) - iq_{jk}(\omega)$, where

$c_{jk}(\omega)$ and $q_{jk}(\omega)$ are real functions of the frequency ω and

$c_{jk}(-\omega) = c_{jk}(\omega)$, $q_{jk}(-\omega) = -q_{jk}(\omega)$ (since $f_{jk}(-\omega) = f_{jk}(\omega)$). Then we can rewrite (4.23) as

(4.24) $B_{jk}(\tau) = \int\cos\omega\tau\, c_{jk}(\omega)d\omega + \int\sin\omega\tau\, q_{jk}(\omega)d\omega.$

The functions $C_{jk}(\omega)$ and $Q_{jk}(\omega)$ are called the *cospectral distribution function* and, respectively, the *quadrature spectral distribution function,* and $c_{jk}(\omega) = C'_{jk}(\omega)$ and $q_{jk}(\omega) = Q'_{jk}(\omega)$ the *cospectral density (function)* and *quadrature spectral density (function)* or, briefly, the *cospectrum* and the *quadrature spectrum* of two real stationarily correlated random functions $X_j(t)$ and $X_k(t)$.[10] Note that $\omega = 0$, (4.24) turns into

(4.25) $B_{jk}(0) = \langle X_j(t)X_k(t)\rangle = \int c_{jk}(\omega)d\omega,$

which shows that the cospectral density determines the frequency distribution of the second mixed moment $\langle X_j(t)X_k(t)\rangle$ of the two stationary and stationarily correlated random functions. However, in distinction to the ordinary spectral density, the cospectral density $c_{jk}(\omega)$ can alternate its sign, so that the contribution of certain frequency bands to the value of $\langle X_j(t)X_k(t)\rangle$ may well be negative.

The first term on the right-hand side of (4.24) is clearly an even function of τ, while the second term is an odd function of τ. Note now that the function $B_{jk}(\tau)$, as any other function, can be uniquely represented as a sum of its even part $B_{jk}^{(e)}(\tau)$ and its odd part $B_{jk}^{(o)}(\tau)$:

$$B_{jk}(\tau) = B_{jk}^{(e)}(\tau) + B_{jk}^{(o)}(\tau) = \frac{1}{2}[B_{jk}(\tau) + B_{jk}(-\tau)]$$

(4.26) $$+ \frac{1}{2}[B_{jk}(\tau) - B_{jk}(-\tau)].$$

Hence, if the spectral density $f_{jk}(\omega)$ exists, the function $B^{(e)}_{jk}(\tau)$ = $[B_{jk}(\tau) + B_{jk}(-\tau)]/2$ coincides with the Fourier cosine transform of the function $c_{jk}(\omega)$, while $B^{(o)}_{jk}(\omega) = [B_{jk}(\tau) - B_{jk}(-\tau)]/2$ is the Fourier sine transform of the function $q_{jk}(\omega)$. Using the known formulae for the inversion of Fourier cosine and sine transforms, we find that, e.g., in the case of continuous time (i.e. for processes $X_j(t)$ and $X_k(t)$)

$$(4.27) \qquad c_{jk}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \cos\omega\tau B^{(e)}_{jk}(\tau)d\tau,$$

$$q_{jk}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \sin\omega\tau B^{(o)}_{jk}(\tau)d\tau.$$

Similar equations, but containing sums, rather than integrals on the right-hand side, can also be obtained in the case of discrete time (i.e. for multidimensional stationary sequences). The general equations for the cospectral and quadrature spectral distribution functions $C_{jk}(\omega)$ and $Q_{jk}(\omega)$ coincide with the result obtained by integrating both sides of (4.27).

Specification of the three spectral densities $f_{jj}(\omega)$, $f_{kk}(\omega)$, and $f_{jk}(\omega)$ is, of course, equivalent to that of four real functions $f_{jj}(\omega)$, $f_{kk}(\omega)$, $c_{jk}(\omega)$, and $q_{jk}(\omega)$. One can also use for the same purpose some other sets of functions of ω. For instance, sometimes instead of the representation $f_{jk}(\omega) = c_{jk}(\omega) - iq_{jk}(\omega)$ the "polar" representation $f_{jk}(\omega) = \alpha_{jk}(\omega)\exp\{i\Theta_{jk}(\omega)\}$ is used where $\alpha_{jk}(\omega) = |f_{jk}(\omega)| = \{c^2_{jk}(\omega) + q^2_{jk}(\omega)\}^{1/2}$, $\Theta_{jk}(\omega) = \tan^{-1}\{-q_{jk}(\omega)/c_{jk}(\omega)\}$). The dimensionless (i.e. independent of the choice of units measuring X_j, X_k, and t) function $\Theta_{jk}(\omega) = \tan^{-1}\{-q_{jk}(\omega)/c_{jk}(\omega)\}$ is then called the *phase spectrum*, and $\alpha_{jk}(\omega)$ is the *cross-amplitude spectrum*. The cross-amplitude spectrum $\alpha_{jk}(\omega)$ is also often replaced by the dimensionless characteristic

$$(4.28) \qquad \gamma_{jk}(\omega) = \frac{\alpha_{jk}(\omega)}{\{f_{jj}(\omega)f_{kk}(\omega)\}^{1/2}} = \frac{\{c^2_{jk}(\omega) + q^2_{jk}(\omega)\}^{1/2}}{\{f_{jj}(\omega)f_{kk}(\omega)\}^{1/2}}$$

which is called the *coherency* of $X_j(t)$ with $X_k(t)$ (or the *coherency spectrum* of the two random functions $X_j(t)$ and $X_k(t)$).[11] It follows immediately from (4.22) that $0 \leqslant \gamma_{jk}(\omega) \leqslant 1$ for all ω.

Almost all the results of Chapter 3 referring to the estimation of the correlation function, spectral density, and spectral distribution function of a stationary function $X(t)$ from a segment of its single realization $x(t)$ can be extended, with slight modification, to the problem of estimating the respective characteristics of the multidimensional stationary function $X(t) = \{X_1(t), ..., X_n(t)\}$ from the observed values of its realization $x(t) = \{x_1(t), ..., x_n(t)\}$ for $0 \leqslant t \leqslant T$ or $t = 1, ..., T$. Let us assume, for definiteness, that the time is continuous (i.e. $X(t)$ is a stationary process). Then, under wide regularity conditions,

$$(4.29) \quad B^*_{jk,T}(\tau) = \begin{cases} \dfrac{1}{T-\tau} \displaystyle\int_0^{T-\tau} x_j(t+\tau)x_k(t)dt & \text{for } 0 \leqslant \tau < T, \\[2ex] \dfrac{1}{T-|\tau|} \displaystyle\int_{-\tau}^{T} x_j(t+\tau)x_k(t)dt & \text{for } -T < \tau < 0, \\[2ex] 0 & \text{for } |\tau| \geqslant T \end{cases}$$

is the unbiased consistent estimate of the cross-correlation function $B_{jk}(\tau)$. However, often a more convenient estimate of $B_{jk}(\tau)$ is

$$(4.30) \quad B^{**}_{jk,T}(\tau) = \begin{cases} \dfrac{1}{T} \displaystyle\int_0^{T-\tau} x_j(t+\tau)x_k(t)dt & \text{for } 0 \leqslant \tau < T, \\[2ex] \dfrac{1}{T} \displaystyle\int_{-\tau}^{T} x_j(t+\tau)x_k(t)dt & \text{for } -T < \tau < 0, \\[2ex] 0 & \text{for } |\tau| \geqslant T. \end{cases}$$

This latter estimate is also consistent (but only asymptotically unbiased) and has the important advantage that the matrix $B^{**}_T(\tau) = \|B^{**}_{jk,T}(\tau)\|$ (as distinct from $B^*_T(\tau) = \|B^*_{jk,T}(\tau)\|$) necessarily belongs to the class of possible correlation matrices of multidimensional stationary processes. It is also sometimes expedient to use, instead of (4.30), the more rapidly falling off estimate

$$(4.31) \quad B^{(a)}_{jk,T}(\tau) = a_T(\tau)B^{**}_{jk,T}(\tau),$$

where $a_T(\tau)$ is a reasonably chosen *lag window* (cf. pp. 243–244). In the Gaussian case also the relay correlation estimate and polarity coincidence correlation estimate (see pp. 235–236) can easily be generalized to the case of cross-correlation function estimation.[12]

It is easy to show that the Fourier transform of the estimate (4.30) satisfies the relation

$$i_{jk,T}(\omega) = \frac{1}{2\pi}\int_{-T}^{T}e^{-i\omega\tau}B_{jk,T}^{**}(\tau)d\tau$$

(4.32)

$$= \frac{1}{2\pi T}\int_{0}^{T}e^{-i\omega t}x_j(t)dt\int_{0}^{T}\overline{e^{-i\omega t}x_k(t)}dt.$$

$i_{jk,T}(\omega)$ (and also the corresponding estimator $I_{jk,T}(\omega)$) is called the *cross-periodogram* between $X_j(t)$ and $X_k(t)$. The cross-periodogram is an asymptotically unbiased estimate of the cross-spectral density $f_{jk}(\omega)$, but this estimate is inconsistent and behaves extremely erratically at large values of T. A consistent estimate of $f_{jk}(\omega)$ can be obtained by forming the Fourier transform of some intelligently selected estimate of the form (4.31) of the cross-correlation function $B_{jk}(\tau)$ or, what is equivalent, by smoothing the cross-periodogram $i_{jk,T}(\omega)$ with the aid of an appropriately selected *spectral window* $A_T(\omega - \omega')$, i.e. by considering the integral

$$(4.33)\ \varphi_{jk,T}^{(A)}(\omega) = \frac{1}{2\pi}\int_{-T}^{T}e^{i\omega\tau}B_{jk,T}^{(a)}(\tau)d\tau = \int_{-\infty}^{\infty}A_T(\omega-\omega')i_{jk,T}(\omega')d\omega',$$

where $A_T(\omega)$ is the Fourier transform of the lag window $a_T(\tau)$. Moreover, the parametric methods of spectral density estimation, considered in Sec. 19 for the one-dimensional case, can also be generalized to the estimation of the spectral density matrix for multidimensional stationary functions $X(t)$. Consistent estimates of the cross-spectral densities $f_{jk}(\omega)$ can, of course, be used to construct consistent estimates of the cospectrum $c_{jk}(\omega)$, quadrature spectrum $q_{jk}(\omega)$, and coherency $\gamma_{jk}(\omega)$. However, we shall not consider all these topics here but shall refer the reader to the available extensive literature on statistical analysis of multidimensional stationary random functions.[13]

21. Homogeneous Random Fields

21.1. One-Dimensional and Multidimensional Homogeneous Random Fields

In this section we generalize the concept of a stationary

random function by introducing the concept of a (statistically) *homogeneous random field* in an n-dimensional space. Recall that on p. 19 and in Note 1 to this chapter we agreed to use the term *random field* to denote a random function $X(t_1, ..., t_n)$ of several real arguments.[14] In applications, one most often encounters cases were $n = 2$ and t_1, t_2 are the planar coordinates or $n = 3$ and t_1, t_2, t_3 are the spatial coordinates in an ordinary three-dimensional space. There are, however, also many situations where one of the variables $t_1, ..., t_n$ (say, t_n) denotes the time, while $t_1, ..., t_{n-1}$ are spatial coordinates; moreover, in some applications the variables $t_1, ..., t_n$ are specialized multidimensional coordinates in some space (e.g., in the velocity space or coordinate and momentum phase space of a physical system). Observational data, forming a realization of a random field, is quite common in the mechanics of turbulence, oceanography (e.g., the heights of a sea surface at different points), seismology, physics (optics, in particular), engineering sciences (e.g., in problems involving rough surfaces), as well as in geography, geology, forestry, agricultural science, and in many other diverse areas of science.[15]

In what follows we consider fields $X(t_1, ..., t_n) = X(\mathbf{t})$ given either in the entire n-dimensional space $\mathbb{R}^n$ of points $\mathbf{t} = (t_1, ..., t_n)$ or on the n-dimensional lattice $\mathbb{Z}^n$ consisting of all points in $\mathbb{R}^n$ with integral coordinates. (Fields on $\mathbb{Z}_n$ will also be sometimes called *discrete parameter random fields.*) The field $X(\mathbf{t})$ in $\mathbb{R}^n$ or on $\mathbb{Z}^n$ is said to be *homogeneous* if its mean value

$$(4.34) \qquad \langle X(\mathbf{t}) \rangle = m$$

is constant, while its correlation function

$$(4.35) \qquad \langle X(\mathbf{t}_1)\overline{X(\mathbf{t}_2)} \rangle = B(\mathbf{t}_1, \mathbf{t}_2)$$

depends only on the vector $\boldsymbol{\tau} = \mathbf{t}_1 - \mathbf{t}_2$, so that

$$(4.36) \qquad B(\mathbf{t}_1, \mathbf{t}_2) = B(\mathbf{t}_1 - \mathbf{t}_2).$$

(More precisely, a random field satisfying conditions (4.34) and (4.36) should be called *wide sense homogeneous,* since it is also easy to define the concept of a *strictly homogeneous random field* related to the concept of a strictly stationary random process or sequence.[16] Further on, however, the concept of strict sense homogeneity will not be used at all, and therefore wide sense

homogeneous fields will simply be called homogeneous fields.) It is clear that the constant m can take any complex value, but not any function of n variables can be the correlation functions of a homogeneous random field. In fact, it follows from (4.35) that $B(-\tau) = \overline{B(\tau)}$ for any correlation function $B(\tau)$; some other, more restrictive conditions on the behavior of $B(\tau)$ will be indicated below. Let us, however, note that $B(-\tau_1, \tau_2, ..., \tau_n)$ is equal neither to $B(\tau_1, \tau_2, ..., \tau_n)$ nor to $\overline{B(\tau_1, \tau_2, ..., \tau_n)}$. In particular, if $n = 2$ and the field $X(t) = X(t_1, t_2)$ is real, then $B(-\tau_1, -\tau_2) = B(\tau_1, \tau_2)$, but, in general, $B(-\tau_1, \tau_2) \neq B(\tau_1, \tau_2)$.

It is clear that the concept of a homogeneous random field in the one-dimensional space $\mathbb{R}^1$ (i.e. on the straight line) coincides with the concept of a (wide sense) stationary random process while the concept of a homogeneous random field on the one-dimensional lattice $\mathbb{Z}^1$ coincides with the concept of a stationary random sequence. Thus, stationary random functions (processes and sequences) are included as a special case in the class of homogeneous random fields. Note also that if $X(t)$ is a homogeneous random field in $\mathbb{R}^n$, then its values at points on an arbitrary straight line $t = at + a_1$, where $-\infty < t < \infty$ and a and a_1 are n-dimensional vectors, form a homogeneous field on the line, i.e., a stationary process of the variable t. Similarly, the values of a discrete parameter homogeneous field on the lattice $\mathbb{Z}^n$ at lattice points along a straight line form a stationary random sequence.

Almost all the results on stationary random functions given in the preceding chapters of the book can be easily generalized to homogeneous random fields in an n-dimensional space. Consider, in particular, the simplest stationary random sequence — a sequence of uncorrelated random variables (discrete parameter white noise) $E(t)$, $t = 0, \pm 1, \pm 2,...$ (see Example 1 on p. 75). An n-dimensional analogue of such a sequence is discrete parameter *field of uncorrelated random variables* — a set of mutually uncorrelated real random variables $E(t) = E(t_1, ..., t_n)$ corresponding to all the vertices of the lattice $\mathbb{Z}^n$ (i.e. to points t with integral coordinates $t_1, ..., t_n$). Suppose, as in the one-dimensional case, that $\langle E(t) \rangle = 0$, $\langle E^2(t) \rangle = \sigma^2 = 1$. Then the correlation function of the field $E(t)$ equals

$$(4.37) \qquad B(\mathbf{\tau}) = \begin{cases} 1 & \text{for } \mathbf{\tau} = (0,...,0), \\ 0 & \text{for } \mathbf{\tau} \neq (0,...,0). \end{cases}$$

The field $E(t)$ can now be used to construct examples of *moving average fields*

$$(4.38) \qquad X(t) = \sum_i b(t_i) E(t + t_i)$$

where t_i, $i = 1, 2,...$ is some finite or infinite set of vectors with integral coordinates, and $b(t_i)$ are numerical coefficients. (If the set $\{t_i\}$ is infinite, the coefficients $b(t_i)$ must satisfy the condition $\sum_i |b(t_i)|^2 < \infty$, which guarantees the

mean square convergence of the sum in the right-hand side of (4.38).) The correlation function of the sequence (4.38) is given by the formula

$$(4.39) \qquad B(\mathbf{\tau}) = \sum_i b(\mathbf{\tau} + t_i)\overline{b(t_i)}.$$

The moving average fields of the form (4.38) cover a large class of discrete parameter homogeneous random fields.[17]

Generalization of the concept of a stationary autoregressive sequence (see Examples 5 and 6 on pp. 77 and 78) to fields in multidimensional spaces is slightly more complicated. If t is a one-dimensional time variable, then the natural distinction of the past and the future plays an important part. This distinction suggests that the value of the sequence at the instant t can depend only on its past (but not future) values. However, there is no such distinction for a multidimensional (or one-dimensional spatial) t; therefore, a spatial generalization of stationary autoregressive sequences must include all the discrete parameter homogeneous fields satisfying difference equations of the form

$$(4.40) \qquad X(t) + \sum_i a(t_i) X(t + t_i) = cE(t).$$

It is obvious that the disappearance of the asymmetry of the past and future generates an important difference between the *autoregressive fields* (satisfying (4.40)) and autoregressive sequences. Nevertheless, models of spatial autoregressions satisfying equations of the form (4.40) are useful for many applications and are treated extensively in the literature.

Among such models particular attention is attracted by so-called *nearest neighbor models*, which, e.g., in two dimensions (i.e. for $n = 2$) are described by an equation of the form

$$(4.41) \quad X(t_1,t_2) = a_1 X(t_1 - 1, t_2) + a_2 X(t_1, t_2 - 1) + a_3 X(t_1 + 1, t_2)$$

$$+ a_4 X(t_1, t_2 + 1) + cE(t_1, t_2).$$

(For $a_1 = a_2 = a_3 = a_4$ the model is called *symmetric*, while for $a_1 = a_3$, $a_2 = a_4$ it is a *laterally symmetric* nearest neighbor model.[18]) The examples $1 - 3$ of stationary random processes considered in Sec. 7 (pp. 82–87) can also be generalized quite easily to the case of homogeneous random fields in $\mathbb{R}^n$, or on $\mathbb{Z}^n$. In particular, it is easy to show that if a random field $X(t)$ in $\mathbb{R}^n$ or on $\mathbb{Z}^n$ is homogeneous and has the form $X(t) = X\Phi(t)$, where X is a random variable and $\Phi(t) = \Phi(t_1, ..., t_n)$ is a complex-valued non-constant function of n variables (continuous if the variables $t_1, ..., t_n$ vary continuously), then $\langle X \rangle = 0$ and

$$\Phi(t) = r \exp\{i(kt + \Theta)\} = r \exp\{i(k_1 t_1 + ... + k_n t_n + \Theta)\}$$

where r and Θ are real numbers, $k = (k_1, ..., k_n)$ is a real n-dimensional vector, and $kt = k_1 t_1 + ... + k_n t_n$ is a scalar product of n-dimensional vectors k and t. (For discrete parameter fields on $\mathbb{Z}^n$ it is, of course, always possible to assume that $-\pi \leqslant k_1 \leqslant \pi, ..., -\pi \leqslant k_n \leqslant \pi$; cf. p. 179.) Since the numerical factor $r \exp \Theta$ can be included in the random variable X, we can write

$$(4.42) \quad X(t) = X e^{ikt}$$

i.e. $X(t)$ is a plane wave with random amplitude and random phase. Moreover, it is easy to show that a random field

$$(4.43) \quad X(t) = \sum_j X_j e^{ik_j t}$$

is also homogeneous provided that $\langle X_1 \rangle = \langle X_2 \rangle = ... = 0$ (or $\langle X_2 \rangle = \langle X_3 \rangle = ... = 0$, while $k_1 = 0 = (0, ..., 0)$) and $\langle X_j \bar{X}_l \rangle = 0$ for $j \neq l$, $\langle |X_j|^2 \rangle = f_j$ where $\sum_j f_j < \infty$. The field (4.43) is a superposition of mutually uncorrelated plane waves with

random amplitudes and phases. The correlation function of this field evidently equals

$$(4.44) \qquad B(\tau) = \langle X(t + \tau)X(t) \rangle = \sum_j f_j e^{ik_j \tau}.$$

By analogy with the special case where $n = 1$ (i.e. with stationary processes and sequences), it is natural to assume that any homogeneous random field $X(t)$ must have a *spectral representation* which can be written in a form similar to (2.58) (or (2.204)) and which indicates that $X(t)$ is either of the form (4.43) or is the limit (in the mean) of a sequence of fields of a form (4.43). The correctness of this assumption can be proved, e.g., in the following manner. As we know from Sec. 2 of this book, the function $B(t_1, t_2)$ is a correlation function of some random function $X(t)$ on the set $T = \{t\}$ if and only if $B(t_1, t_2)$ is a positive definite kernel on T. Let us consider firstly homogeneous random fields in $\mathbb{R}^n$, i.e. assume that the vector parameter t is continuous. The correlation function $B(t_1, t_2)$ of a homogeneous field in $\mathbb{R}^n$ depends only on the difference $\tau = t - t_2$; therefore the class of correlation functions $B(t+\tau, t)$ coincides here with the class of continuous positive definite functions $B(\tau) = B(\tau_1, ..., \tau_n)$ of n variables $\tau_1, ..., \tau_n$. (Recall that we agreed to consider only continuous correlation functions and that the function $B(\tau) = B(\tau_1, ..., \tau_n)$ is said to be positive definite if and only if

$$(4.45) \qquad \sum_{j,l=1}^{m} B(\tau_j - \tau_l) \, c_j \bar{c}_l \geqslant 0$$

for any integer m, n-dimensional vectors $\tau_1, ..., \tau_m$ and complex numbers $c_1, ..., c_m$; cf. p. 58 and Note 8 to the Introduction.) Moreover, it was proved by Bochner in 1933 (see, e.g., Bochner, 1959) that the class of n-dimensional positive definite functions $B(\tau)$ coincides with the class of functions which can be represented as Fourier–Stieltjes integral of the form

$$(4.46) \qquad B(\tau) = \int_{\mathbb{K}^n} e^{i\tau k} F(dk)$$

where $dk = dk_1...dk_n$ is a volume element of the n-dimensional space $\mathbb{K}^n$ of vectors $k = (k_1, ..., k_n)$, while $F(\Delta)$ is a nonnegative bounded measure in $\mathbb{K}^n$. (This means that $F(\Delta)$ is a nonnegative function of the n-dimensional set Δ determined, in

particular, for all n-dimensional intervals $\Delta k = \{\Delta k_1, ..., \Delta k_n\}$ and such that, firstly, $F(\Delta_1 + \Delta_2) = F(\Delta_1) + F(\Delta_2)$ for any nonintersecting sets Δ_1 and Δ_2 and, secondly, $F(\mathbb{K}^n) < \infty$.) The stated result is widely known now as the *multidimensional Bochner theorem*; see Note 8 to the Introduction in Vol. II.[19] The integral on the right-hand side of (4.46) can be defined quite similarly to the definition (2.59)–(2.60) of the one-dimensional integral (2.58). Without claiming maximum generality, one can, e.g., assume that

(4.47)
$$\int_{\mathbb{K}_n} e^{iTk} F(dk) = \lim_{N\to\infty} \int_{\mathbb{K}(N)} e^{iT\kappa} F(dk)$$

$$= \lim_{N\to\infty} \{ \lim_{\max|\Delta_j k|\to 0} \sum_{j=1}^{M_N} e^{iTk'_j} F(\Delta_j k)\}$$

where $\mathbb{K}(N)$ is the n-dimensional cube in $\mathbb{K}^n$ with center at the coordinate origin and with edges of length $2N$, $\Delta_1 k$, $\Delta_2 k$, ..., $\Delta_{M_N} k$ are nonintersecting n-dimensional intervals (i.e.

parallelepipeds with sides parallel to the coordinate axes), which cover the cube $\mathbb{K}(N)$, $|\Delta_j k|$ is the length of the longest edge of the parallelepiped $\Delta_j k$, and k'_j is an arbitrary point in $\Delta_j k$. Instead of the set function (measure) $F(\Delta)$ one can also use the function of a point $k = (k_1, ..., k_n)$

(4.48) $F(k) = F\{\Delta(k)\}$

where $\Delta(k)$ is the set of all the points $k' = (k'_1, ..., k'_n)$ such that $k'_1 \leqslant k_1, ..., k'_n \leqslant k_n$. The function $F(k)$ evidently has the property that its increments on all the n-dimensional intervals $\Delta k = \{\Delta k_1, ..., \Delta k_n\}$ defined by the formula

$$\Delta F(k) = F(k_1 + \Delta k_1, ..., k_n + \Delta k_n)$$

$$- \sum_j F(k_1 + \Delta k_1, ..., k_{j-1} + \Delta k_{j-1}, k_j, k_{j+1} + \Delta k_{j+1}, ..., k_n + \Delta k_n)$$

$$+ \sum_{j<\ell} F(k_1 + \Delta k_1, ..., k_j, ..., k_\ell, ..., k_n + \Delta k_n)$$

$$- ... + (-1)^n F(k_1, k_2, ..., k_n)$$

are nonnegative. Using the function $F(k)$, one can also write formula (4.46) in the form

(4.46a) $B(\tau) = \int_{\mathbb{K}^n} e^{i\tau k} dF(k)$

where the right-hand side contains an n-dimensional Stieltjes integral defined by the relation

(4.47a) $\int_{\mathbb{K}^n} e^{i\tau k} dF(k) = \lim_{N \to \infty} \left\{ \lim_{\max |\Delta_j k| \to 0} \sum_{j=1}^{M_N} e^{i\tau k_j'} \Delta F_j(k) \right\}$

(cf. definition of the one-dimensional improper Stieltjes integral on p. 6). Thus, it follows from the multidimensional Bochner theorem that *the class of correlation functions of homogeneous random fields in* $\mathbb{R}^n$ *coincides with the class of functions of the form* (4.46), *where* $F(\Delta)$ *is a nonnegative bounded measure in the n-dimensional space* $\mathbb{K}^n$ (or, what is the same, of the form (4.46a), where $F(k)$ is a bounded function with nonnegative increments (4.49)).

We now assume for simplicity that $\langle X(t) \rangle = 0$ and note

that formula (4.46) for $B(\tau) = \langle X(t + \tau)\overline{X(t)} \rangle$ can also be rewritten as

(4.50) $B(t_1, t_2) = \langle X(t_1)\overline{X(t_2)} \rangle = \int_{\mathbb{K}^n} e^{ikt_1} e^{-ikt_2} F(dk).$

According to the theorem on the generalized spectral representation given in Note 17 to Chap. 2 (see Vol. II and also pp. 447–448 below), it follows at once from (4.50) that the homogeneous random field $X(t)$ can be represented as

(4.51) $X(t) = \int_{\mathbb{K}^n} e^{ikt} Z(dk).$

where $Z(\Delta)$ is a complex random measure in $\mathbb{K}^n$ determined, in particular, for all the n-dimensional intervals Δk and having the following properties:

(a) $\langle Z(\Delta) \rangle = 0$ for all Δ;

(b) $\langle Z(\Delta_1)\overline{Z(\Delta_2)} \rangle = 0$, if the sets Δ_1 and Δ_2 are nonintersecting;

(c) $Z(\Delta_1 + \Delta_2) = Z(\Delta_1) + Z(\Delta_2)$, if the sets Δ_1 and Δ_2 are nonintersecting.

When $\langle X(t) \rangle = m \neq 0$, we can use the representation $X(t) = \mathring{X}(t) + m$, where $\langle \mathring{X}(t) \rangle = 0$, write $\mathring{X}(t)$ in the form (4.51) and then add the constant m to both sides of (4.51), thereby passing from $\mathring{X}(t)$ back to the field $X(t)$. Further, the constant m may, if desired, be inserted into the integral in the right-hand side of (4.51) by adding the constant m to the

value of the random measure $Z(\{0\})$ corresponding to the set $\{0\}$, which consists of a single zero point $k = 0 = (0, ..., 0)$. Thus, the homogeneous field $X(t)$ with a nonzero mean value $\langle X(t) \rangle = m$ can also be represented in the form (4.51), but now the condition (a) for $Z(\Delta)$ must be replaced by:

(a$_1$) $\langle Z(\Delta) \rangle = 0$ if Δ does not contain the zero point $0 = (0, ...,$ 0)but $\langle Z(\Delta) \rangle = m$ if Δ contains the point 0.

The conditions (b) and (c), however, are valid both in the case where $\langle X(t) \rangle = 0$ and in the case where $\langle X(t) \rangle = m \neq 0$. The relationship between the complex random measure $Z(\Delta)$ and the numerical nonnegative measure $F(\Delta)$ in $\mathbb{K}^n$ or the point function $F(k)$ both for $\langle X(t) \rangle = 0$ and for $\langle X(t) \rangle \neq 0$ is given by the equations

$$(4.52) \qquad \langle |Z(\Delta)|^2 \rangle = F(\Delta), \quad \langle |Z(\Delta k)|^2 \rangle = F(\Delta k) = \Delta F(k)$$

where Δ is a set in $\mathbb{K}^n$, and $\Delta k = (\Delta k_1, ..., \Delta k_n)$ is an n-dimensional interval.

The above proof of the spectral representation (4.51) of the homogeneous random field $X(t)$ in $\mathbb{R}^n$ and the spectral representation (4.46) or (4.46a) of its correlation function $B(\tau)$ can, after some slight modification, be applied also to discrete parameter homogeneous fields on the lattice $\mathbb{Z}^n$. In this case, however, instead of the multidimensional Bochner theorem, we must use the so-called *multidimensional Herglotz theorem*. According to this theorem the function $B(\tau)$ of a vector $\tau = (\tau_1, ..., \tau_n)$ with integral components τ_j, $j = 1,$ ..., n, is positive definite (i.e. such that the inequality (4.45) is always valid) if and only if $B(\tau)$ can be represented as

$$(4.53) \qquad B(\tau) = \int_{\mathbb{K}^n(\pi)} e^{i\tau k} dF(k)$$

where $\mathbb{K}^n(\pi)$ is the n-dimensional cube $\{-\pi \leqslant k_1 \leqslant \pi, ..., -\pi \leqslant k_n \leqslant \pi\}$ and $F(k)$ is a bounded function, whose increments (4.49) are nonnegative for all the n-dimensional intervals $\Delta k = (\Delta k_1, ..., \Delta k_n)$ belonging to the cube $\mathbb{K}^n(\pi)$. The proof of the multidimensional Herglotz theorem is quite similar to the proof of the one-dimensional Herglotz theorem mentioned in Sec. 13 (see p. 181) and in Note 55 to Chap. 2. Note also that at present some more general results are known which include both the multidimensional Bochner theorem and the multidimensional Herglotz theorem as rather simple special cases (see below p. 337).

The multidimensional Herglotz theorem implies that any correlation function of a homogeneous random field on the lattice $\mathbb{Z}^n$ can be represented in the form (4.53) and that any function of the form (4.53) is a correlation function of some discrete parameter homogeneous field. Moreover, we can now use the theorem on the generalized spectral representation to show that any homogeneous field $X(t)$ on the lattice $\mathbb{Z}^n$ can be represented as the Fourier–Stieltjes integral

$$(4.54) \qquad X(t) = \int_{\mathbf{K}^n(\pi)} e^{i\mathbf{k}t} Z(d\mathbf{k})$$

where the random set function (random measure) $Z(\Delta)$ has the same properties as does the random measure $Z(\Delta)$ appearing in the right-hand side of (4.51). The function $Z(\Delta)$ is clearly related to the function $F(\mathbf{k})$ by formulae (4.52) (but now $Z(\Delta)$ is defined only within the cube $\mathbf{K}^n(\pi)$).

The function $F(\mathbf{k})$ with nonnegative increments, which appears in the integrands in (4.46a) and (4.53), is called the *spectral distribution function* of the homogeneous field $X(t)$. It is not hard to show that the knowledge of the correlation function $B(\tau)$ of a homogeneous field $X(t)$ permits one to determine also the corresponding spectral distribution function $F(\mathbf{k})$. The formula expressing $F(\mathbf{k})$ in terms of $B(\tau)$ generalizes the corresponding one-dimensional formula (2.85) (more precisely, formula (0.6') in Note 3 to the Introduction) or formula (2.215); see in this connection also Note 8 to the Introduction in Vol. II of this book.

In the most important special case where the absolute value of $B(\tau)$ falls off rapidly enough at infinity (in particular, if

$$\int_0^\infty \cdots \int_0^\infty |B(\tau_1, \ldots, \tau_n)| d\tau_1 \cdots d\tau_n < \infty \text{ or respectively,}$$

$$\sum_{\tau_1=0}^\infty \cdots \sum_{\tau_n=0}^\infty |B(\tau_1, \ldots, \tau_n)| < \infty) \text{ the } n\text{-dimensional}$$

Fourier–Stieltjes integral (4.46a) or (4.53) can be written as an ordinary n-dimensional Fourier integral of the *spectral density*

$$(4.55) \qquad f(\mathbf{k}) = f(k_1, k_2, \ldots, k_n) = \frac{\partial^n F(k_1, k_2, \ldots, k_n)}{\partial k_1 \partial k_2 \cdots \partial k_n}.$$

In other words, (4.46a) reduces here to

(4.46b) $B(\tau) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{ik\tau}f(k)dk_1...dk_n,$

and (4.53) to

(4.53a) $B(\tau) = \int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi} e^{ik\tau}f(k)dk_1 ... dk_n.$

The spectral density $f(k)$, if it exists, can be obtained from the correlation function by using the usual formula for the inversion of an n-dimensional Fourier integral:

(4.56) $f(k) = \dfrac{1}{(2\pi)^n}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} e^{-ik\tau}B(\tau)d\tau_1...d\tau_n$

or

(4.56a) $f(k) = \dfrac{1}{(2\pi)^n}\ \sum_{\tau_1=-\infty}^{\infty}\cdots\sum_{\tau_n=-\infty}^{\infty} e^{-ik\tau}\ B(\tau).$

Moreover, if the spectral density $f(k)$ exists, then formula (4.52) can be written in the form

(4.52a) $\langle|Z(dk)|^2\rangle = f(k)dk_1 ... dk_n.$

The spectral density $f(k)$ of any homogeneous random field $X(t)$ is clearly a nonnegative function of the wave vector k. Conversely, any nonnegative integrable function $f(k)$ is the spectral density of some homogeneous random field. Hence any function $B(\tau)$, having a Fourier transform which is everywhere nonnegative, is a possible correlation function of a homogeneous random field. In particular, the function (4.37) is, of course, the correlation function of a homogeneous field on $\mathbb{Z}^n$, since the Fourier transform of (4.37) is a positive constant $f(k) = 1/(2\pi)^n$. The function

(4.57)
$$B(\tau_1,...,\tau_n) = C\, a_1^{|\tau_1|} ... a_n^{|\tau_n|} ,$$

$C > 0,\ a_i$ is real and $|a_i| < 1,\quad i = 1,...,n,$

is also the correlation function of a discrete parameter homogeneous n-dimensional field, since to this function there corresponds the everywhere positive spectral density

(4.57a) $f(k_1,...,k_n) = \dfrac{C}{(2\pi)^n}\ \dfrac{(1 - a_1^2)...(1 - a_n^2)}{|e^{ik_1} - a_1|^2 ... |e^{ik_n} - a_n|^2} .$

(This example clearly generalizes Example 3 on p. 204.) Similarly, the function

$$B(\tau_1,...,\tau_n) = Ce^{-\alpha_1|\tau_1|-...-\alpha_n|\tau_n|},$$

(4.58)

$$C > 0, \quad \alpha_1 > 0, \quad i = 1...,n,$$

is the correlation function of a homogeneous random field in $\mathbb{R}^n$ since its Fourier transform $f(\mathbf{k})$ is also everywhere positive:

(4.58a) $$f(\mathbf{k}) = f(k_1,...,k_n) = \frac{C}{\pi^n} \frac{\alpha_1 \cdots \alpha_n}{(k_1^2 + \alpha_1^2) \cdots (k_n^2 + \alpha_n^2)}.$$

(This example generalizes Example 1 on p. 115.) It is also easy to find many other examples of possible correlation functions $B(\tau)$ and corresponding spectral densities $f(\mathbf{k})$.

By analogy with the theory of linear time-invariant transformations of stationary random functions considered in Secs. 11, 12, 14 and 15, one can easily develop a theory of *linear spatially invariant transformations* of homogeneous random fields. We will not dwell on this simple enough theory[20], but will just consider two specific examples of such transformations. We start with the nearest-neighbor model on the two-dimensional lattice $\mathbb{Z}^2$ described by equation (4.41). Let $X(t_1,t_2)$ be a two-dimensional homogeneous random field on $\mathbb{Z}^2$ which has spectral representation (4.54). Then it is easy to see that the linear transformation

$$\mathfrak{L}\{X(t_1,t_2)\} = X(t_1,t_2) - a_1X(t_1-1, t_2) - a_2X(t_1, t_2-1)$$

(4.59)

$$- a_3X(t_1+1, t_2) - a_4X(t_1, t_2+1)$$

converts this field into a new homogeneous field $Y(t_1,t_2) = \mathfrak{L}\{X(t_1,t_2)\}$ which has the spectral representation

$$Y(t_1,t_2) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{i(k_1t_1+k_2t_2)}[1-a_1e^{-k_1}-a_2e^{-k_2}$$

(4.60)

$$-a_3e^{k_1}-a_4e^{k_2}]\, Z(dk_1,dk_2).$$

Since the random field $cE(t_1,t_2)$ has constant spectral density $c^2/4\pi^2$, the homogeneous random field $X(t)$ satisfying the difference equation (4.41) (if such a field exists) must have

spectral density

$$(4.61) \qquad f(k_1, k_2) = \frac{c^2}{4\pi^2 \left|1 - a_1 e^{-k_1} - a_2 e^{-k_2} - a_3 e^{k_1} - a_4 e^{k_2}\right|^2} .$$

In the particular case of the laterally symmetric nearest-neighbor model, where $a_3 = a_1$, $a_4 = a_2$, formula (4.61) takes the form

$$(4.61a) \qquad f(k_1, k_2) = \frac{c^2}{4\pi^2 \left|1 - 2a_1 \cos k_1 - 2a_2 \cos k_2\right|^2} .$$

To calculate the corresponding correlation function $B(\tau_1, \tau_2)$ we now only have to use formula (4.53a).[21]

A related example can also be considered with reference to continuous parameter fields in the plane $\mathbb{R}^2$. The role of the field of uncorrelated random variable $E(t_1, t_2)$ will now be played by the two-dimensional *spatial white noise* $E(t_1, t_2)$ which is a generalized homogeneous field in $\mathbb{R}^2$ with correlation function

$$(4.62) \qquad B(\tau_1, \tau_2) = \delta(\tau_1)\delta(\tau_2)$$

where $\delta(\tau)$ is Dirac's δ-function. (The field $E(t_1, t_2)$ is quite similar to the white noise process $E(t)$; see Example 1a on p. 117. This field can be regarded, e.g., as the limit of the homogeneous fields in $\mathbb{R}^2$ with correlation functions of the form (4.58) as $\alpha_1 \to \infty$, $\alpha_2 \to \infty$, $c \to \infty$, $c/\alpha_1\alpha_2 \to 1/4$. Another, more rigorous approach to such generalized random fields is considered below in Sec. 25.) Formula (4.56) implies that the constant spectral density $f(k_1, k_2) = 1/4\pi^2$ corresponds to the correlation function (4.62). Further, by analogy with Eqs. (2.191a) and (2.197), where $X(t) = cE(t)$, we can consider also homogeneous fields $X(t_1, t_2)$ satisfying partial differential equations whose right-hand sides include a generalized field $cE(t_1, t_2)$. One of the simplest differential equations of such a type has the form

$$(4.63) \qquad \left[\frac{\partial^2}{\partial t_1^2} + \frac{\partial^2}{\partial t_2^2} - a^2\right] X(t_1, t_2) = cE(t_1, t_2).$$

It is easy to show that if the linear operator $\mathfrak{L} = (\partial^2/\partial t_1^2) + (\partial^2/\partial t_2^2) - a^2$ can be applied to the homogeneous random field $X(t_1, t_2)$ which has the spectral density $f(k_1, k_2)$, then it transforms this field into a new homogeneous random field $\mathfrak{L}X(t_1, t_2) = Y(t_1, t_2)$ having the spectral density $f_{YY}(k_1, k_2) = (k_1^2 + k_2^2 + a^2)^2 f(k_1, k_2)$.

Hence the homogeneous solution of Eq. (4.63) (if it exists) must have the spectral density

$$(4.64) \qquad f(k_1,k_2) = \frac{c^2}{4\pi^2(k_1^2 + k_2^2 + a^2)^2}.$$

The correlation function which corresponds to the spectral density (4.64) equals

$$B(\tau_1\tau_2) = \frac{c^2}{4\pi^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp\{i(k_1\tau_1 + k_2\tau_2)\}}{(k_1^2 + k_2^2 + a^2)^2} \, dk_1 dk_2$$

$$(4.65) \qquad = \frac{c^2}{4\pi} \int_0^{\infty}\int_0^{2\pi} \frac{\cos(k\tau\cos\varphi)}{(k^2 + a^2)^2} k \, d\varphi dk = \frac{c^2}{2\pi} \int_0^{\infty} \frac{J_0(k\tau)}{(k^2 + a^2)^2} k \, dk$$

$$= (c^2/2a)\tau K_1(a\tau)$$

where $\tau = |\tau| = (\tau_1^2 + \tau_2^2)^{1/2}$ and $J_0(x)$ and $K_1(x)$ are Bessel functions.[22]

The concept of a homogeneous random field can be generalized further, leading to the concept of a *multidimensional* (or *vector*) *homogeneous random field*. A *multidimensional* (say, m-dimensional) *random field* is an ordered set $\mathbf{X}(t) = \{X_1(t), ..., X_m(t)\}$ of m one-dimensional random fields called its *components*. The field $\mathbf{X}(t)$, where t is a point of the n-dimensional space $\mathbb{R}^n$ or of the lattice $\mathbb{Z}^n$, is said to be *homogeneous* if the mean values of all its components are constant, i.e.

$$(4.66) \qquad \langle X_j(t) \rangle = m_j = \text{const}, \quad j = 1, ..., m$$

and all the elements of its correlation matrix $\mathcal{B}(t,t') = \|\langle X_j(t)\overline{X_1(t')}\rangle\| = \|B_{jl}(t,t')\|$ depend only on the vector $\tau = t - t'$, i.e.

$$(4.67) \qquad \langle X_j(t + \tau)\overline{X_1(t)} \rangle = B_{jl}(\tau), \quad j,l = 1,...,m.$$

(If t is a continuous variable, i.e. $\mathbf{X}(t)$ is the field in $\mathbb{R}^n$, we shall also assume that the dependence of $B_{jl}(\tau)$ on τ is continuous for all j and l.) It is clear that the theory of multidimensional homogeneous fields can be reduced to the theory of one-dimensional such fields in the same manner as was done in the theory of multidimensional stationary

random functions (see Sec. 20). Therefore we shall not consider proofs of the results relating to multidimensional homogeneous fields, but restrict ourselves only to formulation of the main such results.

All the elements $B_{jl}(\tau)$ of the correlation matrix $B(\tau)$ of the m-dimensional homogeneous random field $X(t)$ in R^n or on Z^n can be represented as the n-dimensional Fourier-Stieltjes integrals

$$(4.68) \qquad B_{jl}(\tau) = \int e^{i\mathbf{k}\boldsymbol{\tau}} dF_{jl}(\mathbf{k}), \qquad j,l = 1,...,m,$$

where the integral in the right-hand side is evaluated over the entire wave-number vector space $\mathbf{K}^n$, if the vector parameter t is continuous, and over the cube $\mathbb{K}^n(\pi)$, if the parameter t is discrete. The functions $F_{jl}(\mathbf{k}) = F_{jl}(k, ..., k_n)$ (the cross-spectral distribution functions of the fields $X_j(t)$ and $X_l(t)$) are complex-valued functions of n variables bounded in absolute value and such that the matrix $\Delta F(\mathbf{k}) = \|\Delta F_{jl}(\mathbf{k})\|$, where the increment $\Delta F_{jl}(\mathbf{k})$ of the function $F_{jl}(\mathbf{k})$ is defined similarly to (4.49), is Hermitian and positive definite for any n-dimensional interval $\Delta\mathbf{k} = (\Delta k_1, ..., \Delta k_n)$.   The converse assertion is also true: Any matrix $B(\tau) = \|B_{jl}(\tau)\|$ whose elements have the form (4.68) where the functions $F_{jl}(\mathbf{k})$ satisfy the above conditions, is a correlation matrix of some m-dimensional homogeneous random field. If all the functions $B_{jl}(\tau)$ tend to zero fast enough as $|\tau| \to \infty$ (in particular, if $\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} |B_{jl}(\tau)| d\tau_1 ... d\tau_n < \infty$ or, respectively, $\sum_{\tau_n=-\infty}^{\infty} ... \sum_{\tau_n=-\infty}^{\infty} |B_{jl}(\tau)| < \infty$ for all j and l), then the Fourier–Stieltjes integrals (4.68) can be replaced by the ordinary Fourier integrals

$$(4.69) \qquad B_{jl}(\tau) = \int e^{i\mathbf{k}\boldsymbol{\tau}} f_{jl}(\mathbf{k}) d\mathbf{k}, \; j, l = 1,...,m$$

where

$$(4.69a) \qquad f_{jl}(\mathbf{k}) = f_{jl}(k_1,...,k_n) = \frac{\partial^n F(k_1,...,k_n)}{\partial k_1 ... \partial k_n}$$

$$= \frac{1}{(2\pi)^n} \int e^{-i\mathbf{k}\boldsymbol{\tau}} B_{jl}(\tau) d\tau$$

are complex cross-spectral densities of the fields $X_j(t)$ and

$X_l(t)$. The densities $f_{jl}(k)$, of course, must have the property that the matrix $\|f_{jl}(k)\|$ (the *spectral matrix* of $X(t)$) is Hermitian and positive definite for all k. If the field

$X(t)$ is real, $f_{jl}(-k) = f_{jl}(k)$.

Since all the components $X_j(t)$, $j=1,...,m$ of the multidimensional field $X(t)$ are one-dimensional homogeneous random fields, these components can be written as

$$(4.70) \qquad X_j(t) = \int e^{ikt} Z_j(dk)$$

where $Z_j(\Delta)$ is a random set function (measure) in the n-dimensional space $\mathbb{K}^n$ (or $\mathbb{K}^n(\pi)$) having the properties (a_1), (b), and (c) given on pp. 329–330. The *cross-spectral densities* $f_{jl}(k)$ (if they exist) and the *cross-spectral distribution function* $F_{jl}(k)$ of the multidimensional field $X(t)$ are related to the random set function $Z_j(\Delta)$ by simple formulae

$$(4.71) \qquad \langle Z_j(\Delta k)\overline{Z_l(\Delta k)} \rangle = F_{jl}(\Delta k) = \int_{\Delta k} f_{jl}(k)dk$$

where $F_{kl}(\Delta k) = \Delta F_{kl}(k)$.

It should also be noted that there exists an important generalization of Bochner's theorem used above which has a very general form and includes as special cases the theorems on positive definite functions in the space $\mathbb{R}^n$ and on the lattice $\mathbb{Z}^n$. It is clear that a translation $a \rightarrow a + x$ of the space $\mathbb{R}^n$ or the lattice $\mathbb{Z}^n$ can be associated with any vector x of $\mathbb{R}^n$ or, respectively, $\mathbb{Z}^n$ and that the set of all translations forms a commutative topological group. Consider now positive definite functions defined on a commutative topological group G. (The definition of such functions is quite evident.) Then it can be shown that under very general conditions imposed on G (in particular, in all cases where G is a commutative group with a Haar measure, e.g., an arbitrary commutative locally compact group) the general form of positive definite functions on G is given by the suitably generalized Bochner theorem with the complex exponential functions replaced by the characters of the group G.[23] This immediately implies that the theorems on the spectral representation of stationary random processes and their correlation functions can be generalized to the

case of homogeneous random fields (whose first and second moments are invariant with respect to the application of the group operation to their arguments) defined on a very wide class of commutative topological groups G.[24*]

21.2. Statistical Inference for Homogeneous Fields

We now proceed to the problem of determining the statistical characteristics of a homogeneous random field $X(t)$ from a single observed realization $x(t)$. We restrict ourselves for simplicity to the one-dimensional field $X(t)$ [25] and, moreover, assume this field to be real. Let us begin with the simplest case where the characteristic of interest is the mean value $\langle X(t) \rangle = m$. It has been already noted that a homogeneous random field on a straight line does not differ at all from a stationary random function $X(t)$; therefore only the case where $n \geqslant 2$ requires special consideration. Recall that the values of the field $X(t)$ at points on an arbitrary straight line form a one-dimensional homogeneous random field. Thus, the results of Sec. 16 imply that if the centered correlation function $b(\tau) = \langle [X(t + \tau) - m][X(t) - m] \rangle$ has the property that at least for one direction (unit vector) $\mathbf{a}$ the function $b(\mathbf{a}\tau)$ of τ satisfies Slutsky's condition (3.10a) or (3.10) (e.g., if $|b(\tau)| \to 0$ as $|\tau| \to \infty$ at least along one direction), then the mean value $\langle X(t) \rangle$ can be determined with any degree of accuracy by averaging the values of $x(t)$ over a sufficiently long section of a straight line parallel to $\mathbf{a}$. It is clear, however, that the determination of the mean value $\langle X(t) \rangle$ by averaging the function $x(t)$ along a straight line is hardly expedient, if the values of $x(t)$ are in fact known within some multidimensional region. Assume, e.g., that $n = 2$ (the case of a higher dimension can be treated quite similarly) and that the realization $x(t_1, t_2)$ is observed for all t_1, t_2 from the rectangle $0 < t_1 \leqslant T_1$, $0 < t_2 \leqslant T_2$. Then, if the mean value $\langle X(t_1, t_2) \rangle = m$ is unknown, it is natural to use the following estimate of it:

$$m^*_{T_1 T_2} = \frac{1}{T_1 T_2} \int_0^{T_1} \int_0^{T_2} x(t_1, t_2) dt_1 dt_2 \quad \text{or}$$

(4.72)

$$= \frac{1}{T_1 T_2} \sum_{t_1 = 1}^{T_1} \sum_{t_2 = 1}^{T_2} x(t_1, t_2).$$

The estimate $m^*_{T_1 T_2}$ is clearly unbiased, i.e. the corresponding estimator

$$M^*_{T_1 T_2} = \frac{1}{T_1 T_2} \int_0^{T_1} \int_0^{T_2} X(t_1,t_2) dt_1 dt_2 \quad \text{or}$$

(4.73)

$$= \frac{1}{T_1 T_2} \sum_{t_1=1}^{T_1} \sum_{t_2=1}^{T_2} X(t_1,t_2)$$

always satisfies the condition

$$\langle M^*_{T_1 T_2} \rangle = \langle X(t_1,t_2) \rangle = m.$$

However, the estimate $m^*_{T_1 T_2}$ will be practically useful only if

it is consistent, i.e. if the estimator $M^*_{T_1 T_2}$ tends (in the mean)

to m as $T_1 \to \infty$, $T_2 \to \infty$. If the function $b(\tau,0)$ (or $b(0,\tau)$) of τ
satisfies the Slutsky condition (3.10a) or (3.10), then evidently
$M^*_{T_1 T_2} \to m$ even for $T_1 \to \infty$ and fixed T_2 (or for $T_2 \to \infty$ and

fixed T_1). The general necessary and sufficient condition for
convergence in the mean of $M^*_{T_1 T_2}$ to m as $T_1 \to \infty$, $T_2 \to \infty$ in

the case of a homogeneous field in the plane takes the form

(4.74) $\quad \lim\limits_{T_1 \to \infty, T_2 \to \infty} \dfrac{1}{2T_1 T_2} \int_0^{T_1} \int_{-T_2}^{T_2} b(\tau_1,\tau_2) d\tau_1 d\tau_2 = 0$

or, respectively,

(4.74a) $\quad \lim\limits_{T_1 \to \infty, T_2 \to \infty} \dfrac{1}{2T_1 T_2} \sum\limits_{\tau_1=0}^{T_1-1} \sum\limits_{\tau_2=-T_2+1}^{T_2-1} b(\tau_1,\tau_2) = 0.$

The proof of this condition differs little from the analogous
proof for the one-dimensional case. It can be easily deduced
from the following exact formula for the variance (i.e. the
mean-square error) of $M^*_{T_1 T_2}$:

$$\sigma^2(M^*_{T_1 T_2}) = \langle (M^*_{T_1 T_2} - m)^2 \rangle$$

(4.75)

$$= \frac{1}{T_1^2 T_2^2} \int_0^{T_2} \int_{-T_2'}^{T_2'} \int_0^{T_1} \int_{-T_1'}^{T_1'} b(\tau_1,\tau_2) d\tau_1 dT_1' d\tau_2 dT_2'$$

$$= \frac{1}{T_1^2 T_2^2} \int_{-T_2'}^{T_2} \int_{-T_1}^{T_1} (T_1 - |\tau_1|)(T_2 - |\tau_2|) b(\tau_1,\tau_2) d\tau_1 d\tau_2$$

or, respectively,

$$\sigma^2(M^*_{T_1 T_2}) = \frac{1}{T_1^2 T_2^2} \sum_{T_2'=0}^{T_2-1} \sum_{\tau_2=-T_2'}^{T_2'} \sum_{T_1'=0}^{T_1-1} \sum_{\tau_1=-T_1'}^{T_1'} b(\tau_1, \tau_2)$$

(4.75a)

$$= \frac{1}{T_1^2 T_2^2} \sum_{\tau_2=-T_2+1}^{T_2-1} \sum_{\tau_1=-T_1+1}^{T_1-1} (T_1 - |\tau_1|)(T_2 - |\tau_2|)b(\tau_1 \tau_2)$$

(cf. p. 218–219 and Notes 4 and 5 to Chap. 3 in Vol. II).[26]
Assume now that the argument $t = (t_1, t_2)$ is continuous and that

(4.76) $\quad \dfrac{1}{b(0,0)} \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} b(\tau_1, \tau_2) d\tau_1 d\tau_2 = 4L^2$

is finite and different from zero. (If the field $\overset{\circ}{X}(t_1, t_2) = X(t_1, t_2) - m$ has a spectral density $f(k_1, k_2)$, then obviously $4L^2$

$= 4\pi^2 f(0,0)/b(0,0)$, and hence L^2 is finite and different from zero if and only if $0 < f(0,0) < \infty$.) It is clear that L^2 has the dimension of an area; it is hence natural to call it a *correlation area* (or an *integral area scale*). Assuming the averaging lengths T_1 and T_2 to be sufficiently large, we can obtain the approximate formula

(4.77) $\quad \sigma^2(M^*_{T_1 T_2}) \approx \dfrac{4L^2}{T_1 T_2} b(0,0) = \dfrac{4L^2}{T_1 T_2} \sigma_x^2.$

The derivation of (4.77) is quite similar to that of (3.18) and using (4.77) we may often estimate simply the root-mean-square error of the estimator $M^*_{T_1 T_2}$. A formula of the form

(4.77) will obviously also be valid in the case of the discrete parameter $t = (t_1, t_2)$, but now L^2 is dimensionless and must be defined as follows:

(4.76a) $\quad \dfrac{1}{b(0,0)} \displaystyle\sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} b(\tau_1, \tau_2) = 4L^2.$

Formula (4.77) is, of course, inapplicable if $L^2 = \infty$ or $L^2 = 0$ (i.e., if the integral on the right-hand side of (4.76) or the sum on the right-hand side of (4.76a) diverges or is equal to zero). If $L^2 = 0$, then the variance $\sigma^2(M^*_{T_1 T_2})$ decreases faster than $T_1^{-1} T_2^{-1}$ as $T_1 \to \infty$, $T_2 \to \infty$, while if $L^2 = \infty$, this variance decreases slower than $T_1^{-1} T_2^{-1}$. In particular, if the field $\overset{\circ}{X}(t_1, t_2) = X(t_1, t_2) - m$ has a spectral density $f(k_1, k_2)$, which is

proportional to $k_1^{\beta_1} k_2^{\beta_2}$ in the neighborhood of the point $k_1 = 0$, $k_2 = 0$, then the rate of decrease of $\sigma^2(M^*_{T_1 T_2})$ as $T_1 \to \infty$, $T_2 \to \infty$ is simply expressed through the exponents β_1 and β_2.[27]

*If the parameter **t** is continuous, then, of course, instead of averaging over the entire rectangle $0 \leqslant t_1 \leqslant T_1$; $0 \leqslant t_2 \leqslant T_2$ one may average the data over some discrete set of points. The simplest version is that the segments $0 \leqslant t_1 \leqslant T_1$ and $0 \leqslant t_2 \leqslant T_2$ are divided into N_1 and, respectively, N_2 equal parts by points $t_1^{(1)}, t_1^{(2)}, ..., t_1^{(N_1-1)}$ and $t_2^{(1)}, t_2^{(2)}, ..., t_2^{(N_2-1)}$, and then one uses the estimate

$$(4.78) \qquad m^*_{T_1 T_2 ; N_1 N_2} = \frac{1}{(N_1 + 1)(N_2 + 1)} \sum_{i=0}^{N_1} \sum_{j=0}^{N_2} x(t_1^{(i)}, t_2^{(j)}),$$

where $t_1^{(0)} = t_2^{(0)} = 0$, $t_1^{(N_1)} = T_1$, $t_2^{(N_2)} = T_2$. The estimate (4.78), which is similar to the estimate (3.13b) of the mean value of a stationary random process, is unbiased and consistent (tends to m as $T_1 \to \infty$, $T_2 \to \infty$ and $N_1 \to \infty$, $N_2 \to \infty$).[28] There are, however, also many other finite point sets making up regular patterns different from the rectangular lattice $(iT_1/N_1, jT_2/N_2)$, $i = 0,1, ..., N_1$, $j = 0,1, ..., N_2$, and averaging over any such set can be used to estimate the mean value m. A still greater number of possibilities appears when the dimensionality n of the argument **t** exceeds two. The observed field values at points making up a regular pattern form a *systematic sample*; besides this, *random samples* of a field are also used rather often, i.e. the values of the field are observed at the points $\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_N$ selected randomly in accordance with some fixed probability distribution. In practice it is often very desirable to restrict oneself to the least possible number of observations; therefore, investigation of various systematic and random samples of homogeneous fields yielding small enough mean square errors of the corresponding statistical estimates is of great interest.[29]

Note also that the probability distribution of the mean value estimators $M^*_{T_1 T_2}$ and $M^*_{T_1 T_2 ; N_1 N_2}$ for large values of T_1, T_2 or T_1, T_2 and N_1, N_2 (as well as of many other

estimators of the mean value obtained by averaging the field values over a sufficiently large point set) is usually well approximated by a normal probability distribution. This is a simple consequence of the applicability of the central limit theorem to many homogeneous fields $X(t)$.[30] Moreover, under wide conditions, random functions $M^*_{T_1 T_2}$

and $M^*_{T_1 T_2; N_1 N_2}$ of the arguments T_1, T_2 or T_1, T_2 and N_1, N_2

converge to the mean value $m = \langle X(t) \rangle$ as $T_1 \to \infty, T_2 \to \infty$ or $T_1 \to \infty, T_2 \to \infty$, and $N_1 \to \infty, N_2 \to \infty$ not only in the mean, but also almost surely, i.e. with probability one. (In other words, the functions $m^*_{T_1 T_2}$ and $m^*_{T_1 T_2; N_1 N_2}$ calculated from a given

realization $x(t)$ converge almost surely to m as $T_1 \to \infty, T_2 \to \infty$ or $T_1 \to \infty, T_2 \to \infty$, and $N_1 \to \infty, N_2 \to \infty$; cf. p. 224, where a similar statement for stationary functions $X(t)$ is discussed.) We will not, however, pursue this problem here.[31*]

Consider now the problem of calculating the correlation function $B(\tau) = \langle X(t + \tau) X(t) \rangle$, $\tau = (\tau_1, \tau_2)$ of a homogeneous field $X(t) = X(t_1, t_2)$ from a single realization $x(t_1, t_2)$. If the field $X(t)$ is homogeneous at least up to the fourth order,* this problem is easily reduced to the problem of calculating from a single realization the mean value of a homogeneous field $Y_\tau(t) = X(t + \tau) X(t)$ depending on the vector parameter τ. It is clear that all the considerations applied above to the field $X(t)$ can be applied to this new field as well. We note, however, that if the values of $x(t_1, t_2)$ are known to us only on some limited point set, the values of $y_\tau(t) = x(t + \tau) x(t)$ will be known on another, smaller set depending on τ. Consider, e.g., the case where the values of $x(t_1, t_2)$ are observed in a rectangle (say, at $0 \leqslant t_1 \leqslant T_1, 0 \leqslant t_2 \leqslant T_2$) and first assume that the parameter $t = (t_1, t_2)$ is continuous. The values of $y_\tau(t_1, t_2) = x(t_1 + \tau_1, t_2 + \tau_2) x(t_1, t_2)$ will be known then only for $0 \leqslant t_1 \leqslant T_1 - \tau_1, 0 \leqslant t_2 \leqslant T_2 - \tau_2$ if $0 \leqslant \tau_1 \leqslant T_1, 0 \leqslant \tau_2 \leqslant T_2$; for $0 \leqslant t_1 \leqslant T_1 - \tau_1, -\tau_2 \leqslant t_2 \leqslant T_2$ if $0 \leqslant \tau_1 \leqslant T_1, -T_2 \leqslant \tau_2 \leqslant 0$; for $-\tau_1 \leqslant t_1 \leqslant T_1, 0 \leqslant t_2 \leqslant T_2 - \tau_2$, if $-T_1 \leqslant \tau_1 < 0, 0 \leqslant t_2 \leqslant T_2$; for $-\tau_1 \leqslant t_1 \leqslant T_1, -\tau_2 \leqslant t_2 \leqslant T_2$, if $-T_1 \leqslant \tau_1 < 0, -T_2 \leqslant \tau_2 < 0$; and will not be

---

*I.e., if all the moments $\langle X(t) \rangle, \langle X(t + \tau) X(t) \rangle, \langle X(t + \tau) X(t + \tau_1) X(t) \rangle$, and $\langle X(t + \tau) X(t + \tau_1) X(t + \tau_2) X(t) \rangle$ are independent of t.

known at all if $|\tau_1| > T_1$ or $|\tau_2| > T_2$. Therefore, by analogy with the estimate $B_T^{**}(\tau)$ of the correlation function $B(\tau) = \langle [X(t + \tau)X(t) \rangle$ of a stationary random process $X(t)$ (see (3.33a)) it is expedient to use the following statistics as an estimate of the function $B(\tau) = B(\tau_1, \tau_2)$:

(4.79)
$$B_{T_1 T_2}^{**}(\tau_1, \tau_2) = \begin{cases} \dfrac{1}{T_1 T_2} \int_{\tau_1'}^{T_1'} \int_{\tau_2'}^{T_2'} x(t_1 + \tau_1, t_2 + \tau_2) x(t_1, t_2) dt_1 dt_2 \\ \qquad\qquad \text{for } |\tau_1| \leqslant T_1, |\tau_2| \leqslant T_2, \\[2mm] 0 \qquad\qquad \text{for } |\tau_1| > T_1 \text{ or } |\tau_2| > T_2, \end{cases}$$

where $\tau_1' = \max(0, -\tau_1)$, $\tau_2' = \max(0, -\tau_2)$, $T_1' = \min(T_1, T_1 - \tau_1)$, $T_2' = \min(T_2, T_2 - \tau_2)$. Similarly, in the case of a discrete parameter homogeneous field $X(t_1, t_2)$ observed at $t_1 = 1, ..., T_1$; $t_2 = 1, ..., T_2$, a reasonable estimate of the correlation function $B(\tau_1, \tau_2)$ will be

(4.79a) $$B_{T_1 T_2}^{**}(\tau_1, \tau_2) = \begin{cases} \dfrac{1}{T_1 T_2} \sum_{t_1 = \tau_1'}^{T_1'} \sum_{t_2 = \tau_2'}^{T_2'} x(t_1 + \tau_1, t_2 + \tau_2) x(t_1, t_2) \\ \qquad\qquad \text{for } |\tau_1| \leqslant T_1, |\tau_2| \leqslant T_2, \\[2mm] 0 \qquad\qquad \text{for } |\tau_1| > T_1 \text{ or } |\tau_2| > T_2, \end{cases}$$

where $\tau_1' = \max(1, 1 - \tau_1)$, $\tau_2' = \max(1, 1 - \tau_2)$, $T_1' = \min(T_1, T_1 - \tau_1)$, $T_2' = \min(T_2, T_2 - \tau_2)$. The estimate $B_{T_1 T_2}^{**}(\tau_1, \tau_2)$ is obviously biased, but is it asymptotically unbiased and consistent, i.e. it has the same properties as does the estimate $B_T^{**}(\tau)$ of the function $B(\tau)$ of one variable. It is easy to see that the bias of the estimate $B_{T_1 T_2}^{**}(\tau_1, \tau_2)$ is equal to $\{(T_1 - |\tau_1|)(T_2 - |\tau_2|)/T_1 T_2\} B(\tau_1, \tau_2)$, while its variance may be simply expressed through the fourth moments of the field $X(t_1, t_2)$. If, however, this field is Gaussian, then the variance of $B_{T_1 T_2}^{**}(\tau_1, \tau_2)$ may be expressed through the mean value m and the correlation function $B(\tau_1, \tau_2)$ or $b(\tau_1, \tau_2) = B(\tau_1, \tau_2) - m^2$. [32]

As in the case of the estimate $B_T^{**}(\tau)$ of the function $B(\tau)$, the estimate $B_{T_1 T_2}^{**}(\tau_1, \tau_2)$ is unreliable for relatively large values of $|\tau_1|$ and $|\tau_2|$, which constitute an appreciable part of T_1 and, respectively, T_2. Therefore it is often advisable to use the following modified estimate in place of $B_{T_1 T_2}^{**}(\tau_1, \tau_2)$:

(4.80) $B^{(a)}_{T_1 T_2}(\tau_1, \tau_2) = a_{T_1 T_2}(\tau_1, \tau_2) B^{**}_{T_1 T_2}(\tau_1, \tau_2)$

where $a_{T_1 T_2}(\tau_1, \tau_2)$ is a two-dimensional *lag window*. This lag window can be chosen, e.g., to be of a form $a_{T_1 T_2}(\tau_1, \tau_2) = a_1(\tau_1/k_{T_1}) a_2(\tau_2/k_{T_2})$ where k_{T_1} and k_{T_2} are adjustable positive scale parameters, and $a_1(\tau)$ and $a_2(\tau)$ are some one-dimensional lag window generators considered in Secs. 17 and 19.

Let us now go over to the estimation of the spectral density. Similarly to the situation relating to the case of stationary processes and sequences, the most widely used estimates of the spectral density $f(\mathbf{k}) = f(k_1, k_2)$ of a homogeneous random field $X(\mathbf{t}) = X(t_1, t_2)$ are obtained by applying the Fourier transformation to some appropriate estimates of the correlation function $B(\tau_1, \tau_2)$. Assume that the spectral density of a field $X(\mathbf{t}) = X(t_1, t_2)$ does exist . (It follows from this, in particular, that the mean value of the field $X(\mathbf{t})$ is equal to zero.) It is not hard to show that the two-dimensional Fourier transform of the estimate $B^{**}_{T_1 T_2}(\tau_1, \tau_2)$ can be represented as

$$i_{T_1 T_2}(k_1, k_2) = \frac{1}{4\pi^2} \int_{-T_1}^{T_1} \int_{-T_2}^{T_2} e^{-i(k_1 \tau_1 + k_2 \tau_2)} B^{**}_{T_1 T_2}(\tau_1, \tau_2) d\tau_1 d\tau_2$$

(4.81)

$$= \frac{1}{4\pi^2 T_1 T_2} \left| \int_0^{T_1} \int_0^{T_1} e^{-i(k_1 t_1 + k_2 t_2)} x(t_1, t_2) dt_1 dt_2 \right|^2$$

or, if the parameter $\mathbf{t} = (t_1, t_2)$ is discrete, as

$$i_{T_1 T_2}(k_1, k_2) = \frac{1}{4\pi^2} \sum_{\tau_1 = -T_1 + 1}^{T_1 - 1} \sum_{\tau_2 = -T_2 + 1}^{T_2 - 1} e^{-i(k_1 \tau_1 + k_2 \tau_2)} B^{**}_{T_1 T_2}(\tau_1, \tau_2)$$

(4.81a)

$$= \frac{1}{4\pi^2 T_1 T_2} \left| \sum_{t_1 = 1}^{T_1} \sum_{t_2 = 1}^{T_2} e^{-i(k_1 t_1 + k_2 t_2)} x(t_1, t_2) \right|^2.$$

The non-negativity of the right-hand sides of (4.81) and (4.81a) shows that the function $B^{**}_{T_1 T_2}(\tau_1, \tau_2)$ is a positive definite function of two variables, i.e. it is a possible

correlation function of a homogeneous random field in the plane. It does not imply, however, that its Fourier transform (4.81) or (4.81a) − the *two-dimensional periodogram* − is a satisfactory estimate for the spectral density $f(k_1, k_2)$. Let $I_{T_1 T_2}(k_1, k_2)$ be the *periodogram estimator*, i.e. a random

variable obtained when the realization $x(t_1, t_2)$ in the right-hand sides of (4.81) and (4.81a) is replaced by the random field $X(t_1, t_2)$. Then is is not hard to show that

$$(4.82) \qquad \lim_{T_1 \to \infty, T_2 \to \infty} \langle I_{T_1 T_2}(k_1, k_2) \rangle = f(k_1, k_2),$$

but the variance $\sigma^2\{I_{T_1 T_2}(k_1, k_2)\}$ does not tend to zero as $T_1 \to$

∞ and $T_2 \to \infty$. In fact, under wide conditions imposed on the field $X(t_1, t_2)$ (in particular, in the case where this field is Gaussian)

$$(4.83) \qquad \lim_{T_1 \to \infty, T_2 \to \infty} \sigma^2\{I_{T_1 T_2}(k_1, k_2)\} = f^2(k_1, k_2)$$

for all the values of (k_1, k_2) except the value $(k_1, k_2) = (0,0)$ (and, in the case of discrete $t = (t_1, t_2)$, also the values $(k_1, k_2) = (-\pi, -\pi)$, $(-\pi, 0)$, $(-\pi, \pi)$, $(0, -\pi)$, $(0, \pi)$, $(\pi, -\pi)$, $(\pi, 0)$, and (π, π)), where the right-hand side of (4.83) must be doubled. The proof of formulae (4.82) and (4.83) differs little from the analogous proof for the one-dimensional case (cf. (3.39), (3.39a), and (3.41)).[33]

Formula (4.83) shows that a two-dimensional periodogram is not a consistent estimate of the spectral density. The known results of statistical spectral analysis of stationary sequences and processes (see Sec. 18) suggest that the Fourier transform of the modified estimate $B_{T_1 T_2}^{(a)}(\tau_1, \tau_2)$ of the

correlation function $B(\tau_1, \tau_2)$ would be a consistent estimate of the density $f(k_1, k_2)$ if the two-dimensional lag window $a_{T_1 T_2}(\tau_1, \tau_2)$ is appropriately selected. We assume for

definiteness that $t = (t_1, t_2)$ is discrete (the changes in the formulae necessary in the case of continuous t are quite evident and therefore can be omitted here). Then it is easy to show that the above mentioned estimate of $f(k_1, k_2)$ has the form

$$\varphi^{(A)}_{T_1 T_2}(k_1, k_2)$$

$$(4.84) = \frac{1}{4\pi^2} \sum_{\tau_1 = -T_1+1}^{T_1-1} \sum_{\tau_2 = -T_2+1}^{T_2-1} e^{-i(k_1 \tau_1 + k_2 \tau_2)} a_{T_1 T_2}(\tau_1, \tau_2) B^{**}_{T_1 T_2}(\tau_1, \tau_2)$$

$$= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} A_{T_1 T_2}(k_1 - k_1', \ k_2 - k_2') i_{T_1 T_2}(k_1', k_2') dk_1' dk_2'$$

where

$$(4.85) \quad A_{T_1 T_2}(k_1, k_2) = \frac{1}{4\pi^2} \sum_{\tau_1 = -\infty}^{\infty} \sum_{\tau_2 = -\infty}^{\infty} e^{-i(k_1 \tau_1 + k_2 \tau_2)} a_{T_1 T_2}(\tau_1, \tau_2)$$

is a two-dimensional *spectral window* corresponding to the lag window $a_{T_1 T_2}(\tau_1, \tau_2)$. If the spectral window concentrates more and more closely about the point $k_1 = 0$, $k_2 = 0$ as $T_1 \to \infty$, $T_2 \to \infty$ (i.e. $A_{T_1 T_2}(k_1, k_2)$ has a sharp peak at the point $k_1 = k_2 = 0$ and falls off rapidly with $k = (k_1^2 + k_2^2)^{1/2}$), then under wide conditions

$$(4.86) \quad \langle \Phi^{(A)}_{T_1 T_2}(k_1, k_2) \rangle \approx f(k_1, k_2) \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} A_{T_1 T_2}(k_1, k_2) dk_1 dk_2$$

and

$$(4.87) \quad \sigma^2\{\Phi^{(A)}_{T_1 T_2}(k_1, k_2)\} \approx \frac{4\pi^2}{T_1 T_2} f^2(k_1, k_2) \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} A^2_{T_1 T_2}(k_1, k_2) dk_1 dk_2$$
$$\text{for } k_1 \neq 0, \pm\pi; \ k_2 \neq 0, \pm\pi,$$

where $\Phi^{(A)}_{T_1 T_2}(k_1, k_2)$ is an estimator corresponding to the estimate $\varphi^{(A)}_{T_1 T_2}(k_1, k_2)$, i.e. a random variable obtained by replacement of $i_{T_1 T_2}(k_1, k_2)$ by $I_{T_1 T_2}(k_1, k_2)$ in the right-hand side of (4.84). (In the case where $k_1, k_2 = 0, \pm\pi$, the right-hand side of (4.87) must be doubled.) Formulae (4.86) and (4.87) show that the estimate $\Phi^{(A)}_{T_1 T_2}(k_1, k_2)$ of the spectral density $f(k_1, k_2)$ is indeed consistent provided that

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}A_{T_1T_2}(k_1,k_2)dk_1dk_2 = 1,$$

(4.88)

$$T_1^{-1}T_2^{-1}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}A_{T_1T_2}^2(k_1,k_2)dk_1dk_2 \to 0 \text{ as } T_1 \to \infty, \ T_2 \to \infty.$$

To calculate the estimate $\varphi_{T_1T_2}^{(A)}(k_1,k_2)$ in practice, it is, of

course, more convenient to start from the right-hand side of (4.84) and use the fast Fourier transform (FFT) algorithm to evaluate the periodogram $i_{T_1T_2}(k_1,k_2)$ (see Sec. 19). A number

of specific examples of the spectral density estimation by formula (4.84) for actually observed two-dimensional homogeneous random fields can be found in the available applied literature.[34]

Estimation of the spectral density of a homogeneous random field in the plane by periodogram smoothing occupies up to now the central place in the spectral analysis of such fields. Recently, however, much attention has also been given to the development of other methods for such estimation, which are practically more convenient and/or provide a better resolvability and can be applied successfully even in the case of a relatively small sample (i.e., when only a few observations of the field are available). In particular, a number of authors have studied *spatial autoregressive spectral estimates* based on the assumption that the estimated spectral density $f(k_1,k_2)$ can be approximated accurately enough by a formula of the form

(4.89) $$f(k_1,k_2) = \frac{c^2}{4\pi^2\left|1 + \sum\limits_{j_1=-n_1}^{n_1}\sum\limits_{j_2=-n_2}^{n_2}a_{j_1j_2}e^{-i(j_1k_1+j_2k_2)}\right|^2}$$

where c^2, n_1, n_2 and $a_{j_1j_2}$ are some numerical parameters. (The

form (4.89) of the spectral density obviously corresponds to the spatial autoregressive model described by the partial difference equation

(4.90) $$X(t_1,t_2) + \sum\limits_{j_1=-n_1}^{n_1}\sum\limits_{j_2=-n_2}^{n_2}a_{j_1j_2}X(t_1-j_1, \ t_2-j_2) = cE(t_1,t_2);$$

cf., e.g., (4.41) and (4.61).) The assumption that $f(k_1,k_2)$ has the form (4.89) evidently allows us to reduce the spectral estimation problem to the traditional statistical problem of

estimating unknown parameters of the probabilistic model from the data. An alternative approach to the approximation of an unknown spectral density by a function of the form (4.89) is based on preliminary estimation of a finite (and usually comparatively small) number of correlations $B(\tau_1, \tau_2)$ and subsequent finding of the density $f(k_1, k_2)$ which agrees with the obtained values of $B(\tau_1, \tau_2)$ and maximizes the "entropy"

$$(4.91) \qquad H = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \ln f(k_1, k_2) dk_1 dk_2$$

(the *maximum entropy method* of two-dimensional spectral estimation). A number of recent papers are devoted to the examination of autoregressive (maximum entropy) estimates for two-dimensional spectral density and to the study of some other related multidimensional parametric spectral estimates. These papers contain many additional interesting results; we cannot, however, consider these problems in more detail in this book which, as it is, proves to be rather voluminous.[35]

22. Isotropic Random Fields

22.1. Spectral Representation of Isotropic Correlation Functions and Its Consequences

In Sec. 21 we have already given an example of a homogeneous random field $X(t)$ whose correlation function $B(\tau) = \langle X(t + \tau) X(t) \rangle$ depends only on the length $\tau = |\tau|$ of the vector τ and not on its direction; see (4.65) and also Note 22 to Chap. 4. If the correlation function $B(\tau)$ of the homogeneous random field $X(t)$ in $\mathbb{R}^n$ has this property, so that

$$(4.92) \qquad \langle X(t + \tau) \overline{X(t)} \rangle = B(\tau), \quad \tau = |\tau|,$$

then the field $X(t)$ is said to be an *isotropic random field* in $\mathbb{R}^n$.[36] The corresponding correlation function $B(\tau)$ is called in this case an *isotropic correlation function in* $\mathbb{R}^n$ (or an n-*dimensional isotropic correlation function*). For an isotropic random field in $\mathbb{R}^n$, all directions in space are obviously equivalent. Such fields play an important role in the statistical theory of turbulence, and they are also often encountered in many other applied areas.[37]

Further on, we consider only random fields having a

continuous correlation function; therefore, the funtion $B(\tau)$ will always be assumed a continuous function of τ. Since by definition of the correlation function $B(-\tau) = \overline{B(\tau)}$ for any homogeneous field and $|\tau| = |-\tau| = \tau$, the isotropic correlation function $B(\tau)$ is necessarily real. Let us now consider the values of an isotropic field $X(t)$ at points on some straight line in $\mathbb{R}^n$, say the values $X(t_1, 0, \ldots, 0) = X(t_1)$ at points on the first coordinate axis. These values clearly form a homogeneous random field on the straight line (i.e. a stationary process in the variable t_1) having correlation function $B(\tau)$. Hence, for any n the *n-dimensional isotropic correlation function* $B(\tau)$ *is necessarily also a correlation function of a stationary random process*, i.e., it is a (real) positive definite function of τ which can be represented in the form

$$(4.93) \qquad B(\tau) = \int_{-\infty}^{\infty} e^{ik_1\tau} dF_1(k_1) = \int_0^{\infty} \cos k_1\tau \, dG_1(k_1),$$

where $F_1(k_1)$ and $G_1(k_1)$ are real nondecreasing functions and $dG_1(k_1) = 2dF_1(k_1)$ for any $k_1 > 0$ (cf. (2.52) and (2.53)). If the absolute value of $B(\tau)$ falls off rapidly enough as $\tau \to \infty$, so that

$$(4.94) \qquad \int_0^{\infty} |B(\tau)| d\tau < \infty \text{ or at least } \int_0^{\infty} |B(\tau)|^2 d\tau < \infty,$$

then the functions $F_1(k_1)$ and $G_1(k_1)$ are absolutely continuous (i.e., differentiable and equal to an indefinite integral of their derivatives $f_1(k_1) = dF(k_1)/dk_1$ and $g_1(k_1) = dG_1(k_1)/dk_1 = 2f_1(k_1)$). Therefore, in this case, (4.93) can also be written as

$$(4.95) \qquad B(\tau) = \int_{-\infty}^{\infty} e^{ik_1\tau} f_1(k_1) dk_1 = \int_0^{\infty} \cos k_1\tau \, g_1(k_1) dk_1.$$

The function $F_1(k_1)$ (as well as $G_1(k_1)$) is called the *one-dimensional spectral distribution function* of the isotropic field $X(t)$; similarly, the function $f(k_1)$ (as well as $g_1(k_1) = 2f_1(k_1)$) is called the *one-dimensional spectral density* of $X(t)$.

For $n = 1$, the class of all (one-dimensional) isotropic correlation functions clearly coincides with the class of all real correlation functions of stationary processes, i.e. it differs from the class of all correlation functions of homogeneous fields on a straight line only by an additional requirement that the function $B(\tau)$ be real. If, however, $n > 1$, then not every real positive definite function (i.e. a function of the form (4.93) where $F_1(k_1)$ and $G_1(k_1)$ are

nondecreasing) can be an isotropic correlation function in $\mathbb{R}^n$. Let us begin with the case where $n = 2$. Here $B(\tau) = B(\tau_1, \tau_2) = B(\tau)$, where $\tau = |\tau| = (\tau_1^2 + \tau_2^2)^{1/2}$ while the function $B(\tau)$ can be represented as

$$B(\tau) = B(\tau_1, \tau_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(k_1\tau_1 + k_2\tau_2)} dF(k_1, k_2)$$

(4.96)

$$= \iint_{\mathbb{K}^2} e^{ik\tau\cos\Theta_{k\tau}} dF(k_1, k_2).$$

In the last equation, $F(k_1, k_2) = F(\mathbf{k})$ is a real bounded function of two variables, whose increments on all two-dimensional intervals are nonnegative, $k = |\mathbf{k}|$, $\tau = |\tau|$, and $\Theta_{k\tau}$ is the angle between vectors $\mathbf{k}$ and τ. Since $B(\tau) = B(\tau) = B([\tau_1^2 + \tau_2^2]^{1/2})$,

$$B(\tau\cos\varphi, \tau\sin\varphi) = B(\tau)$$

for any φ and therefore

$$B(\tau) = \frac{1}{2\pi}\int_0^{2\pi} B(\tau\cos\varphi, \tau\sin\varphi) d\varphi.$$

It is clear that, if φ runs through all the values from 0 to 2π, then, for any value of $\mathbf{k}$, the angle $\Theta_{k\tau}$ between $\mathbf{k}$ and τ also runs through all the values from 0 to 2π. Hence,

(4.97) $$B(\tau) = \iint_{\mathbb{K}^2} \{\frac{1}{2\pi}\int_0^{2\pi} e^{ik\tau\cos\Theta_{k\tau}} d\Theta_{k\tau}\} dF(\mathbf{k}) = \iint_{\mathbb{K}^2} J_0(k\tau) dF(\mathbf{k})$$

i.e.,

(4.98) $$B(\tau) = \int_0^{\infty} J_0(k\tau) d\Phi(k),$$

where J_0 is the zero order Bessel function of the first kind

and $\Phi(k) = \iint_{|\mathbf{k}| < k} dF(\mathbf{k})$ is a bounded nondecreasing function on the half-line $0 \leqslant k < \infty$. (Formula 8.411.1 from Gradshteyn and Ryzhik, 1980, is used in deriving (4.98).*) Equation (4.97) also

---

*Formulae from the book by Gradshteyn and Ryzhik (1980) cited in this section and in Notes to it can also be found in the well-known books by Watson (1966), Abramowitz and Stegun (1964), and Erdélyi et al. (1953).

implies that any function $B(\tau) = B(\tau)$ which can be represented in the form (4.98), where $\Phi(k)$ is a bounded nondecreasing function, can be represented in the form (4.96) as well. Indeed, this only requires that the function $F(k_1, k_2) = F(\mathbf{k})$ be chosen in such a way that the corresponding measure in the plane $\mathbb{K}^2$ (i.e., the function $F(\Delta)$ of the two-dimensional set Δ) be rotationally symmetric (i.e. such that $F(\Delta) = F(g\Delta)$ for any Δ and any rotation g about the coordinate origin in $\mathbb{K}^2$) and satisfy the condition $F(S_k) = \Phi(k)$, where $\mathbf{k}$ belongs to the set S_k if and only if $|\mathbf{k}| < k$. (Since both the function $F(\mathbf{k})$, and the measure $F(\Delta)$ are determined uniquely by the correlation function $B(\tau)$, it follows from the above result that the measure $F(\Delta)$ corresponding to any isotropic field in the plane is rotationally symmetric.) The possibility of representing the right-hand side of (4.98) in the form of the right-hand side of (4.96) implies that *the class of all two-dimensional isotropic correlation functions coincides with the class of function* B(τ) *of the form* (4.98), *where* $\Phi(\mathbf{k})$ *is a bounded nondecreasing function of* **k**.

Similar reasoning can also be applied to the three-dimensional isotropic correlation function $B(\tau) = B(\tau_1, \tau_2, \tau_3) = B([\tau_1^2 + \tau_2^2 + \tau_3^2]^{1/2}) = B(\tau)$. Since $B(\tau) = B(\tau_1, \tau_2, \tau_3)$ is the correlation function of a homogeneous field in $\mathbb{R}^3$, it can be written as

(4.99) $\qquad B(\tau) = B(\tau_1, \tau_2, \tau_3) = \iiint\limits_{\mathbb{K}^3} e^{ik\tau\cos\Theta_{\mathbf{k}\tau}} dF(k_1, k_2, k_3)$

where k, τ, and $\Theta_{\mathbf{k}\tau}$ have the same meaning as in (4.96), while the function $F(k_1, k_2, k_3) = F(\mathbf{k})$ differs from the three-dimensional distribution function only by a constant positive factor. Moreover, $B(\tau\sin\Theta\cos\varphi, \tau\sin\Theta\sin\varphi, \tau\cos\Theta) = B(\tau)$ for any Θ and φ and hence

$$B(\tau) = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} B(\tau\sin\Theta\cos\varphi, \tau\sin\Theta\sin\varphi, \tau\cos\Theta)\sin\Theta \; d\Theta \; d\varphi.$$

We now fix the value of the vector **k** and introduce, on the surface of the sphere $|\tau| = \tau$, a new system of spherical coordinates (Θ', φ') such that the new "latitude" Θ' coincides with the angle $\Theta_{\mathbf{k}\tau}$. Then evidently

$$(4.100) \quad B(\tau) = \iiint_{\mathbb{K}^3} \left\{ \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} e^{ik\tau\cos\Theta'} \sin\Theta' d\Theta' d\varphi' \right\} dF(\mathbf{k})$$

$$= \iiint_{\mathbb{K}^3} \frac{\sin k\tau}{k\tau} dF(\mathbf{k})$$

i.e.,

$$(4.101) \quad B(\tau) = \int_0^{\infty} \frac{\sin k\tau}{k\tau} d\Phi(k),$$

where $\Phi(k) = \iiint_{|\mathbf{k}| < k} dF(\mathbf{k})$ is again a bounded nondecreasing function on the half-line $0 \le k < \infty$. Further, it is easy to show, as for $n = 2$, that any function of the form (4.101), where $\Phi(k)$ is a bounded nondecreasing function, is a three-dimensional isotropic correlation function (and that the three-dimensional measure $F(\Delta)$ corresponding to the isotropic field $X(t)$ in $\mathbb{R}^3$ is necessarily rotationally symmetric). Therefore, *formula* (4.101), *where* $\Phi(\mathbf{k})$ *is a bounded nondecreasing function, specifies the class of all three-dimensional isotropic correlation functions.*

The general case of an arbitrary n is only slightly more complicated than the particular cases where $n = 2$ or $n = 3$. If $B(\tau)$ is an n-dimensional isotropic correlation function, then there exists such a correlation function $B(\tau) = B(\tau_1, ..., \tau_n)$ of the homogeneous random field in $\mathbb{R}^n$ that $B(\tau) = B(\tau)$ where $\tau = |\tau|$. Using the spectral representation (4.46a) of the correlation function $B(\tau)$, we easily obtain that

$$(4.102) \quad B(\tau) = \int_{\mathbb{K}^n} \cdots \int e^{ik\tau\cos\Theta_k\tau} dF(k_1,...,k_n)$$

(cf. (4.96) and (4.99)). Since $B(\tau) = B(|\tau|)$, then $B(\tau\mathbf{e}) = B(\tau)$ for any vector $\mathbf{e}$ of unit length. Integrate the last equality over all the values of $\mathbf{e}$ in the unique sphere $|\mathbf{e}| = 1$ of the space $\mathbb{R}^n$ and divide the result by the total area Σ_n of this sphere. Then we get the relation

$$(4.103) \quad B(\tau) = \frac{1}{\Sigma_n} \int_{|\mathbf{e}|=1} B(\tau\mathbf{e}) d\sigma(\mathbf{e})$$

where $d\sigma(\mathbf{e})$ is an element of the $(n-1)$-dimensional area of the sphere $|\mathbf{e}| = 1$. Let us take into account that $\Sigma_n = 2\pi^{n/2}/\Gamma(n/2)$ and pass, in the right-hand side of (4.103), to

spherical coordinates chosen so that the "latitude angle" Θ' coincides with Θ_{kT}. Integrating then over all the spherical coordinates, except for the angle Θ', we can write the $(n-1)$-dimensional area element as $d\sigma(e) = \Sigma_{n-1}(\sin\Theta')^{n-2}d\Theta'$. Now we easily obtain from (4.102) and (4.103) that

$$B(T) = \int_{\mathbb{K}^n}\!\!\!...\int\left\{\frac{\Gamma(n/2)}{\sqrt{\pi}\Gamma((n-1)/2)}\int_0^{\pi}e^{ikT\cos\Theta'}(\sin\Theta')^{n-2}d\Theta'\right\}d\,F(\mathbf{k})$$

(4.104)

$$= \int_{\mathbb{K}^n}\!\!\!...\int\Gamma(\frac{n}{2})(\frac{2}{kT})^{(n-2)/2}J_{(n-2)/2}(kT)dF(\mathbf{k})$$

(see Gradshteyn and Ryzhik, 1980, Eqs. 4.644 and 8.411.5; the formula for Σ_n is given in the same book as Eq. 4.633). Thus,

$$(4.105)\quad B(T) = 2^{(n-2)/2}\Gamma(\frac{n}{2})\int_0^{\infty}\frac{J_{(n-2)/2}(kT)}{(kT)^{(n-2)/2}}\,d\Phi(k),$$

where $\Phi(k) = \int_{|\mathbf{k}|<k}...\int dF(\mathbf{k})$ is a bounded nondecreasing function of k. (The numerical factor $2^{(n-2)/2}\Gamma(n/2)$ could naturally be included in the function $\Phi(k)$, but we will not do that.) It is also easy to show, as for $n = 2$ and $n = 3$, that not only any n-dimensional isotropic correlation function $B(T)$ can be represented in the form (4.105), but, conversely, any function of the form (4.105), where $\Phi(k)$ is a bounded nondecreasing function, is an n-dimensional isotropic correlation function. Hence *formula (4.105) specifies the class of all n-dimensional isotropic correlation functions.*[38] The representation of $B(T)$ in the form (4.105) is called the *spectral representation of an n-dimensional isotropic correlation function.*

If $X(\mathbf{t}) = X(t_1,t_2,t_3)$ is an isotropic random field in the three-dimensional space $\mathbb{R}^3$ and $X(\mathbf{t})$ has a correlation function $B(T)$, then the values of this field at points on any plane of the space $\mathbb{R}^3$ form an isotropic random field in the plane having a correlation function $B(T)$. Similarly, if $X(t)$ is an isotropic random field in the n-dimensional space $\mathbb{R}^n$ and $X(\mathbf{t})$ has a correlation function $B(T)$, then for any $m < n$ the values of this field on an arbitrary m-dimensional linear subspace of the space $\mathbb{R}^n$ form an isotropic random field in $\mathbb{R}^m$ with the correlation function $B(T)$. Thus, *any n-dimensional isotropic correlation function is necessarily also an m-dimensional isotropic correlation function for all integers m less than n.* (It follows that if $m < n$, then any function of the form (4.105), where $\Phi(k)$

is   a   bounded   nondecreasing   function,   can   also   be represented in the form

$$B(\tau) = 2^{(m-2)/2}\Gamma\left[\frac{m}{2}\right] \int_0^\infty \frac{J_{(m-2)/2}(k\tau)}{(k\tau)^{(m-2)/2}} \, d\Phi_1(k)$$

where $\Phi_1(k)$ is another bounded nondecreasing function.[39] Let us denote by $\mathfrak{D}_n$ the set of all n-dimensional isotropic correlation functions and also denote by $\mathfrak{D}_\infty$ the set of all functions $B(\tau)$ which belong to $\mathfrak{D}_n$ for any integer n (i.e. are isotropic correlation functions in $\mathbb{R}^n$, for any value of n). Then, obviously,

$$\mathfrak{D}_1 \supset \mathfrak{D}_2 \supset ... \supset \mathfrak{D}_n \supset ... \supset \mathfrak{D}_\infty$$

where $\mathfrak{D}_1$ is the set of all real positive definite functions, and the symbol $\subset$ denotes inclusion, i.e., $\mathfrak{D}_j \subset \mathfrak{D}_k$ means that every function from the set $\mathfrak{D}_j$ belongs also to the set $\mathfrak{D}_k$. According to the results of Sec. 2, $\mathfrak{D}_n$ coincides also with the set of all positive definite kernels $B(\mathbf{t}',\mathbf{t}'')$ in $\mathbb{R}^n$ which depend only on the distance between $\mathbf{t}'$ and $\mathbf{t}''$, i.e. with the set of such functions $B(\tau)$ that

$$(4.106) \quad \sum_{j,k=1}^m B(|\mathbf{t}_j - \mathbf{t}_k|)c_j\bar{c}_k \geqslant 0$$

for any integer m, vectors $\mathbf{t}_1$, ..., $\mathbf{t}_m$ of the space $\mathbb{R}^n$, and complex   numbers c_1,   ..., c_m.   (In   fact,   just   this   last definition of the class $\mathfrak{D}_n$ was used in the first derivation of formula (4.105); see Note 38.) Similarly, the functions $B(\tau)$ from $\mathfrak{D}_\infty$ are   specified   by   the   validity   of   the inequality (4.106) for all values of m, c_1, ..., c_m and all vectors $\mathbf{t}_1$,   ..., $\mathbf{t}_m$ from   any   Euclidean   space   of   arbitrary dimension.   Note   also   that   the   general   form   of   function $B(\tau)$ belonging to $\mathfrak{D}_\infty$ is given by the formula

$$(4.107) \quad B(\tau) = \int_0^\infty e^{-k^2\tau^2} d\Phi(k)$$

where $\Phi(k)$ is, as usual, an arbitrary bounded nondecreasing function.[40]

By using the definition of $\mathfrak{D}_n$ (where n is an arbitrary integer or ∞) as the class of all n-dimensional isotropic correlation functions, or as the class of all positive definite kernels in $\mathbb{R}^n$ depending only on the distance between two arguments, it is easy to prove the following general properties

of the functions in $\mathcal{D}_n$, which are similar to the known properties of the correlation functions of stationary random processes (see Sec. 4, pp. 58–61):

(i) If the function $B(\tau)$ belongs to $\mathcal{D}_n$ and a and b are two arbitrary positive constants, then the function $aB(b\tau)$ also belongs to $\mathcal{D}_n$; if both the functions $B_1(\tau)$ and $B_2(\tau)$ belong to $\mathcal{D}_n$, then the product $B_1(\tau)B_2(\tau)$ and all the linear combinations $a_1B_1(\tau) + a_2B_2(\tau)$, where a_1 and a_2 are nonnegative constants, also belong to $\mathcal{D}_n$;

(ii) If the functions $B^{(j)}(\tau)$, $j = 1,2, ...$, belong to $\mathcal{D}_n$ and $\lim_{j\to\infty} B^{(j)}(\tau) = B(\tau)$ exists for all τ, then the function $B(\tau)$ also belongs to $\mathcal{D}_n$;

(iii) If the function $B(\tau;a)$ of the argument τ belongs to $\mathcal{D}_n$ for all the values of the real parameter a in some finite or infinite interval A, then the functions

$$(4.108) \quad B_h(\tau) = \int_A B(\tau;a)h(a)da \text{ and } B_H(\tau) = \int_A B(\tau;a)dH(a),$$

where $h(a)$ is any nonnegative function and $H(a)$ is any nondecreasing function of a such that the integrals in the right-hand sides of (4.108) converge, also belong to $\mathcal{D}_n$.

The above properties of $\mathcal{D}_n$ permit one to construct many examples of n-dimensional isotropic correlation functions on the basis of the few known examples of such functions.

Moreover, (4.98), (4.101), (4.105), and (4.107) can also be used for the derivation of a number of further general properties of isotropic correlation functions. Note, to begin with, that the function $\Phi(k)$ in the right-hand sides of these formulae is such that $\Phi(0) = 0$, but $\lim_{\epsilon\to 0} \Phi(\epsilon) = \Phi(+0)$ may be equal either to zero or to some positive number a, i.e. the function $\Phi(k)$ may be either continuous at $k = 0$ or have a jump at the point $k = 0$. The integrand functions $\Lambda_n(ks)$ in these formulae are given by the relations

$$\Lambda_n(x) = 2^{(n-2)/2}\Gamma\left(\frac{n}{2}\right)\frac{J_{(n-2)/2}(x)}{x^{(n-2)/2}}$$

$$(4.109) \qquad = 1 - \frac{x^2}{2n} + \frac{x^4}{2\cdot 4\cdot n(n+2)} - ... \text{ for } n = 2,3, ...,$$

$$\Lambda_n(x) = e^{-x^2} \qquad\qquad \text{ for } n = \infty.$$

Since $\Lambda_n(0) = 1$ for any n, then

(4.110) $B(\tau) = a + \int_0^\infty \Lambda_n(k\tau) d\Phi_1(k) = a + B_1(\tau),$

where a is the jump of the function $\Phi(k)$ at the point $k = 0$, while $\Phi_1(k)$ is a bounded nondecreasing function continuous at the point $k = 0$. It is easy to deduce from the fact that $\Lambda_n(x) \to 0$ as $x \to \infty$ for any n, that $B_1(\tau) \to 0$ as $\tau \to \infty$.[41] Thus, *the multidimensional isotropic correlation function* $B(\tau)$ *necessarily tends either to zero or to some positive constant* a *as* $\tau \to \infty$.

It follows from (4.98), (4.101), (4.105), and (4.107) that for any n

$$B(0) = \int_0^\infty d\Phi(k) = \Phi(\infty).$$

Therefore,

(4.111) $R(\tau) = B(\tau)/B(0) = \int_0^\infty \Lambda_n(k\tau) d\hat{\Phi}(k)$

where $\hat{\Phi}(k) = \Phi(k)/\Phi(\infty)$ is a nondecreasing function satisfying the condition

$$\int_0^\infty d\hat{\Phi}(k) = \hat{\Phi}(\infty) = 1.$$

Equation (4.111) implies that*

(4.112) $R(\tau) \geqslant \min_x \Lambda_n(x).$

Now the tables of the functions $J_0(x) = \Lambda_2(x),(\sin x)/x = \Lambda_3(x)$, and $2J_1(x)/x = \Lambda_4(x)$ and the positiveness of $\exp(-x^2)$ give

$R(\tau) > -0.403$ for $n = 2$,

(4.113) $R(\tau) > -0.218$ for $n = 3$,

$R(\tau) > -0.133$ for $n = 4$,

$R(\tau) > 0$ for $n = \infty$.[42]

---

*Since the function $\Lambda_\infty(x) = \exp(-x^2)$ does not have a minimum, (4.112) for n $= \infty$ must be replaced by the inequality

(4.112a) $R(\tau) > \inf \Lambda_\infty(x),$

where inf stands for infimum (the greatest lower bound).

Note also that $|\Lambda_n(x)|$ decreases as $x^{-(n-1)/2}$ as $x \to \infty$ for $n = 2, 3,$ ... and decreases exponentially for $n = \infty$. Hence, the functions of class $\mathfrak{D}_n$ must be rather smooth for comparatively large values of n. In particular, the known relations $|J_\nu(x)| < C_\nu/x^{1/2}$ and $d\{J_\nu(x)x^{-\nu}\}/dx = -J_{\nu+1}(x)x^{-\nu}$ permit one to prove rather easily that *any function* $B(\tau)$ *of class* $\mathfrak{D}_n$ (*i.e. any* n-*dimensional isotropic correlation function) has, at each* τ*, at least* $[(n-1)/2]$ *derivatives* (where $[(n-1)/2]$ denotes the largest integer which does not exceed $(n-1)/2$).[43] In particular, *an isotropic correlation function in a space of dimension n > 2 is always everywhere differentiable.* As for the functions $B(\tau)$ of class $\mathfrak{D}_\infty$ (i.e. being possible isotropic correlation functions in the space of any dimension), these functions have derivatives of all orders everywhere, i.e. they are infinitely differentiable.[44]

So far we have considered only the general properties of the multidimensional isotropic correlation functions. If, however, a function $B(\tau)$ of the class $\mathfrak{D}_n$ falls off rapidly enough at infinity (in particular, if

$$\int_0^\infty \tau^{n-1} |B(\tau)| d\tau < \infty),$$

then the corresponding isotropic random field $X(t)$ will evidently have a *spectral density* $f(\mathbf{k}) = f(k_1, \dots k_n)$ (cf. p. 331). The spectral density $f(\mathbf{k})$ can be determined from the known correlation function $B(\mathbf{\tau}) = B(\tau)$ with the aid of (4.56). Since the function $B(\mathbf{\tau})$ now depends only on $\tau = |\mathbf{\tau}|$, we can pass to spherical polar coordinates in the right-hand side of (4.56) and then integrate over all the angular coordinates. We then find that the spectral density $f(\mathbf{k})$ depends only on $k = |\mathbf{k}|$ (i.e. it is spherically symmetric) and, moreover, $f(\mathbf{k}) = f(|\mathbf{k}|) = f(k)$ is given by the formula

(4.114) $f(k) = \dfrac{1}{(2\pi)^{n/2}} \displaystyle\int_0^\infty \dfrac{J_{(n-2)/2}(k\tau)}{(k\tau)^{(n-2)/2}} \tau^{n-1} B(\tau) d\tau$

(cf. the derivation of (4.105)). Formula (4.105) for $B(\tau)$ can in this case be rewritten as

(4.115) $B(\tau) = (2\pi)^{n/2} \displaystyle\int_0^\infty \dfrac{J_{(n-2)/2}(k\tau)}{(k\tau)^{(n-2)/2}} k^{n-1} f(k) dk$

since if a spectral density exists, then

(4.116) $\Phi(k) = \Sigma_n \int_0^k (k')^{n-1} f(k') dk' = \dfrac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^k (k')^{n-1} f(k') dk'.$

Note also that by virtue of (4.116) and of the known identity $x^\nu J_{\nu-1}(x) = d\{x^\nu J_\nu(x)\}/dx$ it follows from (4.114) that if a spectral density exists, then

(4.117) $\Phi(k) = \dfrac{1}{2^{(n-2)/2}\Gamma(n/2)} \int_0^\infty J_{n/2}(k\tau)(k\tau)^{n/2} \dfrac{B(\tau)}{\tau} d\tau.$

Using the general formula, which determines the spectral distribution function $F(\mathbf{k})$ of an arbitrary homogeneous random field in terms of the corresponding correlation function $B(\tau)$, and the expression of $\Phi(k)$ in terms of $F(\mathbf{k})$, it is not hard to show that (4.117) is valid for any isotropic correlation function $B(\tau)$ and any $k \geqslant 0$ if the value of $\Phi(k)$ at the point of discontinuity is taken as the half-sum of the left and right limiting values of this function.

In the most important particular cases where $n = 2$ or $n = 3$, equations (4.114) and (4.115) take the form

(4.118) $f(k) = \dfrac{1}{2\pi} \int_0^\infty J_0(k\tau)\tau B(\tau)d\tau, \quad B(\tau) = 2\pi\int_0^\infty J_0(k\tau)kf(k)dk$

for $n = 2$ and

(4.118a) $f(k) = \dfrac{1}{2\pi^2}\int_0^\infty \dfrac{\sin k\tau}{k\tau}\tau^2 B(\tau)d\tau, \quad B(\tau) = 4\pi\int_0^\infty \dfrac{\sin k\tau}{k\tau}k^2 f(k)dk$

for $n = 3$.

The function $f(k)$ is usually called the n-*dimensional spectral density* of the isotropic random field $X(\mathbf{t})$ in $\mathbb{R}^n$; however, in applied works the same term is also used sometimes to denote the function

(4.119) $\varphi(k) = \Phi'(k) = [2\pi^{n/2}/\Gamma(n/2)]k^{n-1}f(k).$

The function $\Phi(k)$ is then called the n-*dimensional spectral distribution function* of an isotropic field $X(\mathbf{t})$.

The n-dimensional spectral density is, of course, a very important statistical characteristic of the random field $X(\mathbf{t})$. However, the determination of this density from observations (or measurements) of $X(\mathbf{t})$ usually involves considerable difficulties. The one-dimensional spectral density $f_1(k_1)$ or $g_1(k_1) = 2f_1(k_1)$ (see (4.95)) is, as a rule, much easier to determine. In fact, for the determination of the

one-dimensional spectral density one only needs to observe (or measure) the values of $X(t)$ at points on an arbitrary straight line and then apply the highly developed technique of spectral analysis for stationary time series (see Secs. 18 and 19). The relative simplicity of determining the one-dimensional spectral densities of isotropic fields stimulates everyone's interest in the question of the relationship between the one-dimensional and n-dimensional spectral densities of the given isotropic field.

Let the isotropic random field $X(t)$ have an n-dimensional spectral density $f(k)$. Then the spectral representation (4.56) of the corresponding correlation function $B(\tau) = B(\tau_1, ..., \tau_n)$ for the case where $\tau = (\tau, 0, ..., 0)$ has the form

$$B(\tau, 0, ..., 0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} e^{ik_1 \tau} f(k_1, k_2, ..., k_n) dk_1 dk_2 ... dk_n.$$

Comparing the last equation with (4.95) and taking into account that $B(\tau) = B(|\tau|)$ and $f(k) = f(|k|)$, we obtain

$$f_1(k_1) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f(k_1, k_2, ..., k_n) dk_2 ... dk_n$$

$$= \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} f([k_1^2 + k_2^2 + ... + k_n^2]^{1/2}) dk_2 ... dk_n$$

$$= \Sigma_{n-1} \int_0^{\infty} f([k_1^2 + \kappa^2]^{1/2}) \kappa^{n-2} d\kappa =$$

$$= \frac{2\pi^{(n-1)/2}}{\Gamma[(n-1)/2]} \int_0^{\infty} f([k_1^2 + \kappa^2]^{1/2}) \kappa^{n-2} d\kappa$$

where $\kappa = (k_2^2 + ... + k_n^2)^{1/2}$. Transforming now to the new variable $k = (k_1^2 + \kappa^2)^{1/2}$ instead of κ and also using the easily derived equation $k dk = \kappa d\kappa$, we finally get

$$(4.120) \quad f_1(k_1) = \frac{2\pi^{(n-1)/2}}{\Gamma[(n-1)/2]} \int_{k_1}^{\infty} f(k)(k^2 - k_1^2)^{(n-3)/2} k dk.$$

It is clear that (4.120) can also be written as

$$(4.120a) \quad g_1(k_1) = \frac{2\Gamma(n/2)}{\pi^{1/2}\Gamma[(n-1)/2]} \int_{k_1}^{\infty} \varphi(k) \frac{(k^2 - k_1^2)^{(n-3)/2}}{k^{n-2}} dk.$$

In the cases where $n = 2$ or $n = 3$ equations (4.120) and (4.120a) have the form

$$(4.121) \quad f_1(k_1) = 2\int_{k_1}^{\infty} \frac{f(k)}{(k^2 - k_1^2)^{1/2}} k dk, \quad g_1(k_1) = \frac{2}{\pi} \int_{k_1}^{\infty} \frac{\varphi(k) dk}{(k^2 - k_1^2)^{1/2}}$$

for $n = 2$, and

(4.122) $\quad f_1(k_1) = 2\pi \int_{k_1}^{\infty} f(k)k\,dk, \quad g_1(k_1) = \int_{k_1}^{\infty} \frac{\varphi(k)dk}{k}$

for $n = 3$.

Note that (4.120) imply, in particular, that the one-dimensional spectral density $f_1(k_1)$ is necessarily differentiable if the corresponding n-dimensional spectral density $f(k)$ exists. This enables us to convert (4.121) and (4.122) and to derive the following equations which express the multidimensional spectral densities $f(k)$ and $\varphi(k)$ in terms of the one-dimensional ones $f_1(k_1)$ and $g_1(k_1)$:

$$f(k) = -\frac{1}{\pi} \int_k^{\infty} \frac{df_1(k_1)}{dk_1} \frac{dk_1}{(k_1^2 - k^2)^{1/2}},$$

(4.123)

$$\varphi(k) = -k \int_k^{\infty} \frac{dg_1(k_1)}{dk_1} \frac{dk_1}{(k_1^2 - k^2)^{1/2}} \quad \text{for} \quad n = 2,$$

(4.124) $\quad f(k) = -\frac{1}{2\pi k} \frac{df_1(k)}{dk}; \quad \varphi(k) = -k \frac{dg_1(k)}{dk} \quad \text{for} \quad n = 3.$

A similar conversion of the general n-dimensional equations (4.120) and (4.120a) is also possible, but, at least for even values of n, the general converse equations expressing $f(k)$ in terms of $f_1(k)$ and $\varphi(k)$ in terms of $g_1(k)$ are rather cumbersome.[45]

*The isotropic random field $X(t)$ is, by definition, also homogeneous. Therefore, to determine the statistical characteristics of $X(t)$ from measured values (referring to some bounded set T of t-values) of a single realization $x(t)$, one can use the techniques considered in Sec. 21, where an analogous problem for homogeneous random fields was studied. It is clear, however, that if the field $X(t)$ is isotropic, then there exist some additional possibilities for estimating its statistical characteristics which cannot be applied to homogeneous fields. Thus, when estimating the correlation function $\langle X(t + \tau)X(t)\rangle = B(\tau)$, $\tau = |\tau|$, we can now use the averaging of the product $x(t_1)x(t_2)$ over a wide set of pairs (t_1, t_2) of points from T satisfying the condition $|t_2 - t_1| = \tau$.[46] Based on the obtained estimate $B^*(\tau)$ of the function $B(\tau)$ we can then also estimate the spectral distribution function $\Phi(k)$ and the spectral density $f(k)$ or $\varphi(k)$ of the

field $X(t)$. In practice, however, it is often more convenient
to measure the values of $X(t)$ only along some straight line
and then estimate the one-dimensional spectral density
$f_1(k_1)$ (or $g_1(k_1)$) and (if needed) the correlation function
$B(\tau)$ with the aid of the techniques described in Secs. 18 and
19. If the values of the corresponding multidimensional
spectral density $f(k)$ (or $\varphi(k)$) are also of interest, then in the
cases where $n = 2$ or $n = 3$, one can, in addition, use equations
(4.123) and (4.124). The spectral densities of turbulent
fluctuations are usually determined just in this way in the
experimental investigations of isotropic turbulence.[47*]

22.2. Examples of Isotropic Correlation Functions

We have already remarked that an n-dimensional isotropic
correlation function, where $n > 1$, is necessarily also a
correlation function of some real stationary random process.
Therefore, in looking for examples of isotropic correlation
functions we can examine only the real functions $B(\tau)$ that are
stationary correlation functions. To check whether or not the
given function $B(\tau)$ (which falls off rapidly enough at infinity)
is an n-dimensional isotropic correlation function, one only
needs to calculate, using (4.114), the corresponding spectral
density $f(k)$ and examine whether or not this function $f(k)$ is
everywhere nonnegative. Another method, which is actually
equivalent to the first, consists in finding, first of all, the
one-dimensional Fourier transform $f_1(k_1)$ of the function $B(\tau)$,
then calculating the corresponding multidimensional spectral
density $f(k)$ from $f_1(k_1)$, and finally checking whether or not the
function $f(k)$ is everywhere nonnegative. This last method is
especially convenient in cases where $n = 2$ or $n = 3$, since one can
use here the comparatively simple equations (4.123) and (4.124).
In particular, formula (4.124) shows that the real stationary
correlation function $B(\tau)$, which has the one-dimensional
Fourier transform $f_1(k_1)$ differentiable at $k_1 \neq 0$, is a
three-dimensional isotropic correlation function if, and only if,
the function $f_1(k_1)$ is nonincreasing on the half-line $0 < k_1 < \infty$.
This also implies, of course, that the lack of increase of $f_1(k_1)$ on
the whole half-line $0 < k_1 < \infty$ is a sufficient (but not a necessary)
condition for the function $B(\tau)$ to be an isotropic correlation
function in the plane. (This last result is also confirmed by
(2.123) which implies that $f(k) \geqslant 0$ for all k if $df_1(k_1)/dk_1 \leqslant 0$ for

all k.)

Many examples of real correlation functions of stationary random processes are given in Sec. 10. However, some of these clearly cannot be examples of isotropic correlation functions in any multidimensional space. For instance, the cosinusoid (2.98) does not satisfy the condition (4.112) for any $n \geqslant 2$ and, hence, it cannot be a member of any set $\mathfrak{D}_n$ with $n \geqslant 2$. Moreover, if the spectral distribution function $F(k)$ of a stationary process $X(t)$ is discontinuous and has a jump at least at one point $k_1 \neq 0$, then the corresponding correlation function cannot be an n-dimensional isotropic correlation function for any $n \geqslant 2$, since this function $B(\tau)$ does not tend to a constant value as $\tau \to \infty$. Moreover, the triangular correlation function (2.123) cannot be an isotropic correlation function in $\mathbb{R}^n$ for $n > 2$, since this function is nondifferentiable at $\tau = T$. By calculating the Fourier–Bessel transform (4.118) for the function (2.123), one can show that the corresponding two-dimensional spectral density $f(k)$ is not everywhere nonnegative; hence, the function (2.123) is not an isotropic correlation function in the plane too.[48] Nevertheless, there are also many examples of stationary correlation functions presented in Sec. 10 which simultaneously provide examples of isotropic correlation functions in multidimensional spaces.

Example 1. The exponential function $B(\tau) = C\exp(-\alpha\tau)$, where $C > 0$, $\alpha > 0$, $\tau \geqslant 0$, is the most common (and most widely used) example of a stationary correlation function; see (2.94). According to (2.95) and Fig. 19, the corresponding spectral density $f_1(k_1)$ (denoted by $f(\omega)$ in Sec. 10) falls off monotonically with k_1 for $0 \leqslant k_1 < \infty$. Hence, the exponential function $B(\tau)$ is an isotropic correlation function in the plane $\mathbb{R}^2$ and in the three-dimensional Euclidean space $\mathbb{R}^3$. It is also easy to show that the function (2.94) is an isotropic correlation function in the space $\mathbb{R}^n$ for any integer n, i.e. that this function belongs to the class $\mathfrak{D}_\infty$. In fact, (4.114) and the known identity

$$(4.125) \qquad \int_0^\infty e^{-\alpha x} J_\mu(kx)x^{\mu+1}dx = \frac{2\alpha(2k)^\mu \Gamma(\mu + 3/2)}{\pi^{1/2}(\alpha^2 + k^2)^{\mu+3/2}}$$

(see e.g., Gradshteyn and Ryzhik, 1980, Eq. 6.623.2) imply that the n-dimensional spectral density

$$(4.126) \qquad f(k) = \frac{C\alpha\Gamma[(n + 1)/2]}{\pi^{(n+1)/2}(\alpha^2 + k^2)^{(n+1)/2}}$$

corresponds to the correlation function (2.94). The function (4.126) is clearly everywhere positive for any integer n. For $n = 1$, (4.126) coincides with (2.95); for $n = 2$ and $n = 3$, we obtain

$$(4.127) \quad f(k) = \frac{C\alpha}{2\pi(\alpha^2 + k^2)^{3/2}} \quad \text{for } n = 2,$$

$$(4.128) \quad f(k) = \frac{C\alpha}{\pi^2(\alpha^2 + k^2)^2} \quad \text{for } n = 3.$$

According to (4.108), the fact that the function (2.94) belongs to the class $\mathfrak{D}_\infty$ implies that any function $B(\tau)$ of the form

$$B(\tau) = \int_0^\infty e^{-\alpha\tau} dH(\alpha),$$

where $H(\alpha)$ is an arbitrary bounded nondecreasing function, also belongs to $\mathfrak{D}_\infty$.

Example 2. The stationary correlation function (2.139) of the form

$$B(\tau) = C(\alpha\tau)^\nu K_\nu(\alpha\tau), \quad C > 0, \ \alpha > 0, \ \nu \geqslant 0, \ \tau \geqslant 0,$$

is a generalization of the exponential correlation function (2.94), since it turns into the function (2.94) if $\nu = 1/2$. The one-dimensional spectral density $f_1(k_1)$ corresponding to the correlation function (2.139) is also a monotonically decreasing function of k_1 for $0 \leqslant k_1 < \infty$ (see (2.140)). Hence, the function (2.139) is an isotropic correlation function in $\mathbb{R}^2$ and $\mathbb{R}^3$. Moreover, it is also easy to show that the function (2.139) is an isotropic correlation function in a space of any dimension. Indeed, according to (4.114), the known identity

$$(4.129) \quad \int_0^\infty x^{\mu+\nu+1} J_\mu(kx) K_\nu(\alpha x) dx = 2^{\mu+\nu} k^\mu \alpha^\nu \frac{\Gamma(\mu + \nu + 1)}{(\alpha^2 + k^2)^{\mu+\nu+1}}$$

(see, e.g., Gradshteyn and Ryzhik, 1980, Eq. 6.576.7) implies that the n-dimensional spectral density

$$(4.130) \quad f(k) = \frac{2^{\nu-1} C\Gamma[\nu + (n/2)]\alpha^{2\nu}}{\pi^{n/2}(\alpha^2 + k^2)^{\nu+n/2}}$$

corresponds to the correlation function (2.139). The function (4.130) is clearly everywhere positive. For $n = 1$, (4.130) turns

into (2.140); for arbitrary n and $\nu = 1/2$, we obtain (4.126). The particular case where $n = 2$ and $\nu = 1$ has already been considered in the preceding section (see (4.64) and (4.65)). Moreover, equations (4.2') in Note 22 to this chapter include (2.139) and (4.130), but in (4.2') it is assumed that $\nu > -n/2$.

In the particular cases where $n = 2$ or $n = 3$, we easily get

$$(4.131) \quad f(k) = \frac{2^{\nu-1} C \Gamma(\nu + 1) \, \alpha^{2\nu}}{\pi (\alpha^2 + k^2)^{\nu+1}} \quad \text{for } n = 2,$$

$$(4.132) \quad f(k) = \frac{2^{\nu-1} C \Gamma[\nu + (3/2)] \alpha^{2\nu}}{\pi^{3/2} (\alpha^2 + k^2)^{\nu+3/2}} \quad \text{for } n = 3.$$

Equation (4.132) is useful, in particular, in the theory of turbulence.[49]

Example 3. One more frequently used form of a real stationary correlation function $B(\tau)$ is given by the "Gaussian curve" $B(\tau) = C\exp(-\alpha\tau^2)$, $C > 0$, $\alpha > 0$, $\tau \geqslant 0$ (see (2.126)). In calculating the n-dimensional spectral density for such a correlation function by (4.114), one should use the density

$$(4.133) \quad \int_0^\infty x^{\nu+1} e^{-\alpha x^2} J_\nu(kx)dx = k^\nu (2\alpha)^{-\nu-1} \exp(-k^2/4\alpha)$$

(see Gradshteyn and Ryzhik, 1980, Eq. 6.631.4). It follows from (4.133) that the n-dimensional spectral density

$$(4.134) \quad f(k) = \frac{C}{2^n (\pi\alpha)^{n/2}} e^{-k^2/4\alpha}$$

corresponds to the correlation function (2.126). The density (4.134) is everywhere positive for any n and α. (Note also that for any n the functions $B(\tau)$ and $f(k)$ have the same shape.) Thus, the function (4.126) belongs to the class $\mathfrak{D}_\infty$. It follows from this at once that the class $\mathfrak{D}_\infty$ also includes all functions of the form (4.107) where $\Phi(k)$ is any bounded nondecreasing function. (Recall that according to the statement on p. 354 functions of the form (4.107) in fact exhaust all the class $\mathfrak{D}_\infty$.)

Example 4. The exponential function (2.132) of the form

$$B(\tau) = C\exp(-\alpha\tau^m), \quad C > 0, \ \alpha > 0, \ \tau \geqslant 0,$$

generalizes both the simplest exponential correlation function (2.94) and the Gaussian correlation function (2.126). It has

already been explained in Sec. 10 (see Example 11 on p. 137) that the function (2.132) is a stationary correlation function if, and only if, $0 \leqslant m \leqslant 2$. (If $m = 0$, then $B(\tau) = C = $ constant is clearly a correlation function which belongs to the class $\mathfrak{D}_\infty$.) Moreover, in Note 35 to Chap. 2 it is indicated that for any m in the interval $0 < m \leqslant 2$ the one-dimensional spectral density, which corresponds to the correlation function (2.132), decreases monotonically with an increase of its argument from zero to infinity. This implies that any function of the form (2.132), where $0 < m \leqslant 2$, can be the correlation function of an isotropic random field in two- and three-dimensional spaces, i.e. that such a function $B(\tau)$ belongs to the functional classes $\mathfrak{D}_2$ and $\mathfrak{D}_3$. And what is more, one can also prove that in fact any such function belongs to the class $\mathfrak{D}_\infty$, i.e. that it can be an isotropic correlation function in a Euclidean space of any dimension.[50]

Example 5. On p. 140 a stationary correlation function of the form

$$(4.135) \quad B(\tau) = \frac{C}{(\alpha^2 + \tau^2)^\nu}, \quad C > 0, \ \alpha > 0, \ \nu > 0,$$

was mentioned. (It was assumed there that $\nu > 1/2$, so that $\int_0^\infty |B(\tau)| d\tau < \infty$, but we now relax this restriction, assuming only that $\nu > 0$.) It is not hard to show that all the functions (4.135) also belong to the class $\mathfrak{D}_\infty$. This follows from the known integral representation of the Γ-function

$$2\int_0^\infty e^{-u^2} u^{2\nu-1} du = \Gamma(\nu)$$

which implies the relation

$$(4.136) \quad \frac{C}{(\alpha^2 + \tau^2)^\nu} = \frac{2C}{\Gamma(\nu)} \int_0^\infty e^{-(\tau^2+\alpha^2)u^2} u^{2\nu-1} du.$$

It follows from (4.136) that for any $\nu > 0$ the function (4.135) can be represented in the form (4.107), where $\Phi(k)$ is some special differentiable nondecreasing function. Using (4.136) and (4.134), it is also easy to calculate the n-dimensional spectral density $f(k)$ corresponding to the correlation function (4.135).[51]

So far we have discussed only isotropic correlation functions, which belong to the class $\mathfrak{D}_\infty$ and therefore,

according to (4.113), must be everywhere positive. However, functions $B(\tau)$, which belong only to certain classes $\mathfrak{D}_n$, where n is finite, can, of course, change their sign.

Example 6. Consider the stationary correlation function (2.101) given by the equation

$$B(\tau) = C\exp(-\alpha\tau)\cos\omega_0\tau, \quad \alpha > 0, \quad \omega_0 > 0, \quad \tau \geqslant 0.$$

It is clear that if α is small enough, this function does not satisfy even the first inequality (4.113), i.e. it does not belong to the class $\mathfrak{D}_2$ of isotropic correlation function in the plane. By evaluating, with the aid of (4.118) (or (2.102) and (4.132)), the two-dimensional spectral density $f(k)$, which corresponds to the correlation function (2.101), and by studying the conditions for the nonnegativity of this $f(k)$, one can show that the *function* (2.101) *is a possible correlation function of the isotropic field in the plane if, and only if,* $\alpha \geqslant \omega_0$.[52] It is still easier to obtain conditions guaranteeing that $B(\tau)$ belongs to $\mathfrak{D}_3$. Indeed, it was noted on pp. 124-125 that a (one-dimensional) spectral density corresponding to the correlation function (2.101) is monotonically nonincreasing on the positive half-axis if, and only if, $\alpha \geqslant \sqrt{3}\omega_0$. It follows that the *function* (2.101) *is a three-dimensional isotropic correlation function in the space* $\mathbb{R}^3$ *if, and only if,* $\alpha \geqslant \sqrt{3}\omega_0$.

Example 7. Formula (4.105) shows that the function

$$(4.137) \quad B(\tau) = C(\alpha\tau)^{-(n-2)/2} J_{(n-2)/2}(\alpha\tau), \quad C > 0, \quad \alpha > 0,$$

is an n-dimensional isotropic correlation function. The spectral distribution function $\Phi(k)$ corresponding to this correlation function is clearly of the form

$$(4.138) \quad \Phi(k) = \begin{cases} 0 & \text{for } k \leqslant \alpha, \\ C/2^{(n-2)/2}\Gamma(\frac{n}{2}) & \text{for } k > \alpha, \end{cases}$$

i.e., the n-dimensional spectral density is given here by the relation

$$f(k) = \Phi'(k)\Gamma(\frac{n}{2})/2\pi^{n/2}k^{n-1} = \{C/(2\pi)^{n/2}k^{n-1}\}\delta(k-\alpha),$$

where $\delta(x)$ is the Dirac δ-function. Using (4.120), we easily obtain an equation of the form (2.138), with $\nu = (n - 2)/2$, for the corresponding one-dimensional spectral density $f_1(k_1)$.

Replacing $(n - 2)/2$ in the right-hand side of (4.137) by an arbitrary real number $\nu > -1/2$, we arrive at a more general stationary correlation function (2.137). The inequality (4.112) clearly implies that the function (2.137) does not belong to the class $\mathfrak{D}_n$ (i.e. cannot be an isotropic correlation function in the n-dimensional space $\mathbb{R}^n$) if $\nu < (n - 2)/2$. On the other hand, it is not hard to show that for $\nu \geqslant (n - 2)/2$ the function (2.137) belongs to the class $\mathfrak{D}_n$. In fact, it is well known that for $\nu > \mu > -1$ the following identity holds:

$$\int_0^\infty J_\mu(kx)J_\nu(\alpha x)x^{\mu-\nu+1}dx = \begin{cases} \dfrac{k^\mu(\alpha^2 - k^2)^{\nu-\mu-1}}{2^{\nu-\mu-1}\alpha^\nu\Gamma(\nu - \mu)} & \text{for } k < \alpha, \\ 0 & \text{for } k > \alpha \end{cases}$$

(see, e.g., Erdélyi et al., 1953, Vol. 2, Eq. 7.14.2(34) or Gradshteyn and Ryzhik, 1980, Eq. 6.575.1). According to (4.114) it follows from this that, for $\nu > (n - 2)/2$, the n-dimensional spectral density $f(k)$, which corresponds to the correlation function (2.137), has the form

$$(4.139) \quad f(k) = \begin{cases} \dfrac{C(\alpha^2 - k^2)^{\nu-n/2}}{2^\nu\pi^{n/2}\alpha^{2\nu}\Gamma[\nu + 1 - (n/2)]} & \text{for } k < \alpha, \\ 0 & \text{for } k > \alpha. \end{cases}$$

The function (4.139) is everywhere nonnegative; hence the function (2.137), where $\nu > (n - 2)/2$, belongs to $\mathfrak{D}_n$.

Some more examples of isotropic correlation function· can also be found in the available literature,[53] but we will not pursue this topic any further here.

22.3. Spectral Representation of Isotropic Random Fields

Since all isotropic random fields are, by definition, also homogeneous, it is clear that any such field $X(t)$ in the n-dimensional space $\mathbb{R}^n$ can be represented as the n-dimensional Fourier–Stieltjes integral (4.51) with respect to some n-dimensional random measure $Z(d\mathbf{k})$. However, the spectral representation (4.51) does not at all use the fact that the field $X(t)$ is not only homogeneous, but also isotropic.

Hence this representation is not specific to isotropic random fields. It is natural to think that there exists also some other spectral representation of the field $X(t)$ which is specific for isotropic random fields and might be simpler in some respects than the Fourier–Stieltjes representation (4.51). It will be shown below that this thought is quite justified.

To begin with, we consider the simplest case of the isotropic field $X(t) = X(t_1, t_2)$ in the plane R^2, i.e. assume that $n = 2$. Then, according to (4.98),

$$(4.140) \quad \langle X(t_1)\overline{X(t_2)} \rangle = B(t_1, t_2) = \int_0^\infty J_0(k|t_1 - t_2|)d\Phi(k),$$

where $\Phi(k)$ is a bounded nondecreasing function. Let now (t, φ) be polar coordinates in R^2. Then, if $t_1 = (t_1, \varphi_1)$, $t_2 = (t_2, \varphi_2)$, $|t_1 - t_2| = [t_1^2 + t_2^2 - 2t_1 t_2 \cos(\varphi_1 - \varphi_2)]^{1/2}$. Further, according to the well-known addition theorem for Bessel functions (see, e.g., Erdélyi et al., 1953, Vol. 2, Eq. 7.15(31), or Gradshteyn and Ryzhik, 1980, Eq. 8.531.1) we have

$$J_0(k|t_1 - t_2|) = \sum_{m=-\infty}^{\infty} J_m(kt_1) J_m(kt_2) e^{im(\varphi_1 - \varphi_2)}.$$

Hence, Eq. (4.140) can also be written as

$$(4.141) \quad \langle X(t_1, \varphi_1)\overline{X(t_2, \varphi_2)} \rangle = \sum_{m=-\infty}^{\infty} \int_0^\infty J_m(kt_1)e^{im\varphi_1} \overline{J_m(kt_2)e^{im\varphi_2}} d\Phi(k).$$

By virtue of the theorem on the generalized spectral representation of random function (see Note 17 to Chap. 2 in Vol. II and pp. 447–448 below), it follows from (4.141) that the field $X(t) = X(t, \varphi)$ can be represented in the form

$$(4.142) \quad X(t, \varphi) = \sum_{m=-\infty}^{\infty} e^{im\varphi} \int_0^\infty J_m(kt) dZ_m(k).$$

The functions $Z_m(k)$, $m = 0, \pm 1, \dots$, in the right-hand side of (4.142) are complex-valued random functions of the nonnegative variable k satisfying the condition

$$(4.143) \quad \langle Z_m(\Delta_1)\overline{Z_p(\Delta_2)} \rangle = \delta_{mp}\Phi(\Delta_1 \cap \Delta_2)$$

where Δ_1 and Δ_2 are intervals of the half-line $[0, \infty)$, $Z_m(\Delta) = \int_\Delta dZ_m(k)$, $\Phi(\Delta) = \int_\Delta d\Phi(k)$, δ_{mp} is the Kronecker symbol (i.e. $\delta_{mp} = 0$ for $m \neq p$ and $= 1$ for $m = p$), and $\Delta_1 \cap \Delta_2$ is the

intersection of the intervals Δ_1 and Δ_2. Of course, (4.143) can be also written as the symbolic relation

(4.143a) $\langle dZ_m(k)d\overline{Z_p(k')} \rangle = \delta_{mp}\delta(k - k')d\Phi(k)dk'$.

If the field $X(\mathbf{t})$ (and __not__ only its correlation function $B(\tau)$) is real, then $Z_{-m}(\Delta) = \overline{Z_m(\Delta)}$, and therefore (4.142) can also be written in the real form:

(4.142a) $X(t,\varphi) = \displaystyle\sum_{m=0}^{\infty} \cos m\varphi \int_0^{\infty} J_m(kt)dZ_m^{(1)}(k)$

$\qquad\qquad + \displaystyle\sum_{m=1}^{\infty} \sin m\varphi \int_0^{\infty} J_m(kt)dZ_m^{(2)}(k)$.

Here $Z_m^{(1)}(k)$, and $Z_m^{(2)}(k)$ are real random functions of k and

(4.143b) $\langle Z_m^{(j)}(\Delta_1)Z_p^{(q)}(\Delta_2) \rangle = \delta_{mp}\delta_{jq}\, d_m\Phi(\Delta_1 \cap \Delta_2)$

where $d_m = 1$ for $m = 0$, and $d_m = 2$ for $m \neq 0$.

Clearly, the existence of a representation of the form (4.142) (or (4.142a)), where the functions $Z_m(k)$ (or $Z_m^{(1)}(k), Z_m^{(2)}(k)$) satisfy (4.143) (or (4.143b)) already guarantees that the field $X(t,\varphi) = X(\mathbf{t})$ is an isotropic random field in the plane. This representation is simpler then the spectral representation (4.51) of the homogeneous random field in the plane, because no multidimensional random functions enter the integrands in (4.142) and (4.142a). The representation (4.142) (or (4.142a)) is sometimes called the *spectral representation* (or the *polar spectral representation*, to distinguish it from the Fourier–Stieltjes spectral representation) of the isotropic random field in the plane.[54]

The polar spectral representation (4.142) can easily be generalized to the case of an isotropic random field in the n-dimensional Euclidean space $\mathbb{R}^n$. In fact, according to (4.105) the correlation function of such a field can be represented as

(4.144) $B(\mathbf{t}_1,\mathbf{t}_2) = 2^{(n-2)/2}\Gamma(n/2)\displaystyle\int_0^{\infty} \dfrac{J_{(n-2)/2}(k|\mathbf{t}_1 - \mathbf{t}_2|)}{(k|\mathbf{t}_1 - \mathbf{t}_2|)^{(n-2)/2}}\, d\Phi(k)$.

Using now the general addition theorem for Bessel functions (Erdélyi et al., 1953, Vol. 2, Eq. 7.15(30), or Gradshteyn and Ryzhik, 1980, Eq. 8.532.1) and the addition theorem for surface harmonics (Erdélyi et al., 1953, Vol. 2, Eq. 11.4(2)), we obtain

$$(4.145) \quad 2^{(n-2)/2}\Gamma(n/2) \frac{J_{(n-2)/2}(k|t_1 - t_2|)}{(k|t_1 - t_2|)^{(n-2)/2}}$$

$$= c_n^2 \sum_{m=0}^{\infty} \sum_{l=1}^{h(m,n)} \frac{J_{[(n-2)/2]+m}(kt_1)}{(kt_1)^{(n-2)/2}} \frac{J_{[(n-2)/2]+m}(kt_2)}{(kt_1)^{(n-2)/2}} S_m^l(t_1^{(0)}) S_m^l(t_2^{(0)}).$$

Here, $c_n^2 = 2^{n-1}\pi^{n/2}\Gamma(n/2)$, $h(n,m) = (2m + n - 2)(m + n - 3)!$ $[m!(n - 2)!]^{-1}$ is the number of surface harmonics of degree m in the n-dimensional space $\mathbb{R}^n$; $t_1 = |t_1|$, $t_2 = |t_2|$; $t_1^{(0)} = t_1/t_1$ and $t_2^{(0)} = t_2/t_2$ are unit vectors corresponding to given n-dimensional vectors t_1 and t_2; $S_m^l(t^0) = S_m^l(\Theta_1, ..., \Theta_{n-2}, \varphi), l = 1, 2, ..., h(m,n)$ are spherical surface harmonics of degree m, and $(\Theta_1, ..., \Theta_{n-2}, \varphi)$ are hyperspherical coordinates of the unit vector $t^{(0)}$ in $\mathbb{R}^n$. Substituting (4.145) into (4.144) and then applying again the theorem on the generalized spectral representation of random functions, we obtain the following *polar spectral representation* of a real isotropic random field $X(t)$ in $\mathbb{R}^n$

$$(4.146) \quad X(t) = c_n \sum_{m=0}^{\infty} \sum_{l=1}^{h(m,n)} S_m^l(t/t) \int_0^{\infty} \frac{J_{[(n-2)/2]+m}(kt)}{(kt)^{(n-2)/2}} \, dZ_m^l(k).$$

The functions $Z_m^l(k)$ in the right-hand side of (4.146) are real random functions of the nonnegative variable k satisfying the relations

$$(4.147) \quad \langle Z_m^l(\Delta_1) Z_p^q(\Delta_2) \rangle = \delta_{mp}\delta_{lq}\Phi(\Delta_1 \cap \Delta_2)$$

where the notation is the same as in (4.143).[55]

22.4. Multidimensional Isotropic Random Fields

The concept of an isotropic random field in $\mathbb{R}^n$ is susceptible to two different multidimensional generalizations. The simpler of them is the following: The multidimensional (say, s-dimensional) random field $X(t) = \{X_0(t), ..., X_s(t)\}$ in the n-dimensional Euclidean space $\mathbb{R}^n$ is said to be *isotropic* if mean values of all the components $X_j(t)$, $j = 1, ..., s$, are constant and all the elements of the correlation matrix $B(t_1, t_2) = \|B_{ji}(t_1, t_2)\|$ depend only on the distance $|t_1 - t_2|$ between the points t_1 and t_2 (i.e. only on the length of the vector $t_1 - t_2$). Clearly, all the components $X_j(t)$, $j = 1, ..., s$, of such multidimensional isotropic

field $X(t)$ are one-dimensional isotropic fields in $\mathbb{R}^n$. The fact that the cross-correlation functions $B_{ji}(t_1,t_2)$) (where $j \neq i$) depend only on $|t_1 - t_2|$ means that isotropic fields $X_j(t)$, $j = 1, ...,$ s, are *isotropically correlated with each other* (e.g., the pressure and temperature fields in isotropic turbulence[56]). It is easy to show that the elements of the correlation matrix of an s-dimensional isotropic random field $X(t)$ (i.e., the correlation and cross-correlation functions of the components of $X(t)$) can be represented in the form

$$(4.148) \quad B_{ji}(\tau) = 2^{(n-2)/2}\Gamma\left(\frac{n}{2}\right)\int_0^\infty \frac{J_{(n-2)/2}(k\tau)}{(k\tau)^{(n-2)/2}} o\, \Phi_{ji}(k),$$

$$j,i = 1,...,s,\ 0 \leqslant \tau < \infty,$$

where $\Phi_{ji}(k)$ are functions of bounded variation having the property that the matrix

$$(4.149) \quad \|\Delta\Phi_{ji}(k)\| = \|\Phi_{ji}(k + \Delta k) - \Phi_{ji}(k)\|$$

is positive definite for any $k \geqslant 0$ and $\Delta k > 0$ (cf. the derivation of a similar result for the matrix (4.13) in Sec. 20). Since all the random fields $X_j(t)$, $j = 1, ..., s$, are isotropic, they can be represented in the form (4.146), i.e.

$$(4.150) \quad X_j(t) = c_n \sum_{m=0}^\infty \sum_{l=1}^{h(m,n)} S_m^l(t/t)\int_0^\infty \frac{J_{[(n-2)/2]+m}(kt)}{(kt)^{(n-2)/2}}\, dZ_{jm}^1(k),$$

where, as is easy to see,

$$(4.151) \quad \langle Z_{jm}^1(\Delta_1)Z_{ip}^q(\Delta_2)\rangle = \delta_{mp}\delta_{lq}\Phi_{ji}(\Delta_1 \cap \Delta_2).$$

We now pass on to the second, more interesting, multidimensional generalization of the concept of isotropic random field, which is often encountered in various applications (e.g., in the theory of turbulence). Let $X(t)$ be an s-dimensional homogeneous field in $\mathbb{R}^n$, the s components of which transform linearly among themselves under spatial rotations and reflections in $\mathbb{R}^n$. (For example, the components of $X(t)$ can form a vector, a polyvector, a second-order tensor, or a tensor of higher order in the n-dimensional space $\mathbb{R}^n$.) Then, it is natural to call the homogeneous random field $X(t)$ an *isotropic field of the quantity* X (e.g., a vector, polyvector, or tensor field) if the mean value $\langle X(t)\rangle = \mathbf{m}$ and the correlation matrix

$\|\langle X_j(t_1)\overline{X_i(t_2)}\rangle\| = B(t_1,t_2)$ do not change when we replace the point t and the pair of point t_1, t_2 by a point t' and a new pair of points t_1', t_2' obtained from t and t_1, t_2 by making a rotation or reflection g in $\mathbb{R}^n$, and simultaneously carry out the linear transformation of the components of X, which corresponds to the rotation or reflection g. The general theory of such isotropic fields X is intimately connected with the mathematical theory of group representations for the group G of spatial rotations and reflections of the n-dimensional Euclidean space $\mathbb{R}^n$.[57] We will not, however, pursue this general theory, but only consider briefly a particular case (quite typical and very important for many applications) of a real isotropic vector field in $\mathbb{R}^n$.

Isotropic Vector Random Fields

Let $X(t) = \{X_1(t), ..., X_n(t)\}$ be a real vector random field in the n-dimensional space $\mathbb{R}^n$ (e.g., $n = 3$ and $X(t) = \{X_1(t), X_2(t), X_3(t)\}$ is the velocity field in a turbulent fluid flow in $\mathbb{R}^3$). If the vector field $X(t)$ is isotropic, then its mean value $\langle X(t)\rangle = m$ must be a constant vector, since an isotropic random field is, by definition, homogeneous. Moreover, the mean value vector m must also be unaffected by spatial rotations and reflections. It follows from this at once that the vector m must be a zero vector. Hence, the *mean value of an isotropic vector field is necessarily equal to zero.*

Let us now go on to the correlation matrix

(4.152) $\qquad \langle X_j(t_1)X_l(t_2)\rangle = B_{jl}(t_1 - t_2) = B_{jl}(\tau), \quad \tau = t_1 - t_2,$

the elements of which obviously form a second-order tensor (the so-called *correlation tensor*). In investigating this tensor, it is convenient to transform, first of all, to a special coordinate system $O'x_1' ... x_n'$ in $\mathbb{R}_n$ instead of the initial set $Ox_1 ... x_n$. The new set $O'x_1' ... x_n'$ is so selected that its origin O' is shifted by the vector t_2 with respect to the origin O (i.e., O' coincides with the terminal point of the vector t_2 referring to the initial coordinates), while the axis $O'x_1'$ lies along the vector $t_1 - t_2 = \tau$. Let us denote by $B_{jl}'(\tau)$ the component of the correlation tensor $B_{jl}(\tau)$ in the new set of coordinates. (The functions $B_{jl}'(\tau)$ clearly depend only on the length $\tau = |\tau|$ of the vector τ, since the direction of τ is fixed in the new set of coordinates.) Note now that in the space $\mathbb{R}^n$ a reflection exists which leaves the points t_1

and t_2 ($= O'$) unmoved, replaces the axis $O'x_1$ by $- O'x_1'$ (where $l \neq 1$ is a fixed number), but does not change the directions of all the other coordinate axes $O'x_k'$, $k \neq l$. It follows immediately from this that

(4.153) $B_{j\,1}'(\tau) = - B_{j\,1}'(\tau) = 0$ for $j \neq l$.

Hence we see that only the diagonal elements $B_{j\,j}'(\tau)$, $j = 1, ...,$ n, of $\mathcal{B}(\tau)$ can differ from zero. Further, if $j \neq 1$ and $l \neq 1$, then the axis $O'x_j'$ can be transformed to the axis $O'x_1'$ by rotation about the axis $O'x_1'$. Hence

(4.154) $B_{2\,2}'(\tau) = B_{3\,3}'(\tau) = ... = B_{n\,n}'(\tau)$.

Thus, we have shown that the tensor $B_{j\,1}'(\tau)$ and, consequently, $B_{jl}(\tau)$ is symmetric, and the components $B_{j\,1}'(\tau)$ can take, at the most, only two nonequal nonzero values: $B_{1\,1}'(\tau)$ (we shall denote this function of τ by $B_{LL}(\tau)$), and $B_{2\,2}'(\tau)$, which is also equal to $B_{3\,3}'(\tau)$, ..., $B_{n\,n}'(\tau)$, and shall be denoted by $B_{NN}(\tau)$. The function $B_{LL}(\tau)$ is called the *longitudinal correlation function* of the isotropic vector field $X(t)$; its normalized value $B_{LL}(\tau)/B_{LL}(0)$ gives the correlation coefficient between the components of the field in the direction of the vector τ at two points t and $t + \tau$. Similarly, the function $B_{NN}(\tau)$ is called the *lateral correlation function* of the field $X(t)$ and determines the correlation coefficient between the components of the field in some direction perpendicular to τ at the points t and $t + \tau$. Thus, $B_{LL}(\tau)$ is a correlation function of the one-dimensional homogeneous field $X_1(t_1, 0, ..., 0) = X_1(t_1)$ on the straight line $t_2 = .. = t_n = 0$ (i.e. on the axis Ox_1), while $B_{NN}(\tau)$ is the correlation function of the one-dimensional homogeneous field $X_2(t_1, 0, ..., 0) = X_2(t_1)$ on the same line. Note also that both $B_{LL}(0)$ and $B_{NN}(0)$ are equal to the mean square of a single component of the random vector $X(t)$. Therefore, the isotropy of $X(t)$ implies that $B_{LL}(0) = B_{NN}(0) = \langle X^2(t) \rangle / n$.

The two functions $B_{LL}(\tau)$ and $B_{NN}(\tau)$ clearly specify uniquely the correlation tensor $B_{jl}(\tau)$. To obtain the explicit form for $B_{jl}(\tau)$, we must resolve the unit vectors of the old coordinate axes $Ox_1, Ox_2, ..., Ox_n$ along the axes of the new system $O'x_1' ... x_n'$ and then represent $B_{jl}(\tau)$ as a linear combination of the functions $B_{i\,k}'(\tau)$. Using (4.153) and

(4.154), we obtain the equation

$$(4.155) \quad B_{jl}(\tau) = [B_{LL}(\tau) - B_{NN}(\tau)] \frac{\tau_j \tau_l}{\tau^2} + B_{NN}(\tau)\delta_{jl},$$

where δ_{jl} is the Kronecker delta, while τ_j, $j = 1, ..., n$ are the components of the vector τ in the reference system $Ox_1 ... x_n$.[58] Equation (4.155) shows again, in particular, that the correlation tensor $B_{jl}(\tau)$ of any isotropic vector field is necessarily symmetric.

According to (4.155), the class of all the possible correlation tensors $B_{jl}(\tau)$ will be specified if the possible forms of the two functions $B_{LL}(\tau)$ and $B_{NN}(\tau)$ are given. Since each of these functions is the correlation function of a homogeneous random field on a line, they are both positive definite, i.e: can be represented in the form

$$(4.156) \, B_{LL}(\tau) = \int_{-\infty}^{\infty} e^{i\tau k_1} dF_{LL}^{(1)}(k_1), \; B_{NN}(\tau) = \int_{-\infty}^{\infty} e^{i\tau k_1} dF_{NN}^{(1)}(k_1),$$

where $F_{LL}^{(1)}(k_1)$ and $F_{NN}^{(1)}(k_1)$ are bounded nondecreasing functions of k_1. Moreover, if the functions $B_{LL}(\tau)$ and $B_{NN}(\tau)$ fall off rapidly enough when $\tau \to \infty$, then

$$(4.157) \, B_{LL}(\tau) = \int_{-\infty}^{\infty} e^{i\tau k_1} f_{LL}^{(1)}(k_1) dk_1, \; B_{NN}(\tau) = \int_{-\infty}^{\infty} e^{i\tau k_1} f_{NN}^{(1)}(k_1) dk_1$$

where $f_{LL}^{(1)}(k_1) \geqslant 0$ and $f_{NN}^{(1)}(k_1) \geqslant 0$. (The functions $F_{LL}^{(1)}(k_1)$ and $F_{NN}^{(1)}(k_1)$ are called the *longitudinal* and *lateral one-dimensional spectral distribution functions* of the field $X(t)$, and the functions $f_{LL}^{(1)}(k_1)$ and $f_{NN}^{(1)}(k_1)$ are, respectively, the *longitudinal* and *lateral one-dimensional spectral densities* of this field.) It is clear, however, that the representability of the functions $B_{LL}(\tau)$ and $B_{NN}(\tau)$ in the form (4.156) does not guarantee that they are possible longitudinal and lateral correlation functions of an isotropic vector field. In order to find the sufficient conditions specifying the possible functions $B_{LL}(\tau)$ and $B_{NN}(\tau)$ we note, first of all, that according to (4.155) the function

$$(4.158) \quad \tau_j \tau_l B_{jl}(\tau) = B_{LL}(\tau)\tau^2$$

depends only on the length $|\tau| = \tau$. (Here and in a number of subsequent equations of this section Einstein's summation convention is adopted, according to which whenever an index occurs twice in a single-term expression, the summation is carried out over n possible values of this index.) Therefore

(4.159) $B_{LL}(\tau)\tau^2 = \dfrac{1}{\Sigma_n}\displaystyle\int_{|e|=1} \tau_j\tau_l B_{jl}(\tau)d\sigma(e)$

where $e = \tau/\tau$ (i.e. we integrate over all the directions of the vector τ; cf. (4.103)). Let us replace now the functions $B_{jl}(\tau)$ in the right-hand side of (4.159) by their spectral representation (4.68) and use the identity

$$\frac{1}{\Sigma_n}\int_{|e|=1} e^{ik\tau}d\sigma(e) = 2^{(n-2)/2}\Gamma\left[\frac{n}{2}\right] J_{(n-2)/2}(k\tau)/(k\tau)^{(n-2)/2}$$

(cf. (4.104)) and the relation $\partial k/\partial k_j = k_j/k$, where $k = |k| = (k_1^2 + ... + k_n^2)^{1/2}$. Then we obtain

$$\tau^2 B_{LL}(\tau) = \int_{-\infty}^{\infty}...\int_{-\infty}^{\infty}\left\{\frac{1}{\Sigma_n}\int_{|e|=1}\tau_j\tau_l e^{ik\tau}d\sigma(e)\right\}F_{jl}(dk)$$

$$= -\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty}\frac{\partial^2}{\partial k_j \partial k_l}\left\{\frac{1}{\Sigma_n}\int_{|e|=1}e^{ik\tau}d\sigma(e)\right\}F_{jl}(dk)$$

(4.160) $= -A_n\displaystyle\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty}\frac{\partial^2}{\partial k_j \partial k_l}\left\{\frac{J_{(n-2)/2}(k\tau)}{(k\tau)^{(n-2)/2}}\right\} F_{jl}(dk)$

$$= -A_n\left\{\int_0^{\infty}\left[\frac{\partial^2}{\partial k^2} - \frac{1}{k}\frac{\partial}{\partial k}\right]\frac{J_{(n-2)/2}(k\tau)}{(k\tau)^{-(n-2)/2}}d\Phi_1(k)\right.$$

$$\left. - \int_0^{\infty}\frac{1}{k}\frac{\partial}{\partial k}\frac{J_{(n-2)/2}(k\tau)}{(k\tau)^{(n-2)/2}}d\Phi_0(k)\right\},$$

where $A_n = 2^{(n-2)/2}\Gamma(n/2)$ and

(4.161) $\Phi_1(k) = \displaystyle\int_{|k'|<k}\frac{k_j'k_l'}{k'^2}F_{jl}(dk'), \quad \Phi_0(k) = \int_{|k'|<k}F_{jj}(dk').$

Since $F_{jl}(\Delta)$ is an Hermitian positive definite matrix for any n-dimensional set Δ, it is clear that $\Phi_1(k)$ and $\Phi_0(k)$ are real bounded nondecreasing functions of k. It is also easy to see that $\Phi_1(0) = \Phi_0(0) = 0$ and that the function

$$\Phi_2(k) = \frac{1}{n-1}\{\Phi_0(k) - \Phi_1(k)\}$$

(4.162)

$$= \frac{1}{n-1}\int_{|k'|<k}\{F_{jj}(dk') - \frac{k_j'k_l'}{k'^2}F_{jl}(dk')\}$$

is also bounded and nondecreasing. (The factor $1/(n-1)$ is added for convenience, while the nonnegativity of the

increments $\Phi_2(k + \Delta k) - \Phi_2(k)$ for $\Delta k > 0$ is implied by the fact that the maximum of the quadratic form $F_{jl}(\Delta)e_je_l$ on the unit sphere $|e| = 1$ cannot exceed the trace $F_{jj}(\Delta)$ of the tensor $F_{jl}(\Delta)$.) Using the identity $d\{z^{-\nu}J_\nu(z)\}/dz = -z^{-\nu}J_{\nu+1}(z)$ (see, e.g., Gradshteyn and Ryzhik, 1980, Eq. 8.472.4), we can now write (4.160) as

$$(4.163) \qquad B_{LL}(\tau) = A_n\left\{\int_0^\infty\left[\frac{J_{n/2}(k\tau)}{(k\tau)^{n/2}} - \frac{J_{(n+2)/2}(k\tau)}{(k\tau)^{(n-2)/2}}\right]d\Phi_1(k)\right.$$

$$\left. + (n-1)\int_0^\infty\frac{J_{n/2}(k\tau)}{(k\tau)^{n/2}}\,d\Phi_2(k)\right\}.$$

It is not hard to verify that arbitrary real bounded functions, monotonically nondecreasing on the half-line $0 \leqslant k < \infty$, can be selected as the functions $\Phi_1(k)$ and $\Phi_2(k)$. Thus, the class of all possible functions $B_{LL}(\tau)$ is specified by formula (4.163), where $\Phi_1(k)$ and $\Phi_2(k)$ satisfy the above-indicated conditions.

To determine the general form of the function $B_{NN}(\tau)$, one has only to use the following corollary of (4.155):

$$(4.164) \qquad B_{jj}(\tau) = B_{LL}(\tau) + (n - 1)B_{NN}(\tau).$$

Since the right-hand side of (4.164) depends only on τ, the following expression can be derived quite similarly to the derivation of (4.160):

$$(4.165) \qquad B_{LL}(\tau) + (n - 1)B_{NN}(\tau) = A_n\int_0^\infty\frac{J_{(n-2)/2}(k\tau)}{(k\tau)^{(n-2)/2}}\,d\Phi_0(k).$$

It follows from (4.161), (4.162), (4.165), and the well-known recurrence equations for Bessel functions that

$$B_{NN}(\tau) = A_n\left\{\int_0^\infty\frac{J_{n/2}(k\tau)}{(k\tau)^{n/2}}\,d\Phi_1(k)\right.$$

$$(4.166)$$

$$\left. + \int_0^\infty\left[\frac{J_{(n-2)/2}(k\tau)}{(k\tau)^{(n-2)/2}} - \frac{J_{n/2}(k\tau)}{(k\tau)^{n/2}}\right]d\Phi_2(k)\right\}$$

which specifies the general form of the function $B_{NN}(\tau)$.[59]

Consider now the particular case where all the correlation functions $B_{jl}(\tau)$, $j, l = 1, ..., n$, fall off rapidly enough at infinity and therefore can be represented as Fourier integrals

$$(4.167) \qquad B_{jl}(\tau) = \int_{\mathbb{R}^n} e^{ik\tau}f_{jl}(k)d\mathbf{k}$$

(see (4.69)). For the isotropic vector field $\mathbf{X}(t)$ we can combine (4.69a) and (4.155) to obtain

$$(4.168) \quad f_{jl}(k) = [f_1(k) - f_2(k)]\frac{k_j k_l}{k^2} + f_2(k)\delta_{jl}.$$

It is also easy to check that the matrix $\|f_{jl}(k)\|$ will be positive definite if, and only if,

$$(4.169) \quad f_1(k) \geqslant 0, \quad f_2(k) \geqslant 0.$$

Thus, both functions $f_1(k)$ and $f_2(k)$ must be everywhere nonnegative, but can be arbitrary in other respects. (These two functions are sometimes called in the literature the *longitudinal* and *lateral* n-*dimensional spectral densities*. Similarly, the functions $\Phi_1(k)$ and $\Phi_2(k)$ can be called *longitudinal* and *lateral* n-*dimensional spectral distribution functions*.)

Replacing $B_{jl}(\tau)$ and $f_{jl}(k)$ on the left- and right-hand sides of (4.167) by their expressions (4.155) and (4.168), and then completing the integration with respect to the angular variables and using some simple identities, we obtain

$$B_{LL}(\tau) = (2\pi)^{n/2}\left\{\int_0^\infty\left[\frac{J_{n/2}(k\tau)}{(k\tau)^{n/2}} - \frac{J_{(n+2)/2}(k\tau)}{(k\tau)^{(n-2)/2}}\right]f_1(k)k^{n-1}dk\right.$$

$$\left. + (n-1)\int_0^\infty \frac{J_{n/2}(k\tau)}{(k\tau)^{n/2}}f_2(k)k^{n-1}dk\right\},$$

$$(4.170)$$

$$B_{NN}(\tau) = (2\pi)^{n/2}\left\{\int_0^\infty\frac{J_{n/2}(k\tau)}{(k\tau)^{n/2}}f_1(k)k^{n-1}dk\right.$$

$$\left. + \int_0^\infty\left[\frac{J_{(n-2)/2}(k\tau)}{(k\tau)^{(n-2)/2}} - \frac{J_{n/2}(k\tau)}{(k\tau)^{n/2}}\right]f_2(k)k^{n-1}dk.\right.$$

Thus, if the sp ectral densities $f_{jl}(k)$ exist, then both the functions $\Phi_1(k)$ and $\Phi_2(k)$ are necessarily absolutely continuous (i.e. are differe n tiable and equal to indefinite integrals of their derivative s), and the derivatives $\Phi_1'(k)$ and $\Phi_2'(k)$ are equal in this case to the functions $k^{n-1}f_1(k)$ and $k^{n-1}f_2(k)$ multiplied by $\Sigma_n = 2\pi^{n/2}/\Gamma(n/2)$.

Let us now u se again the general formula (4.69a), which expresses $f_{jl}(k)$ i n terms of $B_{jl}(\tau)$, and also (4.155) and (4.168). Then, similarly to the derivation of (4.170), we can derive the converse equatio n s

$$f_1(k) = \frac{1}{(2\pi)^{n/2}} \left\{ \int_0^\infty \left[\frac{J_{n/2}(k\tau)}{(k\tau)^{n/2}} - \frac{J_{(n+2)/2}(k\tau)}{(k\tau)^{(n-2)/2}} \right] B_{LL}(\tau)\tau^{n-1}d\tau \right.$$

$$\left. + (n-1) \int_0^\infty \frac{J_{n/2}(k\tau)}{(k\tau)^{n/2}} B_{NN}(\tau)\tau^{n-1}d\tau \right\},$$

(4.171)

$$f_2(k) = \frac{1}{(2\pi)^{n/2}} \left\{ \int_0^\infty \frac{J_{n/2}(k\tau)}{(k\tau)^{n/2}} B_{LL}(\tau)\tau^{n-1}d\tau \right.$$

$$\left. + \int_0^\infty \left[\frac{J_{(n-2)/2}(k\tau)}{(k\tau)^{(n-2)/2}} - \frac{J_{n/2}(k\tau)}{(k\tau)^{n/2}} \right] B_{NN}(\tau)\tau^{n-1}d\tau \right\}.$$

Equations (4.171) are quite similar to (4.170).   When we wish to check whether or not the two given functions $B_{LL}(\tau)$ and $B_{NN}(\tau)$, which fall off rapidly at infinity, can be the longitudinal and lateral correlation functions of an isotropic vector field, we need only substitute these functions in the right-hand sides of (4.171) to see whether or not the obtained functions $f_1(k)$ and $f_2(k)$ are everywhere nonnegative.

In the most important particular cases, where $n = 2$ or $n = 3$, Eqs. (4.170) take the form:

$$B_{LL}(\tau) = 2\pi \left\{ \int_0^\infty \frac{\partial J_1(k\tau)}{\partial(k\tau)} f_1(k)k\,dk + \int_0^\infty \frac{J_1(k\tau)}{k\tau} f_2(k)k\,dk \right\},$$

(4.172)

$$B_{NN}(\tau) = 2\pi \left\{ \int_0^\infty \frac{J_1(k\tau)}{k\tau} f_1(k)k\,dk + \int_0^\infty \frac{\partial J_1(k\tau)}{\partial(k\tau)} f_2(k)k\,dk \right\}$$

for $n = 2$ and

$$B_{LL}(\tau) = 4\pi \left\{ \int_0^\infty \left[\frac{\sin k\tau}{k\tau} + 2\frac{\cos k\tau}{(k\tau)^2} - 2\frac{\sin k\tau}{(k\tau)^3} \right] f_1(k)k^2 dk \right.$$

$$\left. + \int_0^\infty \left[-2\frac{\cos k\tau}{(k\tau)^2} + 2\frac{\sin k\tau}{(k\tau)^3} \right] f_2(k)k^2 dk \right\},$$

(4.173)

$$B_{NN}(\tau) = 4\pi \left\{ \int_0^\infty \left[-\frac{\cos k\tau}{(k\tau)^2} + \frac{\sin k\tau}{(k\tau)^3} \right] f_1(k)k^2 dk \right.$$

$$\left. + \int_0^\infty \left[\frac{\sin k\tau}{k\tau} + \frac{\cos k\tau}{(k\tau)^2} - \frac{\sin k\tau}{(k\tau)^3} \right] f_2(k)k^2 dk \right\}$$

for $n = 3$. Formulae (4.171) for two and three dimensions can

be written quite similarly.

Equation (4.168) and the relation between the functions $f_1(k)$, $f_2(k)$, and $\Phi_1(k)$, $\Phi_2(k)$ imply that if the spectral densities $f_{jl}(k)$ exist, then the tensor $F_{jl}(dk)$ appearing in the spectral decomposition (4.68) of the correlation tensor $B_{jl}(\tau)$ takes the form

$$(4.174) \quad F_{jl}(d\mathbf{k}) = \frac{d\sigma(k)}{\Sigma_n(k)} \left\{ \kappa_j \kappa_l d\Phi_1(k) + (\delta_{jl} - \kappa_j \kappa_l) d\Phi_2(k) \right\},$$

where $d\mathbf{k} = d\sigma(k)dk$ is an element of volume in n-dimensional space $\mathbb{K}^n$ of the vectors $\mathbf{k}$, $d\sigma(k)$ is an element of $(n-1)$-dimensional area of the sphere $|\mathbf{k}| = k$, $\Sigma_n(k) = \Sigma_n k^{n-1}$ is the total area of this sphere, $\kappa_j = k_j/k$, $\kappa_l = k_l/k$, $d\Phi_1(k) = \Sigma_n(k)f_1(k)dk$, and $d\Phi_2(k) = \Sigma_n(k)f_2(k)dk$. It is easy to see that representation (4.174) is, in fact, valid for any isotropic vector field, irrespective of whether the spectral densities $f_{jl}(\mathbf{k})$ exist or not, and the functions $\Phi_1(k)$ and $\Phi_2(k)$ are in all cases related to $F_{jl}(\mathbf{k})$ by (4.161) and (4.162). Conversely, if the spectral matrix $F_{jl}(\mathbf{k})$ in the right-hand side of (4.68) can be represented in the form (4.174), then the corresponding n-dimensional homogeneous field $\mathbf{X(t)}$ in $\mathbb{R}^n$ is an isotropic vector field.

Let us return to the case where the functions $B_{LL}(\tau)$ and $B_{NN}(\tau)$ fall off rapidly enough at infinity, so that spectral densities exist. Comparing (4.69) with (4.157) and taking into account (4.168) we get

$$f_{LL}^{(1)}(k_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ f_1 \left([k_1^2 + \dots + k_n^2]^{1/2} \right) \frac{k_1^2}{k_1^2 + \dots + k_n^2} \right.$$
$$\left. + f_2 \left([k_1^2 + \dots + k_n^2]^{1/2} \right) \left(1 - \frac{k_1^2}{k_1^2 + \dots + k_n^2} \right) \right\} dk_2 \dots dk_n,$$

$$f_{NN}^{(1)}(k_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left\{ f_1 \left([k_1^2 + \dots + k_n^2]^{1/2} \right) \frac{k_2^2}{k_1^2 + \dots + k_n^2} \right.$$
$$\left. + f_2 \left([k_1^2 + \dots + k_n^2]^{1/2} \right) \left(1 - \frac{k_2^2}{k_1^2 + \dots + k_n^2} \right) \right\} dk_2 \dots dk_n.$$

Transforming to spherical polar coordinates in the $(n - 1)$-dimensional space of variables $k_2, \dots, k_n$, we have, after integration over all the angular variables,

$$f_{LL}^{(1)}(k_1) = \frac{2\pi^{(n-1)/2}}{\Gamma[(n-1)/2]} \int_{k_1}^{\infty} \left\{ k_1^2 f_1(k) \right.$$

$$\left. + (k^2 - k_1^2) f_2(k) \right\} (k^2 - k_1^2)^{(n-3)/2} \frac{dk}{k} ,$$

(4.175)

$$f_{NN}^{(1)}(k_1) = \frac{\pi^{(n-1)/2}}{\Gamma[(n+1)/2]} \int_{k_1}^{\infty} \left\{ (k^2 - k_1^2) f_1(k) \right.$$

$$\left. + [(n-2)k^2 + k_1^2] f_2(k) \right\} (k^2 - k_1^2)^{(n-3)/2} \frac{dk}{k} .$$

These formulae express the one-dimensional spectral densities $f_{LL}^{(1)}(k)$ and $f_{NN}^{(1)}(k)$ in terms of the three-dimensional densities $f_1(k)$ and $f_2(k)$.[60] They show, in particular, that if $f_1(k) \geqslant 0$ and $f_2(k) \geqslant 0$, then necessarily also $f_{LL}^{(1)}(k) \geqslant 0$ and $f_{NN}^{(1)}(k) \geqslant 0$, i.e. they confirm once again that both the functions $B_{LL}(\tau)$ and $B_{NN}(\tau)$ are positive definite.

Note in conclusion that isotropic vector fields $X(t)$ also admit of a quite specific spectral representation, similar to the polar spectral representation (4.146) of a one-dimensional (i.e. scalar) isotropic random field $X(t)$ in R^n. As well as (4.146), this new representation expresses the field $X(t)$ in terms of Bessel functions, some functions related to spherical harmonics (so-called *vector spherical harmonics*) and a countable number of mutually uncorrelated one-dimensional random measures.[61]

*In turbulence mechanics and also in some other applied areas, two special classes of isotropic vector fields $X(t)$ are of considerable interest. These classes are the class of *isotropic solenoidal vector fields* (i.e. fields having zero divergence) and the class of *isotropic potential vector fields* (having a zero curl, i.e. equal to the gradient of a scalar field $\psi(t)$). If the field $X(t)$ is solenoidal, i.e. $\partial X_j(t)/\partial t_j = 0$, then clearly $\partial B_{jl}(\tau)/\partial \tau_j = 0$ according to the results of Sec. 4. Substituting (4.155) in the last relation, we easily obtain that

$$B_{NN}(\tau) = B_{LL}(\tau) + \frac{\tau}{n-1} \frac{dB_{LL}(\tau)}{d\tau} ,$$

(4.176)

$$\text{i.e., } B_{LL}(\tau) = \frac{n-1}{\tau^{n-1}} \int_0^{\tau} B_{NN}(x) x^{n-2} dx.$$

Similarly, if the field $X(t)$ is a potential one, then $X_j(t) = \partial \psi(t)/\partial t_j$ and $\partial B_{ji}(\tau)/\partial \tau_l = \partial B_{li}(\tau)/\partial \tau_j$. Then (4.155) implies that

$$B_{LL}(\tau) = B_{NN}(\tau) + \tau \frac{dB_{NN}(\tau)}{d\tau},$$

(4.177)

i.e., $B_{NN}(\tau) = \frac{1}{\tau} \int_0^\tau B_{LL}(x)dx.$

Let us now use the spectral representations (4.68) and (4.69) of the correlation tensor $B_{jl}(\tau)$ and also equations (4.174) or (4.168). Then we easily obtain that in the case of an isotropic solenoidal vector field $X(t)$

(4.178) $d\Phi_1(k) = 0,$ and hence $f_1(k) = 0$

if there exist spectral densities. Similarly, in the case of an isotropic potential vector field $X(t)$

(4.179) $d\Phi_2(k) = 0,$ and hence $f_2(k) = 0$

if there exist spectral densities. Thus, the correlation tensor $B_{jl}(\tau)$ of an isotropic solenoidal vector field is completely specified by one function of a single variable (namely, by the function $B_{LL}(\tau)$, or $B_{NN}(\tau)$, or $\Phi_2(k)$, or, if the spectral densities exist, $f_2(k)$). Similarly, the correlation tensor of an isotropic potential vector field is specified uniquely by the function $B_{LL}(\tau)$, or $B_{NN}(\tau)$, or $\Phi_1(k)$, or, if spectral densities exist, $f_1(k)$. It is also clear from the above results that the correlation tensor of an arbitrary isotropic vector field can always be represented as a sum of the correlation tensors of one solenoidal and one potential isotropic vector fields.[62]
Many important specific results can be derived with the aid of (4.176) − (4.179). For instance, it follows from (4.176) and (4.177) that for a solenoidal isotropic field

(4.180) $\int_0^\infty B_{NN}(\tau)\tau^{n-2}d\tau = 0$

while for a potential isotropic field

(4.181) $\int_0^\infty B_{LL}(\tau)d\tau = 0,$

if the integrals on the left-hand sides of (4.180) and

(4.181) are convergent. Equations (4.180) and (4.181) imply, in particular, that the function $B_{NN}(\tau)$, which corresponds to an isotropic solenoidal field $X(t)$, cannot be everywhere positive, while for an isotropic potential field $X(t)$ the function $B_{LL}(\tau)$ must change sign. Moreover, according to (4.163) and (4.166), if $d\Phi_1(k) = 0$, then

$$B_{LL}(\tau) = 2^{(n-2)/2}\Gamma(n/2)(n-1)\int_0^\infty \frac{J_{n/2}(k\tau)}{(k\tau)^{n/2}}\, d\Phi_2(\tau),$$

(4.182) $B_{LL}(\tau) + (n-1)B_{NN}(\tau)$

$$= 2^{(n-2)/2}\Gamma(n/2)(n-1)\int_0^\infty \frac{J_{(n-2)/2}(k\tau)}{(k\tau)^{(n-2)/2}}d\Phi_2(k),$$

and if $d\Phi_2(k) = 0$, then

$$B_{NN}(\tau) = 2^{(n-2)/2}\Gamma(n/2)\int_0^\infty \frac{J_{n/2}(k\tau)}{(k\tau)^{n/2}}\, d\Phi_1(k),$$

(4.183) $B_{LL}(\tau) + (n-1)B_{NN}(\tau)$

$$= 2^{(n-2)/2}\Gamma(n/2)\int_0^\infty \frac{J_{(n-2)/2}(k\tau)}{(k\tau)^{(n-2)/2}}\, d\Phi_1(k).$$

Note that, by virtue of (4.176) and (4.177), $B_{LL}(\tau) + (n - 1)B_{NN}(\tau) = nB_{LL}(\tau) + \tau B'_{LL}(\tau)$ for any isotropic solenoidal field while $B_{LL}(\tau) + (n - 1)B_{NN}(\tau) = nB_{NN}(\tau) + \tau B'_{NN}(\tau)$ for any isotropic potential field. Equations (4.182) and (4.183) evidently imply that the *class of possible longitudinal correlation functions of isotropic solenoidal vector fields in* $\mathbb{R}^n$ *coincides with the class of possible lateral correlation functions of isotropic potential vector fields in* $\mathbb{R}^n$. Moreover, *both these classes are also identical to the class* $\mathfrak{D}_{n+2}$ *of all the (n + 2)-dimensional isotropic correlation functions* (i.e., correlation functions of one-dimensional isotropic fields in $\mathbb{R}^{n+2}$) *and to the class of functions* $B(\tau)$ *such that* $nB(\tau) + \tau B'(\tau)$ *belongs to* $\mathfrak{D}_n$ (i.e., is a possible correlation function of a one-dimensional isotropic field in $\mathbb{R}^n$).

According to the results stated above, examples of isotropic correlation functions presented in subsection 22.2 also yield examples of functions $B_{LL}(\tau)$ corresponding to isotropic solenoidal vector fields and of functions $B_{NN}(\tau)$ corresponding to potential vector fields. In particular, all the functions (2.94), (2.139), (2.126), (2.132) where $0 < m \leqslant 2$, and (4.135) are, for any

integer n, possible longitudinal correlation functions of an isotropic solenoidal vector field and possible lateral correlation functions of an isotropic potential vector field in $\mathbb{R}^n$.   It also easily follows from (4.105), (4.182), (4.126), (4.129), and (4.134) that the spectral density $f_2(k)$ of an isotropic solenoidal field $X(t)$ in $\mathbb{R}^n$, having a longitudinal correlation function of the form $B_{LL}(\tau) = C\exp(-\alpha\tau)$, or $B_{LL}(\tau) = C(\alpha\tau)^\nu K_\nu(\alpha\tau)$, $\nu \geqslant 0$, or, finally, $B_{LL}(\tau) = C\exp(-\alpha\tau^2)$, is given, respectively, by the formulae

$$(4.184) \quad f_2(k) = \frac{2C\alpha\Gamma[(n+3)/2]}{(n-1)\pi^{(n+1)/2}} \frac{k^2}{(\alpha^2 + k^2)^{(n+3)/2}},$$

or

$$(4.185) \quad f_2(k) = \frac{2^\nu C\Gamma[\nu + (n/2) + 1]\,\alpha^{2\nu}}{(n-1)\pi^{n/2}} \frac{k^2}{(\alpha^2 + k^2)^{(n/2)+\nu+1}},$$

or

$$(4.186) \quad f_2(k) = \frac{Ck^2}{2^{n+1}(n-1)\pi^{n/2}\alpha^{1+n/2}} e^{-k^2/4\alpha}.$$

If, however, we consider the same functions as models of the lateral correlation function $B_{NN}(\tau)$ for an isotropic potential field in $\mathbb{R}^n$, then, according to (4.183), the densities $f_1(k)$, which correspond to three above-mentioned forms of the correlation function, will exceed the functions (4.184)−(4.186) by the factor of $(n-1)$.[63*]

22.5. Homogeneous Fields on Spheres and Other Homogeneous Spaces

The concept of a homogeneous (or an isotropic) random field in $\mathbb{R}^n$ can be generalized in such a way that it includes a wide collection of different classes of random functions. Namely, instead of the n-dimensional Euclidean space $\mathbb{R}^n$, we can consider an arbitrary *homogeneous space* $\mathbb{R}$, i.e., a space $\mathbb{R}$ with a transitive "group of motions" G defined in it. (A group of motions G is said to be transitive if for each pair of points t_1 and t_2 of the space $\mathbb{R}$ there is a "motion" g transforming t_1 into t_2, i.e., a g such that $gt_1 = t_2$.) Then, a random field $X(t)$ defined on the space $\mathbb{R}$ is called *homogeneous* if the mean value $\langle X(t) \rangle = m(t)$ of this field is constant (so that $m(gt) =$

$m(t)$ for all motions g), and if the correlation functions $\langle X(t_1)\overline{X(t_2)}\rangle = B(t_1,t_2)$ of $X(t)$ satisfied the condition

(4.187) $B(gt_1,gt_2) = B(t_1,t_2)$ for all g

i.e. it is the same for any two pairs of points which can be carried into each other by some motion of the group G. (The term "isotropic" is also applied sometimes to such fields in place of "homogeneous".)   It is clear that this definition contains the definitions of a homogeneous random field in $\mathbb{R}^n$ or on $\mathbb{Z}^n$ and of an isotropic random field in $\mathbb{R}^n$ as special cases. In the first case, the space $\mathbb{R}$ coincides with $\mathbb{R}^n$ or $\mathbb{Z}^n$, the translation group being the group of motions G. In the second case, $\mathbb{R}$ corresponds to the space $\mathbb{R}^n$ and the ordinary Euclidean group of isometric transformations consisting of all translations, rotations, and reflections is the group of motions.
    The problem of finding the general form of the correlation function of a homogeneous random field in the space $\mathbb{R}$ can evidently be reduced to the problem of finding the general form of a positive definite kernel $B(t_1,t_2)$ defined on $\mathbb{R}$ which is invariant under the group of motions in $\mathbb{R}$ (i.e. satisfies (4.187)). This last problem is treated in many mathematical works.[64] It has been shown, for instance, that for a particular (but a rather wide and most commonly encountered in applications) class of homogeneous spaces $\mathbb{R}$ the general form of a positive definite kernel $B(t_1,t_2)$ which satisfies (4.187) is the following:

(4.188) $B(t_1,t_2) = \int_\Lambda H^{(\lambda)}(t_1,t_2)F(d\lambda),$

where $H^{(\lambda)}(t_1,t_2)$ are some special functions (so-called *zonal spherical functions* or *zonal spherical harmonics* in the space $\mathbb{R}$) depending only on the "composite distance between t_1 and t_2," (i.e. are such that $H^{(\lambda)}(gt_1,gt_2) = H^{(\lambda)}(t_1,t_2)$ for all t_1,t_2 and g), integration is extended over the set $\Lambda = \{\lambda\}$ of all the zonal spherical functions, and $F(d\lambda)$ is a nonnegative measure on Λ such that the integral in the right-hand side of (4.188) converges.[65] Equation (4.188) generalizes the spectral representations (4.46) and (4.105) of the correlation functions for homogeneous and for isotropic random fields in $\mathbb{R}^n$. The spectral representation (4.188) of the correlation function $B(t_1,t_2)$ implies also the spectral representation of the homogeneous field $X(t)$, of the form

(4.189) $X(t) = \sum\limits_{m} \int_{\Lambda_m} \sum\limits_{j=0}^{h(m)} H_{jm}^{(\lambda)}(t) Z_j^{(m)}(d\lambda),$

where $H_{jm}^{(\lambda)}(t)$ are special functions more general than $H^{(\lambda)}(t_1, t_2)$ (so-called *spherical functions* or *spherical harmonics of order* m in the space $\mathbb{R}$), $H_{00}^{(\lambda)}(t) = H^{(\lambda)}(t, t_0)$, where t_0 is a fixed point, are zonal spherical functions, Λ_m (where $m = 0$, 1, ... and $\Lambda_0 = \Lambda$) is a set of all spherical functions of order m, $h(m)$ is the number of such spherical functions, and $Z_j^{(m)}(d\lambda)$ is a family of random measures in Λ_m (i.e. random functions of the point set in Λ_m) such that

(4.190) $\langle Z_j^{(m)}(S_1) \overline{Z_i^{(n)}(S_2)} \rangle = \delta_{mn} \delta_{ji} F(S_1 \cap S_2)$

where $S_1 \cap S_2$ is the intersection of the sets S_1 and S_2 in Λ_m.[66] The spectral representation (4.189) is a generalization of the representations (4.51) and (4.146), which hold for homogeneous fields on $\mathbb{Z}^n$ and in $\mathbb{R}^n$ and for isotropic fields in $\mathbb{R}^n$. When the space $\mathbb{R}$ is compact (i.e. has a finite volume) the sets Λ and Λ_m are discrete, so that the integrals in the right-hand sides of (4.188) and (4.189) reduce to sums. In the literature cited in Note 66, more general spectral representations for homogeneous random fields defined on wider classes of homogeneous spaces and on topological groups are also studied, but we will not consider them here.

Let us now consider a few specific examples.

Example 1. We begin by discussing homogeneous random fields on the sphere $\mathbb{S}^2$ of the ordinary three-dimensional Euclidean space $\mathbb{R}^3$. Then random field $X(\Theta, \varphi)$ (where Θ is the latitude, and φ is the longitude, so that $0 \leqslant \Theta \leqslant \pi$, $0 \leqslant \varphi < 2\pi$), on the sphere $\mathbb{S}^2$ is homogeneous if $\langle X(\Theta, \varphi) \rangle = \text{const}$ (without loss of generality, we can assume that $\langle X(\Theta, \varphi) \rangle = 0$), while

$\langle X(\Theta_1, \varphi_1) \overline{X(\Theta_2, \varphi_2)} \rangle = B(\Theta_{12})$, where Θ_{12} is the spherical (angular) distance between the points (Θ_1, φ_1) and (Θ_2, φ_2). The function $B(\Theta_{12})$ is a positive definite kernel on $\mathbb{S}^2$, which depends only on the spherical distance between its two arguments. It is known that the general form of such kernel is the following:

(4.191) $B(\Theta) = \sum\limits_{m=0}^{\infty} f_m P_m(\cos\Theta)$

where P_m are Legendre's polynomials, and f_m are nonnegative

constants satisfying the condition

(4.192) $\sum\limits_{m=0}^{\infty} f_m < \infty$

(see, e.g., Note 69, where the proof of an even more general statement is outlined). By using the well-known addition theorem for spherical harmonics (see, e.g., Gradshteyn and Ryzhik, 1980, Eq. 8.814, or Erdélyi et al., 1953, vol. 2, Eq. 11.4(8)), we can also write (4.191) as

(4.193) $\langle X(\Theta_1,\varphi_1)\overline{X(\Theta_2,\varphi_2)}\rangle = \sum\limits_{m=0}^{\infty} \sum\limits_{l=-m}^{m} \frac{4\pi}{2m + 1} f_m S_m^l (\Theta_1,\varphi_1)\overline{S_m^l(\Theta_2,\varphi_2)}$

where $S_m^l(\Theta,\varphi) = P_m^l(\cos\Theta)e^{il\varphi}$ are ordinary (three-dimensional) spherical harmonics, and P_m^l are the associated Legendre functions. Using the theorem on the generalized spectral representation of random functions, it is easy to deduce from (4.193) that any homogeneous random field $X(\Theta,\varphi)$ on the sphere $\mathbb{S}^2$ can be represented as

(4.194) $X(\Theta,\varphi) = \sum\limits_{m=0}^{\infty} \sum\limits_{l=-m}^{m} S_m^l(\Theta,\varphi)Z_m^l$

where Z_m^l are random variables such that

(4.195) $\langle Z_m^l \overline{Z_n^k} \rangle = \frac{4\pi}{2m + 1} \delta_{mn}\delta_{lk} f_m.$

Formulae (4.191) and (4.194) − (4.195) are particular cases of (4.188) − (4.190) relating to homogeneous random fields on the sphere $\mathbb{S}^2$. Formula (4.194) gives the spectral representation of such a homogeneous field, while (4.191) is the spectral representation of the corresponding correlation function $B(\Theta_{12})$. The role which is played by the spectral distribution function $F(\lambda)$ in the theory of stationary random functions $X(t)$ is in this case played by the set of nonnegative numbers f_m satisfying the condition (4.192).[67]

Let now $\mathbf{X}(\Theta,\varphi) = \{X_1(\Theta,\varphi), ..., X_s(\Theta,\varphi)\}$ be a multidimensional (s-dimensional) homogeneous random field on the sphere $\mathbb{S}^2$, which has constant mean value vector $\langle \mathbf{X}(\Theta,\varphi)\rangle = \mathbf{m}$ and the correlation matrix $\|\langle X_j(\Theta_1,\varphi_1)X_l(\Theta_2,\varphi_2)\rangle \| = \|B_{jl}(\Theta_{12})\| = \mathcal{B}(\Theta_{12})$ which depends only on the spherical distance, Θ_{12}, between the points (Θ_1,φ_1) and (Θ_2,φ_2). Then, the general form of the correlation matrix $\mathcal{B}(\Theta)$ is given by

$$(4.191a) \qquad B(\Theta) = \sum_{m=0}^{\infty} \hat{f}_m P_m(\cos\Theta).$$

which is quite similar to (4.191). Now, however, the

coefficients $\hat{f}_m = \| f_{mij} \|$ are no longer nonnegative constants, but positive definite ($s \times s$)-matrices. Note that all the components $X_j(\Theta,\varphi)$, $j=1,...,s$, of the field $\mathbf{X}(\Theta,\varphi)$ are one-dimensional homogeneous fields on the sphere. It is clear, therefore, that all these components can be represented in the form (4.194), where the random variables Z_m^l must, of course, be replaced by Z_{mj}^l while the condition (4.195) now takes the form

$$(4.195a) \qquad \langle Z_{mj}^l \overline{Z_{ni}^k} \rangle = \frac{4\pi}{2m+1} \delta_{mn} \delta_{lk} f_{mji}.$$

More complicated spectral representations, which include some generalizations of the ordinary spherical functions $S_m^l(\Theta,\varphi)$, can be obtained for correlation matrices $B(\Theta_1,\varphi_1;\Theta_2,\varphi_2)$

and components $X_i(\Theta,\varphi)$ of homogeneous fields on $\mathbb{S}^2$ of various multidimensional quantities $\mathbf{X}(\Theta,\varphi)$, e.g. vectors, polyvectors, tensors, etc. (Recall that the components of such quantities $\mathbf{X}$ transform linearly among themselves under the rotations of the sphere.) The problem of the spectral representation of homogeneous fields of quantities $\mathbf{X}$ is also of interest for some applications, but we cannot pursue this topic further.[68]

Example 2. The spectral theory of homogeneous random fields $X(\Theta_1,...,\Theta_{n-2},\varphi) = X(\mathbf{t})$ on the sphere $\mathbb{S}^{n-1}$ of the n-dimensional Euclidean space $\mathbb{R}^n$ is almost as simple. (We are using here hyperspherical coordinates $\Theta_1, ..., \Theta_{n-2}, \varphi$ on the sphere $\mathbb{S}^{n-1}$, and denote by $\mathbf{t}$, where $|\mathbf{t}| = 1$, the unit vector in $\mathbb{R}^n$ corresponding to the point of $\mathbb{S}^{n-1}$ with coordinates $(\Theta_1, ... \Theta_{n-2},\varphi)$.) Here, too, the mean value $\langle X(\mathbf{t}) \rangle = m$ must be constant,

while the correlation function $\langle X(\mathbf{t}_1)\overline{X(\mathbf{t}_2)} \rangle = B(\Theta_{12})$ may depend only on the spherical distance Θ_{12} between the points $\mathbf{t}_1$ and $\mathbf{t}_2$ (i.e. on the angle between the vectors $\mathbf{t}_1$ and $\mathbf{t}_2$). The general form of the positive definite kernel $B(\Theta_{12})$ on the sphere $\mathbb{S}^{n-1}$, which depends only on the spherical distance Θ_{12} between its two arguments, is given by

$$(4.196) \qquad B(\Theta) = \sum_{m=0}^{\infty} f_m C_m^{(n-2)/2}(\cos\Theta)$$

where $C_m^{(n-2)/2}(x)$ are Gegenbauer's polynomials of degree m and order $(n - 2)/2$ (i.e. n-dimensional zonal surface harmonics), while f_m are nonnegative constants satisfying condition (4.192).[69] Moreover, the general addition theorem for n-dimensional surface harmonics (see Erdélyi et al., 1953, Vol. 2, Sec. 11.4), permits one to write (4.196) as

$$(4.197) \qquad B(\Theta_{12}) = \sum_{m=0}^{\infty} \sum_{l=1}^{h(m,n)} c_{m,n}^2 f_m S_m^l(t_1) S_m^l(t_2),$$

where $c_{m,n} = 4\pi^{n/2}[(2m + n - 2)\Gamma([n - 2]/2)]^{-1}$, and $h(m,n)$ is the number of surface harmonics of degree m in the n-dimensional space (cf. (4.145) above). Equation (4.197) and the theorem on generalized spectral representation of random functions now imply that any homogeneous random field $X(t)$ on the sphere $\mathbb{S}^{n-1}$ admit the following spectral representation:

$$(4.198) \qquad X(t) = \sum_{m=0}^{\infty} \sum_{l=1}^{h(m,n)} S_m^l(t) Z_m^l,$$

where Z_m^l are random variables such that

$$(4.199) \qquad \langle Z_m^l Z_n^k \rangle = \delta_{mn} \delta_{lk} c_{m,n} f_m.$$

Equations (4.196) − (4.199), which determine the spectral representations for homogeneous random fields on $\mathbb{S}^{n-1}$ and for their correlation functions, are also particular cases of the general equations (4.188) − (4.190). The role of the spectral distribution function here is again played by the set of nonnegative numbers f_m satisfying the condition (4.192).[70]

Equation (4.196) enables one to construct many examples of possible correlation functions $B(\Theta)$ of homogeneous fields on multidimensional spheres. Note also that if $X(t)$ is a homogeneous random field on the sphere $\mathbb{S}^{n-1}$, then the values of this field at points belonging to the intersection of $\mathbb{S}^{n-1}$ with a k-dimensional linear subspace $\mathbb{R}^k$ of the space $\mathbb{R}^n$, where $k < n$, will form a homogeneous random field on the sphere $\mathbb{S}^{k-1}$ of the k-dimensional Euclidean space $\mathbb{R}^k$. It follows from this that *any correlation function* $B(\Theta)$ *of a homogeneous random field on the sphere* $\mathbb{S}^{n-1}$ *is also a possible correlation function of a homogeneous random field on the sphere* $\mathbb{S}^{k-1}$, *where* $k < n$. Moreover, if $X(t)$ is an isotropic random field in the whole space $\mathbb{R}^n$, then the values of $X(t)$ on the sphere $\mathbb{S}_r^n$ of this space of radius r clearly form a homogeneous random field on $\mathbb{S}_r^n$. Let the spherical (i.e. angular) distance between the points t_1 and t_2

of the sphere $\mathbb{S}_r^n$ be Θ; then, of course, the Euclidean distance between the points $\mathbf{t}_1$ and $\mathbf{t}_2$ of the space $\mathbb{R}^n$ is equal to $2r\sin(\Theta/2)$. Therefore, *if the function* $B_1(\tau)$ *belongs to the class* $\mathfrak{D}_n$ (i.e. is a possible correlation function of an isotropic random field in $\mathbb{R}^n$), *then the function* $B(\Theta) = B_1(2r\sin\Theta/2)$ *is a possible correlation function of a homogeneous random field on the sphere* $\mathbb{S}^{n-1}$. Thus, any function of the form

$$(4.200) \quad B(\Theta) = 2^{(n-2)/2}\Gamma(n/2)\int_0^\infty \frac{J_{(n-2)/2}(2kr\sin\Theta/2)}{(2kr\sin\Theta/2)^{(n-2)/2}}d\Phi(k),$$

where $\Phi(k)$ is a bounded nondecreasing function, is a possible correlation function of a homogeneous random field on $\mathbb{S}^{n-1}$, i.e. it can also be represented in the form (4.196), where f_m are nonnegative constants.[71]

Example 3.   A number of specific spectral representations of homogeneous random fields can be obtained by considering the situation where $X(\mathbf{t}) = X(t_1, ..., t_m, t_m+1, ..., t_n)$, while the group of motions G acting in $\mathbb{R}^n$ is a direct product of some group G_1 acting in the space $\mathbb{R}^m$ of the vectors $(t_1, ..., t_m) = \mathbf{t}^{(1)}$ and of another group G_2 acting in the space $\mathbb{R}^{n-m}$ of vectors $(t_{m+1}, ..., t_n) = \mathbf{t}^{(2)}$. Thus, for instance, in many applications (e.g., in the theory of turbulence) one has to deal with random fields $X(\mathbf{s},t)$ depending on the point $\mathbf{s} = (s_1, ..., s_n)$ of the n-dimensional Euclidean space (where usually $n = 2$ or $n = 3$) and on the time t. Moreover, it is often natural to assume that

$$\langle X(\mathbf{s},t)\rangle = m = \text{const},$$
(4.201)
$$\langle X(\mathbf{s}_1,t_1)X(\mathbf{s}_2,t_2)\rangle = B(|\mathbf{s}_1 - \mathbf{s}_2|, t_1 - t_2),$$

where $|\mathbf{s}_1 - \mathbf{s}_2|$ is the Euclidean distance between the points $\mathbf{s}_1$ and $\mathbf{s}_2$. In this case the field $X(\mathbf{s},t)$ is homogeneous under all isometric transformations (translations, rotations, and reflections) of vectors $\mathbf{s}$ and all shifts of the time argument t. It is then easy to derive, for the case where $n = 2$, the following general form of the correlation function $B(\sigma,\tau)$

$$(4.202) \quad B(\sigma,\tau) = \int_{-\infty}^\infty\int_0^\infty e^{i\tau\omega}J_0(k\sigma)\,F(d\omega,dk),$$

where $F(d\omega,dk)$ is a nonnegative measure in the half-plane

$-\infty < \omega < \infty,\ 0 \leqslant k < \infty$. The corresponding spectral representation of the field $X(\mathbf{s},t) = X(s,\varphi,t)$ itself has the form

$$(4.203) \quad X(s,\varphi,t) = \sum_{m=-\infty}^{\infty} e^{im\varphi} \int_{-\infty}^{\infty}\int_{0}^{\infty} e^{it\omega} J_m(ks) Z_m(d\omega,dk)$$

where

$$(4.204) \quad \langle Z_m(d\omega,dk) Z_n(d\omega',dk') \rangle$$

$$= \delta_{mn}\delta(k - k')\delta(\omega - \omega')F(d\omega,dk)d\omega'dk'.$$

Similar formulae can readily be given for an arbitrary value of n.[72]

Many additional examples of spectral representations for homogeneous fields on various homogeneous spaces can be found in the literature cited in Note 66.[73]

23. Random Processes with Stationary Increments

Now we return to random functions $X(t)$ of one real variable t. Consider the examples of the Poisson random process (Fig. 10(b) on p. 33) and of the Brownian motion (i.e., Wiener) process $W(t)$ (Fig. 4(b) and pp. 172–173 and 175–176). Both processes are obviously not stationary, since it is clear that $\langle |X(t)|^2 \rangle$ in both cases increases without limit as t increases. However, there is still something resembling stationarity involved here: For example, the set of increments of $X(t)$ during consecutive and equal time intervals, i.e., the set of random variables

$$(4.205) \quad \Delta_k X = X((k + 1)\Delta) - X(k\Delta), \quad k = \ldots,-2,-1,0,1,2, \ldots ,$$

where Δ is an arbitrary fixed number, form a stationary random sequence. Note that the Poisson and Brownian motion processes have, in fact, an additional rather special property: Their increments $\Delta_k X$ are mutually uncorrelated (and even independent). However it is not hard to construct other examples of processes $X(t)$ where this special property is not valid, but the sequence of increments (4.205) is nevertheless stationary. For example, instead of the Poisson process one may consider a more general counting process $X(t)$ corresponding to the arbitrary stationary point process $\{t_k\}$ (a random point system whose probability distributions do not

change when the time origin is arbitrarily shifted). Similarly, instead of Wiener's model of the Brownian motion one can consider a refined model of the one-dimensional Brownian motion (namely, Ornstein–Uhlenbeck's model; see pp. 174–176). One more example of the same type is provided by a phase modulated oscillation of the form

(4.206) $Y(t) = A\cos(\int_0^t [\omega_0 + \xi(t)]dt + \Theta_0) = A\cos[\omega_0 t + X(t)]$

where ω_0 and Θ_0 are constants and $\xi(t)$ is a stationary random

process. In this case the phase $X(t) = \Theta_0 + \int_0^t \xi(t')dt'$ is a

nonstationary random process, the increments of which (given by (4.205)) also form a stationary sequence. All these examples suggest that we try to generalize the definition of stationarity in such a way as to include the examples mentioned above. The present section is devoted to this generalization.

We are interested now in processes $X(t)$, where only the increment $X(u) - X(t)$ is stationary but the process $X(t)$ itself it not. Because of this, in studying such $X(t)$ it is appropriate to choose as the basic characteristics of $X(t)$ the mean value of the increment

(4.207) $\langle X(u) - X(t) \rangle = m(t,u)$

and the correlation function of the increment

(4.208) $\langle [X(u) - X(t)][\overline{X(v) - X(s)}] \rangle = D(t,s;u,v)$

and not the mean value $\langle X(t) \rangle = m(t)$ and the correlation

function $\langle X(t)\overline{X(s)} \rangle = B(t,s)$ of the process $X(t)$. We shall always suppose below that $m(t,u)$ and $D(t,s;u,v)$ are continuous functions of their arguments. (This condition is, of course, valid, if $X(t)$ is a mean square continuous process.) The function (4.208) is called the general *structure function* of the process $X(t)$.[74]

The mean value of the increment $m(t,u)$ and the structure function $D(t,s;u,v)$ obviously tell us less about the process $X(t)$ than the mean value $m(t)$ and the correlation function $B(t,s)$. In fact, if we know $m(t)$ and $B(t,s)$, we can easily determine

$m(t,u)$ and $D(t,s;u,v)$ from the equations

(4.209)
$$m(t,u) = m(u) - m(t),$$
$$D(t,s;u,v) = B(u,v) - B(u,s) - B(t,v) + B(t,s),$$

but if we only know $m(t,u)$ and $D(t,s;u,v)$ we cannot, in general uniquely determine $m(t)$ and $B(t,s)$. However, one practically important case will be indicated below, where $B(t,s)$ can be determined from $D(t,s;u,v)$. (It should also be noted that in the case of a process $X(t)$ having finite mean value of the increment and structure function, the mean value $\langle X(t) \rangle$ and correlation function $\langle X(t)X(s) \rangle$ need not exist in general.)

It follows at once from (4.207) and (4.208) that

(4.210) $m(t,u) = - m(u,t), \; m(t,u) + m(s,t) = m(s,u),$

(4.211) $D(t,s;u,v) = \overline{D(s,t;v,u)},$

(4.212) $D(t,s;u,v) + D(u,s;w,v) = D(t,s;w,v),$

(4.213) $D(t,s;u,v) = - D(u,s;t,v).$

Equation (4.210) implies, in particular, that instead of the function $m(t,u)$ of two variables we need only consider the function $m_1(u) = m(u,0) = \langle X(u) - X(0) \rangle$ of one variable: $m(t,u) = m_1(u) - m_1(t)$. Similarly equations (4.212) and (4.213) imply that instead of the general structure function (4.208) we need only consider the function

(4.214) $D(t;u,v) = D(t,t;u,v) = \langle [X(u) - X(t)][\overline{X(v) - X(t)}] \rangle$

which depends on three variables. In fact, in terms of $D(t;u,v)$ the function $D(t,s;u,v)$ is given by

(4.215) $D(t,s;u,v) = D(t;u,v) - D(t;u,s).$

In the real case, we can go even further in this direction by considering the function

(4.216) $D_1(t,u) = D(t;u,u) = \langle |X(u) - X(t)|^2 \rangle$

which depends on two variables. In fact, using the

elementary identity

(4.217) $(a - b)(c - d) = \dfrac{1}{2}\{(a - d)^2 + (b - c)^2 - (a - c)^2 - (b - d)^2\}$

and substituting $a = X(u)$, $b = X(t)$, $c = X(v)$, $d = X(s)$, we obtain that

(4.218) $D(t,s;u,v) = \dfrac{1}{2}\{D_1(s,u) + D_1(t,v) - D_1(v,u) - D_1(t,s)\}$

in the real case (and therefore $D(t;u,v) = (1/2)\{D_1(t,u) + D_1(t,v) - D_1(v,u)\}$). However, in the general case of complex $X(t)$ the relation (4.218) is no longer true and from a knowledge of just the function (4.216) we cannot uniquely determine the more general functions (4.208) and (4.214). From now on, we shall not use the general function $D(t,s;u,v)$ and we shall reserve the term "structure function" to designate the simpler functions $D(t;u,v)$ and $D_1(t,u)$.

We shall call the random process $X(t)$ a *process with stationary increments,* if the mean value of its increment $m(t,u)$ depends only on the length $u - t$ of the interval $[u,t]$:

(4.219) $\langle X(u) - X(t) \rangle = m_1(u - t)$

and if the structure function $D(t;u,v)$ of the process $X(t)$ depends only on the differences $u - t$ and $v - t$:

(4.220) $\langle [X(u) - X(t)][X(v) - X(t)] \rangle = D(u{-}t, v{-}t).$

(Of course, the term "a process with wide sense stationary increments" would be more appropriate for such processes. However this refinement is unnecessary in this book where more special processes with strictly stationary increments will not be considered at all.[75]) Thus a process with stationary increments is characterized by one function of one variable

(4.221) $m_1(\tau) = \langle X(t + \tau) - X(t) \rangle$

and by one function of two variables

(4.222) $D(\tau_1,\tau_2) = \langle [X(t + \tau_1) - X(t)][\overline{X(t + \tau_2) - X(t)}] \rangle.$

In the real case, instead of (4.222) we can consider a simpler function of one variable

(4.223) $D(\tau) = D(\tau,\tau) = \langle |X(t + \tau) - X(t)|^2 \rangle$

since according to (4.218)

(4.224) $D(\tau_1,\tau_2) = \dfrac{1}{2} \{ D(\tau_1) + D(\tau_2) - D(|\tau_1 - \tau_2|) \}$

for real $X(t)$. Recall that we have agreed to consider only the case where all the functions $m(\tau)$, $D(\tau)$, and $D(\tau_1,\tau_2)$ are continuous. The determination of the general form of these three functions will be our next task.

Note first of all that the concept of a process with stationary increments can also be applied to random functions of a discrete parameter, i.e., to random sequences $X(t)$, $t = ..., -1, 0, 1, ...$. The random sequence $X(t)$ is called a *sequence with stationary increments*, if $\langle X(t + \tau) - X(t) \rangle = m(\tau)$ and $\langle [X(t + \tau_1) - X(t)][\overline{X(t + \tau_2) - X(t)}] \rangle = D(\tau_1,\tau_2)$ (i.e., in particular, $\langle |X(t + \tau) - X(t)|^2 \rangle = D(\tau)$) for any integers t, τ, τ_1 and τ_2. It is clear, however, that the study of such sequences is of little theoretical interest, since it can always be reduced to the study of the stationary sequence $Y(t) = X(t+1) - X(t)$ which consists of the differences of neighboring elements in the original sequence. Let

$$\langle Y(t) \rangle = c_1, \quad \langle Y(t + \tau)\overline{Y(t)} \rangle = B_y(\tau) = \int^{\pi} e^{i\tau\omega} dF_y(\omega),$$

$$Y(t) = \int_{-\pi}^{\pi} e^{i\tau\omega} dZ_y(\omega),$$

where m_1 is a constant, $F_y(\omega)$ is a bounded nondecreasing function, and $Z_y(\omega)$ is a random function with uncorrelated increments such that $\langle |dZ_y(\omega)|^2 \rangle = dF_y(\omega)$. Since $X(t + \tau) - X(t) = \sum\limits_{s=t}^{t+\tau-1} Y(s)$, we obtain

(4.225) $m_1(\tau) = c_1\tau,$

(4.226)

$$D(\tau_1,\tau_2) = \int_{-\pi}^{\pi} \frac{(e^{i\tau_1\omega} - 1)(e^{-i\tau_2\omega} - 1)}{|e^{i\omega} - 1|^2} \, dF_y(\omega),$$

$$D(\tau) = \int_{-\pi}^{\pi} \frac{|e^{i\tau\omega} - 1|^2}{|e^{i\omega} - 1|^2} \, dF_y(\omega),$$

and

$$(4.227) \quad X(t + \tau) - X(t) = \int_{-\pi}^{\pi} e^{it\omega} \frac{e^{iT\omega} - 1}{e^{i\omega} - 1} \, dZ_y(\omega).$$

Now note that $\omega^2/|e^{i\omega} - 1|^2$ is a strictly positive bounded continuous function in the interval $-\pi \leqslant \omega \leqslant \pi$. Therefore, if

$$(4.228) \quad F(\omega) = \int_0^{\omega} \frac{\lambda^2}{|e^{i\lambda} - 1|^2} \, dF_y(\lambda), \quad Z(\omega) = \int_0^{\omega} \frac{i\lambda}{e^{i\lambda} - 1} \, dZ_y(\lambda),$$

then the function $F(\omega)$ is also a bounded nondecreasing function and $Z(\omega)$ is a random function with uncorrelated increments satisfying the relation $\langle |dZ(\omega)|^2 \rangle = dF(\omega)$. Hence the representations (4.226) and (4.227) can also be written as

$$D(\tau_1, \tau_2) = \int_{-\pi}^{\pi} \frac{(e^{iT\omega} - 1)(e^{-iT\omega} - 1)}{\omega^2} \, dF(\omega),$$

(4.226a)

$$D(\tau) = 4 \int_{-\pi}^{\pi} \frac{\sin^2(T\omega/2)}{\omega^2} dF(\omega),$$

$$(4.227a) \quad X(t + \tau) - X(t) = \int_{-\pi}^{\pi} e^{it\omega} \frac{e^{iT\omega} - 1}{i\omega} \, dZ(\omega)$$

where $F(\omega)$ is a bounded nondecreasing function and

$$(4.229) \quad \langle dZ(\omega)\overline{dZ(\omega')} \rangle = \delta(\omega - \omega')dF(\omega)d\omega'.$$

It will be explained below that similar formulae are also valid for any process with stationary increments which depends on continuous time.

Consider now again the process $X(t)$ where t takes any real value. It is clear that any stationary process is at the same time a process with stationary increments. For a stationary process, $m(\tau) = \langle X(t - \tau) - X(t) \rangle = 0$ and the functions $D(\tau_1, \tau_2)$ and $D(\tau)$ can be expressed in terms of the correlation function $B(\tau)$ as follows:

$$D(\tau_1, \tau_2) = B(\tau_1 - \tau_2) - B(\tau_1) - B(-\tau_2) + B(0),$$

(4.230)

$$D(\tau) = 2B(0) - B(\tau) - B(-\tau).$$

Therefore, in this case

$$(4.231) \quad D(\tau_1, \tau_2) = \int_{-\infty}^{\infty} (e^{i\tau_1 \omega} - 1)(e^{-i\tau_2 \omega} - 1) dF(\omega),$$

$$D(\tau) = 2 \int_{-\infty}^{\infty} (1 - \cos \omega \tau) dF(\omega)$$

where $F(\omega)$ is the spectral distribution function of the process $X(t)$. If $X(t)$ is real, we need only consider the structure function $D(\tau)$ and according to (4.230) in this case

$$(4.232) \quad D(\tau) = 2\{B(0) - B(\tau)\}.$$

It follows from this equation that, in the special case under consideration, the structure function is necessarily bounded: $|D(\tau)| \le 4B(0)$. Moreover (4.232) implies that, if we know the correlation function $B(\tau)$ for a real stationary process, this means that we know its structure function $D(\tau)$ also. The converse statement is not, in general, correct: The function $B(\tau)$ cannot be uniquely determined from $D(\tau)$. If, however, it is clear from physical considerations that $\langle X(t) \rangle = 0$ and the correlation between $X(t)$ and $X(t + \tau)$ vanishes as $\tau \to \infty$, then $B(\tau) \to 0$ as $\tau \to \infty$ and by (4.232) we have

$$(4.233) \quad \lim_{\tau \to \infty} D(\tau) = D(\infty) = 2B(0), \quad B(\tau) = \{D(\infty) - D(\tau)\}/2.$$

Hence, for real stationary random processes with a correlation function $B(\tau)$ which vanishes at infinity, the function $B(\tau)$ and $D(\tau)$ are mutually interchangeable: Each can be determined from the other. (In the case of real stationary processes $X(t)$ with $\langle X(t) \rangle = m \ne 0$ and a centered correlation function $b(\tau) = \langle [X(t + \tau) - m][X(t) - m)] \rangle = B(\tau) - m^2$ which vanishes at infinity, the constant $m = \lim_{\tau \to \infty} \{B(\tau)\}^{1/2}$ and the functions $b(\tau) = B(\tau) - m^2$

and $D(\tau) = 2\{B(0) - B(\tau)\}$ can be determined from $B(\tau)$; however, only $b(\tau)$, but not m and not $B(\tau)$, can be determined from $D(\tau)$.)

In the experimental determination of the structure function $D(\tau)$ the statistical ("ensemble") average is usually replaced by the time average of the quantities $\Delta_\tau X(t) = X(t + \tau) - X(t)$ (where τ is fixed) over a sufficiently long time interval T. It then frequently turns out that for a stationary process $X(t)$ the structure function $D(\tau)$ found by averaging over a given time interval T is obtained with a smaller error than in the case of the correlation function $B(\tau)$.[76] Therefore, even in those cases where the correlation function $B(\tau)$ is of our main interest, it is frequently useful to find from the data the

values of $D(\tau)$ and also $B(0) = \langle X^2(t) \rangle$, and then use the second formula (4.233).

Let us now consider general processes with stationary increments which are not necessarily stationary. It follows from (4.210) that, in general, the function $m_1(\tau)$ must satisfy the equation

(4.234) $m_1(\tau_1) + m_1(\tau_2) = m_1(\tau_1 + \tau_2).$

Since $m_1(\tau)$ is a continuous function of τ, (4.234) implies that

(4.235) $m_1(\tau) = c_1 \tau$

where c_1 is a constant. If the mean value $\langle X(0) \rangle = c_0$ exists, $\langle X(t) \rangle = \langle [X(t) - X(0)] + X(0) \rangle$ must have the form

(4.236) $\langle X(t) \rangle = c_0 + c_1 t,$

i.e., the mean value $\langle X(t) \rangle = m(t)$ is a linear function of t.

Next, we study the structure function $D(\tau) = \langle |X(t + \tau) - X(t)|^2 \rangle$, again restricting ourselves to the case where $X(t)$ is real. The function $D(\tau)$ by definition is nonnegative, even, and such that $D(0) = 0$. Since

$$\frac{\langle [X(t + \tau_1 + \tau_2) - X(t + \tau_2)][X(t + \tau_2) - X(t)] \rangle}{\langle [X(t + \tau_1 + \tau_2) - X(t + \tau_2)]^2 \rangle^{1/2} \langle [X(t + \tau_2) - X(t)]^2 \rangle^{1/2}} \leqslant 1$$

(see 0.21)) and $D(\tau_1 + \tau_2) = \langle \{[X(t + \tau_1 + \tau_2) - X(t + \tau_2)] + [X(t + \tau_2) - X(t)]\}^2 \rangle$,

(4.237) $D(\tau_1 + \tau_2) \leqslant [\{D(\tau_1)\}^{1/2} + \{D(\tau_2)\}^{1/2}]^2$

for any $\tau_1 \geqslant 0$ and $\tau_2 \geqslant 0$. It follows easily from this that for any $\tau_0 > 0$ a constant A exists such that

(4.238) $D(\tau) \leqslant A\tau^2$

for $\tau \geqslant \tau_0$.[77] It is clear that the function $D(\tau)$ can increase quadratically with τ: e.g., $D(\tau) = A\tau^2$, $A = \langle |X_1|^2 \rangle$, for the process $X(t) = X_0 + X_1 t$ (where X_0 and X_1 are random variables), which evidently has stationary increments. Moreover, according to (0.10)

$$D(\tau) = \langle X(t + \tau) - X(t) \rangle^2 + \sigma^2 \{X(t + \tau) - X(t)\}$$

(4.239)

$$= m_1^2(\tau) + D_1(\tau) = c_1^2 \tau^2 + D_1(\tau),$$

where $D_1(\tau) = \sigma^2 \{X(t + \tau) - X(t)\} \geqslant 0$. Therefore $D(\tau)$ includes a term proportional to τ^2 in all the cases where $c_1 \neq 0$. Thus we see that the function $D(\tau)$ is, in general, not bounded. However, according to (4.238), this function cannot increase too rapidly as $\tau \to \infty$.

Looking for a general form of the functions $D(\tau)$ and $D(\tau_1, \tau_2)$ it is convenient to begin with the case of *differentiable* processes $X(t)$. This case is rather simple, since if $X(t)$ is a process with stationary increments and its mean square derivative $X'(t)$ exists, this derivative clearly is a stationary process. Therefore the study of differentiable processes with stationary increments can always be reduced to the study of stationary processes $Y(t) = X'(t)$. Let

$$Y(t) = \int_{-\infty}^{\infty} e^{it\omega} dZ_y(\omega), \quad B_y(\tau) = \int_{-\infty}^{\infty} e^{i\tau\omega} dF_y(\omega)$$

be spectral representations of the process $Y(t)$ itself and of its correlation function $B_y(\tau) = \langle Y(t + \tau)\overline{Y(t)} \rangle$. Then it is readily seen that

$$X(t) - X(0) = \int_0^t Y(s)ds = \int_{-\infty}^{\infty} \left\{ \int_0^t e^{is\omega}ds \right\} dZ_y(\omega)$$

(4.240)

$$= \int_{-\infty}^{\infty} \frac{e^{it\omega} - 1}{i\omega} dZ_y(\omega).$$

Hence

(4.241) $$X(t) = \int_{-\infty}^{\infty} \frac{e^{it\omega} - 1}{i\omega} dZ_y(\omega) + X_0,$$

where $X_0 = X(0)$ is a constant random variable, and

$$D(\tau_1; \tau_2) = \langle [X(\tau_1) - X(0)][\overline{X(\tau_2) - X(0)}] \rangle$$

(4.242)

$$= \int_{-\infty}^{\infty} \frac{(e^{i\tau_1\omega} - 1)(e^{-i\tau_2\omega} - 1)}{\omega^2} dF_y(\omega),$$

$$D(\tau) = D(\tau,\tau) = 2\int_{-\infty}^{\infty} \frac{1 - \cos\omega\tau}{\omega^2}\, dF_y(\omega)$$

(4.243)

$$= 4\int_{-\infty}^{\infty} \frac{\sin^2\omega\tau/2}{\omega^2}\, dF_y(\omega).$$

Note that, if the point $\omega = 0$ is a jump discontinuity of $Z_y(\omega)$, i.e., $\lim_{\epsilon \to 0}[Z_y(\epsilon) - Z_y(-\epsilon)] = X_1 \neq 0$, where X_1 is a

random variable (which can, in particular, be constant), then $\lim_{\epsilon \to 0}[F_y(\epsilon) - F_y(-\epsilon)] = \langle|X_1|\rangle = A_1^2 > 0$, i.e., the point $\omega = 0$ is a

jump discontinuity of $F_y(\omega)$ also.   The contributions of this discontinuity to the integrals on the right-hand parts of (4.241), (4.242), and (4.243) are evidently equal to $X_1 \lim_{\omega \to 0} (e^{it\omega} - 1)/i\omega =$

$X_1 t$, $A_1^2 \lim_{\omega \to 0} (e^{iT_1\omega} - 1)(e^{-iT_2\omega} - 1)/\omega^2 = A_1^2 T_1 T_2$ and $2A_1^2 \lim_{\omega \to 0} (1 - $

$\cos\omega\tau)/\omega^2 = A_1^2 \tau^2$ respectively.

It is easy to see that formulae (4.241) - (4.243) can also be written as

(4.244) $X(t) = \int_{-\infty}^{\infty} (e^{it\omega} - 1)dZ(\omega) + X_0 + X_1 t,$

(4.245) $D(\tau_1,\tau_2) = \int_{-\infty}^{\infty} (e^{iT_1\omega} - 1)(e^{-iT_2\omega} - 1)dF(\omega) + A_1^2 \tau_1 \tau_2,$

(4.246) $D(\tau) = 2\int_{-\infty}^{\infty} (1 - \cos\omega\tau)dF(\omega) + A_1^2 \tau^2.$

To do so we need only introduce the functions

(4.247) $Z(\omega_2)-Z(\omega_1) = \int_{\omega_1}^{\omega_2} \frac{dZ_y(\omega)}{i\omega}$, $F(\omega_2)-F(\omega_1) = \int_{\omega_1}^{\omega_2} \frac{dF_y(\omega)}{\omega^2}$

for $0 < \omega_1 < \omega_2$ or $\omega_1 < \omega_2 < 0$ and then interpret the integral from $-\infty$ to ∞ in (4.244) $-$ (4.246) as the limit

(4.248) $\int_{-\infty}^{\infty} = \lim_{T \to \infty, \epsilon \to 0} \left\{ \int_{-T}^{-\epsilon} + \int_{\epsilon}^{T} \right\}.$

(Note that the integrals on the right-hand sides of (4.247) can be nonexistent for intervals $[\omega_1,\omega_2]$ containing the point $\omega = 0$ and therefore $Z(\omega)$ and $F(\omega)$ are not defined at the point $\omega = 0$.) Formulae (4.244) $-$ (4.246) specify *spectral representations* of differentiable processes $X(t)$ with

stationary increments and of their structure functions $D(\tau_1, \tau_2)$ and $D(\tau)$. The function $F(\omega)$ is the *spectral distribution function* of a process $X(t)$, and $Z(\omega)$ is a random function with uncorrelated increments satisfying (4.229). By virtue of (4.247) $F(\omega)$ is a nondecreasing function on the half-lines $(-\infty, 0)$ and $(0, \infty)$ such that

$$(4.249) \quad \int_{-\infty}^{\infty} \omega^2 dF(\omega) < \infty.$$

If the correlation function $B_y(\tau)$ of $Y(t) = X'(t)$ falls off rapidly enough with $|\tau|$ (say satisfies the condition (2.66) or (2.68)), then $A_1 = 0$ and $dF(\omega)$ can be replaced in (4.245), (4.246), and (4.249) by $f(\omega)d\omega$, where $f(\omega) = F'(\omega) \geqslant 0$ is the *spectral density* of a process $X(t)$. Now suppose that the

process $X(t)$ is real; then $dZ(-\omega) = \overline{dZ(\omega)}$, $dF(-\omega) = dF(\omega)$ and it is sufficient to consider only the function $D(\tau)$ (but not $D(\tau_1, \tau_2)$). Therefore two equations, (4.245) and (4.246), can be replaced in the real case by a single equation

$$(4.250) \quad D(\tau) = 4\int_0^{\infty} (1 - \cos\omega\tau)dF(\omega) + A_1^2 \tau^2$$

or, if the spectral density exists,

$$(4.250a) \quad D(\tau) = 4\int_0^{\infty} (1 - \cos\omega\tau)f(\omega)d\omega$$

where $\int_0^{\infty} \omega^2 dF(\omega) < \infty$ or, respectively, $\int_0^{\infty} \omega^2 f(\omega)d\omega < \infty$.

It can be shown that spectral representations similar to (4.244) - (4.246) also exist for any nondifferentiable process $X(t)$ with stationary increments. The only difference as compared with the differentiable case is that, in the general case, the spectral distribution function $F(\omega)$ can increase so rapidly as $|\omega| \to \infty$, that the integral in (4.249) becomes infinite. Instead of (4.249) it is only necessary that, for any $\omega_0 > 0$,

$$(4.251) \quad \int_{-\infty}^{-\omega_0} dF(\omega) + \int_{-\omega_0}^{\omega_0} \omega^2 dF(\omega) + \int_{\omega_0}^{\infty} dF(\omega) < \infty$$

where the integral from $-\omega_0$ to ω_0 is interpreted as $\lim_{\epsilon \to 0} \{\int_{-\omega_0}^{-\epsilon} + \int_{\epsilon}^{\omega_0}\}$. If a process $X(t)$ is real and its spectral distribution function $F(\omega)$ is absolutely continuous (i.e., the spectral

density $f(\omega)$ exists), then $f(\omega) = f(-\omega)$ and, for any $\omega_0 > 0$,

$$(4.251a) \quad \int_0^{\omega_0} \omega^2 f(\omega)d\omega + \int_{\omega_0}^{\infty} f(\omega)d\omega < \infty.$$

In other words, near zero frequency the integral $\int \omega^2 dF(\omega)$ (or $\int \omega^2 f(\omega)d\omega$) must converge while at infinity the integral $\int dF(\omega)$

(or $\int f(\omega)d\omega$) must converge. However, the integral $\int_{-\infty}^{\infty} dF(\omega)$

(or $\int_{-\infty}^{\infty} f(\omega)d\omega$), which is usually interpreted as the total

power (or total energy) in the case of stationary processes, may become infinite, if $X(t)$ has stationary increments but is not stationary. This fact shows that, for the processes with stationary increments, the mentioned integral will either have a different physical meaning or it will be infinite only because of the mathematical idealization of the process $X(t)$ which cannot really be used to describe the true spectral behavior at the lowest frequencies (as $\omega \to 0$). Note also that

the divergence of the integral $\int dF(\omega)$ (or $\int f(\omega)d\omega$) near zero frequency is compensated by the fact that the functions $[\exp(i\tau_1\omega) - 1][\exp(-i\tau_1\omega) - 1]$ and $1 - \cos\omega\tau$ tend to zero as ω

$\to 0$ in proportion to ω^2. Therefore, condition (4.251) ensures that the integrals on the right-hand sides of (4.245) and (4.246) are always convergent.

Equations (4.244) − (4.246) and (4.229) give general *spectral representations* of a process $X(t)$ with stationary increments and of its structure functions $D(\tau_1,\tau_2)$ and $D(\tau)$. It is also easy to show that, if the functions $F(\omega)$ and $Z(\omega)$ satisfy (4.251) and (4.229), then any random process $X(t)$ which can be represented in the form (4.244) is a process with stationary increments while any functions $D(\tau_1,\tau_2)$ and $D(\tau)$, which can be represented in the form (4.245) and (4.246), are structure functions of such a process. The proofs of the existence of the spectral representations (4.244) − (4.246) and the converse statement, that (4.244) − (4.246) with the conditions (4.251) and (4.229) specify a process with stationary increments and its structure functions, were first given by Kolmogorov in 1940. In another connection, unrelated to the theory of random processes, the same proof was also obtained by von

Neumann and Schoenberg in 1941. At present, several proofs of all these statements can be found in literature; one of them (applied to a wider class of random processes) will be outlined in Sec. 24.[78] Note also that, in fact, only the possibility of the representation (4.245) must be proved, since (4.246) is a special case of (4.245) while (4.244) follows at once from (4.245) by virtue of the theorem on the generalized spectral representations of random processes (given in Note 17 to Chap. 2 and formulated also on pp. 447–448).

In the special case when not only condition (4.251), but also the more restrictive condition $\int_{-\infty}^{\infty} dF(\omega) < \infty$ are satisfied and also $A_1 = 0$, the process $X(t)$ will differ little from the ordinary stationary process. In fact, under this condition the integral on the right-hand side of (4.244) can be written as the difference between two convergent integrals, the first of which is the usual spectral representation of a stationary process and the second is a constant random variable. Of course, the considered process $X(t)$ is nevertheless not necessarily stationary (e.g., if $X_0 = 0$, then $X(0) = 0$ and hence $X(t)$ cannot be stationary if not all its values are equal to zero), but it is easy to see that it is asymptotically stationary and can also be made stationary by adding an appropriately chosen constant random variable to $X(t)$. Therefore the structure functions $D(\tau_1, \tau_2)$ and $D(\tau)$ are in this case bounded and can be written in the form (4.230).

In the other special case when condition (4.249) is satisfied, the process $X(t)$ is differentiable and $Y(t) = X'(t)$ is a stationary process. Its spectral distribution function $F_y(\omega)$ has then a jump (equal to A_1^2) at $\omega = 0$ and satisfies the second equation (4.247) for $0 < \omega_1 < \omega_2$ or $\omega_1 < \omega_2 < 0$.

Consider now a few simple examples of processes with stationary increments.

Example 1. *Brownian motion (i.e., Wiener) and Poisson processes.* Let $X(t)$ be a Brownian motion process, or centered Poisson process shown in Fig. 16, or any other mean square continuous process with stationary and independent (or, at least, uncorrelated) increments which have mean value zero. Then it is readily seen that

$$\langle [X(\tau_1 + \tau_2) - X(0)]^2 \rangle = \langle [X(\tau_1 + \tau_2) - X(\tau_2)]^2 + [X(\tau_2) - X(0)]^2 \rangle$$

for any $\tau_1 \geqslant 0, \tau_2 \geqslant 0$ and hence

$$(4.252) \quad D(\tau_1 + \tau_2) = D(\tau_1) + D(\tau_2) \text{ for } \tau_1 \geqslant 0, \quad \tau_2 \geqslant 0.$$

Since the function $D(\tau)$ is even and continuous, (4.252) implies that

$$(4.253) \quad D(\tau) = C|\tau|,$$

where C is a positive constant. If the process $X(t)$ is real, then according to (4.224)

$$(4.253a) \quad D(\tau_1, \tau_2) = \begin{cases} C\min\{|\tau_1|, |\tau_2|\} & \text{for } \tau_1\tau_2 \geqslant 0, \\ 0 & \text{for } \tau_1\tau_2 < 0. \end{cases}$$

If $X(t)$ is real and $X(0) = 0$, then $B(t,s) = \langle [X(t) - X(0)][X(s) - X(0)] \rangle = D(t,s)$. Therefore, if $X(t)$ is a Brownian motion process or centered Poisson process or any other real process with stationary uncorrelated increments such that $X(0) = 0$ and $\langle X(u) - X(t) \rangle = 0$ for any u and t, then

$$(4.253b) \quad B(t,s) = \begin{cases} C\min\{|t|, |s|\} & \text{for } ts \geqslant 0, \\ 0 & \text{for } ts < 0 \end{cases}$$

where C is a positive constant (cf. pp. 47-48 and 172).

It is easily shown that the function (4.253) can be represented in the form (4.250a) where

$$(4.254) \quad f(\omega) = \frac{C}{2\pi\omega^2}$$

is the spectral density corresponding to the structure function (4.253). It follows from this that in the case under consideration $X_1 = 0$ and $\langle |dZ(\omega)|^2 \rangle = Cd\omega/2\pi\omega^2$. The representation (4.244) can also be written similarly to (4.241):

$$(4.255) \quad X(t) - X(0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{it\omega} - 1}{i\omega} \, dZ^*(\omega),$$

where $dZ^*(\omega) = \sqrt{2\pi} \, i\omega \, dZ(\omega)$. It is clear that $\langle |dZ^*(\omega)|^2 \rangle$

$= Cd\omega$, $\langle |Z^*(\omega + \lambda) - Z^*(\omega)|^2 \rangle = C|\lambda|$, i.e., the structure function of $Z^*(\omega)$ is of the same form as that of $X(t)$. It can be easily proved that the process $Z^*(\omega)$ can be expressed in terms of $X(t)$ as

(4.255a) $Z^*(\omega) - Z^*(0) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{\infty} \dfrac{e^{-it\omega} - 1}{-it}\, dX(t).$

The processes $X(t)$ and $Z^*(\omega)$, which both have uncorrelated increments, are sometimes called the *Fourier transforms of each other*.[79]

The form (4.254) of the spectral density $f(\omega)$ is clearly connected with the fact that the Brownian motion process $X(t) = W(t)$ is an indefinite integral of the white noise process $cE(t)$ and hence $X'(t) = cE(t)$ (see p. 172). We know (see Sec. 10, Example 1a) that the spectral density of $cE(t)$ is constant and equal to $C/2\pi$, where $C = c^2$. Recall that differentiation leads to the multiplication of the spectral density by ω^2 (see Sec. 12, Example 1); hence integration must lead to the multiplication of the spectral density by ω^{-2}. Therefore it is

natural that the process $X(t) = c\displaystyle\int_0^t E(t')dt'$ has the spectral density (4.254).

Formula (4.254) also has a rather simple physical interpretation. This interpretation is connected with the possibility of using condition (4.251a) to construct examples of nonstationary processes with stationary increments by taking the limit of a sequence of stationary processes whose low frequency components have a spectral density which increases without limit. As a typical example we consider the one-dimensional Brownian motion of a harmonic oscillator (i.e., of a small particle whose motion is restrained by an elastic spring). It is shown on p. 176 that, if the inertia of the oscillator is neglected (i.e., in the framework of the Einstein–Smoluchowski approximation), then the oscillator coordinate $X(t)$ satisfies the differential equation $w\,dX(t)/dt + \lambda X(t) = c_1 E(t)$, i.e.,

(4.256) $\dfrac{dX(t)}{dt} + aX(t) = cE(t),\quad a = \lambda/w,\ c = c_1/w,$

where w is the proportionality constant between the force of friction and velocity, λ is the proportionality constant between the force due to the spring and coordinate, and $c_1 = (2wkT)^{1/2}$, i.e., $c = (2kT/w)^{1/2}$ (k is Boltzmann's constant and

T is the absolute temperature).   It follows from (4.256) that the oscillator displacement $X(t)$ is, in the stationary case, a stationary random process with the spectral density

$$(4.257) \quad f(\omega) = \frac{C}{2\pi(\omega^2 + a^2)}, \quad C = c^2.$$

Now let the stiffness of the spring "binding" the oscillator go to zero.   Then $a \to 0$ and the spectral density. at $\omega = 0$ increases without limit so that $\int_{-\omega_0}^{\omega_0} f(\omega)d\omega \to \infty$ for any $\omega_0 > 0$. It is clear, therefore, that as $a \to 0$, $X(t)$ does not converge in the mean to any random variable.   However, from the fact that $\omega^2 f(\omega)$ remains bounded as $a \to 0$, it is easily deduced that the difference

$$X(t) - X(0) = \int_{-\infty}^{\infty}(e^{it\omega} - 1)dZ(\omega)$$

approaches a definite limit as $a \to 0$.   The obtained limit specifies a random process with stationary increments which describes the Brownian motion of a free particle, and hence is no longer stationary.   The spectral density of the limiting Brownian motion process can be obtained by replacing a by zero in (4.257), i.e., it is given by (4.254).   Equation (4.253) in this case can be written as

$$\langle |X(t + \tau) - X(t)|^2 \rangle = C|\tau| = (2kT/w)|\tau|,$$

i.e., it coincides with the famous main formula of Einstein's theory of the Brownian motion (cf. p. 175).

Example 2. *The Ornstein–Uhlenbeck theory of Brownian motion.* The above arguments leading to the derivation of (4.254) from (4.257) can also be applied to the more precise theory of Brownian motion developed by Uhlenbeck and Ornstein in 1930.   In this theory the inertia (i.e., mass m) of a Brownian particle is not neglected.   Therefore the equation describing the one-dimensional Brownian motion of the oscillator now has the form

$$m\frac{d^2X}{dt^2} + w\frac{dX}{dt} + \lambda X = c_1 E(t),$$

i.e.,

$$(4.258) \quad \frac{d^2X}{dt^2} + 2a\,\frac{dX}{dt} + b^2X = cE(t),$$

where $a = w/2m$, $b^2 = \lambda/m$, $c = (2wkT)^{1/2}/m$. It follows from (4.258) that the stationary Brownian motion of an oscillator is described, with the Ornstein–Uhlenbeck approximation, by a stationary random process having the spectral density

$$(4.259) \quad f(\omega) = \frac{c^2}{2\pi[(\omega^2 - b^2)^2 + 4a^2\omega^2]}.$$

Letting again the stiffness of the spring go to zero, i.e., assuming that $b \to 0$, we obtain the following form of the limiting spectral density:

$$(4.260) \quad f(\omega) = \frac{c^2}{2\pi\omega^2(\omega^2 + 4a^2)} = \frac{wkT}{\pi m^2\omega^2(\omega^2 + 4a^2)}.$$

The integral of this spectral density diverges but the density satisfies condition (4.251a). Hence the limiting random process is a process with stationary increments but not a stationary process. This process describes the Brownian motion of a free particle in the Ornstein–Uhlenbeck theory. Combining (4.260) and (4.250a) we obtain the formula for the mean square displacement of a free Brownian particle during the time τ:

$$(4.261) \quad D(\tau) = \langle|X(t + \tau) - X(t)|^2\rangle = \frac{2mkT}{w^2}\Big\{e^{-w\tau/m} + \frac{w\tau}{m} - 1\Big\}.$$

(Equation 3.826.1 from Gradshteyn and Ryzhik's book, 1980:

$$\int_0^\infty \frac{\sin^2\beta x}{x^2(x^2 + \alpha^2)}\,dx = \frac{\pi}{4\alpha^3}(e^{-2\beta\alpha} + 2\beta\alpha - 1),$$

which can be easily deduced with the aid of the theory or residues, is used in the derivation of (4.261).) Formula (4.261) is one of the main results of the Ornstein–Uhlenbeck theory.[80] It generalizes Einstein's formula for the mean square displacement: For values of τ large compared to the time scale $\tau_0 = m/w$ both formulae become identical.

Example 3. *Self-similar processes with stationary increments: Fractional Brownian motions.* The structure function (4.253) is a special example of a power structure function of the form

$$(4.262) \quad D(\tau) = C|\tau|^m,$$

where C is a positive constant. Since $D(0) = 0$ and the function $D(\tau)$ must satisfy (4.238), it is clear that the condition $0 < m \leqslant 2$ must be fulfilled. The case where $m = 2$ corresponds to a limiting process $X(t) = X_0 + X_1 t$, where X_0 and X_1 are constant random variables and $\langle |X_1|^2 \rangle = C$. Moreover, it is readily seen that, for $0 < m < 2$, the function (4.262) can be written in the form (4.250a), where

$$
f(\omega) = \frac{C_1}{\omega^{m+1}}, \quad C_1 = \frac{C}{4\int_0^\infty (1 - \cos x) x^{-m-1} dx}
$$

(4.263)

$$
= \frac{\Gamma(m + 1)\sin(\pi m/2)}{2\pi} C.
$$

Since the function (4.236) is everywhere positive and satisfies (2.251a), the function (4.262) is a possible structure function of a process with stationary increments for any m in the interval $0 < m \leqslant 2$.

If the structure function $D(\tau)$ of a process $X(t)$ is given by (4.262) where $0 < m < 2$, then necessarily $m_1(t) = \langle X(t + \tau) - X(t) \rangle = 0$ by virtue of (4.239). If the process $X(t)$ is real, then according to (4.224)

(4.264) $D(\tau_1, \tau_2) = \dfrac{C}{2} \left\{ \tau_1^m + \tau_2^m - |\tau_1 - \tau_2|^m \right\}$

for any $\tau_1 \geqslant 0$, $\tau_2 \geqslant 0$. If in addition $X(0) = 0$, then $B(t,s) = \langle X(t)X(s) \rangle = D(t,s)$ and hence

(4.264a) $B(t,s) = \dfrac{C}{2} \{ t^m + s^m - |t - s|^m \}$ for $t \geqslant 0$, $s \geqslant 0$.

The processes with stationary increments which have a structure function of the form (4.262) were first studied by Kolmogorov in 1940. They were also later considered by many authors and it appears that such processes play an important role in several applied fields.[81] The Gaussian processes with stationary increments which have a structure function of the form (4.262) are sometimes called *fractional Brownian motions*, since they can be obtained from an ordinary Brownian motion (the Wiener process $W(t)$) by means of a fractional derivation.[82]

The structure functions (4.262) have the following special (and rather interesting) property: Their form is invariant under a group of similarity transformations $t \to ht$, $X \to a(h)X$,

where $a(h)$ is some function of h. In fact, it is easy to see that in this case $D(\tau) = a^2 D(h\tau)$ for $a = h^{-m/2}$ and any $h > 0$. The random processes $X(t)$, whose structure functions have the stated property, are called *self-similar*. They are such that no characteristic scale can be associated with their structure function. It is easy to show that the stationary random processes $X(t)$ cannot be self-similar (i.e., no correlation function $B(\tau)$ can satisfy the relation $B(\tau) = a^2 B(h\tau)$) and that the structure functions of the form (4.262) are the only self-similar structure functions.[83]

If $X(t)$ is a random process with stationary increments, then, for any fixed τ, $\Delta_\tau X(t) = X(t + \tau) - X(t)$ is a stationary random process which depends on the real parameter τ. Therefore the results of Chap. 3 can be applied to the study of the experimental determination of the main statistical characteristics of the process $\Delta_\tau X(t)$ from a single realization $\Delta_\tau x(t) = x(t + \tau) - x(t)$ (i.e., from the measured values of $x(t)$). The determination of the mean value $\langle \Delta_\tau X(t) \rangle = m_1(\tau) = c_1\tau$ allows us to estimate c_1 and the determination of $\langle |\Delta_\tau X(t)|^2 \rangle$ for a number of τ values gives the structure function $D(\tau)$. (In the real case the functions $D(\tau_1, \tau_2)$ and $D(s; \tau_1, \tau_2)$ can be easily obtained from $D(\tau)$.) The spectral density $f_\tau(\omega)$ of $\Delta_\tau X(t)$ is evidently given by the relationship

$$(4.265) \quad f_\tau(\omega) = |e^{i\omega\tau} - 1|^2 f(\omega) = 4\sin^2(\omega\tau/2) f(\omega)$$

where $f(\omega)$ is the spectral density of $X(t)$. Therefore the determination of $f_\tau(\omega)$ allows us to estimate also the spectral density $f(\omega)$ of the process with stationary increments $X(t)$. Another method for the estimation of $f(\omega)$ is based on the calculation of the sample value (realization) of the modified periodogram

$$(4.266) \quad J_T(\omega) = \frac{1}{2\pi T} \left| \int_0^T e^{-i\omega t} dX(t) \right|^2$$

which depends only on the increments of $X(t)$. It is readily seen that the smoothed periodogram $J_T(\omega)$ of the form

$$(4.267) \quad \psi_T^{(A)}(\omega) = \int_{-\infty}^{\infty} A_T(\omega - \omega') J_T(\omega') d\omega',$$

where $A_T(\omega)$ is an appropriately chosen spectral window, is, under wide conditions, an asymptotically unbiased and

consistent estimator of the function $\omega^2 f(\omega)$. Moreover, all the properties of this estimator are quite similar to the properties of the estimator (3.43a) of the spectral density $f(\omega)$ of a stationary random process.[84]

Let us now briefly consider *multidimensional random processes with stationary increments*, i.e., multidimensional processes $X(t) = \{X_1(t), ..., X_m(t)\}$ such that all their components are processes with stationary increments and, in addition, the functions

$$(4.268) \quad \langle [X_j(t + s + \tau_1) - X_j(t + s)][\overline{X_l(t + \tau_2) - X_l(t)}] \rangle = D_{jl}(s;\tau_1,\tau_2),$$

where $j, l = 1, ..., m$, depend continuously on s, τ_1, and τ_2, but do not depend on t.[85] (The function $D_{jl}(s;\tau_1,\tau_2)$ is called a general *cross-structure function* of the processes $X_j(t)$ and $X_l(t)$.) The spectral representations of multidimensional processes with stationary increments and of their cross-structure functions can be obtained by the same method which was applied in Sec. 20 to multidimensional stationary processes. Since all the processes $X_j(t)$, $j = 1, ..., m$, are one-dimensional processes with stationary increments, it is clear that

$$(4.269) \quad \langle \mathbf{X}(t + \tau) - \mathbf{X}(t) \rangle = \mathbf{c}_1 \tau,$$

and

$$(4.270) \quad \mathbf{X}(t) = \int_{-\infty}^{\infty} (e^{i\omega t} - 1)d\mathbf{Z}(\omega) + \mathbf{X}_0 + \mathbf{X}_1 t,$$

where $\mathbf{c}_1$ is a constant vector, $\mathbf{X}_0$ and $\mathbf{X}_1$ are random vectors, and $\mathbf{Z}(\omega) = \{Z_1(\omega), ..., Z_m(\omega)\}$ is a random vector function of ω. The function $\mathbf{Z}(\omega)$ is such that

$$(4.271) \quad \langle dZ_j(\omega_1)\overline{dZ_j(\omega_2)} \rangle = \delta(\omega_1 - \omega_2)dF_{jj}(\omega_1)d\omega_2$$

where $F_{jj}(\omega)$, $j = 1, ..., m$, are real nondecreasing functions on the half-lines $(-\infty, 0)$ and $(0, \infty)$ satisfying the conditions

$$(4.272) \quad \int_{-\infty}^{-\omega_0} dF_{jj}(\omega) + \int_{-\omega_0}^{\omega_0} \omega^2 dF_{jj}(\omega) + \int_{\omega_0}^{\infty} dF_{jj}(\omega) < \infty$$

for any $\omega_0 > 0$. Note that the integral on the right-hand side of (4.270) must be understood in accordance with (4.248).

Moreover, it follows from (4.268) that the symbolic relationship

(4.271a) $\langle dZ_j(\omega_1)d\overline{Z_l(\omega_2)}\rangle = \delta(\omega_1 - \omega_2)dF_{jl}(\omega_1)d\omega_2$,

generalizing (4.271), must also be valid. The functions $F_{jl}(\omega)$, $j, l = 1, ..., m$, are complex functions of bounded variation on the half-lines $(-\infty, 0)$ and $(0, \infty)$ and they are such that the matrix

(4.273) $\| F_{jl}(\omega_2) - F_{jl}(\omega_1) \|$

is Hermitian and positive definite for any ω_1 and ω_2, where either $\omega_2 > \omega_1 > 0$ or $\omega_1 < \omega_2 < 0$. In the important special case, where all the functions $F_{jl}(\omega)$ are absolutely continuous (i.e., equal to the indefinite integral of their derivatives $f_{jl}(\omega)$), (4.271a) takes the form

(4.272b) $\langle dZ_j(\omega_1)d\overline{Z_l(\omega_2)}\rangle = \delta(\omega_1 - \omega_2)f_{jl}(\omega_1)d\omega_1 d\omega_2$

and the matrix $\| f_{jl}(\omega) \|$ is Hermitian and positive definite for any $\omega \neq 0$.

It follows from (4.270) and (4.271a) that the cross-structure function (4.268) can be represented in the form

(4.274) $D_{jl}(s;\tau_1,\tau_2) = \int_{-\infty}^{\infty} e^{is\omega}(e^{i\tau_1\omega} - 1)(e^{-i\tau_2\omega} - 1)dF_{jl}(\omega)$.

In particular, the simpler function

(4.275)
$$D_{jl}(\tau) = D_{jl}(0;\tau,\tau)$$
$$= \langle [X_j(t + \tau) - X_j(t)][\overline{X_l(t + \tau) - X_l(t)}]\rangle,$$

which is also called the cross-structure function of the processes $X_j(t)$ and $X_l(t)$, is given by

(4.276) $D_{jl}(\tau) = 2\int_{-\infty}^{\infty}(1 - \cos\tau\omega)dF_{jl}(\omega)$.

It is clear that the matrix $\| D_{jl}(\tau) \|$ is Hermitian and positive definite for any τ. We note, however, that in the multidimensional case, if even the process $\mathbf{X}(t)$ is real, the functions $D_{jl}(s;\tau_1,\tau_2)$ cannot, in general, be expressed in terms of the functions $D_{jl}(\tau)$:   The identity (4.217) cannot be applied to

the cross-structure functions. The identity

$$
\begin{aligned}
(a_j - b_j)(c_1 - d_1) + (a_1 - b_1)(c_j - d_j) &= (a_j - d_j)(a_1 - d_1) \\
&\quad + (b_j - c_j)(b_1 - d_1) - (a_j - c_j)(a_1 - c_1) \\
&\quad - (b_j - d_j)(b_1 - d_1)
\end{aligned}
$$

(4.277)

is more useful in this case. It is readily seen that (4.277) implies the following result: In the real case the sum $D_{jl}(s;\tau_1,\tau_2) + D_{1j}(s;\tau_1,\tau_2)$ can always be expressed as

(4.278)
$$
\begin{aligned}
&D_{jl}(s;\tau_1,\tau_2) + D_{1j}(s;\tau_1,\tau_2) \\
&= D_{jl}(s + \tau_2) + D_{jl}(s - \tau_1) - D_{jl}(s - \tau_1 + \tau_2) - D_{jl}(s).
\end{aligned}
$$

However, only in the special case when $X(t)$ is real and

(4.279) $\quad D_{jl}(s;\tau_1,\tau_2) = D_{1j}(s;\tau_1,\tau_2)$

does it follow from (4.278) that the general cross-structure function $D_{jl}(s;\tau_1,\tau_2)$ can be expressed in terms of the simpler function $D_{jl}(\tau)$. According to (4.274) the conditions $dF_{jl}(\omega) = dF_{1j}(\omega)$ must be valid for the validity of (4.279), i.e., the matrix (4.273) must be not only Hermitian but real.

24. Generalized Stationary Processes. Processes with Stationary Increments of Order n

24.1. Generalized Random Processes

In Chapters 1 - 3 stationary random processes $X(t)$ with a correlation function $\langle X(t + \tau)X(t)\rangle = B(\tau)$ were studied. It was shown there that such processes and their correlation functions can always be represented as

$$
X(t) = \int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega), \qquad B(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau} dF(\omega)
$$

where $F(\omega)$ is a bounded nondecreasing spectral distribution function of the process $X(t)$, and $Z(\omega)$ is a random function with uncorrelated increments (i.e., $Z(d\omega) = dZ(\omega)$ is a random measure on the frequency axis) which satisfied condition

(2.78). (In this section, the time t will always be assumed continuous.) Later on, in Sec. 23, we introduced the generalization of the concept of stationarity, which led us to a wider class of spectral distribution functions which are nondecreasing on the half-lines $(-\infty, 0)$ and $(0, \infty)$, but are not necessarily bounded and can diverge as $\omega \to 0$. This subsection will discuss another generalization of the concept of a stationary process which justifies the use of another type of unbounded spectral distribution functions diverging not at zero but at infinity.

Recall that one special example of the function $F(\omega)$ tending to infinity as $\omega \to \infty$ has already been mentioned in this book. Namely, as noted in Example 1a, Sec. 10, in applications it is often convenient to use the concept of a stationary "white noise" with constant spectral density

(4.280) $f(\omega) = f_0 = \text{const.}$

Since the spectral density (4.280) corresponds to an unbounded spectral distribution function of the form $F(\omega) = f_0\omega + \text{const.}$, i.e., to an infinite mean power of the process $X(t)$, it is clear that white noise can be meaningful only as a mathematical idealization. As was explained in Sec. 11 (see pp. 160–161), the usefulness of this idealization stems from the fact that the frequency response of any actual physical device is always effectively zero outside some finite frequency interval Λ. Hence, when a random process $X(t)$, the spectral density $f(\omega)$ of which is effectively constant over Λ, is applied to such a device, there is no need to describe in detail how $f(\omega)$ falls off at high frequencies (as it must), but one can regard $X(t)$ as being white noise, with constant spectral density $f(\omega) = f_0$. This last assumption is usually quite convenient because it greatly simplifies any analysis involving the process $X(t)$.

Moreover, it is intuitively clear for the same reason that the concept of the "value $X(t)$ of the process X at the time t" is also merely a mathematical idealization which simplifies the actual physical situation.   The point is that in practice one must always use some physical device to measure $X(t)$. Since any such device always has "inertia" (or "memory"), corresponding to its nonzero "time constant", the input process $X(t)$ will inevitably be subjected to some time averaging. Consequently, "the value $X(t)$ of a process X at time t" cannot be determined absolutely precisely, i.e, it is a mathematical idealization from the physical

point of view. This suggests that a change in the very definition of a random process is desirable. It will be shown below that with a suitable new definition, the case of white noise is no "worse" than thè case of ordinary random processes.

A "suitable new definition" is such that it does not include at all the values $X(t)$ of the process X at time t, but involves only the actually measurable "values of the process X obtained with the aid of some measuring device." It is reasonable to restrict ourselves to linear measuring devices, which are much simpler than all the others. Then the given device is completely characterized by its *weighting function* or *impulse response $h(t)$* (cf. p. 143), and the result of measuring $X(t)$ is just the quantity

$$(4.281) \quad X(h) = \int_{-\infty}^{\infty} X(t)h(t)dt.$$

(For simplicity, measurements made with one and the same device at different times, which correspond to the weighting functions $h_1(t)$ and $h_2(t) = h_1(t + t_0)$, will be regarded below as made with different devices. This accounts for the difference between formula (4.281) and formulae (2.144) and (2.147) used in Sec. 11.) In fact, in practice a physicist can only deal with quantities like (4.281), rather than with the process $X(t)$ itself. These considerations suggest the following new definition of a random process: By a *generalized (random) process* we mean a linear functional $X(h)$, the values of which are random variables and the arguments $h = h(t)$ belong to some specified function space K. Thus, for every function h in the function space K, the quantity $X(h)$ is a random variable, and moreover

$$(4.282) \quad X(\alpha_1 h_1 + \alpha_2 h_2) = \alpha_1 X(h_1) + \alpha_2 X(h_2)$$

for all (real) numbers α_1, α_2, and all elements φ_1, φ_2, of K. (The condition (4.282) presupposes, of course, that the function space K is linear, i.e. contains, together with any pair (h_1, h_2), all linear combinations $\alpha_1 h_1 + \alpha_2 h_2$. This supposition will always be assumed to be true below.)

The condition (4.282) is completely natural if we think of K as the space of weighting function corresponding to all possible linear measuring devices, and of the $X(h)$ as the results obtained when various devices are used to measure a time-variable quantity $X(t)$. In other words, a generalized random process $X(h)$ is a random function on a function

space $\mathcal{K}$ satisfying the linearity condition (4.282). The word *generalized* appearing in the above definition is explained by the fact that any *ordinary* random process $X(t)$ evidently gives rise to a generalized random process $X(h)$ of a special kind, which is related to $X(t)$ by the formula (4.281) and is said "to have the point values (at time t) $X(t)$." However, there also exist generalized processes $X(h)$ which are not defined by the integral formula (4.281) for any $X(t)$ at all. These kinds of generalized processes are said "to have no point values". Nevertheless, the processes having no point values do not differ from the ordinary random processes in the following sense: By repeatedly measuring such a generalized process with the same linear device, one obtains sample values (realizations) of a random variable with a perfectly definite probability distribution.

The concept of a generalized random process was introduced by K. Itô in Japan and, independently, by I.M. Gel'fand in the USSR.[86] To specify a generalized random process (i.e., a random function $X(h)$ on a function $\mathcal{K}$) we must give, for any finite set of functions $h_1(t)$, ..., $h_n(t)$ in $\mathcal{K}$, the distribution function

(4.283)
$$F_{h_1,h_2,...,h_h}(x_1,x_2,...,x_n)$$
$$= P\{X(h_1) < x_1, X(h_2) < x_2,...,X(h_n) < x_n\}$$

of the n-dimensional random vector

(4.284) $$\mathbf{X}_{h_1,h_2,...,h_n} = \{X(h_1), X(h_2),..., X(h_n)\}$$

(cf. the treatment given on pp. 39–41 for ordinary random processes). As in the case of ordinary random processes, the system of distribution functions (4.283) must satisfy the usual symmetry and compatibility conditions (cf. (1.5a) and (1.6a)). However, in the case of a generalized random process $X(h)$ the random vectors (4.284) are not completely arbitrary but must also satisfy the important linearity condition (4.282). It is possible to show that this condition allows us to reconstruct uniquely the n-dimensional distribution function (4.283) from a knowledge of all the one-dimensional distribution functions of the random variables

$$X(\alpha_1 h_1 + \alpha_2 h_2 + \dots + \alpha_n h_n) = \alpha_1 X(h_1)$$

$$+ \alpha_2 X(h_2) + \dots + \alpha_n X(h_n)$$

for arbitrary $\alpha_1, \dots, \alpha_n$.[87] Thus, the generalized random process $X(h)$ can actually be specified by giving only the one-dimensional distributions of random variables $X(h)$ for all $h(t)$ in K.

From the standpoint of correlation theory, a generalized random process $X(h)$ is characterized by its *mean value functional*

(4.285) $\langle X(h) \rangle = m(h)$

and by its *correlation functional*

(4.286) $\langle X(h_1)X(h_2) \rangle = B(h_1,h_2).$

The functional $m(h)$ is evidently linear in h, whereas $B(h_1,h_2)$ is a bilinear functional (i.e., linear both in h_1 and h_2). In the important particular case when the generalized process $X(h)$ is Gaussian (i.e., when $X(h)$ has a Gaussian probability distribution for any h), the functionals $m(h)$ and $B(h_1,h_2)$ uniquely specify the generalized random process $X(h)$. If (as often proves convenient) one considers more general complex generalized random processes (i.e., $X(h)$ are complex random variables), it is natural to assume that the weighting functions $h(t)$ and the constants α_1,α_2 in (4.282) can also be complex and the correlation functionals $B(h_1,h_2)$ are defined as[88]

(4.286a) $B(h_1,h_2) = \langle X(h_1)\overline{X(h_2)} \rangle.$

We must now discuss how the linear function space   is chosen. In actual practice this space can be chosen in different ways, and corresponding to this freedom of choice there are various kinds of generalized random processes $X(h)$. Naturally, the larger the space K, the narrower the corresponding class of generalized random processes, in the following sense: If K' is contained in K, then all random processes $X(h)$ defined on K are also defined on K', but in general there are random processes defined on K' which cannot be defined on the larger space K. Perhaps the most general space K for which generalized processes can be considered is the space

K_0 of piecewise continuous functions $h(t)$ vanishing outside a finite interval of the axis $-\infty < t < \infty$. Then, given any finite interval Δt, we can define a random interval function $X(\Delta t) = X(h_{\Delta t})$, where $h_{\Delta t}$ is the function equal to 1 on Δt and 0 elsewhere. This leads us to the class of random interval functions $X(\Delta t)$ satisfying condition (c) on p. 99, but not, in general, conditions (a) and (b).

In practice, however, it is often more convenient to consider the smaller function spaces K_n consisting of functions $h(t)$ which are differentiable at least n times and vanish outside the finite interval (it is also stipulated that $h(t)$ and its first n derivatives $h'(t), h''(t), ..., h^{(n)}(t)$ all vanish at the end points of this interval). Gel'fand and Itô, who were the first to introduce the concept of a generalized random process, both followed the works of Schwartz on the theory of ordinary (nonrandom) generalized functions and chose K to be the space K_∞ of infinitely differentiable functions which vanish, together with all their derivatives, outside a finite interval (and also at its end points).[89] They also assumed the generalized process $X(h)$ to be continuous in the following sense: Let $h_1, h_2, ...$ be a sequence of functions in K_∞, which tends, in the space K_∞, to a function h. (This means that all functions h_k, $k = 1, 2, ...$, vanish outside the same finite interval and $\lim_{k \to \infty} h_k = h$, $\lim_{k \to \infty} h_k^{(n)} = h^{(n)}$ for all $n = 1, 2,$

..., where $h^{(n)}$ is the nth derivative of h.) Then $X(h_k)$ converges (in the mean) to $X(h)$ as $k \to \infty$.   From now on we shall always assume, unless otherwise specified, that $K = K_\infty$, and that the process $X(h)$ is continuous in the above sense. (Clearly, if the process $X(h)$ is representable as (4.281), i.e., has point values, $X(t)$, then this implies that $X(t)$ is a mean continuous process.) It follows from the continuity of $X(h)$ that the functionals $m(h)$ and $B(h_1, h_2)$ are continuous, i.e., $\lim_{k \to \infty} m(h_k) = m(h)$, $\lim_{k \to \infty} B(h_k, h_0) =$

$B(h, h_0)$ for any h_0 as $h_k \to h$ in K_∞.

Of course, one may also choose K to be another space quite different from K_0, K_n, and K_∞. For instance, one may choose K to be a space of entire analytic functions of some class which fall off sufficiently rapidly at infinity.[90] One corollary of this remark will be mentioned below.

It should be noted that in some respects, generalized random processes $X(h)$ are even simpler than ordinary processes $X(t)$. For instance, if all the functions h in K are differentiable (e.g., for $K = K_\infty$), we can always define the *derivative* $X'(h)$ of a

generalized process $X(h)$ by means of the formula

(4.287) $X'(h) = - X(h')$.

It is clear that if our generalized process $X(h)$ has point values $X(t)$ (i.e., is of a form (4.281)) and if $X(t)$ is differentiable in the sense of Sec. 4, then the appropriate definition of $X'(h)$ is

(4.288) $X'(h) = \int_{-\infty}^{\infty} X'(t)h(t)dt = - \int_{-\infty}^{\infty} X(t)h'(t)dt,$

which justifies the definition (4.287). However, unlike ordinary processes, generalized processes are *always* differentiable. Therefore, an ordinary process $X(t)$ may also be regarded as *always differentiable* except that its derivative may be a generalized process instead of an ordinary process. This shows the great convenience of the space K_∞ since when $K = K_\infty$, all generalized processes are *infinitely* differentiable.

Let us now consider *stationary generalized processes*. It is natural to say that the generalized process $X(h)$ is *stationary in the strict sense* (or *strictly stationary*) if for every real τ, the random variable $X(S_\tau h)$ has the same probability distribution as $X(h)$. Here, $h = h(t)$ is an arbitrary element of K, and S_τ is the *translation* (or *shift*) *operator*, defined by

$$S_\tau h(t) = h(t + \tau)$$

(cf. p. 142). Similarly, we say that the generalized process $X(h)$ is *stationary in the wide sense* (or *widely stationary*) if it has a mean value functional $m(h)$ and a correlation functional $B(h_1,h_2)$ such that for every real τ

(4.289) $m(S_\tau h) = m(h),$

(4.290) $B(S_\tau h_1, S_\tau h_2) = B(h_1,h_2),$

where h, h_1, h_2 are arbitrary elements of K (cf. pp. 51 and 56). As in the theory of ordinary processes $X(t)$ studied in Chap. 2, we will consider below only wide-sense stationary generalized processes, which will simply be called stationary.

Further on we will use some simple results from the theory of ordinary (nonrandom) generalized functions without

adducing their strict proofs. It is not hard to show that all continuous linear functionals $m(h)$ in the space $K = K_\infty$, which satisfy the condition (4.289), are necessarily of the form

$$(4.291) \quad m(h) = m \int_{-\infty}^{\infty} h(t)dt$$

where m is a numerical constant.[91] Therefore the mean value of the generalized stationary process $X(h)$ is specified by a single constant m. This constant will be called below the *mean value* of $X(h)$. (If $X(h)$ has point values $X(t)$, then obviously $m = \langle X(t) \rangle$.)

Now we consider the correlation functionals $B(h_1, h_2)$ of generalized stationary processes $X(h)$. Note, first of all, that the condition (4.290) is satisfied, if

$$(4.292) \quad B(h_1, h_2) = B(h_1 * \hat{h}_2) = B(\int_{-\infty}^{\infty} h_1(s)\overline{h_2(s - t)}ds)$$

(where $*$ denotes a convolution, and $\hat{h}(t) = \overline{h(-t)}$), i.e., if $B(h_1, h_2)$ is a continuous linear functional of the convolution

of $h_1(t)$ and $\hat{h}_2(t) = \overline{h_2(-t)}$. In fact, if

$$h_1^{(T)}(t) = S_T h_1(t) = h_1(t + \tau), \quad h_2^{(T)}(t) = S_T h_2(t) = h_2(t + \tau),$$

then clearly

$$h_1^{(T)} * \hat{h}_2^{(T)} = \int_{-\infty}^{\infty} h_1(s + \tau)\overline{h_2(s + \tau - t)}ds =$$

$$= \int_{-\infty}^{\infty} h_1(s)\overline{h_2(s - t)}ds = h_1 * \hat{h}_2$$

and therefore (4.290) is valid for the functional (4.292). On the other hand, it is proved in the theory of generalized nonrandom functions that any continuous bilinear functional in the space $K = K_\infty$ satisfying condition (4.290) for any τ can be represented as (4.292), i.e., is a linear functional of a single function $h_1 * \hat{h}_2$.[92] Recall that the linear functionals $B(h)$ in the space K are called generalized functions; thus the *correlation functional* $B(h_1, h_2)$ *of the generalized stationary process* $X(h)$ *is uniquely specified by a* (nonrandom) *generalized function* $B(h)$. The generalized function $B(h)$ will be called a *correlation function* of a generalized stationary random process $X(h)$. It is easy to see that if $X(h)$ has the

point values $X(t)$, then

(4.293) $B(h) = \int_{-\infty}^{\infty} B(\tau)h(\tau)d\tau,$

where $B(\tau) = \langle X(t + \tau)\overline{X(t)} \rangle$ is the correlation function of a process $X(t)$.[93] Hence in this case the generalized function $B(h)$ also has point values, i.e., $B(h)$ can be regarded as an ordinary (nongeneralized) function $B(\tau)$.

Note that the correlation functional $B(h_1,h_2)$ always satisfies the following important condition:

(4.294) $B(h,h) \geqslant 0$

for any h in $\mathcal{K}$ (since $B(h,h) = \langle |X(h)|^2 \rangle$).   Bilinear functionals $B(h_1,h_2)$ satisfying condition (4.294) are said to be *positive definite*. Thus, *the correlation functional of a generalized process is always positive definite*.   It is not hard to show that, conversely, *every continuous positive definite bilinear functional* $B(h_1,h_2)$ *in* $\mathcal{K}$ *is a correlation functional of some generalized random process* $X(h)$ (cf. Sec. 2, p. 47). If a generalized random process $X(h)$ is stationary, then according to (4.292) the positive definiteness condition (4.294) takes the form

(4.295) $B(h * \hat{h}) \geqslant 0$ for any h.

A generalized function $B(h)$ satisfying (4.295) is said to be a *positive definite generalized function*. It is easy to see that for generalized functions $B(h)$ of the form (4.293), condition (4.295) becomes the ordinary condition of positive definiteness of $B(\tau)$.

Recall that according to Bochner's theorem the class of (nongeneralized) positive definite functions $B(\tau)$ coincides with the class of the Fourier–Stieltjes integrals (2.52), where $F(\omega)$ is a bounded nondecreasing function of ω (see p. 106 and Note 3 to Introduction.) Therefore, if $B(h)$ is a generalized positive definite function having point values (i.e., representable as (4.293)), then

$$B(h) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i\tau\omega}h(\tau)dF(\omega)d\tau = \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} e^{i\tau\omega}h(\tau)d\tau\right\}dF(\omega)$$

and hence

$$(4.296) \quad B(h) = \int_{-\infty}^{\infty} \overset{\smile}{h}(\omega)dF(\omega), \quad \overset{\smile}{h}(\omega) = \int_{-\infty}^{\infty} e^{i\tau\omega}h(\tau)d\tau$$

where $F(\omega)$ is a bounded nondecreasing function. The Bochner theorem of 1932, which implies (4.296), was generalized by Schwartz in 1950. He showed that the *class of generalized positive definite functions* (i.e., of all continuous linear functionals in $K = K_\infty$ satisfying (4.295)) *coincides with the class of functions representable as* (4.296), *where* $F(\omega)$ *is a nondecreasing function which may be unbounded but must satisfy the condition*

$$(4.297) \quad \int_{-\infty}^{\infty} \frac{dF(\omega)}{(1 + \omega^2)^p} < \infty$$

for some integer p.[94] This generalization of the classical Bochner theorem is often called the *Bochner–Schwartz theorem*. If $B(h_1 * \overset{\smile}{h}_2)$ is the correlation functional of a generalized stationary process $X(h)$, then the function $F(\omega)$ is called the *spectral distribution function* of this process. If the spectral distribution function is absolutely continuous (i.e., coincides with the indefinite integral of its derivative), the derivative $F'(\omega) = f(\omega)$ is called the *spectral density* of the process $X(h)$. According to (4.297), a spectral distribution function can diverge at infinity, but not faster than some power of $|\omega|$. Similarly, the spectral density $f(\omega)$ must satisfy the condition

$$(4.297a) \quad \int_{-\infty}^{\infty} \frac{f(\omega)d\omega}{(1 + \omega^2)^p} < \infty$$

for some integer p, i.e., it also can diverge at infinity not faster than some power of $|\omega|$. Formula (4.296) describes the *spectral representation of the correlation function* B(h) *of a generalized stationary process* X(h). Moreover, (4.292) and (4.296) imply the following *spectral representation for the correlation functional* $B(h_1, h_2)$:

$$(4.298) \quad B(h_1, h_2) = \int_{-\infty}^{\infty} \overset{\smile}{h}_1(\omega)\overline{\overset{\smile}{h}_2(\omega)}dF(\omega),$$

where $\overset{\smile}{h}(\omega)$ is the Fourier transform of $h(t)$. Note that by choosing the space K to be not K_∞, but a specific space of rapidly decreasing entire analytic functions, one may obtain stationary generalized processes with exponentially (or even faster than exponentially) increasing spectral distribution functions and spectral densities.[95]

Let us now apply a slightly modified form of reasoning which proves the theorem on the generalized spectral representation of random functions (see Note 17 to Chap. 2 in Vol. II). Then one can readily derive from (4.298) the following *spectral representation of the generalized stationary process* X(h):

$$(4.299) \quad X(h) = \int_{-\infty}^{\infty} \hat{h}(\omega) Z(d\omega).$$

Here $Z(\Delta\omega)$ is a random interval function with uncorrelated increments, i.e., $Z(\Delta\omega)$ satisfies conditions (a_1), (b), and (c) on p. 99, as well as the condition

$$(4.300) \quad \langle |Z(\Delta\omega)|^2 \rangle = F(\Delta\omega)$$

where $\Delta\omega$ is the interval $[\omega, \omega + \Delta\omega]$, $F(\Delta\omega) = F(\omega + \Delta\omega) - F(\omega)$, while the integral on the right-hand side of (4.299) is defined in the same way as the integral (2.58).[96]
The definition (4.287) of the derivative of $X(h)$ implies that $Z(d\omega)$ and $F(\omega)$ must be replaced by $i\omega Z(d\omega)$, and

$$(4.301) \quad F_1(\omega) = \int_0^\omega \omega'^2 dF(\omega')$$

in the spectral representations of $X'(h)$ and its correlation functional (cf. (2.173) and (2.174)). It is clear that $F_1(\omega)$ satisfies (4.297) for any $F(\omega)$ satisfying this condition. This explains why the derivative $X'(h)$ always exists. The simplest example of a generalized stationary process which has no point values (i.e., cannot be represented as (2.281)) corresponds to the correlation function $B(h)$ proportional to the Dirac δ-function: $B(h) = c^2\delta(h) = c^2 h(0)$ (where c^2 is an arbitrary positive constant). By the Bochner–Schwartz theorem the function $c^2\delta(h)$ is a positive definite generalized function since it can evidently be represented as

$$(4.302) \quad c^2\delta(h) = c^2 h(0) = \frac{c^2}{2\pi} \int_{-\infty}^{\infty} \hat{h}(\omega) d\omega.$$

Thus, the functional

$$B(h_1, h_2) = c^2 \delta(h_1 * \hat{h}_2) = c^2 \int_{-\infty}^{\infty} h_1(t) \overline{h_2(t)} dt$$

is a correlation functional of the generalized stationary random process $X(h)$, which has constant spectral density $f(\omega) = c^2/2\pi = \text{const}$. This generalized process is, of course, a

white noise, which was studied from a different point of view in Sec. 10, Example 1a.

By using (4.281) and (4.287) it is easy to show that the white noise $X(h)$ coincides with the derivative of the Wiener process $W(t)$, which has mean value zero and correlation function of a form (2.195). (Recall that, according to the theory developed in this section, any random process is differentiable, but its derivative may be a generalized, rather than an ordinary random process.) This conclusion agrees well with the representation (2.194) of the process $W(t)$ and with formula (4.254), where $C = c^2$, for the spectral density of $W(t)$.

The derivatives $X'(h)$, $X''(h)$, ..., $X^{(m)}(h)$ of the white noise $X(h)$ are also generalized stationary random processes which do not have point values. According to (4.301) the spectral density of the process $X^{(m)}(h)$ is equal to $c^2 \omega^{2m}/2\pi$ while its correlation function has the form $B(h) = (-1)^m c^2 \delta^{(2m)}(h)$, where $\delta^{(2m)}(h) = h^{(2m)}(0)$. It follows from this, in particular, that the generalized function $(-1)^m \delta^{(2m)}(h)$ is positive definite. Note also that any generalized positive definite function is a correlation function of some generalized stationary process.[97]

24.2. A Novel Approach to Processes with Stationary Increments

It was mentioned on pp. 394–395 that the study of random sequences with stationary increments is of little theoretical interest, since it can always be reduced to the study of ordinary stationary sequences by passing from the original sequence $X(t)$ to the stationary sequence of its differences $Y(t) = X(t+1) - X(t)$. Using the concept of a generalized stationary process we can treat the theory of processes with stationary increments in a similar manner, passing from the original process $X(t)$ to its stationary (but possibly generalized) derivative $Y(h) = X'(h)$. Since the derivative $X'(h)$ always exists, we may now *define* processes with stationary increments as processes X whose derivative $X' = Y$ is a stationary random process.

Now let $X = X(h)$ be an arbitrary generalized random process (maybe having no point values) with stationary increments. Then, according to (4.287) and (4.299),

$$(4.303) \quad X(h') = -X'(h) = -\int_{-\infty}^{\infty} \hat{h}(\omega) dZ_y(\omega).$$

If $h(t)$ belongs to the space $K = K_\infty$, then, naturally, $h'(t)$ belongs to the same space, but not every function in K is a derivative of some other function in K. It is clear that the function $h'(t)$ satisfies the condition

$$\int_{-\infty}^{\infty} h'(t)dt = 0$$

which may not be fulfilled for an arbitrary function in K. If, however, the function $h(t)$ in K satisfies the condition

(4.304) $\alpha_0(h) = \int_{-\infty}^{\infty} h(t)dt = 0,$

then $h(t)$ coincides with the derivative of the function $h_1(t) = \int_{-\infty}^{t} h(t')dt'$ which also belongs to K.   Thus, *the class of derivatives of the functions in K coincides with the linear subspace of the function space K consisting of functions which satisfy the condition* (4.304).   Note now that

$$\widetilde{h}(\omega) = \int_{-\infty}^{\infty} e^{i\omega\tau}h(t)dt = - \int_{-\infty}^{\infty} \frac{e^{i\omega t} - 1}{i\omega} h'(t)dt.$$

Therefore, denoting h' by h and using (4.304), we can write (4.303) as

(4.305) $X(h) = \int_{-\infty}^{\infty} \widetilde{h}(\omega)dZ(\omega) + X_1\alpha_1(h) .$

Here $h(t)$ is a function from K satisfying (4.304), $X_1 = \lim_{\epsilon \to 0}[Z_y(\epsilon) - Z_y(- \epsilon)] = Z_y(+ 0) - Z_y(- 0)$ is the jump of the function $Z_y(\omega)$ at $\omega = 0$, $\alpha_1(h) = \int_{-\infty}^{\infty} th(t)dt$, the function $Z(\omega)$ is related to $Z_y(\omega)$ by (4.247), and the integral from $- \infty$ to ∞ is interpreted as the limit (4.248).   If follows clearly from (4.305) that

(4.306) $m(h) = \langle X(h)\rangle = m_1\alpha_1(h),$

(4.307)
$$B(h_1,h_2) = \langle X(h_1)\overline{X(h_2)}\rangle = \int_{-\infty}^{\infty} \widetilde{h}_1(\omega)\overline{\widetilde{h}_2(\omega)}dF(\omega)$$
$$+ A_1^2\alpha_1(h_1)\overline{\alpha_1(h_2)}$$

for any h, h_1, and h_2 from K satisfying (4.304).   Here m_1

$= \langle X_1 \rangle$, $A_1^2 = \langle |X_1|^2 \rangle$, the function $F(\omega)$ satisfies the relation

(4.308) $\langle dZ(\omega)\overline{dZ(\omega')} \rangle = \delta(\omega - z')dF(\omega)d\omega'$

and is determined by (4.247) where $F_y(\omega)$ is the spectral distribution function of the stationary process $X'(h)$, while the integral on the right-hand side of (4.307) is again interpreted in accordance with (4.248). The function $F(\omega)$ is evidently a nondecreasing function on the half-lines $(-\infty, 0)$ and $(0, \infty)$ such that

(4.309) $\int_{-\infty}^{-A} \frac{dF(\omega)}{\omega^{2p}} + \int_{-A}^{A} \omega^2 dF(\omega) + \int_{A}^{\infty} \frac{dF(\omega)}{\omega^{2p}} < \infty$

for some integer p and any $A > 0$. The convergence of the integrals on the right-hand sides of (4.305) and (4.307) at the point $\omega = 0$ is ensured by condition (4.304), which guarantees that $\hat{h}(0) = \hat{h}_1(0) = \hat{h}_2(0) = 0$.

Thus, in formulae (4.305) – (4.307), $F(\omega)$ is an arbitrary function nondecreasing on the half-lines $(-\infty, 0)$ and $(0, \infty)$ and satisfying (4.309), $Z(\omega)$ is an arbitrary random interval function which satisfies (4.308), X_1 is an arbitrary random variable, $m_1 = \langle X_1 \rangle$, and $A_1^2 = \langle |X_1|^2 \rangle$. These formulae determine the values of all possible generalized processes with stationary increments and of their first and second moments on the subspace $\mathfrak{K}^{(1)}$ of the function space $\mathfrak{K}$ consisting of functions that satisfy (4.304). Recall that when studying ordinary (nongeneralized) processes with stationary increments $X(t)$, we have already noted that for such processes the mean values $\langle X(t) \rangle$ and $\langle X(t)X(u) \rangle$ and even the one-dimensional and multidimensional probability distributions for $X(t)$ may not exist at all, since the very definition of processes with stationary increments involves only the probability distributions for the differences $X(t) - X(u)$ (see p. 392 and Note 75 to this chapter in Vol. II). So, in the case of a random process $X(t)$ with stationary increments, only the differences $X(t) - X(u)$ and their statistical characteristics must be considered. Similarly, the definition of a generalized process with stationary increments $X(h)$ involves only the random variable $X(h)$ where h satisfies the condition (4.304), i.e., belongs to $\mathfrak{K}^{(1)}$. Therefore in the theory of such processes one can restrict oneself to the study of random variables $X(h)$ for $h = h(t)$ belonging to the subspace $\mathfrak{K}^{(1)}$. If we wish, nevertheless, also to determine the random variables $X(h)$ for all h in $\mathfrak{K}$ (i.e., to

extend the random function $X(h)$, given in $K^{(1)}$, to the wider space K), then we must specify the value of $X'(h_0)$ for one arbitrary function h_0 in K which does not belong to $K^{(1)}$ (i.e., does not satisfy the condition (4.304)). Let, e.g., $X(h_0) = X_0$,

where $\alpha_0(h_0) = \int_{-\infty}^{\infty} h_0(t)dt \neq 0$. Let us assume, for

definiteness, that $\alpha_0(h_0) = 1$. (This can always be achieved by replacing the original function h_0 by $h_0/\alpha_0(h_0)$.) Then $X(h)$ for any h in K can be determined as follows:

$$X(h) = X(h - \alpha_0(h)h_0 + \alpha_0(h)h_0) = X(h - \alpha_0(h)h_0)$$

(4.310)

$$+ X(\alpha_0(h)h_0) = X(h - \alpha_0(h)h_0) + \alpha_0(h)X_0$$

where $X(h - \alpha_0(h)h_0)$ is given by (4.305), since clearly $\alpha_0(h - \alpha_0(h)h_0) = 0$. Thus, the specification of a random function $X(h)$ for all h in K requires knowledge of the random measure $Z(d\omega)$ (which satisfies (4.308) and (4.309)), the random variable X_1, and the random variable $X_0 = X(h_0)$ corresponding to some function h_0 in K with $\alpha_0(h_0) \neq 0$.[98]

It is easy to show that the process $X(h)$ has point values, if and only if the inequality (4.309) is satisfied for $p = 0$, i.e., in this case $X(h)$ can be represented in the form (4.281), where $X(t)$ is an ordinary process with stationary increments. Then (4.305) − (4.307) can be used to obtain a very simple derivation of (4.244), (4.245), and (4.235) (or (4.236)), i.e., to prove all the main results of the theory of ordinary processes with stationary increments.

24.3. Random Processes with Stationary Increments of Order n

Using generalized stationary processes for a novel definition of the concept of a process with stationary increments, we can then easily take still another step in the same direction and define a *(generalized) process with stationary increments of order n* as a (generalized) process whose nth derivative is stationary. Thus, a (generalized) process $X(h)$ is said to be a process with stationary increments of order n, if

$$(4.311) \quad X^{(n)}(h) = (-1)^n X(h^{(n)}) = \int_{-\infty}^{\infty} \hat{h}(\omega) dZ_y(\omega)$$

where $dZ_y(\omega) = Z_y(d\omega)$ satisfies the conditions (a$_1$), (b), and (c) on p. 99.

Note now that the functions in $K = K_\infty$ which are nth derivatives of some other functions in the same space, form a linear subspace $K^{(n)}$ of functions $h(t)$ in K which satisfy the conditions

$$(4.312) \quad \alpha_0(h) = \int_{-\infty}^{\infty} h(t)dt = 0, \quad \alpha_1(h) = \int_{-\infty}^{\infty} th(t)dt = 0,$$

$$...,\alpha_{n-1}(h) = \int_{-\infty}^{\infty} t^{n-1}h(t)dt = 0.$$

Therefore it is natural to consider from the outset a process with stationary nth increments as a random linear functional $X(h)$ given on the linear subspace $K^{(n)}$ of the space K. Moreover

$$\widehat{h}(\omega) = \int_{-\infty}^{\infty} e^{i\omega t}h(t)dt$$

$$(4.313)$$

$$= (-1)^n \int_{-\infty}^{\infty} \frac{\{e^{i\omega t} - 1 - i\omega t - ... - (i\omega t)^{n-1}/(n-1)!\}}{(i\omega)^n} h^{(n)}(t)dt$$

for any h in K. Now we may denote the function $h^{(n)}(t)$ by $h(t)$ and, taking into account (4.312), write (4.311) as

$$(4.314) \quad \dot{X}(h) = \int_{-\infty}^{\infty} \widehat{h}(\omega)dZ(\omega) + X_n\alpha_n(h).$$

In (4.314) h belongs to $K^{(n)}$,

$$(4.315) \quad Z(\omega) = \begin{cases} \int_{-\infty}^{\omega} \dfrac{dZ_y(\lambda)}{(i\lambda)^n} & \text{for } \omega < 0, \\[3mm] -\int_{\omega}^{\infty} \dfrac{dZ_y(\lambda)}{(i\lambda)^n} & \text{for } \omega > 0, \end{cases}$$

$\alpha_n(h) = \int_{-\infty}^{\infty} t^n h(t)dt$, $X_n = [Z_y(+0) - Z_y(-0)]/n$, and the integral from $-\infty$ to ∞ is again interpreted as the limit (4.248). It follows clearly from (4.314) that for any h, h_1, and h_2 in $K^{(n)}$

$$(4.316) \quad m(h) = \langle X(h) \rangle = m_n\alpha_n(h),$$

$$(4.317) \quad B(h_1,h_2) = \langle X(h_1)\overline{X(h_2)} \rangle = \int^\infty \overset{\smile}{h}_1(\omega)\overline{\overset{\smile}{h}_2(\omega)}dF(\omega)$$

$$+ A_n^2\alpha_n(h_1)\overline{\alpha_n(h_2)},$$

where $m_n = \langle X_n \rangle$, $A_n^2 = \langle |X_n|^2 \rangle$ while the function $F(\omega)$ is nondecreasing on the half-lines $(-\infty, 0)$ and $(0, \infty)$ and is related to $F_y(\omega)$ by an expression similar to (4.315). It is easy to see that $F(\omega)$ satisfies equation (4.308) and condition

$$(4.318) \quad \int_{-\infty}^A \frac{dF(\omega)}{\omega^{2p}} + \int_{-A}^A \omega^{2n}dF(\omega) + \int_A^\infty \frac{dF(\omega)}{\omega^{2p}} < \infty$$

where p is some integer and $A > 0$ is arbitrary. Convergence of the integrals on the right-hand sides of (4.314) and (4.317) at $\omega = 0$ is guaranteed by the fact that according to (4.312)

$$\overset{\smile}{h}(0) = \overset{\smile}{h}{}'(0) = \dots = \overset{\smile}{h}{}^{(n-1)}(0) = \overset{\smile}{h}_1(0) = \overset{\smile}{h}{}'(0) = \dots$$

$$= \overset{\smile}{h}_1^{(n-1)}(0) = \overset{\smile}{h}_2(0) = \overset{\smile}{h}_2'(0) = \dots = \overset{\smile}{h}_2^{(n-1)}(0) = 0.$$

Having defined $X(h)$ for all h in $K^{(n)}$, one can extend the process $X(h)$ with stationary increments of order n to the whole space K, but we will not pursue this topic here. [99]

Equation (4.318) shows that the spectral distribution function $F(\omega)$ of a process with stationary increments of order n can diverge not only at infinity, but also at the origin. If the function $F(\omega)$ diverges only at $\omega = 0$, but not at infinity, i.e., if (and only if) one can put $p = 0$ in (4.318), the process $X(h)$ has point values and can be represented in the form (4.281), where $X(t)$ is an ordinary (nongeneralized) random process. This ordinary process will then be an *ordinary process with* (wide-sense) *stationary increments of order* n. This means that nth differences of the process $X(t)$

$$(4.319) \quad \Delta_\tau^{(n)}X(t) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} X(t - k\tau)$$

(where $\binom{n}{k}$, $k = 0,1, \dots, n$, are binomial coefficients) are such that the mean values

$$(4.320) \quad \langle \Delta_\tau^{(n)}X(s) \rangle = m^{(n)}(\tau)$$

$$(4.321) \quad \langle \Delta_{\tau_1}^{(n)}X(t + s) \overline{\Delta_{\tau_2}^{(n)}X(s)} \rangle = D^{(n)}(t;\tau_1,\tau_2)$$

exist and are continuous functions of the indicated

arguments which do not depend on s. It is easy to deduce from (4.281) and (4.314) $-$ (4.317) that

$$(4.322) \qquad \Delta_\tau^{(n)} X(t) = \int_{-\infty}^{\infty} e^{it\omega}(1 - e^{-i\tau\omega})^n dZ(\omega) + (n!)X_n\tau^n,$$

$$(4.323) \qquad m^n(\tau) = (n!)m_n\tau^n,$$

$$(4.324) \qquad D^{(n)}(t;\tau_1,\tau_2) = \int_{-\infty}^{\infty} e^{it\omega}(1 - e^{-i\tau_1\omega})^n(1 - e^{-i\tau_2\omega})^n dF(\omega)$$
$$+ (n!)^2 A_n^2 \tau_1^n \tau_2^n ,$$

where $Z(\omega)$, $F(\omega)$, X_n, m_n, A_n^2, and the integrals from $-\infty$ to ∞ have the same meaning as in (4.315), (4.316), and (4.317). (Note, that the integrals on the right-hand sides of (4.322) and (4.324) will converge if, and only if, the function $F(\omega)$ satisfies (4.318) with $p = 0$.)

It should be remarked that in studying processes with stationary increments of order n one can, in general, restrict oneself to considering only nth differences (4.319) but not the quantities $X(t)$. If, however, it is reasonable to assume (as is often the case in applications) that $X(t)$ for fixed t is a random variable (i.e., has probability distribution) and the set of variables $X(t)$ form a random process, then, using (4.322), one can obtain the following *spectral representation of the process* X(t) *with stationary increments of order* n

$$X(t) = \int_{-\infty}^{-1} e^{it\omega} dZ(\omega) + \int_{-1}^{1} \left\{ e^{t\omega} - 1 - it\omega - \ldots \right.$$

$$(4.325) \qquad \left. - \frac{(it\omega)^{n-1}}{(n-1)!} \right\} dZ(\omega) + \int_{1}^{\infty} e^{it\omega} dZ(\omega)$$

$$+ X_0' + X_1't + \ldots + X_{n-1}'t^{n-1} + X_n t^n,$$

where X_0', X_1', ..., X_{n-1}', X_n are some random variables. This spectral representation evidently generalizes the spectral representation (4.244) of processes with stationary first-order increments.[100]

The derivative X' of an ordinary (nongeneralized) process $X(t)$ with stationary increments of order n which has the spectral distribution function $F(\omega)$ is clearly a

(usually generalized) process with stationary increments of order $n - 1$ which has spectral distribution function $F^{(1)}(\omega)$ such that $dF^{(1)}(\omega) = \omega^2 dF(\omega)$. (This derivative will be an ordinary process $X'(t)$ if, and only if, the condition (4.318) holds for $p = -1$.) Similarly, the second derivative X'' of $X(t)$ (which is usually a generalized process) will be a process with stationary increments of order $n - 2$, while the nth derivative $X^{(n)}$ of the process $X(t)$ will be a stationary process (usually a generalized process) having a spectral distribution function which satisfies condition (4.297) for $p = n$. It is not hard to prove that any generalized stationary process whose spectral distribution function satisfies condition (4.297) for $p = n$ coincides with the nth derivative of some ordinary (nongeneralized) process $X(t)$ with stationary increments of order n; see Note 100. Thus, the set of all generalized stationary processes coincides with the set of all nth derivatives of nongeneralized processes with stationary increments of order n corresponding to all integers n.

Example: *Self-similar processes with stationary increments of order n.* A random process $X(t)$ with stationary increments of order n is said to be *self-similar* if the first and second moments of its nth increments are invariant under a similarity transformation group $t \to kt$, $X \to a(k)X$, where $a(k)$ is some function of k (see pp. 407–408 for the case where $n = 1$). In this case the functions $m^{(n)}(\tau)$ and $D^{(n)}(t;\tau_1,\tau_2)$ must evidently satisfy the relations

(4.326)
$$m^{(n)}(\tau) = a(k)m^{(n)}(k\tau),$$

$$D^{(n)}(t;\tau_1,\tau_2) = a^2(k)D^{(n)}(kt;k\tau_1,k\tau_2).$$

According to (4.323), if $m_n \neq 0$ (i.e., $\langle \Delta_\tau^{(n)} X(t) \rangle \neq 0$), then $m^{(n)}(\tau) \sim \tau^n$ and, hence $a(k) = m^{(n)}(\tau)/m^{(n)}(k\tau) \sim k^{-n}$. Therefore in this case

(4.327) $$\langle |\Delta_\tau^{(n)} X(t)|^2 \rangle = D^{(n)}(0;\tau,\tau) = \frac{D^{(n)}(0;1,1)}{a^2(\tau)} = C\tau^{2n}.$$

It is easy to show, however, that if $t = 0$, $\tau_1 = \tau_2 = \tau$, then for $\tau \to \infty$ the increase of the term expressed by the integral on the right-hand side of (4.324) is always slower than $|\tau|^{2n}$ (cf. Note 77 to Chap. 4 in Vol. II). Thus, if $m_n \neq 0$, then $F(\omega) = \text{const.}$ and $X(t) = X_0 + X_1 t + ... X_n t^n$, $D^{(n)}(t;\tau_1,\tau_2)$

$= C\tau_1^n \tau_2^n$ where $C = (n!)^2 \langle |X_n|^2 \rangle$.

Suppose now that $m_n = 0$ (i.e., $\langle \Delta_\tau^{(n)} X(t) \rangle = 0$). Then $a^2(k)$

$= D^{(n)}(0;1,1)/D(0;k,k)$ is an even function of k (i.e., $a^2(k) = $

$a^2(|k|))$, and $a^2(k_1 k_2) = a^2(k_1) a^2(k_2)$; therefore, $a^2(k) = |k|^\alpha$ (cf. Note 83 to this chapter in Vol. II). This implies that

$$(4.328) \quad D^{(n)}(0;\tau,\tau) = \frac{D(0;1,1)}{a^2(\tau)} = C|\tau|^m,$$

where $m = -2\alpha$ and $m > 0$, since $D(0) = 0$. We have already noted above that if $t = 0$, $\tau_1 = \tau_2 = \tau$, then the integral term on the right-hand side of (4.324) increases slower than $|\tau|^{2n}$ at infinity. Therefore, $0 < m < 2n$, and if $m = 2n$, then $F(\omega) = $ const. If, however, $0 < m < 2n$, then

$$D^{(n)}(0;\tau,\tau) = C|\tau|^m = 2^n \int_{-\infty}^{\infty} (1 - \cos\omega\tau)^n dF(\omega).$$

The last equality is evidently valid if the spectral distribution function is absolutely continuous (i.e., $dF(\omega) = f(\omega)d\omega$) and the spectral density $f(\omega)$ is proportional to $|\omega|^{-m-1}$, i.e., if

$$(4.329) \quad f(\omega) = \frac{C_1}{|\omega|^{m+1}}, \quad C_1 = \frac{C}{2^n \int_{-\infty}^{\infty} (1 - \cos x)^n x^{-m-1} \, dx}.$$

It is also easy to see that if the spectral density $f(\omega)$ has the form (4.329), then the second equation (4.326) is fulfilled for all t, τ_1, τ_2 (i.e., the process $X(t)$ is a self-similar process with stationary increments of order n). The converse statement is also true: A process with stationary increments of order n is self-similar only if either $F(\omega) = $ const. or $dF(\omega) = f(\omega)d\omega$, $A_n = 0$, and $f(\omega)$ is proportional to $|\omega|^{-m-1}$, where $0 < m < 2n$.[101] Note that if $m > 2$, then the process $X(t)$ is differentiable and its derivative $X'(t)$ is a self-similar process with stationary increments of order $n - 1$; if $m > 4$, then $X(t)$ is twice differentiable and its second derivative $X''(t)$ is a self-similar process with stationary increments of order $n - 2$, and so on.

25. Generalized Homogeneous Fields. Locally Homogeneous and Locally Isotropic Fields

25.1. Generalized Homogeneous Fields

All the results of the preceding section concerning generalized stationary processes can easily be extended to generalized homogeneous fields in the n-dimensional space $\mathbb{R}^n$. A *generalized random field* is by definition a random linear functional $X(h)$ which depends on an infinitely differentiable complex-valued function $h = h(t)$ of $t = (t_1, ..., t_n)$ such that $h(t)$ vanishes outside some n-dimensional interval. (The space of all such functions $h(t)$ will again be denoted by K_∞ or simply by K.) A generalized field $X(h)$ is said to be *homogeneous* if its mean value functional $m(h) = \langle X(h) \rangle$ and

the correlation functional $B(h_1, h_2) = \langle X(h_1)\overline{X(h_2)} \rangle$ satisfy (4.289) and (4.290), where S_T is now replaced by S_T and denotes the translation in $\mathbb{R}^n$ by vector $\tau = (\tau_1, ..., \tau_n)$ (i.e., $S_T h(t) = h(t + \tau)$). If the generalized field $X(h)$ can be represented as

$$(4.330) \quad X(h) = \int_{\mathbb{R}^n} h(t)X(t)dt$$

(i.e. $X(h)$ has point values and can therefore be reduced to an ordinary field $X(t)$), then the homogeneity of the field $X(h)$ clearly implies that the field $X(t)$ is homogeneous in the ordinary sense of Sec. 21.

The general form of the mean value functional $m(h)$ and correlation functional $B(h_1, h_2)$ of a homogeneous random field $X(h)$ is given by

$$(4.331) \quad m(h) = m \int_{\mathbb{R}^n} h(t)dt = m\widehat{h}(0)$$

and

$$(4.332) \quad B(h_1, h_2) = \int_{\mathbb{K}^n} \widehat{h}_1(k)\overline{\widehat{h}_2(k)}F(dk).$$

Here m is a numerical constant, $\widehat{h}(k) = \int_{\mathbb{R}^n} e^{ikt}h(t)dt$ is the Fourier transform of $h(t)$, and $F(dk)$ is the spectral measure of the field $X(h)$, which is such a nonnegative measure in the

n-dimensional space $\mathbb{K}^n$ of wave vectors $\mathbf{k}$ that

$$(4.333) \quad \int_{\mathbb{K}^n} \frac{F(d\mathbf{k})}{(1 + k^2)^p} < \infty, \quad k = |\mathbf{k}| = (k_1^2 + \dots + k_n^2)^{1/2},$$

for some integer p (i.e., $F(d\mathbf{k})$ is a "slowly increasing measure"). Conversely, every linear functional of the form (4.331), where m is a constant, is the mean value functional of some generalized homogeneous random field, and every bilinear functional of the form (4.332), where $F(d\mathbf{k})$ is a slowly increasing measure in $\mathbb{K}^n$, is the correlation functional of some such field. Equations (4.331) and (4.332) are simply generalizations of equations (4.291) and (4.298) valid in the case where $n = 1$, and their proof does not differ, in fact, from that of (4.291) and (4.298).[102] It is also clear that if the mean value functional (4.331) is different from zero (i.e., $m \neq$

0), then $B(h_1,h_2) = b_1(h_1,h_2) + |m|^2 \widehat{h}_1(0)\widehat{h}_2(0)$ where $b_1(h_1,h_2)$ is the correlation functional of the "centered" generalized field $\widehat{X}(h) = X(h) - m(h)$ with mean value zero. Hence, if $m(h) = m\widehat{h}(0)$, then $F(0) \geqslant |m|^2$, where $F(0)$ is the F-measure of the point $\mathbf{0} = (0, \dots, 0)$.

It is shown in the works cited in Note 102 that for a generalized homogeneous random field $X(h)$ (in a similar way as for a generalized stationary process) the *spectral representation* (4.332) of $B(h_1,h_2)$ also implies easily the existence of the *spectral representation for the field* $X(h)$ itself which has the form

$$(4.334) \quad X(h) = \int_{\mathbb{K}^n} \widehat{h}(\mathbf{k})Z(d\mathbf{k}).$$

Here $Z(d\mathbf{k})$ is a complex random measure in $\mathbb{K}^n$ satisfying (4.52) and conditions (a_1), (b), and (c) on pp. 329-330.

The partial derivative

$$(4.335) \quad X^{(m_1,\dots,m_n)}(h) = \frac{\partial^M X(h)}{\partial t_1^{m_1} \dots \partial t_n^{m_n}}, \quad M = m_1 + \dots + m_n,$$

of a generalized homogeneous random field $X(h)$ can be naturally defined by the relation

$$X^{(m_1,\dots,m_n)}(h) = (-1)^M X\big[h^{(m_1,\dots,m_n)}\big],$$

(4.336)
$$h^{(m_1,\dots,m_n)} = \frac{\partial^M h(t)}{\partial t_1^{m_1} \dots \partial t_n^{m_n}}.$$

It is easy to see that if $X(h)$ has point values $X(t)$ (i.e., is representable in the form (4.330)) and $X(t)$ is a homogeneous field having Mth derivative (in the mean) $\partial X^M(t)/\partial t_1^{m_1} \dots \partial t_n^{m_n} = X^{(m_1,\dots,m_n)}(t)$, then $X^{(m_1,\dots,m_n)}(h)$ also has point values coinciding with $X^{(m_1,\dots,m_n)}(t)$.   However, unlike ordinary (nongeneralized) homogeneous fields $X(t)$, generalized homogeneous fields $X(h)$ always have partial derivatives of all orders and all these derivatives are also generalized homogeneous fields.   Since $\widetilde{h}^{(m_1,\dots,m_n)}(k) = (-i)^M k_1^{m_1} \dots k_n^{m_n} \widetilde{h}(k)$ the nonnegative spectral measure $F^{(m_1,\dots,m_n)}(\Delta)$ and the random spectral measure $Z^{(m_1,\dots,m_n)}(\Delta)$ of the generalized homogeneous field $X^{(m_1,\dots,m_n)}(h)$ are related to the spectral measures $F(\Delta)$ and $Z(\Delta)$ of the original field $X(h)$ by the equations

$$F^{(m_1,\dots,m_n)}(\Delta) = \int_\Delta k_1^{2m_1} \dots k_n^{2m_n}\, F(d\mathbf{k}),$$

(4.337)
$$Z^{(m_1,\dots,m_n)}(\Delta) = i^M \int_\Delta k_1^{m_1} \dots k_n^{m_n} Z(d\mathbf{k}).$$

25.2. Locally Homogeneous Fields

A generalization of a random process with stationary increments which resembles the generalization leading from a stationary random process to a homogeneous field in the n-dimensional space $\mathbb{R}^n$ is a *locally homogeneous random field* (or, what is the same, a *random field with homogeneous increments*).   By this we mean a random field

$X(\mathbf{h})$ for which the mean values of both the field difference in an arbitrary pair of points and the product of two such differences related to two pairs of points remain unchanged under any translation of all the points under consideration.[103]

If $X(\mathbf{t})$ is a locally homogeneous random field, then the mean value of its increment

$$\langle X(\mathbf{t} + \boldsymbol{\tau}) - X(\mathbf{t}) \rangle = \Delta_{\boldsymbol{\tau}} X(\mathbf{t}) = m_1(\boldsymbol{\tau})$$

clearly depends only on the vector $\boldsymbol{\tau}$ and satisfies the relation

(4.338) $m_1(\boldsymbol{\tau}_1 + \boldsymbol{\tau}_2) = m_1(\boldsymbol{\tau}_1) + m_1(\boldsymbol{\tau}_2).$

Let us assume that $m_1(\boldsymbol{\tau})$ is a continuous functions of $\boldsymbol{\tau}$; then (4.338) implies that $m_1(\boldsymbol{\tau})$ is a linear function of the vector $\boldsymbol{\tau}$:

(4.339) $m_1(\boldsymbol{\tau}) = \mathbf{c}_1 \boldsymbol{\tau} = c_{11} \tau_1 + \dots + c_{1n} \tau_n$

where $\mathbf{c}_1 = (c_{11}, \dots, c_{1n})$ is a constant n-dimensional vector. The mean value of the field $X(\mathbf{t})$ itself, if it exists, is given by

(4.340) $\langle X(\mathbf{t}) \rangle = c_0 + \mathbf{c}_1 \mathbf{t}$

where $c_0 = \langle X(0) \rangle$ is a constant.

The general second moment of field differences at two pairs of points depends, in the case of a locally homogeneous field $X(\mathbf{t})$, on three vector arguments:

(4.341)
$$\langle \{X(\mathbf{t}_1 + \boldsymbol{\tau}_1) - X(\mathbf{t}_1)\} \overline{\{X(\mathbf{t}_2 + \boldsymbol{\tau}_2) - X(\mathbf{t}_2)\}} \rangle$$
$$= D(\mathbf{t}_1 - \mathbf{t}_2, \boldsymbol{\tau}_1, \boldsymbol{\tau}_2).$$

It is easy to see, however, that instead of the function (4.341) we need only consider the simpler function

(4.342) $D(\boldsymbol{\tau}_1, \boldsymbol{\tau}_2) = \langle \{X(\mathbf{t} + \boldsymbol{\tau}_1) - X(\mathbf{t})\} \overline{\{X(\mathbf{t} + \boldsymbol{\tau}_2) - X(\mathbf{t})\}} \rangle$

which is equal to $D(0; \boldsymbol{\tau}_1, \boldsymbol{\tau}_2)$ and depends on only two vectors $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_2$. In fact, (4.341) and (4.342) imply that

(4.343) $D(s;\tau_1,\tau_2) = D(s + \tau_1,\tau_2) - D(s,\tau_2)$.

If a field $X(t)$ is real, we can also use identity (4.217), which allows us to express the function (4.341) through the still simpler function

(4.344) $D(\tau) = D(0;\tau,\tau) = \langle |X(t + \tau) - X(t)|^2 \rangle$

which depends on only one vector τ. Both functions $D(\tau_1,\tau_2)$ and $D(\tau)$ will be called the *structure functions* of a locally homogeneous field $X(t)$ (cf. Sec. 23).

The main result of the spectral theory of locally homogeneous random fields concerns the existence of spectral representations for the structure functions $D(\tau_1,\tau_2)$ and $D(\tau)$ of the form

(4.345) $D(\tau_1,\tau_2) = \int\limits_{\mathbb{K}^n_-} (e^{i\tau_1 k} - 1)(e^{-\tau_2 k} - 1)F(dk) + \tau_1 \cdot \tau_2,$

(4.346) $D(\tau) = 2\int\limits_{\mathbb{K}^n_-} (1 - \cos k\tau)F(dk) + \tau \cdot \tau .$

Here $\mathbb{K}^n_-$ is the n-dimensional space of all vectors k except for the "zero vector" $0 = (0, ..., 0)$, $F(dk)$ is the *spectral measure* of $X(t)$ which is a nonnegative measure in the space $\mathbb{K}^n_-$ such that

(4.347) $\int\limits_{\mathbb{K}^n_-} \dfrac{k^2 F(dk)}{1 + k^2} < \infty, \quad k = |k|,$

$A = \|a_{jk}\|$ is a constant Hermitian positive definite $(n \times n)$–matrix (i.e., $\sum\limits_{j,k=1}^{n} a_{jk} b_j \overline{b_k} \geq 0$ for any complex numbers $b_1, ..., b_n$), and $\tau_1 \cdot \tau_2 = \tau_1 \tau_2$ is the scalar product of the vectors τ_1 and τ_2 in $\mathbb{R}^n$. Conversely, any functions $D(\tau_1,\tau_2)$ and $D(\tau)$, which admit of a representation in the form (4.345) and, respectively, (4.346), where $F(dk)$ and have the meaning indicated above, are structure functions of some locally homogeneous random field $X(t)$ in the n-dimensional space $\mathbb{R}^n$.[104]

Consider now the random function $\Delta_T X(t) = X(t + \tau) - X(t)$ depending on two vector arguments t and τ. According to (4.341), (4.343), and (4.345) the correlation function of the random function $\Delta_T X(t)$ has the form

$$\langle \Delta_{T_1} X(t_1) \, \overline{\Delta_{T_2} X(t_2)} \rangle = D(t_1 - t_2 ; \tau_1, \tau_2)$$

(4.348)
$$= \int_{\mathbb{K}^n_-} e^{it_1 k} (e^{iT_1 k} - 1) \, \overline{e^{it_2 k} (e^{iT_2 k} - 1)} \, F(d\mathbf{k})$$

$$+ A\tau_1\tau_2 \, .$$

It is not hard to show that according to the theorem on the generalized spectral representation of random functions (see Note 17 to Chap. 2 in Vol. II and also pp. 447–448) equation (4.348) implies the possibility of representing the random function $\Delta_T X(t)$ in the form

(4.349) $\Delta_T X(t) = \int_{\mathbb{K}^n_-} e^{itk} (e^{iTk} - 1) \, Z(d\mathbf{k}) + \mathbf{X}_1 \tau.$

Here $Z(d\mathbf{k})$ is a complex random measure in the space $\mathbb{K}^n$ (i.e., a random set function in $\mathbb{K}^n$ defined for all n-dimensional intervals Δk not containing the origin) which satisfies conditions (a), (b), and (c) on p. 329 and the relation (4.52), while $\mathbf{X}_1 = (X_{11}, ..., X_{1n})$ is a random vector such that

(4.350) $\langle X_k \overline{Z(\Delta)} \rangle = 0, \quad \langle X_k \rangle = c_{1k}, \quad \langle X_k \bar{X}_l \rangle = a_{kl}$

for any set Δ in $\mathbb{K}^n_-$ and any $k, l = 1, ..., n$.   Since $X(t) = \Delta_t X(0) + X(0)$, the following spectral representation of the locally homogeneous field $X(t)$ itself is implied by (4.349):

(4.351) $X(t) = \int_{\mathbb{K}^n_-} (e^{itk} - 1)Z(d\mathbf{k}) + \mathbf{X}_1 t + X_0$

where $X_0 = X(0)$ is a constant random variable.

Along with ordinary (nongeneralized) locally homogeneous fields $X(t)$ one can consider also generalized locally homogeneous fields $X(h)$. A generalized random field $X(h)$ defined on the subspace $K^{(1)}$ of functions in $K = K_\infty$ satisfying

the relation

(4.352) $\int_{\mathbb{R}^n} h(t)dt = 0$

is said to be a *generalized locally homogeneous field*, if for any h, h_1, h_2 in $K^{(1)}$ and a vector τ in $\mathbb{R}^n$ the functionals $m(h)$

$= \langle X(h) \rangle$ and $B(h_1,h_2) = \langle X(h_1)\overline{X(h_2)} \rangle$ satisfy the conditions (4.289) and (4.290) with S_τ replaced by S_τ.   If the field $X(h)$ has point values $X(t)$ (i.e., is representable as (4.330)), then condition (4.352) implies that only random variables of the form $\Sigma a_i X(t_i)$ are considered where $\Sigma a_i = 0$ (i.e., only linear combinations of field differences $X(t) - X(u)$ and the limits of such linear combinations).   In this case conditions (4.289) and (4.290) show that the first and second moments of field differences at pairs of points are unaffected by any translations of all the points involved.   Another definition (fully equivalent to the first) of a generalized locally homogeneous field is as follows: A generalized random field $X(h)$ is said to be locally homogeneous if all its first-order partial derivatives $X^{(1)}(h)$, ..., $X^{(n)}(h)$ (where $X^{(j)}(h) = \partial X(h)/\partial t_j$) are homogeneous random fields.

Let us base our considerations on the first definition of a locally homogeneous field.   Then the condition (4.352) implies that $\widecheck{h}(0) = 0$ and similarly $\widecheck{h}_1(0) = \widecheck{h}_2(0) = 0$ (where $h(k)$ is, as usual, the Fourier transform of $h(t)$).   The following formulae for $m(h)$, $B(h_1,h_2)$, and $X(h)$ are analogous to (4.339), (4.345), and (4.349):

(4.353) $m(h) = ic_1 \cdot \nabla\widecheck{h}(0),$

(4.354) $B(h_1,h_2) = \int_{\mathbb{K}^n} \widecheck{h}_1(k)\widecheck{h}_2(k)F(dk) + A\nabla\widecheck{h}_1(0) \cdot \nabla\widecheck{h}_2(0),$

(4.355) $X(h) = \int_{\mathbb{K}^n} \widecheck{h}(k)Z(dk) + iX_1\nabla\widecheck{h}(0).$

Here $c_1 = (c_{11}, ..., c_{1n})$ is a constant vector, $\nabla = (\partial/\partial k_1, ..., \partial/\partial k_n)$ is

the gradient operator, $F(d\mathbf{k})$ is a measure in the space $\mathbb{K}^n_-$ such that

(4.356) $$\int_{\mathbb{K}^n_-} \frac{k^2 F(d\mathbf{k})}{(1+k^2)^{p+1}} < \infty$$

for some integer p, $A = \|a_{ik}\|$ is a constant Hermitian positive definite matrix, $Z(d\mathbf{k})$ is a random measure in $\mathbb{K}^n_-$ satisfying conditions (a), (b), and (c) on p. 329 and (4.52), and $\mathbf{X}_1$ is a random vector satisfying (4.350).   Note also that if h, h_1, h_2, $\mathbf{c}_1$, A, $\mathbf{X}_1$, $F(d\mathbf{k})$, and $Z(d\mathbf{k})$ are such as indicated above, then any generalized random field of the form (4.355) is a generalized locally homogeneous field, and any functionals $m(h)$ and $B(h_1,h_2)$ of the form (4.353) and (4.354) are a mean value functional and a correlation functional of such a field.[105]

A generalized locally homogeneous field $X(h)$ has point values if, and only if, its spectral measure $F(d\mathbf{k})$ satisfies (4.356) for $p = 0$. Then the field $X(\mathbf{t})$ on the right-hand side of (4.330) is an ordinary (nongeneralized) locally homogeneous field satisfying the relations (4.349), (4.339), and (4.345).

Similar results can be obtained for more general *fields with homogeneous increments of order* M. It is simpler to assume from the outset that we are considering a generalized field $X(h)$ with homogeneous increments of order M and to define such a field as a generalized random field $X(h)$ whose partial derivatives $X^{(m_1,\ldots,m_n)}(h)$, where $m_1 + \cdots + m_n = M$, are generalized homogeneous fields. We will not, however, consider further this class of generalized random fields.[106]

25.3. Locally Isotropic Fields

A locally homogeneous random field $X(\mathbf{t})$ in the space $\mathbb{R}^n$ is said to be a *locally isotropic field* (or, what is the same, a random *field with isotropic increments*) if the mean value of a field difference at any pair of points and the second moment of field differences at two pairs of points are unaffected by arbitrary rotations and reflections of all the points involved. Locally isotropic random fields (both one-dimensional, i.e., *scalar*, and *n*-dimensional *vector fields* in $\mathbb{R}^n$) play an important part in the modern statistical theory of turbulence, and this is the main reason for the attention given to this

class of random functions.[107]

Let us consider, for simplicity, only real locally isotropic random fields $X(t)$. (This limitation is in fact inessential, since locally isotropic fields are usually real or, at least, possess some properties similar to those of real fields.[108]) Recall that if $X(t)$ is a locally homogeneous field, the mean value $m_1(\tau)$ of a field difference $X(t + \tau) - X(t)$ is given by (4.339), where c_1 is a constant vector. If a field $X(t)$ is a locally isotropic one, then $m_1(\tau)$ must be unaffected by any rotation of the vector τ. Since a rotation can always make the vector τ orthogonal to c_1, the mean value of a field difference must be identically zero for any locally isotropic field $X(t)$. Thus, if $X(t)$ is a locally isotropic field, then necessarily

$$(4.357) \quad \langle X(t + \tau) - X(t) \rangle = 0.$$

Hence the mean value of the field $X(t)$ itself, if it exists, must be constant.

Let now $D(\tau)$ be the structure function (4.344) of a locally isotropic field $X(t)$. Clearly, this function depends only on the length $\tau = |\tau|$ of the vector τ, i.e., $D(\tau) = D(\tau)$. It follows from this, firstly, that for a locally isotropic field the term $A\tau \cdot \tau$ on the right-hand side of (4.346) must be of the form $A_1 \tau^2$, where A_1 is a nonnegative constant (i.e., $A = A_1 \mathcal{E}$, where $\mathcal{E}$ is a unit matrix, and $a_{ik} = A_1 \delta_{ik}$). Secondly, the measure $F(dk)$ in the space $\mathbb{K}^n_-$ must be spherically symmetric (i.e., $F(g\Delta) = F(\Delta)$ for any set Δ in $\mathbb{K}^n_-$ and any spatial rotation g about the origin). If we now transform the integral on the right-hand side of (4.346) to spherical polar coordinates and integrate with respect to all angular variables, we arrive at the formula

$$(4.358) \quad D(\tau) = \int_0^\infty \{1 - \Lambda_n(k\tau)\} d\Phi(k) + A_1 \tau^2.$$

Here, $\Lambda_n(x)$ is the function (4.109), A_1 is a nonnegative constant, and $\Phi(k + \Delta k) - \Phi(k)$ is equal to the F-measure of the spherical shell $k < |k| \leqslant k + \Delta k$ so that according to (4.347) $\Phi(k)$ is a nondecreasing function on the half-line $0 < k < \infty$ which satisfies the condition

$$(4.359) \quad \int_0^\infty \frac{k^2 d\Phi(k)}{1 + k^2} < \infty$$

(cf. derivation of (4.105) in Sec. 22). Conversely, any function of the form (4.358), where A_1 is a nonnegative

constant and $\Phi(k)$ is a nondecreasing function on the half-line $0 < k < \infty$, satisfying condition (4.359), is a structure function of some locally isotropic random field.[109] In the important particular case where $A_1 = 0$ and the function $\Phi(k)$ is absolutely continuous (i.e., $d\Phi(k) = \varphi(k)dk$, where $\varphi(k) = \Phi'(k)$) formulae (4.358) and (4.359) take the form

(4.358a) $D(\tau) = \int_0^\infty \{1 - \Lambda_n(k\tau)\}\varphi(k)dk$

and

(4.359a) $\varphi(k) \geqslant 0, \quad \int_0^\infty \frac{k^2\varphi(k)dk}{1 + k^2} < \infty.$

If $X(t)$ is a locally isotropic random field in $\mathbb{R}^n$, then the values of this field at points on any straight line form a locally isotropic field on the straight line (i.e., a process with stationary increments). It follows from this that the structure function of a locally isotropic field in $\mathbb{R}^n$, where $n \geqslant 1$, can always be also represented in the form (4.246), where $F(\omega + \Delta\omega) - F(\omega) = F(-\omega) - F(-\omega-\Delta\omega)$ (since a locally isotropic field is invariant under reflections). In other words, each function of the form (4.358) can also be represented as

(4.360) $D(\tau) = 2\int_0^\infty (1 - \cos k\tau) \, d\Phi_1(k) + A_1\tau^2, \quad A_1 \geqslant 0,$

where $\Phi_1(k)$ is another nondecreasing function on the half-line $(0, \infty)$, which also satisfies condition of the form (4.359) by virtue of (4.251). In the particular case where both functions $\Phi(k)$ and $\Phi_1(k)$ are absolutely continuous, the *one-dimensional spectral density* $g_1(k) = \Phi_1'(k)$ is expressed via the n-*dimensional spectral density* $\varphi(k) = \Phi'(k)$ with the aid of the same relation (4.120a) which holds for isotropic random fields.

The values of a locally isotropic random field in $\mathbb{R}^n$ at points t belonging to a k-dimensional linear subspace $\mathbb{R}^k$, where $k < n$, form a locally isotropic field in $\mathbb{R}^k$. Let us now denote by $\mathfrak{D}_n^!$ the set of all structure functions $D(\tau)$ of locally isotropic random fields in the n-dimensional space $\mathbb{R}^n$ and introduce also the set $\mathfrak{D}_\infty^!$ of functions $D(\tau)$ which belong to $\mathfrak{D}_n^!$ for any integer n. Then evidently

$$\mathfrak{D}_1^! \supset \mathfrak{D}_2^! \supset \ldots \supset \mathfrak{D}_n^! \subset \ldots \subset \mathfrak{D}_\infty^!$$

where, just as on p. 354, $\mathfrak{D}_j^! \subset \mathfrak{D}_k^!$ means that the function set

$\mathfrak{D}_j'$ is part of the set $\mathfrak{D}_k'$. Moreover, any isotropic random field $X(t)$ in $\mathbb{R}^n$ is also a locally isotropic field in $\mathbb{R}^n$, and in this case

$$(4.361) \quad D(\tau) = 2\{B(0) - B(\tau)\}$$

where $B(\tau)$ is the correlation function of the field $X(t)$. Therefore the function class $\mathfrak{D}_n'$ includes, in particular, all the functions of the form $B(0) - B(\tau)$, where $B(\tau)$ belongs to the class $\mathfrak{D}_n$ of all n-dimensional isotropic correlation functions.

Example. *Self-similar locally isotropic fields.* Self-similar locally isotropic (in the wide sense) random fields in $\mathbb{R}^n$ form a natural generalization of self-similar processes with stationary increments examined in Example 3 of Sec. 23 (p. 406). In both, the structure function $D(\tau)$ is a power function of the form (4.262). Since the structure function of a locally isotropic field in $\mathbb{R}^n$ must, for any n, also be the structure function of a process with stationary increments, it is clear that the exponent m in (4.262) for any n must satisfy the condition $0 < m \leqslant 2$. On the other hand, the power function $D(\tau) = C\,|\tau|^m$ for $m = 2$ is evidently representable in the form (4.358), where $A_1 = C$ and $\Phi(k) = $ const., while for $0 < m < 2$ it is representable as (4.358a), where

$$(4.362) \quad \varphi(k) = Ak^{-m-1}, \quad A = \frac{2^m m\Gamma[(n+m)/2]}{\Gamma(n/2)\Gamma[1-(m/2)]}\,C.$$

(The last result can be derived, e.g., from formula 6.561.14 of Gradshteyn and Ryzhik's book, 1980, by means of analytic continuation; for $n = 1$, it agrees with equation (4.263) in Sec. 23.) Clearly, if $C > 0$ and $0 < m < 2$, then the function (4.362) satisfies conditions (4.359a). Therefore, the functions $D(\tau) = C\tau^m$, where $C > 0$ and $0 < m \leqslant 2$ are structure functions of some locally isotropic fields in $\mathbb{R}^n$ for any n, i.e., all of them belong to the class $\mathfrak{D}_\infty'$.

Let $X(t)$ be a locally isotropic random field in the space $\mathbb{R}^n$, which has a structure function of the form (4.262) where $C > 0$, $0 < m \leqslant 2$ and, besides, $X(0) = 0$. Then, according to (4.217), the correlation function $B(t,s) = \langle X(t)X(s)\rangle = \langle\{X(t) - X(0)\}\{X(s) - X(0)\}\rangle$ of the field $X(t)$ is given by

$$(4.363) \quad B(t,s) = \frac{C}{2}\{|t|^m + |s|^m - |t-s|^m\}.$$

Thus, for $C > 0$, $0 < m \leqslant 2$ the function (4.363) is a correlation function in the space $\mathbb{R}^n$ of arbitrary dimension n, i.e., it is a positive definite kernel in $\mathbb{R}^n$. [110]

Clearly, the structure function (4.262) of a locally isotropic random field in $\mathbb{R}^n$ is self-similar in the sense introduced on pp. 407–408 for the case where $n = 1$ (i.e., for random processes with stationary increments). Let $t \rightarrow hgt + t_0$ be an arbitrary similarity transformation in $\mathbb{R}^n$ (here, g is a rotation or a rotation and/or reflection, h is a positive numerical dilatation factor, and t_0 is a vector in $\mathbb{R}^n$), and $X \rightarrow a(h)X$ is a simultaneous similarity transformation of the field X. Then for $a = h^{-m/2}$ the structure function (4.262) will be invariant under a group of such joint transformations of t and X. It is also easy to show that no structure function $D(\tau)$ different from the power function (4.262) is self-similar in the indicated sense (cf. Note 83 to this chapter in Vol. II).

Since any locally isotropic field $X(t)$ is also locally homogeneous, it is clear that such a field is always representable as (4.349). However, the spectral representation (4.349) does not use the local isotropy of the field $X(t)$ and therefore it is not specific to the class of locally isotropic fields. It seems obvious that for locally isotropic fields there must also exist a more special spectral representation convertible into the polar spectral representation (4.146) in the case where the field $X(t)$ is not only locally isotropic, but also simply isotropic. However, such a "polar spectral representation of locally isotropic fields" has never been considered so far, and we also shall not investigate it here.

[*]In studying locally homogeneous random fields in multidimensional spaces $\mathbb{R}^n$, we restricted ourselves to the case of a one-dimensional (i.e., scalar) field $X(t)$, since the theory of locally homogeneous multidimensional fields $\mathbf{X}(t)$ is trivial and uninteresting — it may be developed quite similarly to the theory of multidimensional stationary processes, homogeneous fields, and processes with stationary increments. However, in the case of locally isotropic fields in $\mathbb{R}^n$ the situation is somewhat more complicated since there exist two different multidimensional generalizations of the concept of a locally isotropic field (cf. presentation of the theory of multidimensional isotropic fields in Sec.

22).   The simplest of these generalizations deals with a multidimensional locally isotropic field $X(t) = \{X_1(t), ..., X_s(t)\}$ which has the property that the mean value of any field difference $X_j(t_2) - X_j(t_1)$, and the second moment of any two such differences $X_j(t_2) - X_j(t_1)$ and $X_k(t_4) - X_k(t_3)$ are unaffected by any translation, rotation, and reflection of the points t_1 and t_2 or, respectively, t_1, t_2, t_3, and t_4.   Such a field $X(t)$ is composed of s one-dimensional locally isotropic components which are locally isotropically correlated with each other. Clearly, in this case all the mean values $\langle X_j(t + \tau) - X_j(t)\rangle = m_j(\tau)$, $j = 1, ..., s$, are identically zero, while all the *cross-structure functions*

$$D_{jl}(\tau) = \langle \{X_j(t + \tau) - X_j(t)\}\{X_l(t + \tau) - X_l(t)\}\rangle,$$
(4.364)

$$j, l = 1,...,s,$$

are representable as

(4.365) $D_{jl}(\tau) = \int_0^\infty \{1 - \Lambda_n(k\tau)\}d\Phi_{jl}(k) + A_{jl}\tau^2,$

where A_{jl} are real constants, $\Phi_{jl}(\tau)$ are real functions of bounded variation such that the matrices $\|A_{jl}\|$ and $\|\Delta\Phi_{jl}(k)\| = \|\Phi_{jl}(k + \Delta k) - \Phi_{jl}(k)\|$, where $\Delta k > 0$, are positive definite, and the trace $\Phi(k) = \Sigma\Phi_{jj}(k)$ of $\|\Phi_{jl}(k)\|$ satisfies condition (4.359). Note that in this case the more general structure functions

$$D_{jl}(s; \tau_1,\tau_2)$$
(4.366)

$$= \langle \{X_j(t + s + \tau_1) - X_j(t + s)\}\{X_l(t + \tau_2) - X_l(\tau_2)\}\rangle,$$

depending on three vectors s, τ_1, and τ_2, can always be expressed via functions $D_{jl}(\tau)$ of a single scalar variable $\tau = |\tau|$:

$$D_{jl}(s; \tau_1,\tau_2) = \frac{1}{2} \{D_{jl}(|s + \tau_1|) + D_{jl}(|s - \tau_2|)$$
(4.367)

$$- D_{jl}(|s + \tau_1 - \tau_2|) - D_{jl}(|s|)\}.[111]$$

The second, more complicated but also more interesting, multidimensional generalization of the

concept of a one-dimensional locally isotropic field stems from the assumption that $\mathbf{X}(t) = \{X_1(t), ..., X_s(t)\}$ is an *s-dimensional quantity whose* s *components transform linearly under spatial rotations and reflections.* Here we restrict ourselves to few remarks relating to the case of a *vector locally isotropic field* $\mathbf{X}(t) = \{X_1(t), ..., X_n(t)\}$ in the space $\mathbb{R}^n$, i.e. we assume that $\mathbf{X}$ is an n-dimensional vector (see pp. 372–383, where vector isotropic fields in $\mathbb{R}^n$ are considered[112]). It is easy to show that if $\mathbf{X}(t)$ is a vector locally isotropic field in $\mathbb{R}^n$, then

$$(4.368) \quad \langle \mathbf{X}(t + \tau) - \mathbf{X}(t) \rangle = m_1(\tau) = m_1\tau$$

where m_1 is a constant. Moreover, the structure functions

$$(4.369) \quad D_{jl}(\tau) = \langle \{X_j(t + \tau) - X_j(t)\}\{X_l(t + \tau) - X_l(t)\} \rangle$$

here form a symmetric tensor (the *structure tensor* of $\mathbf{X}(t)$) which has the form

$$(4.370) \quad D_{jl}(\tau) = [D_{LL}(\tau) - D_{NN}(\tau)] \frac{\tau_j \tau_l}{\tau^2} + D_{NN}(\tau)\delta_{jl},$$

where δ_{jl} is the Kronecker symbol, τ_j are the components of the vector τ, and $D_{LL}(\tau)$, $D_{NN}(\tau)$ are two scalar functions of $\tau = |\tau|$. (It is easy to see that those functions are equal to the mean squares of the differences $X_L(t + \tau) - X_L(t)$ and $X_N(t + \tau) - X_N(t)$, where X_L and X_N denote the projection of the vector $\mathbf{X}$ on to the direction of τ and, correspondingly, on to any direction perpendicular to τ; cf. the similar equation (4.155).) The functions $D_{LL}(\tau)$ and $D_{NN}(\tau)$ are said to be the *longitudinal* and *lateral structure functions* of the vector locally isotropic field $\mathbf{X}(t)$. The general form of these two functions is given by the formulae

$$D_{LL}(\tau) = \int_0^\infty \left\{1 - \Lambda_{n+2}(k\tau) + \frac{(k\tau)^2}{n + 2} \Lambda_{n+4}(k\tau)\right\} d\Phi_1(k)$$

$$(4.371) \qquad + (n - 1) \int_0^\infty \left\{1 - \Lambda_{n+2}(k\tau)\right\} d\Phi_2(k) + A_L\tau^2,$$

$$D_{NN}(\tau) = \int_0^\infty \left\{1 - \Lambda_{n+2}(k\tau)\right\} d\Phi_1(k) +$$

$$(n - 1)\int_0^\infty \left\{1 - \frac{n}{n - 1} \Lambda_n(k\tau) + \frac{1}{n - 1}\Lambda_{n+2}(k\tau)\right\} d\Phi_2(k) + A_N\tau^2.$$

Here, $\Lambda_n(x)$ is the function (4.109), $\Phi_1(k)$ and $\Phi_2(k)$ are two nondecreasing functions on the half-line $0 < k < \infty$ satisfying conditions

(4.372) $\displaystyle\int_0^\infty \frac{k^2 d\Phi_j(k)}{1 + k^2} < \infty, \quad j = 1, 2,$

while A_L and A_N are two arbitrary nonnegative constants. Under wide conditions (in particular, in all cases where $A_L = A_N = 0$) the *general structure tensor* (4.366) is also symmetric (i.e., $D_{jl}(s;\tau_1,\tau_2) = D_{lj}(s;\tau_1,\tau_2)$) and it is representable in terms of the tensor (4.369) via the relation

(4.373) $$D_{jl}(s\,;\,\tau_1,\tau_2) = \frac{1}{2}\Big\{ D_{jl}(s + \tau_1) + D_{jl}(s - \tau_2) \\ - D_{jl}(s + \tau_1 - \tau_2) - D_{jl}(s)\Big\}.$$

In the important particular case where the vector locally isotropic field $X(t)$ is *solenoidal* (i.e., $\partial X_1/\partial t_1 + \dots + \partial X_n/\partial t_n = 0$, where the derivatives may commonly be generalized random fields),

(4.374) $\Phi_1(k) = \mathrm{const}, \quad D_{NN}(\tau) = D_{LL}(\tau) + \dfrac{\tau}{n-1} \dfrac{dD_{LL}(\tau)}{d\tau}.$

Similarly, if $X(t)$ is a locally isotropic *potential* field (i.e., $\partial X_j/\partial t_l = \partial X_l/\partial t_j$ for any j and l), then

(4.375) $\Phi_2(k) = \mathrm{const}, \quad D_{LL}(\tau) = D_{NN}(\tau) + \tau\, \dfrac{dD_{NN}(\tau)}{d\tau}.$

Thus, formulae (4.371) and (4.358) imply that *the class of longitudinal structure functions* $D_{LL}(\tau)$ *of solenoidal vector locally isotropic fields in* $\mathbb{R}^n$ *coincides with the class of lateral structure functions* $D_{NN}(\tau)$ *of locally isotropic potential vector fields in* $\mathbb{R}^n$ *and also with the class* $\mathfrak{D}'_{n+2}$ *of structure functions* $D(\tau)$ *of scalar locally isotropic random fields in the* (n + 2)-*dimensional space* $\mathbb{R}^{n+2}$.*

In conclusion we shall also say a few words about generalized locally isotropic fields in the *n*-dimensional space $\mathbb{R}^n$. The generalized locally homogeneous field $X(h)$ in $\mathbb{R}^n$ given in the subspace $K^{(1)}$ of functions $h(x)$ in K which satisfy the condition (4.352) is said to be a *generalized locally isotropic field* if its mean value functional $m(h)$ and

correlation functional $B(h_1, h_2)$ satisfy the conditions

(4.376) $m(gh) = m(h), \quad B(gh_1, gh_2) = B(h_1, h_2)$

for any h, h_1, and h_2 in $\mathbb{K}^{(1)}$ and for any rotation or reflection g in the space $\mathbb{R}^n$. It is not hard to show that a locally homogeneous field having a mean value functional (4.353) and a correlation functional (4.354) will be locally isotropic if, and only if, $c_1 = 0$ (i.e., $m(h) \equiv 0$), the measure $F(dk)$ is spherically symmetric (i.e., $F(g\Delta) = F(\Delta)$ for any g and for any F-measurable set Δ in $\mathbb{K}^n_{\cdot}$), and $A = A_1 \mathcal{E}$, where A_1 is a nonnegative constant, $\mathcal{E}$ is a unit $(n \times n)$-matrix. Under these conditions, (4.354) can be written as

$$B(h_1, h_2) = \int_0^\infty\!\!\!\int_{\mathbb{R}^n}\!\!\int_{\mathbb{R}^n} h_1(t_1)\overline{h_2(t_2)}\,\Lambda_n(|t_1 - t_2|k)\,dt_1 dt_2\, d\Phi(k)$$

(4.377)

$$+ A_1\!\int_{\mathbb{R}^n}\!\!\int_{\mathbb{R}^n} t_1 \cdot t_2\, h_1(t_1)\overline{h_2(t_2)}\,dt_1 dt_2,$$

where $\Lambda_n(x)$ is the function (4.109), A_1 is a nonnegative constant, and $\Phi(k)$ is a nondecreasing function on the half-line $0 < k < \infty$ satisfying the condition

(4.378) $\displaystyle\int_0^\infty \frac{k^2 d\Phi(k)}{(1 + k^2)^{p+1}} < \infty$

for some integer p. Conversely, each bilinear functional of the form (4.377), where A_1 and $\Phi(k)$ are such as indicated above, is a correlation functional of a generalized locally isotropic field. An equation similar to (4.377) (but a more cumbersome one) can also be written out for the tensor functional $B_{jl}(h_1, h_2)$ corresponding to the vector generalized locally isotropic field $X(h) = \{X_1(h), ..., X_n(h)\}$.[113]

If condition (4.378) holds for $p = 0$, the locally isotropic field $X(h)$ has point values, i.e. is representable as (4.330), where $X(t)$ is an ordinary (nongeneralized) locally isotropic field. However, the whole class of generalized locally isotropic fields is much wider than the class of ordinary locally isotropic fields. In particular, the condition (4.378) shows that the power function

(4.379) $\varphi(k) = Ak^{-m-1}, \quad 0 < k < \infty,$

can be the spectral density of a generalized locally isotropic
field, provided that $m < 2$. When $0 < m < 2$ the corresponding
field $X(h)$ has point values, but when $m \leqslant 0$ it is essentially a
generalized field.   It is easy to show that this generalized
field   is   also   self-similar,   i.e,   its   correlation   functional
$B(h_1,h_2)$, where h_1 and h_2 belong to $K^{(1)}$, is invariant under a
group   of   similarity   transformations   in   the   space $\mathbb{R}^n$
accompanied by multiplication of the quantities X by an
appropriate factor; however, we shall not pursue this topic
further.   We shall only note that in fact the function (4.379)
can be the spectral density of a self-similar generalized
random field for $m > 2$ as well, but in this case the field $X(h)$
will no longer be locally isotropic, but will have isotropic
increments of some higher order.[114]

26. Further Examples of Various Spectral Representations

In the previous sections of this chapter we repeatedly used
Karhunen's theorem on the *generalized spectral representation*
(or *orthogonal representation*) *of random functions* which is
stated in Note 17 to Chap. 2.   Recall that this theorem
formulates conditions guaranteeing that the random function
$X(t)$ <u>on</u> the set $T = \{t\}$ with a correlation function $B(t,s) =$
$\langle X(t)X(s) \rangle$ is representable as

$$(4.380) \quad X(t) = \int_A \varphi(t,a)dZ(a)$$

where A is some set of elements a, $\varphi(t,a)$ is a complex-valued
function of two variables t and a, and Z is an *orthogonal
random measure* on A.   (This means that Z is a random
function of the subset Δa of A which is determined for a
wide enough class $\mathfrak{a}$ of subsets Δa which includes, together
with any two disjoint subsets $\Delta_1 a$ and $\Delta_2 a$, their union $\Delta_1 a +$
$\Delta_2 a$; moreover, $Z(\Delta a)$ is such that

$$(4.381a) \quad Z(\Delta_1 a) + Z(\Delta_2 a) = Z(\Delta_1 a + \Delta_2 a),$$

$$(4.381b) \quad \langle Z(\Delta_1 a)\overline{Z(\Delta_2 a)} \rangle = 0$$

for any two disjoint subsets $\Delta_1 a$ and $\Delta_2 a$ from class $\mathfrak{a}$.)
According to Karhunen's theorem, the random function $X(t)$
can be represented in the form (4.380) if, and only if, its
correlation function $B(t,s)$ can be represented in the form

$$(4.382) \quad B(t,s) = \int_A \varphi(t,a)\overline{\varphi(s,a)}F(da),$$

where F is a measure on A (i.e., a nonnegative additive function of the subset Δa determined for all Δa from $\mathfrak{a}$). The measure F is, of course, related to the random measure Z by the equation

$$(4.383) \quad \langle |Z(\Delta a)|^2 \rangle = F(\Delta a).$$

In the final section of the book we will consider some additional applications of the theorem on the generalized spectral representation. For definiteness, we will, as a rule, speak only of functions $X(t)$ of a continuous parameter t (i.e. of random processes), although many of the results cited can also be readily formulated for functions of a discrete argument t (random sequences) or an argument t belonging to $\mathbb{R}^n$ or $\mathbb{Z}^n$ (random fields). Moreover, in conclusion we will also study briefly another type of spectral representations of the form (4.380), where the random measure Z is no longer orthogonal (i.e., does not satisfy condition (4.381b)) and we will discuss one generalization of the concepts of the spectral distribution function and spectral density applicable to a wide class of nonstationary random processes.

26.1. Karhunen – Loève Expansion

Let $X(t)$ be an arbitrary mean square continuous random process given on the interval $a \leqslant t \leqslant b$ of the t-axis, and let $B(t,s) = \langle X(t)\overline{X(s)} \rangle$ be a correlation function of $X(t)$. Consider the integral equation

$$(4.384) \quad \int_a^b B(t,s)\psi(s)ds = \lambda\psi(t), \quad a \leqslant t \leqslant b.$$

Equation (4.384) is a linear integral equation with a symmetric (or, in the complex case, Hermitian) positive definite kernel. According to the general theory of such integral equations, it has a denumerable set of nonnegative eigenvalues λ_j, $j = 1, 2, \ldots$, which are matched by a denumerable set of eigenfunctions $\psi_j(t)$, $j = 1, 2, \ldots$, forming a complete orthonormal set of functions on $[a,b]$. Hence, any square integrable function can be expanded into a series in functions $\psi_j(t)$, and

(4.385) $\displaystyle\int_a^b \psi_j(t)\overline{\psi_k(t)}dt = \delta_{jk}.$

Moreover, by virtue of Mercer's theorem,

(4.386) $\displaystyle B(t,s) = \sum_{j=1}^{\infty} \lambda_j \psi_j(t)\overline{\psi_j(s)}$

where the series converges uniformly in both variables.[115]

The representation of the correlation function $B(t,s)$ in the form (4.386) is clearly a particular case of representation (4.382), where the set A is discrete and consists of integers $j = 1$, $2, \ldots$. Hence, (4.386) implies the possibility of representing the random process $X(t)$ as

(4.387) $\displaystyle X(t) = \sum_{j=1}^{\infty} \sqrt{\lambda_j}\, \psi_j(t)Z_j$

where, according to (4.381b) and (4.383),

(4.388) $\langle Z_j \overline{Z_k} \rangle = \delta_{jk}.$

Conversely, if the random process $X(t)$, $a \leqslant t \leqslant b$, can be represented in the form (4.387), where the random variables Z_j and function $\psi_j(t)$ satisfy conditions (4.388) and (4.385), then evidently

$$B(t,s) = \langle X(t)\overline{X(s)} \rangle = \langle \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} (\lambda_j\lambda_k)^{1/2}\, \psi_j(t)\overline{\psi_k(s)}\, Z_j\overline{Z_k} \rangle$$

$$= \sum_{k=1}^{\infty} \lambda_j\psi_j(t)\overline{\psi_j(s)}.$$

Multiplying both sides of this relation by $\psi_k(t)$ and integrating the result with respect to t from a to b, we immediately find that the functions $\psi_k(t)$ and the nonnegative constants λ_k must be eigenfunctions and eigenvalues of the integral equation (4.384). Thus, the orthonormal family of functions $\psi_j(t)$ and the constant coefficients λ_j appearing in the representation (4.387) of the process $X(t)$ coincide with the eigenfunctions and eigenvalues of integral equations (4.380). We see that the correlation function $B(t,s)$ determines uniquely the orthonormal functions $\psi_j(t)$ and constants λ_j, provided that the random variables Z_j satisfy (4.388). It follows from this that the random variables Z_j are determined uniquely by the process $X(t)$: equations (4.387) and (4.385) clearly imply that

$$(4.389) \qquad Z_j = \lambda_j^{1/2} \int_a^b X(t)\overline{\psi_j(t)}dt.$$

If, however, the validity of the orthonormality condition (4.385) is not required, then the representation of the process $X(t)$ in the form (4.387), where Z_j satisfy (4.388), is determined with a high degree of arbitrariness.[116] Note also that if $X(t)$ is a Gaussian random process, then, according to (4.389), the random variables Z_j also have a Gaussian probability distribution. In particular, if $X(t)$ is a real Gaussian process with mean value zero, then the coefficients Z_n are independent Gaussian random variables.

Representation (4.387) of a random process $X(t)$ as a series in eigenfunctions of the correlation kernel $B(t,s) = \langle X(t)\overline{X(s)} \rangle$ is, of course, very natural and theoretically simple. It is no wonder therefore that such a representation has been introduced independently by a number of scientists.[117] At present the mathematicians commonly call it the *Karhunen–Loève expansion* after the names of two of these scientists. In the applied literature, however, it is also sometimes used under the name of "proper orthogonal decomposition" or "canonical decomposition", or else "the expansion in empirical orthogonal functions". Note that the variances λ_n of the random variables Z_n in most practical applications decrease rapidly with the number n (if, as is usually done, the eigenvalues λ_n are arranged in the decreasing order, i.e., $\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3 \geqslant \dots$). Therefore, the few first terms of expansion (4.387) usually determine almost all the variability of $X(t)$, so that in practice all the terms of the series, with the exception of the first few terms, may often be neglected.[118]

Example. Let $B(t,s) = \min(t,s)$, $0 \leqslant t,s \leqslant 1$ (cf. (1.20) and (4.253b)). In this case (4.384) has the form

$$(4.390) \qquad \int_0^t s\psi(s)ds + t\int_t^1 \psi(s)ds = \lambda\psi(t), \qquad 0 \leqslant t \leqslant 1.$$

The left-hand side of (4.390) is evidently differentiable with respect to t; hence, the right-hand side may also be differentiated. Differentiating both sides, we obtain

$$(4.391) \qquad \int_t^1 \psi(s)ds = \lambda\psi'(t), \qquad 0 \leqslant t \leqslant 1.$$

Differentiation of both sides of the obtained equation now yields

(4.392) $-\psi(t) = \lambda\psi''(t),$ i.e., $\psi(t) = A\sin\dfrac{t}{\sqrt{\lambda}} + B\cos\dfrac{t}{\sqrt{\lambda}}$

for $\lambda \neq 0$. If $\lambda = 0$, then the first equation (4.392) implies that $\psi(t) = 0$ for all t; therefore $\lambda = 0$ is not an eigenvalue of (4.390). (In fact, the last conclusion is of no importance, since the eigenvalue zero, if it exists, contributes nothing to the expansion (4.387).) Putting $t = 0$ in (4.390), we obtain that $\psi(0) = 0$; consequently, $B = 0$. Moreover, (4.391) for $t = 1$ yields $\psi'(1) = 0$. This implies that

$$\cos\frac{1}{\sqrt{\lambda}} = 0, \quad \frac{1}{\sqrt{\lambda}} = \frac{(2n-1)\pi}{2}, \; n = 1,2,\ldots,$$

i.e.,

$$\lambda_n = \frac{4}{(2n-1)^2\pi^2}, \; \psi_n(t) = \sqrt{2}\sin\left(n - \frac{1}{2}\right)\pi t, \; n = 1,2,\ldots.$$

Thus the Karhunen–Loève expansion of the process $X(t)$, where $0 \leqslant t \leqslant 1$, with the correlation function $B(t,s) = \min(t,s)$ has the form

$$(4.393) \quad X(t) = \sqrt{2}\sum_{n=1}^{\infty} \frac{\sin[n-(1/2)]\pi t}{[n-(1/2)]\pi} Z_n.$$

If the process $X(t)$ is real, then the random variables Z_n are also real by virtue of (4.389), so that in this case $\langle Z_n Z_m \rangle = 0$ for $m \neq n$, $\langle Z_n^2 \rangle = 1$ for all n.[119]

It is proved similarly that if $B(t,s) = C\min(t,s)$ and $0 \leqslant t \leqslant T$, then

$$(4.393a) \quad X(t) = \sqrt{2CT}\sum_{n=1}^{\infty} \frac{\sin[n-(1/2)]\pi t/T}{[n-(1/2)]\pi} Z_n$$

where the random variables Z_n have the same properties as in (4.393).

A number of other examples of Karhunen–Loève expansions for random processes on finite intervals (and also for random fields on bounded two-dimensional domains; cf. Note 115) can be found in the available literature, but we shall not consider them here.[120]

26.2. Moving Average Representations of Stationary Random Processes

In Sec. 6, Examples 2–4, we considered a class of stationary

random sequences $X(t)$ formed by moving averages of a sequence of uncorrelated random variables $E(t)$. In the case of continuous t (i.e. when $X(t)$ is a stationary random process) the role of a sequence of uncorrelated random variables is evidently played by a white noise $E(t)$, $-\infty < t < \infty$. Therefore the moving averages of $E(t)$ must here be replaced by integrals of the form $\int b(t - s)E(s)ds = \int b(t - s)dW(s)$ where $b(t)$ is a numerical weighting function, and $W(t)$ is a standard Wiener process, which is an indefinite integral of a white noise $E(t)$ (see (2.194a)) and has the properties

$$(4.394) \quad \langle[W(t + s) - W(t)]^2\rangle = s, \quad \langle dW(s)dW(s')\rangle = \delta(s - s')dsds'.$$

(Equations (4.394) differ from (2.195a,b) by the absence of the factor c^2, since now it is more convenient to consider that

$W(t) = \int_0^t E(s)ds$, i.e. to incorporate the constant c in the

weighting function $b(t)$.) Processes of the form

$$(4.395) \quad \begin{aligned} X(t) &= \int_{-\infty}^{\infty} b(t - s)dW(s) \\ &= \int_{-\infty}^{\infty} b(\tau)dW(t - \tau), \quad \int_{-\infty}^{\infty}|b(t)|^2dt < \infty, \end{aligned}$$

are clearly analogous to two-sided moving average sequences, while the role of one-sided moving average sequences is played by random processes

$$(4.396) \quad X(t) = \int_0^{\infty} b(t - s)dW(s) = \int_0^{\infty}b(\tau)dW(t - \tau), \quad \int_0^{\infty}|b(t)|^2dt < \infty.$$

Consider firstly *two-sided moving average processes* (4.395). Since $|b(t)|^2$ is integrable over $(-\infty,\infty)$, the Fourier transform of $b(t)$ does exist, and therefore [121]

$$(4.397) \quad b(t) = \frac{1}{(2\pi)^{1/2}}\int_{-\infty}^{\infty} e^{i\omega t}g(\omega)d\omega, \quad g(\omega) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty}e^{-i\omega t}b(t)dt.$$

Moreover, (4.395) and (4.394) imply that

$$(4.398) \quad \begin{aligned} B(\tau) &= \langle X(t + \tau)\overline{X(t)}\rangle = \int_{-\infty}^{\infty} b(t + \tau - s)\overline{b(t - s)}ds \\ &= \int_{-\infty}^{\infty}b(\tau + s')\overline{b(s')}ds'. \end{aligned}$$

(The last equation is, of course, a continuous time analog of

(2.9).) Since the Fourier transform of $b(\tau + s')$ (considered as a function of s') is equal to $\exp(i\omega\tau)g(\omega)$, the Plancherel theorem implies that

$$(4.399) \quad B(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau}|g(\omega)|^2 d\omega.$$

We see that a two-sided moving average process always has the spectral density $f(\omega) = |g(\omega)|^2$, where $g(\omega)$ is the Fourier transform of the weighting function $b(t)$.

Conversely, let $X(t)$ be a stationary random process having a spectral density $f(\omega)$ and let $g(\omega)$ be any function satisfying the relation

$$(4.400) \quad |g(\omega)|^2 = f(\omega).$$

(The function $g(\omega)$ is clearly determined by (4.400) only to within a factor of the form $\exp\{i\Theta(\omega)\}$, where $\Theta(\omega)$ is an arbitrary real function of ω.) Then (2.67) implies (4.399). Since $|g(\omega)|^2 = f(\omega)$ is integrable over $(-\infty,\infty)$, the Fourier transform of the function $g(\omega)$ does exist. Denote this Fourier transform by $b(u)$; then $b(u + \tau)$ is the Fourier transform of $\exp(i\omega\tau)g(\omega)$. Applying the Plancherel theorem again, we obtain

$$(4.398a) \quad \begin{aligned} B(t - s) &= \langle X(t)\overline{X(s)} \rangle = \int_{-\infty}^{\infty} b(t - s + u)\overline{b(u)} du = \\ &\int_{-\infty}^{\infty} b(t - u')\overline{b(s - u')} du' \end{aligned}$$

where $u' = s - u$ (this result is equivalent to (4.398)). According to the theorem on the generalized spectral representation of random functions as given on pp. 447–448, (4.398a) implies that the random process $X(t)$ can be represented in the form (4.395), where $W(t)$ satisfies (4.394). Thus we see that the class of two-sided moving average processes coincides with the class of stationary random processes having a spectral density.[122]

Now we turn to *one-sided moving average processes*, i.e., to stationary processes admitting the representation (4.396), where $W(t)$ is a Wiener process. According to the foregoing, the stationary process $X(t)$ is representable in the form (4.396) if, and only if, the following two conditions are met: (i) the process $X(t)$ has a spectral density $f(\omega)$ (otherwise $X(t)$ could not even be represented as (4.395)), and (ii) there is, among the functions $g(\omega)$ satisfying (4.400), a function $g_0(\omega)$ such that the Fourier transform of $g_0(\omega)$ vanishes for $t \leqslant 0$. The

problem of finding the conditions (imposed on $f(\omega)$) which guarantee the existence of such a function $g_0(\omega)$ is a challenging problem of Fourier analysis, which was solved by Paley and Wiener.[123] According to the result of these two authors, the condition (ii) is fulfilled if, and only if, the spectral density $f(\omega)$ satisfies the condition

$$(4.401) \quad \int_{-\infty}^{\infty} \frac{|\log f(\omega)|}{1 + \omega^2} \, d\omega < \infty.$$

Thus, the class of stationary random processes $X(t)$ representable as a one-sided moving average (4.396) coincides with the class of processes which have a spectral density $f(\omega)$ satisfying the condition (4.401).[124]

It is clear that condition (4.401) is not fulfilled if the spectral density $f(\omega)$ vanishes (i.e., $\log f(\omega)$ becomes infinite) on some subinterval (of any length) of the frequency axis $-\infty < \omega < \infty$. (The same is true, of course, if the density $f(\omega)$ vanishes on an arbitrary set of positive measure.) Condition (4.401) is not fulfilled also in the cases where the density $f(\omega)$ vanishes at just a single point $\omega = \omega_0$ but "sticks very close" to the ω-axis at ω_0 (e.g., is proportional to $\exp\{-(\omega - \omega_0)^{-2}\}$ in the vicinity of $\omega = \omega_0$). The point ω_0 can, of course, also be infinite; e.g., in the case of the "Gaussian correlation function" (2.126), where the spectral density is proportional to $\exp\{-\omega^2/4\alpha\}$, condition (4.401) is not fulfilled since $f(\omega)$ falls off too rapidly as $\omega \to \infty$. However, if the spectral density $f(\omega)$ is everywhere positive (or has only zeros of finite algebraic orders at some isolated points) and $f(\omega)$ tends to zero not too rapidly at infinity (e.g., only as some negative power of $|\omega|$), then condition (4.401) is clearly fulfilled. Consequently, in all these cases the process $X(t)$ is a one-sided moving average process.

Example. Let $B(\tau) = C\exp(-\alpha|\tau|)$ and hence $f(\omega) = A/(\omega^2 + \alpha^2)$, $A = C\alpha/\pi$ (see Sec. 10, Example 1). For the function $g(\omega)$ satisfying (4.400) it is most natural to choose the function

$$(4.402) \quad g(\omega) = + \sqrt{f(\omega)} = A^{1/2}/\sqrt{\omega^2 + \alpha^2}.$$

It is well known that the Fourier transform of the function $(\omega^2 + \alpha^2)^{-1/2}$ is equal to $(2/\pi)^{1/2}K_0(\alpha t)$, where $K_0(x)$ is the modified Bessel function of the third kind and zero order (see, e.g., Gradshteyn and Ryzhik, 1980, Eq. 3.754.2). Therefore the following moving average representation of the process $X(t)$

corresponds to such a choice of $g(\omega)$,

(4.403) $\quad X(t) = A_1 \int_{-\infty}^{\infty} K_0(\alpha\tau)dW(t - \tau), \quad A_1 = (2C\alpha)^{1/2}/\pi.$ [125]

If, however, we use, instead of the function (4.402), the function

(4.404) $\quad g_0(\omega) = \dfrac{A^{1/2}}{i\omega + \alpha}$

which also satisfies (4.400), then the Fourier transform of $g_0(\omega)$ has the form

(4.405) $\quad b(t) = \dfrac{A^{1/2}}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} \dfrac{e^{it\omega}d\omega}{i\omega + \alpha} = \begin{cases} (2\pi A)^{1/2}e^{-\alpha t} & \text{for } t > 0, \\ \\ 0 & \text{for } t < 0. \end{cases}$

Hence

$$X(t) = (2C\alpha)^{1/2}\int_0^{\infty} e^{-\alpha\tau}dW(t - \tau) = (2C\alpha)^{1/2}\int_{-\infty}^{t} e^{-\alpha(t-s)}dW(s).$$

This one-sided moving average representation of the process under consideration has already appeared before (see (2.194a) on p. 172).

It is easy to understand why the choice (4.404) of the function $g_0(\omega)$ leads to one-sided moving average representation of $X(t)$. Let us continue the function (4.404) to the complex plane (i.e., regard ω as a complex variable). Then this function will have a single pole at the point $\omega = i\alpha$, but it will be an analytic function of ω in the lower half-plane. By virtue of Cauchy's theorem, this implies that the Fourier transform $b(t)$ of the function $g_0(\omega)$ vanishes for $t < 0$.[126] In the general case of an arbitrary spectral density $f(\omega)$ satisfying (2.401) the determination of the function $g_0(\omega)$ such that $|g_0(\omega)|^2 = f(\omega)$ and the Fourier transform of $g_0(\omega)$ vanishes on the negative half-axis also reduces, in fact, to finding a function with a given absolute value $|g_0(\omega)| = [f(\omega)]^{1/2}$ which can be analytically continued to the lower half-plane. Such a function can be found most easily in the case of a rational spectral density of the form (2.129). This circumstance implies a special simplicity of the linear extrapolation problem (i.e., of finding the best prediction of a future value of the process from its past values) for stationary processes $X(t)$ with rational spectral densities.[127]

26.3. Oscillatory Processes and Evolutionary Spectra

We know that if the correlation function $B(t,s)$ of a random process $X(t)$ can be represented in the form (4.382), then there exists the generalized spectral representation of $X(t)$ of the form (4.380).   It is clear, however, that the given correlation function $B(t,s)$ very often admits a multitude of different representations of the form (4.382).   (E.g., if $-\infty < a \leqslant t \leqslant b < \infty$, then there exist many different representations of the function $B(t,s)$ in the form of the sum (4.386), which is a very particular case of (4.382); see Note 116 and cf. also Note 129.) Thus, the given random process $X(t)$ often has a lot of widely differing generalized spectral representations.   Therefore, it is not surprising that not every generalized spectral representation is interesting for applications, but only special classes of such representations possessing some useful property or other.

We have considered above two special classes of generalized spectral representations: Karhunen–Loève expansions of mean square continuous processes $X(t)$ on a finite interval, and moving average representations of stationary processes $X(t)$. We now turn to one more such special class, namely the class of so-called *evolutionary spectral representations*. For these new generalized spectral representations, the parametric set A coincides with the frequency axis $-\infty < \omega < \infty$, but they differ from ordinary spectral representations (2.61) of stationary processes $X(t)$ in that now both the spectral representation itself and the corresponding spectral distribution function (or spectral density) can change slowly over time. Correspondingly, a random processes $X(t)$ which has an evolutionary spectral representation is, in general, weakly nonstationary.

Evolutionary spectral representations were first introduced by Granger and Hatanaka and (in more detail) by Priestley in the mid-1960s.[128] These representations have the form

$$(4.406) \quad X(t) = \int_{-\infty}^{\infty} e^{it\omega} A(t,\omega) dZ(\omega)$$

where $Z(\omega)$ is a random function with uncorrelated increments, which satisfies the ordinary relation (2.78):

$$\langle dZ(\omega)\overline{dZ(\omega')}\rangle = \delta(\omega - \omega')dF(\omega)d\omega',$$

and $A(t,\omega)$ is a "slowly varying" function of t satisfying the condition

(4.407) $\int_{-\infty}^{\infty}|A(t,\omega)|^2 dF(\omega) < \infty$

for all t. According to Priestley, strict restrictions of the "permissible variability" of the function $A(t,\omega)$ of t can be formulated as follows: $A(t,\omega)$ *must admit the generalized Fourier representation*

(4.408) $A(t,\omega) = \int_{-\infty}^{\infty} e^{it\theta} dK(\theta;\omega)$

(where $dK(\theta;\omega)$ is the differential with respect to θ) *with* $|dK(\theta;\omega)|$ *having an absolute maximum at* $\theta = 0$ *for any fixed* ω. The correlation function $B(t,s)$ of the process $X(t)$, which has the evolutionary spectral representation (4.406), is clearly of the form

(4.409) $B(t,s) = \int_{-\infty}^{\infty} e^{i(t-s)\omega} A_t(\omega)\overline{A_s(\omega)} dF(\omega)$

which is a special case of (4.382). In particular,

(4.410) $\langle|X(t)|^2\rangle = B(t,t) = \int_{-\infty}^{\infty}|A_t(\omega)|^2 dF(\omega)$

so that condition (4.407) is necessary for $\langle|X(t)|^2\rangle$ to be finite at all t, i.e. for the existence of a correlation function $B(t,s)$. The representation (4.406) may be interpreted as representation of the process $X(t)$ in the form of a superposition of sinusoidal oscillations with different frequencies ω and random amplitudes $A_t(\omega)dZ(\omega)$ varying over time (i.e., $X(t)$ is a superposition of *amplitude modulated random oscillations*). According to this interpretation, (4.410) describes the distribution of the "total power" $\langle|X(t)|^2\rangle$ of the process $X(t)$ at time t over the frequencies ω; hence, the contribution from the frequency ω is $|A_t(\omega)|^2 dF(\omega)$. Therefore the function $F_t(\omega)$ defined by the relation

(4.411) $dF_t(\omega) = |A_t(\omega)|^2 dF(\omega)$

is called the *evolutionary spectral distribution function* of the process $X(t)$. If $dF(\omega) = f(\omega)d\omega$ (i.e., if $F(\omega)$ is an absolutely continuous function), then $|A_t(\omega)|^2 f(\omega)$ is said to be the *evolutionary spectral density* of $X(t)$. (Note that, in general, the given random process $X(t)$ may have a number of different evolutionary spectral representations (4.406)).   Therefore both the evolutionary spectral distribution function and the

spectral density depend on the specific choice of the family of functions $A_t(\omega)$ appearing on the right-hand side of (4.406).[129])   A random process $X(t)$ having at least one representation of the form (4.406), where the functions $A_t(\omega)$ and the random measure $dZ(\omega)$ satisfy the above-indicated conditions, is called an *oscillatory process*. Since the functions $A(t,\omega) \equiv 1$ clearly satisfy the conditions to be imposed on $A(t,\omega)$, *the class of oscillatory processes certainly includes all the stationary processes having a finite variance.*[130]

Note that any generalized spectral representation (4.380), where the parametric set A may be taken to be a straight line (or a part of a straight line), may be written as (4.406). In fact, in such a case the set A may be considered to be the real line $-\infty < \omega < \infty$ (or a part of it), while $\varphi(t,a) = \varphi(t,\omega)$ may be written as $\exp(it\omega)A(t,\omega)$, where $A(t,\omega) = \exp(-it\omega)\varphi(t,\omega)$. Of course, it would not be physically meaningful to interpret ω as the frequency in all cases, and to consider the product $A(t,\omega)\exp(it\omega)$ as an amplitude modulated sinusoidal oscillation. In the physical theory of oscillations the function $A(t)\exp(it\omega_0)$ is said to describe the amplitude modulated oscillation of frequency ω_0 only if the "amplitude" $A(t)$ is a slowly varying (compared to $\exp(it\omega_0)$) function, i.e., if the Fourier transform of $A(t)$ includes mainly frequencies much lower than ω_0; it is even often assumed that this transform must be concentrated in a neighborhood of zero frequency. This is the reason for requiring that $|dK(\Theta,\omega)|$ be maximal at $\Theta = 0$ for any fixed ω.

Since $A(t,\omega)$ is a slowly varying function of time, it is clear that the process (4.406) may be regarded as being "approximately stationary" throughout rather long time intervals. If, however, we examine the behavior of $X(t)$ during two such intervals which are sufficiently far apart, we shall find that although $X(t)$ is practically stationary in both intervals, the spectral distribution function of the two "portions" of $X(t)$ will, in general, be different (i.e., the spectral distribution of the power of $X(t)$ varies slowly over time).[131]

Example. *Modulated Stationary Processes.* A class of nonstationary random processes which are often encountered in applications is the class of *modulated* (or *weighted*) *stationary processes* X(t). These processes are obtained when a stationary random process $X_0(t)$ is multiplied by some nonrandom modulating (or weighting) function $A(t)$:

(4.412) $X(t) = A(t)X_0(t)$.

This class of random processes has been studied by a number of authors.[132] It is clear that the *random process* (4.412) *is oscillatory if the function* A(t) (which can always be so normalized that $A(0) = 1$) *has a generalized Fourier transform* (i.e. admits the representation similar to (4.408)) *and the modulus of this transform has an absolute maximum at zero.* (In particular, $A(t)$ may be any nonnegative function whose Fourier transform exists.)   Process $X(t)$ can clearly be represented as

(4.413) $X(t) = \int_{-\infty}^{\infty} e^{i\omega t} A(t) dZ_0(\omega)$,

where $Z_0(\omega)$ is a random function with uncorrelated increments, which appears in the spectral representation of the stationary process $X_0(t)$.   Therefore, if $\langle |dZ_0(\omega)|^2 \rangle = dF_0(\omega)$, then the evolutionary spectral distribution function of the process $X(t)$ is given by

(4.414) $dF(\omega) = A(t)dF_0(\omega)$.

We see that in this case the intensities of different frequency components of $X(t)$ are varying over time in exactly the same way, so that the shape of the frequency distribution of the average power of the process does not change at all.   The oscillatory   processes $X(t)$ possessing   this   property   are sometimes called *uniformly modulated random processes.*[133]

The class of oscillatory random processes encompasses many processes dealt with by workers in various applied fields (e.g., by engineers, applied physicists, or economists).   A very important advantage of this class is that many widely used results of the theory of stationary random processes (in particular,   results   relating   to   the   effect   of   linear transformations,   the   theory   of   linear   extrapolation   and filtering,   and   spectral   analysis,   i.e.   methods   for   the determination of spectral densities and distribution functions from   experimental   data)   can   readily   be   generalized   to oscillatory processes.   We, however, cannot examine all these topics in this book.[134]   Nor shall we consider a number of attempts, differing from the present one, at generalization of the concept of power spectrum (more precisely, spectral density

and spectral distribution function) to broad classes of nonstationary random processes[135] except for one more approach to this problem yet to be discussed at the end of this section.

26.4. Harmonizable Random Processes

In accordance with the commonly accepted terminology, in Chaps. 2 and 3 the spectral representation of a random process $X(t)$ was always interpreted as its representation in the form of superposition of complex exponential functions $\exp(i\omega t)$ (i.e. sine and cosine waves of different frequencies ω). However, in the present chapter, a number of examples of generalized spectral representations were considered which relate to the decomposition of the random process $X(t)$ in terms of some other functions $\varphi(t,a)$ depending on the argument t and an additional parameter a. Now we revert to the ordinary interpretation of the spectral representation as a decomposition in terms of complex exponential functions, but abandon the requirement adopted in Chaps. 2 and 3 that the function $Z(\omega)$ in the integrand of (2.61) necessarily has uncorrelated increments. In this case the process $X(t)$ is in general not stationary, and hence now we consider the Fourier–Stieltjes representations of nonstationary random processes.

Thus, assume that the random process $X(t)$ is representable as a Fourier–Stieltjes integral

$$(4.415) \quad X(t) = \int_{-\infty}^{\infty} e^{i\omega t} dZ(\omega)$$

where $Z(\omega)$ is a random function of ω. As before, we interpret the integral on the right-hand side of (4.415) as the mean square limit of the corresponding integral sums, i.e. we define this integral by (2.62). It is easy to see that then the correlation function $B(t,s) = \langle X(t)\overline{X(s)} \rangle$ is given by the formula

$$(4.416) \quad B(t,s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(t\omega - s\omega')} d^2 F(\omega,\omega')$$

where

$$(4.417) \quad F(\omega,\omega') = \langle Z(\omega)\overline{Z(\omega')} \rangle.$$

It is clear that if $F(\omega,\omega')$ is a function of bounded variation

in the plane, i.e.[136]

$$(4.418) \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|d^2F(\omega,\omega')| = V_F < \infty,$$

then the integral (4.416) is necessarily convergent. It readily follows from this that the Fourier–Stieltjes integral (4.415) (i.e., the double limit (2.62)) also exists in this case and determines the mean square continuous random process $X(t)$.

The random processes $X(t)$ representable as the Fourier–Stieltjes integral (4.115), where the random function $Z(\omega)$ has the correlation function (4.417) of bounded variation, were introduced by Loève in the mid-1940s.[137] Such processes are called *harmonizable random processes*. Clearly, in the special case where the complex measure in the plane

$$(4.419) \quad F(\Delta,\Delta') = \int_{\Delta}\int_{\Delta'} d^2F(\omega,\omega')$$

corresponding to the function $F(\omega,\omega')$ is concentrated on the diagonal $\omega = \omega'$, formula (4.416) becomes the ordinary formula (2.52) for the correlation function of a stationary process. Therefore, the representation (4.415) coincides in this case with the ordinary spectral representation (2.61). Thus, any mean square continuous stationary process is necessarily harmonizable, so that the *concept of a harmonizable random process generalizes the concept of a stationary process*. Note also that condition (4.418) imposed on the correlation function $F(\omega,\omega')$ of the random function $Z(\omega)$ can in fact be considerably weakened by using the more sophisticated definition of the integral in the right-hand side of (4.415). In other words, the harmonizability concept may be made more inclusive by introducing a refined definition of the Fourier–Stieltjes integral (4.415).[138] However, we will not dwell on the possible generalization of the harmonizability concept here, but will, for simplicity, restrict ourselves to the case where condition (4.418) is satisfied.

It has already been noted above that the correlation function $B(t,s)$ of a harmonizable process $X(t)$ is always representable as the double Fourier–Stieltjes integral (4.416), where $F(\omega,\omega')$ is a complex function of bounded variation. It is not hard to show that the converse statement is also true: *each random process having a correlation function of the form* (4.416), *where* $F(\omega,\omega')$ *is a function of bounded variation, is harmonizable*, i.e. such a process is representable as the Fourier–Stieltjes integral (4.415), where

$Z(\omega)$ is a random function whose correlation function coincides with $F(\omega,\omega')$. The proof of this last statement is quite similar to the proof of the theorem on the spectral representation of stationary random processes which is based on the use of Khinchin's formula (2.52) for the correlation function $B(\tau) =$

$\langle X(t + \tau)\overline{X(t)}\rangle$.[139]

The complex function of bounded variation $F(\omega,\omega')$ is called the (two-dimensional) *spectral distribution function* of the harmonizable random process $X(t)$, while the complex measure in the plane $F(\Delta,\Delta')$ corresponding to the function $F(\omega,\omega')$ (see (4.419)) is said to be the *spectral measure* of the process $X(t)$. In the particular case where

$$(4.420)\quad F(\omega,\omega') = \int_{-\infty}^{\omega}\int_{-\infty}^{\omega'} f(\lambda,\lambda')d\lambda d\lambda', \quad f(\omega,\omega') = \frac{\partial^2 F(\omega,\omega')}{\partial\omega\partial\omega'}$$

(i.e., $F(\omega,\omega')$ is an absolutely continuous function of two variables) the function $f(\omega,\omega')$ is called the (two-dimensional) *spectral density* of the process $X(t)$. According to (4.416), if the spectral distribution function $F(\omega,\omega')$ is given, the correlation function $B(t,s)$ of the harmonizable random process $X(t)$ can be determined uniquely. It is not hard to show that, conversely, specification of the correlation function $B(t,s)$ of a harmonizable process $X(t)$ determines the corresponding spectral distribution function $F(\omega,\omega')$ to within an arbitrary additive constant and an exact definition of the values of $F(\omega,\omega')$ at the points of discontinuity of this function, which do not affect the value of the integral on the right-hand side of (4.416).[140] Indeed, it is easy to show that the values of the two-dimensional increments $\Delta\Delta'F(\omega,\omega') = F(\omega+\Delta, \omega'+\Delta') - F(\omega+\Delta, \omega') - F(\omega, \omega'+\Delta') + F(\omega,\omega')$ where the points $(\omega + \Delta, \omega' + \Delta')$, $(\omega + \Delta, \omega')$, $(\omega, \omega' + \Delta')$ and (ω,ω') are any four continuity points of the function $F(\omega,\omega')$, can be determined uniquely from $B(t,s)$ with the aid of a simple inversion formula. This inversion formula coincides, of course, with the inversion formula for two-dimensional characteristic functions and for spectral distribution functions of homogeneous random fields in the plane (see p. 332 and Note 8 to the Introduction in Vol. II).[141]

The ordinary formula for inversion of one-dimensional Fourier–Stieltjes integrals (see, e.g., formulae (0.6') and (2.9') in Vol. II) may also be applied to the random process (4.415) themselves. According to this formula

$$(4.421) \quad Z(\omega_2) - Z(\omega_1) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \frac{e^{-i\omega_2 t} - e^{-i\omega_1 t}}{- it} X(t)dt$$

where lim denote the mean square limit, and if ω_2 and/or ω_1 are discontinuity points of $Z(\omega)$, then $Z(\omega_2)$ and/or $Z(\omega_1)$ must be interpreted as $[Z(\omega_2 + 0) + Z(\omega_2 - 0)]/2$ and/or $[Z(\omega_1 + 0) + Z(\omega_1 - 0)]/2$. Moreover, we can easily show, as in the derivation of formula (2.32') for a stationary random process $X(t)$ (see Note 22 to Chap. 2 in Vol. II), that for any harmonizable process the following relation is valid:

$$(4.422) \quad \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} e^{-i\omega t} X(t)dt = Z(\omega + 0) - Z(\omega - 0)$$

which generalizes (2.32'). If $\omega = 0$, then (4.422) coincides with (2.32') and can be interpreted as the *law of large numbers* (i.e., the law which was called the *ergodic theorem* in the case of stationary processes $X(t)$; see Sec. 16) for harmonizable random processes $X(t)$. In fact, it is easy to see that (4.422) for $\omega = 0$ implies that $\lim_{T \to \infty} \int_{-T/2}^{T/2} X(t)dt$ *does exist* (as a limit in the mean) *for any harmonizable process* X(t) *and that*

$$(4.423) \quad \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} X(t)dt = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} m(t)dt$$

where $m(t) = \langle X(t) \rangle$ *if, and only if, the point* $\omega = 0$, $\omega' = 0$ *provides a zero contribution to the spectral measure* $\hat{F}(\Delta,\Delta')$ *of the centered process* $\overset{\circ}{X}(t) = X(t) - \langle X(t) \rangle$. (In the particular case of a stationary process $X(t)$ this result coincides with Slutsky's theorem; see p. 220.)[142]

We now proceed to a more detailed study of the class of harmonizable random processes. Since the function $\exp\{i(t\omega - s\omega')\}$ in the integrand of (4.416) is a continuous function of t and s which does not exceed 1 in absolute value, it follows from (4.418) that the correlation function $B(t,s)$ of a harmonizable random process is a bounded (by the constant V_F) continuous function of two variables t and s. (Hence any harmonizable process $X(t)$ is mean square continuous.[143]) Thus, the class of correlation functions $B(t,s)$ of harmonizable random processes (these functions are often called *harmonizable correlation functions*) is a subclass of the class of all bounded and continuous (in both variables) positive definite kernels in the plane.

It is clear that this subclass is very wide. Among other things, it evidently includes all positive definite kernels $B(t,s)$ which can be represented in the form

$$B(t,s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(t\omega - s\omega')} f(\omega,\omega')d\omega d\omega'$$

(and, in particular, all positive definite kernels satisfying the condition

$$(4.424) \qquad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |B(t,s)|dtds < \infty;$$

cf. (2.66)), as well as all positive definite kernels depending only on the difference of the two arguments $t - s$ (in the last case $B(t,s)$ is a stationary correlation function). The available mathematical literature contains several different formulations of the necessary and sufficient conditions for the given function (of one or, sometimes, also of two variables) to be representable as a Fourier–Stieltjes integral, but all these conditions are rather complicated and verifiable only in some exceptional cases.[144] It seems very probable, however, that the requirement of the representability of a function $B(t,s)$ as a Fourier–Stieltjes integral is not a severe restriction and that any bounded and continuous positive definite kernel satisfying some mild regularity conditions may be represented as such an integral. We note in this connection that some time ago it was even suggested by some authors that any bounded and continuous positive definite kernel in the plane may be represented as (4.416), i.e. it is a harmonizable correlation function. Later on, however, some specific examples of bounded and continuous correlation functions $B(t,s)$ were given, which are not harmonizable (but all these functions are rather complicated and have some unusual, even pathological, properties).[145]

Let us now consider some particular examples of harmonizable random processes.

Example 1. *Modulated Stationary Processes.* Let

$$(4.425) \qquad X(t) = A(t)X_0(t)$$

where $X_0(t)$ is a stationary random process having the ordinary spectral representation (2.61), while $A(t)$ is a nonrandom function of the form

(4.426) $A(t) = \int_{-\infty}^{\infty} e^{it\theta}\, dK(\Theta)$

where $K(\Theta)$ is a complex function of bounded variation. It is easy to see that in this case the random process $X(t)$ is representable as the Fourier–Stieltjes integral (4.415), where

(4.427) $Z(\omega) = \int_{-\infty}^{\infty} K(\omega - v)\, dZ_0(v)$

and $Z_0(v)$ is the random function with uncorrelated increments appearing in the spectral representation of the stationary process $X_0(t)$. According to (4.227),

(4.428) $\langle Z(\omega)\overline{Z(\omega')}\rangle = F(\omega,\omega') = \int_{-\infty}^{\infty} K(\omega - v)\overline{K(\omega' - v)}\, dF_0(v)$

where $F_0(v)$ is the spectral distribution function of the stationary process $X_0(t)$. The function (4.428) clearly is a function of bounded variation in the plane; hence, the process (4.425) is harmonizable.

Example 2.   *Output Processes of Linear Systems.*   Let $\mathfrak{L}$ be a time-invariant linear system specified by its impulse response function $h(u)$ or by its transfer function $H(\omega)$. (The function $H(\omega)$ is the Fourier transform of $h(u)$; see (2.153)). Assume that a harmonizable random process $X(t)$ having a correlation function (4.416) is applied to the input of the system $\mathfrak{L}$. Then the output process $Y(t)$ is given by (2.147a) and has the correlation function

$B_y(t,s) = \langle Y(t)\overline{Y(s)}\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(u)\overline{h(v)}B(t - u,\ s - v)\,du\,dv$

(4.429) $= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(u)\overline{h(v)}\,e^{i\{(t-u)\omega-(s-v)\omega'\}}\,d^2F(\omega,\omega')\,du\,dv$

$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(t\omega - s\omega')}H(\omega)\overline{H(\omega')}\,d^2F(\omega,\omega').$

Thus, if

(4.430) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|H(\omega)||H(\omega')||d^2F(\omega,\omega')| < \infty,$

then the output process $Y(t)$ is also harmonizable and its spectral distribution function $F_y(\omega,\omega')$ is related to the spectral distribution function $F(\omega,\omega')$ of the input process by

the simple relation

(4.431) $d^2F_y(\omega,\omega') = H(\omega)\overline{H(\omega')}d^2F(\omega,\omega')$

which generalizes (2.164). It is also easy to see that the spectral representation of the output process $Y(t) = \mathfrak{L}\{X(t)\}$ has the form

(4.432) $Y(t) = \int_{-\infty}^{\infty}e^{it\omega}H(\omega)dZ(\omega)$

where $Z(\omega)$ is the random function appearing in the spectral representation (4.415) of the input process. Equation (4.432) clearly coincides with (2.162); hence, the spectral theory of harmonizable processes preserves the simplicity of description of linear time-invariant transformation, which is one of the main features of the spectral theory of stationary random processes making this theory so useful for practical applications.[146] Note that condition (4.430) will clearly be valid for any function $F(\omega,\omega')$ of bounded variation in the plane if the absolute value of the function $H(\omega)$ is uniformly bounded (i.e., $|H(\omega)| < C < \infty$ for all ω). Hence, the application of any harmonizable random process $X(t)$ to the input of a linear system $\mathfrak{L}$ with the bounded transfer function always produces a harmonizable output process $Y(t)$.

Since harmonizable processes are, generally speaking, nonstationary, it is clear that it is no longer reasonable to restrict oneself to time-invariant systems $\mathfrak{L}$. Let $\mathfrak{L}$ be a time-varying linear system described by the equation

(4.433) $\mathfrak{L}X(t) = \int_{-\infty}^{\infty}h(t,u)X(t - u)du$

where $h(t,u)$ is a time-varying impulse response function having the Fourier transform

(4.434) $\int_{-\infty}^{\infty}h(t,u)e^{-i\omega u}du = H(t,\omega)$.

Consider the case where $X(t)$ is a harmonizable process having the spectral distribution function $F(\omega,\omega')$ that satisfies the condition

(4.435) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|H(t,\omega)||H(t,\omega')||d^2F(\omega,\omega')| < \infty$

for all t. (This condition obviously guarantees that the output

mean square $\langle |Y(t)|^2 \rangle$ is finite for all t.) Then it is easy to see that the output process $Y(t)$ = $\mathfrak{L}\{Y(t)\}$ is also necessarily harmonizable.[147]

Moreover, it is even unnecessary to require that the input process $X(t)$ be harmonizable. Actually, it is possible to prove that if $X(t)$ is a random process having a bounded mean square and the output process $Y(t)$ = $\mathfrak{L}\{X(t)\}$ also has a bounded mean square, then, under some wide conditions (imposed on the impulse response function $h(t,u)$ and the correlation function $B(t,s)$ of the process $X(t)$) the process $Y(t)$ will be harmonizable even if the input process $X(t)$ is nonharmonizable.[148] This result shows that the harmonizable processes are the most general processes that need be considered in the analysis of a wide class of linear systems $\mathfrak{L}$.

Example 3. *Oscillatory Processes.* An oscillatory process $X(t)$ is given by (4.406) − (4.408), where $Z(\omega)$ satisfies (2.78), while $K(\Theta;\omega)$ as a function of Θ has bounded variation. (The additional condition which was imposed on p. 457 on $K(\Theta;\omega)$ may now be neglected.) The correlation function of such a process $X(t)$ clearly has the form

$$(4.436)\; B(t,s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i\{t(\omega+\theta)-s(\omega+\theta')\}} dK(\Theta;\omega)d\overline{K(\Theta';\omega)}dF(\omega)$$

It is easy to see that if

$$(4.437)\quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |dK(\Theta;\omega)||dK(\Theta';\omega)||dF(\omega)| < \infty,$$

then the function (4.436) is necessarily a harmonizable correlation function.[149]

Example 4. *Generalized Moving Average Processes.* The moving average process has been defined above in this section as a process $X(t)$ which admits a representation of the form

$$(4.438)\quad X(t) = \int_{-\infty}^{\infty} A(t - s)dZ(s)$$

where the real random function $Z(s)$ satisfies the condition $\langle dZ(s)dZ(s')\rangle = \delta(s - s')dsds'$; see (4.394) and (4.395). Let us now consider a much wider class of processes $X(t)$ representable as (4.438), where $Z(s)$ is an arbitrary complex random function having the correlation function $F(s,s') = \langle Z(s)\overline{Z(s')}\rangle$ of

bounded variation in the plane. It is not hard to see that if the function $A(t)$ is the Fourier–Stieltjes integral (4.426) with respect to some function $K(\Theta)$ of bounded variation, then the process (4.438) is necessarily harmonizable.[150]

Example 5. *Amplitude Modulated Harmonic Oscillation.* Consider a random process of the form

(4.439) $X(t) = X_0(t)\cos\omega_0 t,$

where $X_0(t)$ is a stationary random process and ω_0 is a fixed frequency. The process (4.439), which is a special case of (4.425), is clearly nonstationary (since it vanishes at fixed points $t = (2k+1)\pi/2, k = 0, \pm 1, ...$). It is harmonizable, however:

$$X(t) = \frac{1}{2}\left[e^{i\omega_0 t} + e^{-i\omega_0 t}\right]\int_{-\infty}^{\infty}e^{i\omega t}dZ_0(\omega)$$

(4.440)
$$= \frac{1}{2}\int_{-\infty}^{\infty}\left\{e^{(i\omega+\omega_0)t} + e^{i(\omega-\omega_0)t}\right\}dZ_0(\omega)$$

$$= \frac{1}{2}\int_{-\infty}^{\infty}e^{i\omega t}\left\{dZ_0(\omega-\omega_0) + dZ_0(\omega+\omega_0)\right\} = \int_{-\infty}^{\infty}e^{i\omega t}dZ(\omega)$$

where

(4.441) $Z(\omega) = \dfrac{1}{2}\{Z_0(\omega - \omega_0) + Z_0(\omega + \omega_0)\}.$

The last equation is, of course, a special case of (4.427). Since $\langle dZ_0(\omega)\overline{dZ_0(\omega')}\rangle = \delta(\omega - \omega')dF_0(\omega)d\omega'$, we have

$$\langle dZ(\omega)\overline{dZ(\omega')}\rangle = \frac{1}{4}\Big\{\delta(\omega - \omega')\left[dF_0(\omega - \omega_0) + dF_0(\omega + \omega_0)\right]d\omega'$$

(4.442)
$$+ \delta(\omega - \omega' - 2\omega_0)dF_0(\omega - \omega_0)d\omega'$$

$$+ \delta(\omega - \omega' + 2\omega_0)dF_0(\omega + \omega_0)d\omega'\Big\}.$$

Thus, the spectral measure $F(d\omega, d\omega') = \langle dZ(\omega)\overline{dZ(\omega')}\rangle$ is concentrated on three parallel lines: $\omega = \omega'$, $\omega = \omega' + 2\omega_0$ and $\omega = \omega' - 2\omega_0$. Such a particular spectral distribution of the process (4.439) is related to its special property which will be considered in the next subsection.

The available literature also contains many other examples of harmonizable random processes $X(t)$ and harmonizable correlation functions $B(t,s)$.[151] We will not, however, pursue this subject any further.

26.5. Periodically Correlated Processes

Consider now the class of random random processes $X(t)$ with finite second moments, whose mean values $m(t) = \langle X(t) \rangle$ and correlation functions $B(t,s) = \langle X(t)\overline{X(s)} \rangle$ have the properties that

(4.443) $m(t + T_0) = m(t)$, $B(t + T_0,\ s + T_0) = B(s,t)$

for some fixed T_0 and every t and s. In this case

(4.443a) $m(t + kT_0) = m(t)$, $B(t + kT_0,\ s + kT_0) = B(t,s)$

for any integer k, i.e. the functions $m(t)$ and $B(t,s)$ exhibit periodicity with period T_0. Random processes having the stated property are often said to be *periodically correlated* (with period T_0). There are also some other terms used in the available literature to denote the same class of random processes.[152]

Periodically correlated processes are very common in radiophysics and radioengineering, meteorology and oceanography, astronomy, biology, economics, and many other applied sciences.[153] They are covered by an extensive literature, which includes much material supplementing the results presented below.[154]

It is clear that any stationary random process $X(t)$ is a stationary correlated process with period T_0 for any T_0. Thus, the *class of periodically correlated random processes includes all stationary processes*. Note also that if $X(t)$ is a periodically correlated random process with period T_0, then the process

(4.444) $X^{(0)}(t) = X(t + \Phi)$

is stationary if Φ is a random variable which is independent of $X(t)$, takes on the values in the interval $0 \leqslant \Phi \leqslant T_0$, and is uniformly distributed over this interval. In fact, it is easy to

see that

$$(4.445) \quad \langle X^{(0)}(t) \rangle = m_0(t) = \frac{1}{T_0} \int_0^{T_0} m(t + \varphi)d\varphi,$$

$$\langle X^{(0)}(t)X^{(0)}(s) \rangle = B_0(t,s) = \frac{1}{T_0} \int_0^{T_0} B(t + \varphi, \ s + \varphi)d\varphi \ .$$

Hence by virtue of (4.443) $m_0(t)$ is a constant, while $B_0(t,s)$ depends only on $t - s$. (See in this connection Note 7 to Chap. 1 in Vol. II). The process (4.444) is called the *stationarizable version* of the periodically correlated process $X(t)$.

Some useful models of periodically correlated processes are listed below.

Example 1. *Amplitude Modulated Periodic Oscillation.* Let $P(t)$ be an arbitrary bounded periodic function with period T_0 and $X_0(t)$ be a stationary random process with mean value $\langle X_0(t) \rangle = m$ and correlation function $\langle X_0(t)X_0(s) \rangle = B(t-s)$. If

$$(4.446) \quad X(t) = P(t)X_0 (t),$$

then obviously

$$(4.447) \quad m(t) = \langle X(t) \rangle = mP(t),$$

$$B(t,s) = \langle X(t)\overline{X(s)} \rangle = B(t - s)P(t)P(s).$$

It follows from (4.447) that the process (4.446) is periodically correlated with period T_0.

In the particular case where $P(t)$ is a sine wave, $X(t)$ is an amplitude modulated harmonic oscillation which has already been discussed (see (4.439)).

Example 2. *Process with Fixed Equidistant Discontinuity Points and Adjoined Random Variables.* Assume that the random process $X(t)$ takes on a constant value X_n for $nT_0 \leqslant t < (n + 1)T_0$ and that X_n, $n = 0,\pm1, \ldots$ is a sequence of independent identically distributed real random variables with mean value $\langle X_n \rangle = m_0$ and mean square $\langle X_n^2 \rangle = m_0^2 + \sigma_0^2$ (cf. the definition of the Poisson point process with adjoined random variables on pp. 32-33). In this case the one-dimensional probability distributions of $X(t)$ for all t are clearly the same. However, the process $X(t)$

is nevertheless nonstationary; it has fixed discontinuity points at $t = nT_0$, where n are integers, and its two-dimensional probability distributions $F_{t_1,t_2}(x_1,x_2) = P\{X(t_1)<x_1, \ X(t_2)<x_2\}$

depend on whether or not the interval $[t_1,t_2]$ includes at least one point of the form nT_0. It is easy to see that if $t = kT_0 + t'$, $s = k_1T_0 + s'$, where k and k_1 are integers, $0 \leqslant t' < T_0$, $0 \leqslant s' < T_0$, then $m(t) = \langle X(t) \rangle$ and $B(t,s) = \langle X(t)\,X(s) \rangle$ are given by

$$(4.448) \quad m(t)= m_0, \ B(t,s) = \begin{cases} m_0^2 + \sigma_2^2 & \text{for } k = k_1, \\ \\ m_0^2 & \text{for } k \neq k_1. \end{cases}$$

The functions $m(t)$ and $B(t,s)$ clearly satisfy (4.443), and hence $X(t)$ is a periodically correlated process.

In the particular case where random variables X_n take on only two values (say 0 and 1 or $-1/2$ and $1/2$), the process $X(t)$ may be regarded as a model of a telegraphic message, which is a sequence of current pulses and pauses (when the current is cut off) or any other binary transmission.[155]

Example 3. *Pulse-Amplitude-Modulated Signal.* Let

$$(4.449) \quad X(t) = \sum_{n=-\infty}^{\infty} X_n \Gamma(t - nT_0)$$

where $\Gamma(t)$ is a fixed real function, while the amplitudes X_n form a sequence of independent and identically distributed real random variables with mean value m_0 and variance σ_0^2. Then

$$(4.450) \quad \begin{aligned} m(t) &= m_0 \sum_{n=-\infty}^{\infty} \Gamma(t - nT_0), \\ B(t,s) &= \sigma_0^2 \sum_{n=-\infty}^{\infty} \Gamma(t - nT_0)\Gamma(s - nT_0) + m(t)m(s). \end{aligned}$$

Equations (4.450) imply that the process (4.449) is periodically correlated. In the particular case where $\Gamma(t) = 1$ for $0 \leqslant t < T_0$ and $\Gamma(t) = 0$ for $t < 0$ or $t \geqslant T_0$, the process (4.449) coincides with that of Example 2.

Note that the sequence of pulse amplitudes $\{X_n\}$, $n = 0,\pm 1$, ..., need not necessarily be a sequence of independent identically distributed random variables; it will suffice if $\{X_n\}$ is an arbitrary stationary random sequence. If this sequence is real and $\langle X_n \rangle = m_0$, $\langle X_{n+k}X_n \rangle = b_0(k) + m_0^2$, then

$$m(t) = m_0 \sum_{n=-\infty}^{\infty} \Gamma(t - nT_0),$$

(4.451)

$$B(t,s) = \sum_{n=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} b_0(k)\Gamma(t - nT_0 - kT_0)\Gamma(s - nT_0) + m(t)m(s).$$

It is easy to show that the functions (4.451) also satisfy conditions (4.443). Moreover, the pulse shape (e.g., its width) may also vary randomly with n, i.e. the following more general form of the pulse-modulated signal may be considered instead of (4.449):

$$(4.452) \quad X(t) = \sum_{n=-\infty}^{\infty} \Gamma(t - nT_0, \mathbf{X}_n),$$

where $\mathbf{X}_n$ is a random variable (maybe multidimensional; e.g., it can include both the amplitude X_n and the width W_n of the nth pulse). It is not hard to show that if the modulating sequence $\{\mathbf{X}_n\}$ is stationary, then the process (4.452) is periodically correlated with period T_0. In fact, even asynchronous pulse sequences that result from including random epoch jitters in the model (4.452), by replacing nT_0 with $nT_0 + \epsilon_n$, where ϵ_n is a random sequence, are periodically correlated if the multidimensional sequence $\{\mathbf{X}_n, \epsilon_n\}$ is stationary.[156]

Many other examples of periodically correlated random processes can also be found in the literature.[157]

Passing over to the general theory of periodically correlated processes, we begin for simplicity with consideration of *periodically correlated sequences* (i.e., processes in discrete time). In other words, we first assume that the time t in (4.443) and (4.443a) takes on only integral values and that T_0 is also an integer. The theory of such periodically correlated sequences can be developed in a very simple way by reducing them to multidimensional (namely, T_0-dimensional) stationary random sequences. Indeed, if $X(t)$ is a periodically correlated sequence with period T_0, then evidently

$$(4.453) \quad \mathbf{Y}(k) = \{X(kT_0), X(kT_0 + 1), ..., X(kT_0 + T_0 - 1)\}, k = 0, \pm 1, ...,$$

is a T_0-dimensional stationary sequence. Conversely, if $\mathbf{Y}(k) = \{Y_0(k), ..., Y_{T_0-1}(k)\}$ is a T-dimensional stationary sequence and $X(kT_0 + s) = Y_s(k)$, $k = 0, \pm 1, ..., s = 0, 1, ..., T_0 - 1$, then $X(t)$ is a periodically correlated sequence with period T_0. It is clear, therefore, that the correlation function $B(t,s)$ of any

periodically correlated sequence with period T_0 can be expressed in terms of the correlation matrix of some T_0-dimensional stationary sequence. It is, however, more convenient to accomplish such a reduction to the multidimensional stationary case without direct utilization of the stationary sequence (4.453), but by applying another method, which may be rather easily developed also for the case of continuous time t.

It is clear that any periodic nonrandom sequence $m(t)$, $t = 0, \pm 1, ...$, with period T_0 may be the mean value sequence of $X(t)$. Since the mean value $m(t) = \langle X(t) \rangle$ can always be subtracted from $X(t)$, one can assume from the outset, without loss of generality, that $\langle X(t) \rangle = 0$. It is also convenient to change the variables by using, instead of $B(t,s) = \langle X(t)X(s) \rangle$, the function

$$(4.454) \quad C(t,\tau) = B(t + \tau, t) = \langle X(t + \tau)\overline{X(t)} \rangle$$

which will also be called the correlation function of $X(t)$. According to (4.443) the function $C(t,\tau)$ of the integral arguments t and τ is periodic in t with period T_0. Therefore $C(t,\tau)$ may be represented as

$$(4.455) \quad C(t,\tau) = \sum_{k=0}^{T_0-1} C_k(\tau)e^{ik\omega_0 t},$$

where $\omega_0 = 2\pi/T_0$ and

$$(4.456) \quad C_k(\tau) = \frac{1}{T} \sum_{s=0}^{T_0-1} C(t,\tau)e^{-ik\omega_0 t}.$$

To determine the class of all correlation functions $B(t,s) = C(s, t - s)$ of periodically correlated sequences $X(t)$ one should specify the general form of the functions $C_k(\tau)$, $k = 0, 1, ..., T_0 - 1$. Assume that the functions $C_k(\tau)$, which are defined only for $k = 0, 1, ..., T_0 - 1$, are determined for other values of k with the aid of the equality $C_{k+T_0}(\tau) = C_k(\tau)$. Then one can show that *a function* $C(t,\tau)$ *of the form* (4.455) *is the correlation function of a periodically correlated sequence* $X(t)$ *if, and only if, the matrix*

$$(4.457) \quad B(\tau) = \|B_{jk}(\tau)\|_{j,k=0,...,T_0-1}, \quad B_{jk}(\tau) = C_{k-j}(\tau)e^{ij\omega_0 \tau},$$

is the correlation matrix of some T_0-*dimensional stationary sequence.*[158] Recall that the necessary and sufficient

conditions for the matrix $B(\tau)$ to be the correlation matrix of a multidimensional stationary sequence are given in Sec. 20. Hence these conditions uniquely determine the class of the collections $C_k(\tau)$, $k = 0,1, ..., T_0 - 1$.

The conditions obtained are, of course, not very convenient since their verification is rather difficult. However, these conditions imply some simple and important consequences. According to (4.457) the upper left corner of the matrix $B(\tau)$ is occupied by the function $C_0(\tau)$; therefore, $C_0(\tau)$ must be the correlation function of a stationary random sequence. (It is easy to show with the aid of (4.456) that $C_0(\tau)$ is the correlation function of the stationary sequence

$$(4.458) \quad X^{(0)}(t) = [X(t) + X(t + 1) + ... + X(t + T_0 - 1)]/T_0.$$

This sequence may be called the *stationarizable version* of the periodically correlated sequence $X(t)$.) Similarly, the functions $C_k(\tau)$, $k = 1, ..., T_0 - 1$, in the first column of the matrix $B(\tau)$ are cross-correlation functions for some pairs of stationary and stationarily correlated sequences. Hence all functions $C_k(\tau)$ are representable as the Fourier–Stieltjes integrals

$$(4.459) \quad C_k(\tau) = \int_{-\pi}^{\pi} e^{i\omega\tau} dF_k(\omega), \quad k = 0,1, ..., T_0 - 1,$$

where $F_0(\omega)$ is a real bounded nondecreasing function and $F_1(\omega), ..., F_{T_0-1}(\omega)$ are complex functions of bounded variation.

Equations (4.459) and (4.455) make it possible to represent the correlation function $B(t,s) = C(s, t - s) = \langle X(t)\overline{X(s)} \rangle$ as

$$(4.460) \quad B(t,s) = \sum_{k=0}^{T_0-1} \int_{-\pi}^{\pi} e^{i\{\omega t - (\omega - k\omega_0)s\}} dF_k(\omega).$$

Since $\omega_0 = 2\pi/T_0$ (i.e. $T_0\omega_0 = 2\pi$), we have

$$\exp\{i[\omega t - (\omega - k\omega_0)s]\} = \exp\{i[\omega t - (\omega + (T_0 - k)\omega_0)s]\}.$$

Therefore (4.460) can also be written as
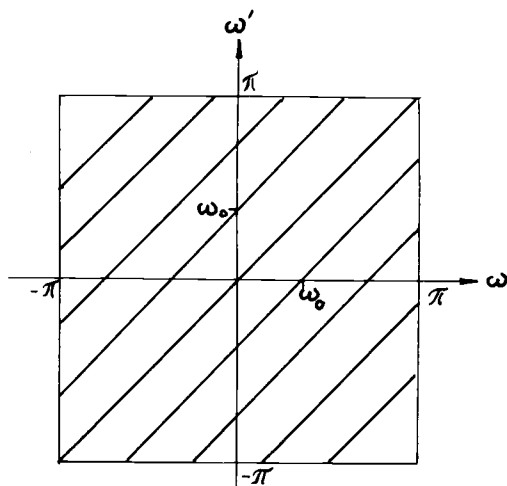
Fig. 48.

$$B(t,s) = \sum_{k=-T_0+1}^{T_0-1} \int_{\Pi_K} e^{i\{\omega t - (\omega-k\omega_0)s\}} dF_k(\omega)$$

(4.461)

$$= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{i(\omega t - \omega' s)} \sum_{k=-T_0+1}^{T_0-1} \delta(\omega - \omega' - k\omega_0) dF_k(\omega) d\omega',$$

where Π_k is the intersection of the intervals $-\pi \leqslant \omega < \pi$ and $-\pi + k\omega_0 \leqslant \omega \leqslant \pi + k\omega_0$, while $F_k(\omega) = F_{k+T_0}(\omega)$ for $k = -T_0+1$,

$-T_0+2, ..., -1$. Thus, *the correlation function* $B(t,s) = \langle X(t)\overline{X(s)}\rangle$ *of a periodically correlated sequence with period* T_0 *is always harmonizable, and its spectral measure* $F(\Delta,\Delta')$ *is concentrated on* $2T_0 - 1$ *segments of the lines* $\omega - \omega' = k\omega_0$, $k = -T_0+1, -T_0+2, ..., T_0$ $- 1$, *located inside the square* $-\pi \leqslant \omega, \omega' \leqslant \pi$ (see Fig. 48). It is also clear that any harmonizable correlation function of the form (4.461) (i.e., corresponding to a spectral measure which is concentrated on the segments of the lines $\omega - \omega' = kT_0$, $k = -T_0+1$, $-T_0+2, ..., T_0 - 1$, satisfies the second condition (4.443), i.e., it is the correlation function of a periodically correlated sequence $X(t)$. Note also that the harmonizability of the correlation function $B(t,s) = \langle X(t)\overline{X(s)}\rangle$ always implies the harmonizability of the random

sequence $X(t)$ (cf. pp. 461-462, where this result is formulated for random processes depending on continuous t). Therefore each periodically correlated random sequence $X(t)$, $t = 0, \pm 1, ...,$ is harmonizable, i.e. is representable as the Fourier–Stieltjes integral

$$(4.462) \quad X(t) = \int_{-\pi}^{\pi} e^{i\omega t} dZ(\omega)$$

where, according to (4.461) and the assumption that $\langle X(t) \rangle = 0$,

$$(4.463) \quad \langle dZ(\omega) \rangle = 0, \langle dZ(\omega)\overline{dZ(\omega')} \rangle = \sum_{k=-T_0+1}^{T_0-1} \delta(\omega - \omega' - k\omega_0) dF_k(\omega) d\omega'.$$

Thus *the class of periodically correlated sequences with mean value zero coincides with the class of harmonizable sequences* (4.462) *satisfying conditions* (4.463).[159]

In the most important case in practice where $|C(t,\tau)|$ falls off rapidly enough with $|\tau|$ for all t, the functions $C_k(\tau)$ are representable as Fourier integrals, and therefore (4.461) takes the form

$$(4.464) \quad B(t,s) = \sum_{k=-T_0+1}^{T_0-1} \int_{\Pi_K} e^{i\{\omega t - (\omega - k\omega_0)s\}} f_k(\omega) d\omega.$$

If the sequence $X(t)$ is real (and hence $B(t,s) = B(s,t)$), then clearly

$$(4.465) \quad f_0(-\omega) = f_0(\omega), \quad f_k(-\omega) = \overline{f_{-k}(\omega)} \quad \text{for } k \neq 0.$$

The restrictions on the behavior of the functions $F_k(\omega)$ and $f_k(\omega)$ will be considered in more detail later for the case of continuous t.

We now proceed to periodically correlated random processes $X(t)$, assuming that the time t is continuous. As before, we restrict ourselves for simplicity to the case where $\langle X(t) \rangle = 0$. We again define the function $C(t,\tau)$ by (4.454). Since $C(t,\tau)$ is a periodic function in t with period T_0, it is natural to associate with this function its Fourier series

$$(4.466) \quad C(t,\tau) \sim \sum_{k=-\infty}^{\infty} C_k(\tau) e^{ik\omega_0 t}$$

where, as usual, $\omega_0 = 2\pi/T_0$, the symbol $\sim$ means "corresponds to", and for $k = 0, 1, ...$

$$(4.467) \quad C_k(\tau) = \frac{1}{T_0} \int_0^{T_0} C(t,\tau) e^{-ik\omega_0 t} dt, \quad C_{-k}(\omega) = \overline{C_k(\tau)} .$$

In all previous sections of this book we considered only continuous correlation functions.   Therefore it may seem reasonable to assume here, too, from the outset that the function $B(t,s)$ is a continuous function of t and s, and hence $C(t,\tau) = B(t+\tau, t)$ is a continuous function of t and τ. In this case the Fourier coefficients (4.467) will clearly be continuous functions of τ for all k, while the Fourier series on the right-hand side of (4.466) will converge to the function on the left-hand side for any t under wide regularity conditions (but not always).[160] However, in the case of periodically correlated processes $X(t)$ the restriction to continuous correlation functions $B(t,s)$ and $C(t,\tau)$ cannot be considered appropriate, because it excludes many very natural and practically useful models of such processes. (In particular, in Example 1 given above, the correlation function (4.447) is discontinuous if the function $P(t)$ is a discontinuous function; the correlation function (4.448) in Example 2 is always discontinuous, and so is the correlation function of pulse-amplitude-modulated signals in the case of rectangular pulses.) Therefore, in the theory of periodically correlated processes $X(t)$ it is desirable to use a weaker restriction of the class of the considered correlation functions, which does not exclude the possibility that the functions $B(t,s)$ and $C(t,\tau)$ have jump discontinuities (i.e., that some inequalities of the form $B(t - 0, s) \neq B(t + 0, s)$ or $B(t, s + 0) \neq B(t, s - 0)$ are valid).

If the correlation function $B(t,s)$ has only jump discontinuities (i.e., the values $B(t \pm 0, s \pm 0)$ exist for all t,s) then $C(t,\tau)$ also has only such discontinuities, and hence the Fourier coefficients (4.467) clearly exist.   It can be shown that then the following statement is true, which is similar to the result stated on p. 473 for periodically correlated sequences $X(t)$:   *A function* $C(t,\tau)$ *periodic* <u>*in t*</u> *(with period* T_0) *is the correlation function* $\langle X(t + \tau)X(t) \rangle$ *of some periodically correlated random process* X(t) *if, and only if, the following inequality is true for any integer* M *and any collections of real numbers* $t_1, ..., t_M$, *complex numbers* $c_1, ..., c_M$ *and indices* $k_1, ..., k_M$:

$$(4.468) \quad \sum_{p,q=1}^{M} B_{k_p k_q}(t_p - t_q) c_p \overline{c_q} \geq 0$$

where

(4.469) $B_{jk}(\tau) = C_{k-j}(\tau)e^{ij\omega_0\tau}$

and $C_k(\tau)$ are the Fourier coefficients (4.467) of $C(t,\tau)$. The inequality (4.468) implies, of course, that the matrix $B(\tau) = \|B_{jk}(\tau)\|$ is the correlation matrix of some infinite-dimensional stationary sequence.[161]

Assuming $k_1 = ... = k_M = 0$ in (4.468), we find at once that the function $C_0(\tau)$ is the correlation function of a stationary random process. This result has already been indicated above; see pp. 469-470 where it is shown that $C_0(\tau)$ coincides with the correlation function of the stationarizable version $X^{(0)}(t)$ of $X(t)$. Similarly, all the functions $C_k(\tau)$, $k = \pm1,\pm2, ...$, are cross-correlation functions for certain pairs of stationary and stationarily correlated processes. This does not imply, however, that all the functions $C_k(\tau)$ can be represented as Fourier–Stieltjes integrals, since if the function $C(t,\tau)$ has discontinuities, then, in general, its Fourier coefficients $C_k(\tau)$ may also prove to be discontinuous. For all the functions $C_k(\tau)$, $k = 0,\pm1, ...$, to be continuous one should impose some additional restrictions on the class of the correlation functions being considered. Clearly, a sufficient condition for each $C_k(t)$ to be everywhere continuous is that the function $C(t,\tau)$ be mean continuous in τ in the following sense:

(4.470) $\lim\limits_{\tau-\tau'\to0} \int_0^{T_0} |C(t,\tau) - C(t,\tau')|dt = 0.$

Indeed, since

$$|C_k(\tau) - C_k(\tau')| \leqslant \frac{1}{T_0}\int_0^{T_0}|C(t,\tau) - C(t,\tau')|dt,$$

(4.470) implies that $C_k(\tau)$ is everywhere continuous for each k. Thus, all the functions $C_k(\tau)$ are stationary continuous cross-correlation functions if (4.470) is valid. Hence, the functions $C_k(\tau)$ are representable in this case as

(4.471) $C_k(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau}dF_k(\omega), \quad k = 0,\pm1, ... ,$

where $F_0(\omega)$ is a bounded nondecreasing function and $F_k(\omega)$ for $k \neq 0$ is a complex function of bounded variation.[162] Note that condition (4.470) does not rule out the possibility that $B(t,s) = C(s,t-s)$ is a discontinuous function, and this condition is fulfilled for the most important examples of discontinuous correlation functions of periodically correlated processes.

In the case of a periodically correlated sequence $X(t)$, it follows from (4.459) (which is similar to (4.471)) that (4.461) is valid.   Thus, (4.459) implies that any periodically correlated sequence is harmonizable and has a spectral measure, which is concentrated on $2T_0 - 1$ segments shown in Fig. 48.   However, if the time t is continuous, then from (4.466) and (4.471) it only follows that

$$
\begin{aligned}
(4.472) \quad B(t,s) &= C(s,t{-}s) \sim \sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i\{\omega t - (\omega - k\omega_0)s\}} dF_k(\omega) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i(\omega t - \omega' s)} \sum_{k=-\infty}^{\infty} \delta(\omega - \omega' - k\omega_0) dF_k(\omega) d\omega'.
\end{aligned}
$$

It is clear that the symbol $\sim$ in (4.472) cannot, in the general case, be replaced by the equality sign.   Indeed, the equality sign would mean that a periodically correlated process $X(t)$, the correlation function $C(t,\tau) = B(t + \tau, t)$ of which satisfies (4.470), is necessarily harmonizable; meanwhile we know that in this case the function $C(t,\tau)$ may be discontinuous, and this rules out the possibility of $B(t,s)$ being harmonizable. Moreover, even if the correlation function $B(t,s)$ of a periodically correlated random process $X(t)$ is bounded and continuous, it may not be harmonizable; see Note 145.   To guarantee the harmonizability of the correlation function $B(t,s)$ of a periodically correlated process, one must impose some additional regularity conditions on $B(t,s)$ besides the continuity condition.   (In particular, it can be shown that it will suffice to require the existence of a continuous second derivative $\partial^2 B(t,s)/\partial t^{2}.$[163]) It is clear, however, that (4.472) may also be interpreted as a relation demonstrating that the correlation function $B(t,s)$ is harmonizable in some generalized sense.   The precise meaning of this generalized harmonizability is associated with the precise meaning of the correspondence (4.466) between the function $C(t,\tau)$ and its Fourier series.   If some mode of convergence of the Fourier partial sums $S_N(t,\tau) = \sum_{k=-N}^{N} C_k(\tau)\exp(ik\omega_0 t)$ to $C(t,\tau)$ holds, then there is a similar mode of convergence of the integral sums corresponding to the Fourier–Stieltjes integral on the right-hand side of (4.472) to the correlation function $B(t,s)$. We cannot, however, pursue further this topic here.[164]

Assume now for simplicity that a periodically correlated random process is ordinary sense harmonizable, i.e., that the

correlation function $B(t,s)$ can be represented in the form
(4.416), where $F(\omega,\omega')$ is a function of bounded variation in
the plane. (The results obtained under this assumption will
actually often be also true for wider classes of periodically
correlated processes $X(t)$ if an appropriate generalized
interpretation of (4.416) and of some other equation is used;
see Note 163.) Relation (4.472) suggests that the spectral
measure $F(\Delta,\Delta')$ of the harmonizable periodically correlated
process $X(t)$ must be concentrated on lines $\omega - \omega' = k\omega_0$, $k =$
$0, \pm 1, \ldots$, (see Fig. 49), i.e. $F(\Delta,\Delta')$ must satisfy the symbolic
equation

$$(4.473) \quad F(d\omega, d\omega') = \sum_{k=-\infty}^{\infty} \delta(\omega - \omega' - k\omega_0) dF_k(\omega) d\omega'.$$

Substituting this equation in (4.416), we obtain (4.472) with
the equality sign between the two sides

$$(4.472a) \quad B(t,s) = \sum_{k=-\infty}^{\infty} e^{ik\omega_0 s} \int_{-\infty}^{\infty} e^{i\omega(t-s)} dF_k(\omega).$$

Note that (4.472a) clearly implies the validity of the second
equation (4.443), and therefore the random process $X(t)$ with
$\langle X(t) \rangle = 0$ and with a correlation function of the form
(4.472a) is necessarily periodically correlated. Conversely, if
$X(t)$ is a harmonizable periodically correlated process, then we
have for any N
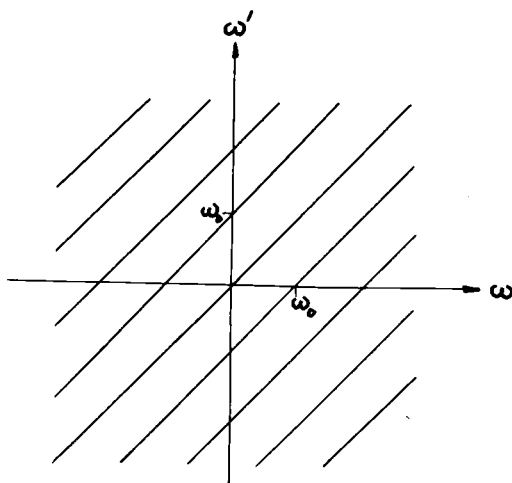
Fig. 49. The lines of non-zero spectral measure for
         periodically correlated processes.

$$B(t,s) = \frac{1}{2N+1} \sum_{n=-N}^{N} B(t + nT_0, \, s + nT_0)$$

(4.474)

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i(\omega t - \omega' s)} D_N(\omega,\omega') d^2 F(\omega,\omega')$$

where

(4.475) $D_N(\omega,\omega') = \frac{1}{2N+1} \sum_{n=-N}^{N} e^{i(\omega-\omega')nT_0}$.

It is easy to see that $|D_N(\omega,\omega')| \leqslant 1$ for any $\omega, \, \omega'$ and N, and that $D_N(\omega,\omega') = 1$ if $\omega - \omega' = 2\pi k/T_0 = k\omega_0$ where k is an integer, but $D_N(\omega,\omega') \rightarrow 0$ as $N \rightarrow \infty$, provided the equality $\omega - \omega' = k\omega_0$ holds for no integer k. Hence, for $N \rightarrow \infty$ we obtain from (4.474)

$$B(t,s) = \sum_{k=-\infty}^{\infty} \iint_{\omega-\omega'=k\omega_0} e^{i(\omega t - \omega' s)} d^2 F(\omega,\omega')$$

which is equivalent to (4.472a). Thus, *the class of harmonizable periodically correlated processes* $X(t)$ *coincides with the class of harmonizable processes whose spectral measure is concentrated on lines* $\omega - \omega' = k\omega_0, \, k = 0, \pm 1, \ldots$.[165] The boundedness of the two-dimensional variation of $F(\omega,\omega')$ is here clearly equivalent to condition

(4.476) $\sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} |dF_k(\omega)| < \infty$.

Note that $d^2 F(\omega,\omega') = \langle dZ(\omega)\overline{dZ(\omega')} \rangle$ where $Z(\omega)$ is the random function appearing in the spectral representation (4.415) of a process $X(t)$. Therefore $F_0(\omega)$ is always a real nondecreasing function (naturally, if the arbitrary constant, which can be added to $F_0(\omega)$, is chosen appropriately), while $F_k(\omega)$ for $k=\pm 1, \pm 2$ are complex functions of bounded variation such that

(4.477) $\sum_{p,q=1}^{M} \Delta F_{k_p - k_q}(\omega + k_p \omega_0) \, c_p \overline{c_q} \geqslant 0$

(where $\Delta F_k(\omega) = F_k(\omega + \Delta) - F_k(\omega))$ for any integer M, real number $\omega, \, \Delta > 0$, and any collections of indices $k_1, \ldots, k_M$ and complex numbers $c_1, \ldots, c_M$. Conversely, any function $B(t,s)$ of the form (4.472a), where the functions $F_k(\omega), \, k = 0, \pm 1, \pm 2, \ldots$, satisfy conditions (4.476) and (4.477), is the correlation function of a harmonizable periodically correlated process. Condition (4.477) implies, in particular, that

(4.478) $\Delta F_k(\omega) = \overline{\Delta F_{-k}(\omega - k\omega_0)}, \ k = 0, \pm 1, \ \ldots \ .$

Moreover, if the process $X(t)$ is real, then, evidently,

(4.479) $\Delta F_k(\omega) = \overline{\Delta F_{-k}(-\omega - \Delta)} \ .$

We also note that in the important case where $B(t + \tau, t) = C(t, \tau)$ falls off rapidly enough with $|\tau|$ (e.g., if

(4.480) $\int_{-\infty}^{\infty} |B(t + \tau, \ t)| d\tau < \infty$

for any t), all the function $F_k(\omega)$ are absolutely continuous. In this case (4.472a) takes the form

(4.481) $B(t,s) = \sum_{k=-\infty}^{\infty} e^{ik\omega_0 s} \int_{-\infty}^{\infty} e^{i\omega(t-s)} f_k(\omega) d\omega$

where the functions $f_k(\omega), \ k = 0, \pm 1, \ldots,$ satisfy conditions

(4.482) $\sum_{k=-\infty}^{\infty} \int_{-\infty}^{\infty} |f_k(\omega)| d\omega < \infty$

and

(4.483) $\sum_{p,q=1}^{M} f_{k_p - k_q}(\omega + k_p \omega_0) c_p \overline{c_q} \geqslant 0$

for any $M, \omega, k_1, \ldots, k_M,$ and $c_1, \ldots, c_M$ such as those indicated above (cf. (4.477)). From (4.483) (as well as from (4.478) and (4.479)) it follows that in the case of a real process $X(t)$ the functions $f_k(\omega)$ have the properties that

(4.484) $f_0(\omega) \geqslant 0, \ f_0(-\omega) = f_0(\omega), \ f_k(\omega) = \overline{f_{-k}(\omega - k\omega_0)} = f_{-k}(-\omega),$

$\text{for } k = \pm 1, \pm 2, \ldots \ .$

The functions $F_k(\omega), \ k = 0, \pm 1, \ldots,$ are called *spectral distribution components* (or *spectral distribution functions*), while $f_k(\omega), \ k = 0, \pm 1, \ldots,$ are called *spectral density components* (or simply *spectral densities*) of the periodically correlated random process $X(t)$.[166]

Since we assume that the process $X(t)$ is harmonizable, it is representable as the Fourier–Stieltjes integral (4.415). Some other representations of periodically correlated processes $X(t)$ can also be found in the literature and are sometimes used in special applications; we shall not consider them, however.[167]

If $\langle X(t)\rangle = 0$, as was suggested above, then naturally, $\langle Z(\omega_2) - Z(\omega_1)\rangle = 0$ for any ω_2 and ω_1. If, however, $\langle X(t)\rangle = m(t) \neq 0$, then $\langle Z(\omega)\rangle = M(\omega) \neq 0$ and

$$(4.485) \quad m(t) = \langle X(t)\rangle = \int_{-\infty}^{\infty} e^{i\omega t} dM(\omega).$$

Thus, the mean value of harmonizable process $X(t)$ is always representable as the Fourier–Stieltjes integral (4.485). According to the first equation (4.443), for periodically correlated $X(t)$ the function $m(t)$ is periodic with period T_0, and therefore the measure $M(d\omega)$ must be concentrated on a set of points of the form $\omega = k\omega_0$, $\omega_0 = 2\pi/T_0$, $k = 0, \pm 1, \ldots$. Thus, in this case (4.485) turns into an ordinary Fourier series:

$$(4.486) \quad m(t) = \sum_{k=-\infty}^{\infty} M_k e^{ik\omega_0 t}$$

where $M_k = M(k\omega_0+0) - M(k\omega_0-0) = (1/T_0)\int_0^{T_0} m(t)\exp(-ik\omega_0 t)dt$. Therefore

$$(4.487) \quad \begin{aligned} B(t,s) &= \langle X(t)\overline{X(s)}\rangle = b(t,s) + m(t)\overline{m(s)} \\ &= b(t,s) + \sum_{k=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} M_k \overline{M_j} e^{i(kt-js)\omega_0} \end{aligned}$$

where $b(t,s) = \langle \{X(t) - m(t)\}\{\overline{X(s)} - \overline{m(s)}\}\rangle$ is the correlation function (of the form (4.472a)) of the centered process $\mathring{X}(t) = X(t) - \langle X(t)\rangle$. We see that in the case of a harmonizable periodically correlated process with nonzero mean value, the spectral measure $F(\Delta, \Delta')$ corresponding to the centered process $X(t) - \langle X(t)\rangle$ must be supplemented by nonzero contributions ("point masses" equal to $M_k \overline{M_j}$) at points of the two-dimensional lattice $\omega = k\omega_0$, $\omega' = j\omega_0$, $k, j = 0, \pm 1, \ldots$.

If $X(t)$ is a periodically correlated process with period T_0, then the sequence $X_t(k) = X(t + kT_0)$, $k = 0, \pm 1, \ldots$, is clearly a stationary random sequence, which has mean value $\langle X_t(k)\rangle = m(t)$. If, in addition, $X(t)$ is periodic up to the fourth order (so that $B^{(4)}(t,s,u,v) = B^{(4)}(t+T_0, s+T_0, u+T_0, v+T_0)$ where $B^{(4)}(t,s,u,v) = \langle X(t)\overline{X(s)})X(u)\overline{X(v)}\rangle$), then the sequence $X_{t,s}(k) = X(t+kT_0)\overline{X(s+kT_0)}$, $k = 0, \pm 1, \ldots$, is also stationary and has mean value $\langle X_{t,s}(k)\rangle = B(t,s)$. In accordance with this, if just one

one realization $x(t)$ of the periodically correlated process $X(t)$ with period T_0 is known from observations for $0 \leqslant t \leqslant T$, where T greatly exceeds T_0, then the mean value $\langle X(t) \rangle = m(t)$ and the correlation function $\langle X(t)\overline{X(s)} \rangle = B(t,s)$ may, under wide conditions, be calculated to a good accuracy by using the formulae

$$(4.488) \quad m(t) \approx m_T^*(t) = \frac{1}{N+1} \sum_{k=0}^{N} x(t + kT_0)$$

and

$$(4.489) \quad \begin{aligned} B(t + \tau, t) &\approx B_T^*(t + \tau, t) \\ &= \frac{1}{N_1 + 1} \sum_{k=0}^{N_1} x(t + \tau + kT_0)\overline{x(t + kT_0)} \end{aligned}$$

where $0 \leqslant t < T_0$, $0 \leqslant \tau \leqslant T$, $NT_0 + t \leqslant T < (N+1)T_0 + t$, $N_1 T_0 + t + \tau \leqslant T < (N_1+1)T_0 + t + \tau$. The statistical errors of estimates (4.488) and (4.489) can easily be calculated with the aid of error formulae for the estimates of the mean value of stationary random sequences given in Sec. 16.[168] Proceeding from the estimates of $m(t)$ and $B(t,s)$, one can also develop general methods for approximate determinations of the spectral characteristics $f_k(\omega)$ and $F_k(\omega)$ (and also, if desired, M_k), where $k = 0, \pm 1, \ldots$, from a single realization of the process $X(t)$. Such *statistical spectral analysis of periodically correlated random processes* (and *sequences*) is studied in a number of sources.[169] The particular problem of determining M_0, $f_0(\omega)$, and $F_0(\omega)$ will also be additionally discussed below.

26.6. Processes Having Time Average Mean Value and Correlation Function

The correlation theory of stationary random functions is widely applicable to many scientific and engineering problems, first of all because in the stationary case the mean value $\langle X(t) \rangle = m$ and the correlation function $\langle X(t)X(s) \rangle = B(t - s)$ can be determined rather easily and accurately from a single realization of the random function $X(t)$; see Secs. 16 and 17. (We will consider only real random processes in this subsection; therefore, we write $\langle X(t)X(s) \rangle$ instead of $\langle X(t)\overline{X(s)} \rangle$.) In dealing with nonstationary processes $X(t)$ the situation is much less favorable in this respect. In fact, here

the determination of $\langle X(t) \rangle = m(t)$ and $\langle X(t)X(s) \rangle = B(t,s)$ from observations is usually highly complicated and costly (and sometimes altogether impossible, since more than one realization of $X(t)$ cannot, in principle, be observed; cf. pp. 26–29). Therefore, of great interest is the question of finding some simplified statistical characteristics of nonstationary random processes and sequences $X(t)$ describing only some comparatively special properties of $X(t)$, but determinable easily enough from a single realization $x(t)$ observed during a finite (but long enough) time interval $0 \leqslant t \leqslant T$.

We begin with characteristics depending on the mean value $\langle X(t) \rangle = m(t)$ of the process $X(t)$. It is clear that only in some exceptional cases can one hope to determine rather accurately the evolutionary (i.e. time-varying) mean value $m(t)$ for all t from just one realization $x(t)$. We have seen, however, that if the process $X(t)$ is harmonizable, then

$$(4.490) \quad \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} m(t)dt = M^{(0)}$$

exists and can, under wide conditions, be determined rather accurately from a single lengthy realization $x(t)$ (see p. 463 and Note 142 to Chap. 4 in Vol. II). Note also that the limit (4.490) coincides with

$$(4.491) \quad \lim_{T \to \infty} \frac{1}{T} \int_0^T m(t)dt = M^{(0)}$$

for any harmonizable process $X(t)$.[170] Therefore it is reasonable to consider the class $\mathfrak{M}_1$ of random processes $X(t)$ for which the limit (4.491), where $m(t) = \langle X(t) \rangle$, exists and is bounded. This limit clearly exists if the mean value $m(t)$ is periodic in t, or almost periodic, or if it is representable as the Fourier–Stieltjes integral (4.485) with respect to some function $M(\omega)$ of bounded variation. The limit (4.491) also exists for any function $m(t)$ tending to a limit as $t \to \infty$, and also for all sums of the form $a_1 m_1(t) + a_2 m_2(t)$, where a_1 and a_2 are constants, while both $m_1(t)$ and $m_2(t)$ are the mean values of some random processes of class $\mathfrak{M}_1$. Thus, we can see that the class $\mathfrak{M}_1$ is wide enough and includes many of the nonstationary processes $X(t)$ which are encountered in practical applications. We will call the processes of class $\mathfrak{M}_1$ *random processes having a time average mean value*, and $M^{(0)}$, the *average mean value* of $X(t)$.

The average mean value $M^{(0)}$ of any random process $X(t)$, if it exists, is clearly a constant. If $X(t)$ is a stationary process (or, at least, it has a constant mean value $\langle X(t)\rangle = m$), then $M^{(0)} = m = \langle X(t)\rangle$. However, in all other cases $M^{(0)}$ is a poorer statistical characteristic than $\langle X(t)\rangle$, determining only the time average (or asymptotic) behavior of the true mean value $m(t)$. (In particular, if $X(t)$ is a periodically correlated process, then $M^{(0)}$ coincides with the zero Fourier coefficient M_0 of the periodic function $m(t)$, but it contains no information about the other Fourier coefficients of $m(t)$.) However, the constant $M^{(0)}$ has an important advantage over $m(t)$, namely, it can be, under wide conditions, easily determined from a single realization $x(t)$ of the process $X(t)$. The formulation of the conditions which guarantee the possibility of such determination of $M^{(0)}$ constitutes the content of the *ergodic theorem* (i.e. the *law of large numbers*) for nonstationary random processes of class $\mathfrak{M}_1$.

The obvious relation

$$(4.492) \qquad \frac{1}{T}\int_0^T X(t)dt = \frac{1}{T}\int_0^T m(t)dt + \frac{1}{T}\int_0^T \{X(t) - m(t)\}dt$$

shows that the left-hand side of (4.492) tends to $M^{(0)}$ as $T \to \infty$ if, and only if,

$$(4.493) \qquad \lim_{T\to\infty} \frac{1}{T}\int_0^T \mathring{X}(t)dt = 0$$

where $\mathring{X}(t) = X(t) - \langle X(t)\rangle = X(t) - m(t)$ is the centered process $X(t)$. Let

$$(4.494) \qquad b(t,s) = \langle \mathring{X}(t)\mathring{X}(s)\rangle = B(t,s) - m(t)m(s)$$

be the correlation function of the process $\mathring{X}(t)$. (As usual, we assume that the process $X(t)$ is mean continuous and the functions $B(t,s)$ and $b(t,s)$ are continuous in t and s.) According to (1.55)

$$(4.495) \qquad \left\langle \left|\frac{1}{T}\int_0^T \mathring{X}(t)dt\right|^2 \right\rangle = \frac{1}{T^2}\int_0^T\int_0^T b(t,s)dtds = \frac{2}{T^2}\int_0^T\int_0^t b(t,s)dsdt.$$

Therefore condition (4.493), where lim denotes the limit in the mean, and hence also its equivalent

$$(4.496) \qquad \frac{1}{T}\int_0^T X(t)dt \to M^{(0)} \quad \text{for } T \to \infty$$

is fulfilled if, and only if,

(4.497) $\lim\limits_{T\to\infty} \dfrac{1}{T^2} \int_0^T\int_0^t b(t,s)\,ds\,dt = 0.$

This is the simplest and most general ergodic theorem for random processes of class $\mathfrak{M}_1$. (In the particular case where the process $X(t)$ is stationary, condition (4.497) becomes Slutsky's condition (3.10a).)[171] Thus, (4.497) is a necessary and sufficient condition for the estimator M_T^* (see (3.6)) to be a consistent estimator of $M^{(0)}$.

Based on (4.495), one may also derive the *strong law of large numbers* for random processes $X(t)$ of class $\mathfrak{M}_1$, i.e. find the conditions guaranteeing the convergence of the estimator of M_T^* to $M^{(0)}$ with probability one (cf. p. 224).[172] Note also that for some special subclasses of class $\mathfrak{M}_1$ condition (4.497) may often be modified and sometimes simplified. Thus, for instance, it has already been pointed out above that if $X(t)$ is a harmonizable process, it necessarily belongs to class $\mathfrak{M}_1$, and then (4.496) is valid if, and only if, the point $\omega = \omega' = 0$ contributes nothing to the spectral measure $F(\Delta,\Delta')$. Another form of a necessary and sufficient condition for the validity of (4.496), which is applicable to a subclass of class $\mathfrak{M}_1$ wider than the subclass of harmonizable processes, will be indicated below.

We now proceed to the correlation characteristics of random processes $X(t)$ and assume, for simplicity, from the outset that $\langle X(t)\rangle = 0$. (This means, of course, that we consider the centered process $\overset{\circ}{X}(t) = X(t) - m(t)$ instead of $X(t)$.) The correlation function $\langle X(t)X(s)\rangle = B(t,s)$ of $X(t)$ depends, in general, on two parameters t and s and cannot, as a rule, be determined from a single realization $x(t)$ of the process $X(t)$. It seems reasonable, therefore, to introduce into consideration the *average correlation function* $B^{(0)}(\tau)$ of $X(t)$ defined for $\tau \geqslant 0$ as the limit

(4.498) $\lim\limits_{T\to\infty} \dfrac{1}{T} \int_0^{T-\tau} B(t+\tau,t)\,dt = B^{(0)}(\tau)\,(\tau)$

and extended symmetrically to negative values of τ (so that $B^{(0)}(-\tau) = B^{(0)}(\tau)$). Random processes $X(t)$ for which the limit (4.498) exists and is a nonzero continuous function of τ, will be called *processes having an average correlation function*, while the class of all such processes will be denoted as $\mathfrak{M}_2$. It is easy to see that the function $B^{(0)}(\tau)$ is always positive

definite[173]. Therefore it follows from Bochner's theorem (see p. 106 and Note 3 to the Introduction) that $B^{(0)}(\tau)$ can be represented as a Fourier–Stieltjes integral

$$(4.499) \quad B^{(0)}(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau} dF^{(0)}(\omega)$$

where $F^{(0)}(\omega)$ is a bounded nondecreasing function. The function $F^{(0)}(\omega)$ is called the *average spectral distribution function* of the process $X(t)$. Thus *any process of class $\mathfrak{M}_2$ also has an average spectral distribution function.* In the most important case in practice, where the function $B^{(0)}(\tau)$ falls off rapidly enough with $|\tau|$ (e.g., if $|B^{(0)}(\tau)|$ is integrable from zero to infinity), the function $F^{(0)}(\omega)$ is absolutely continuous, and (4.499) can be rewritten as

$$(4.500) \quad B^{(0)}(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau} f^{(0)}(\omega) d\omega$$

where $f^{(0)}(\omega) = dF^{(0)}(\omega)/d\omega$ is the *average spectral density* of $X(t)$.

In the particular case of a stationary process $X(t)$ we have $B(t+\tau, t) = B(\tau)$. Therefore *all* stationary processes belong to the class $\mathfrak{M}_2$, and the average correlation function, spectral distribution function and spectral density of such processes coincide with the ordinary functions $B(\tau)$, $F(\omega)$ and $f(\omega)$ studied in Chaps. 1–3 of this book. The concept of the average correlation function of a nonstationary process (but not the given term for it) was introduced in the books by Bunimovich and Blanc–Lapierre and Fortet published in the early 1950s; later this concept was also studied repeatedly (together with averaged spectral characteristics) by a number of other authors.[174] Note that in the engineering literature the average spectral density is often called simply the *average power spectrum* of a nonstationary process $X(t)$.

If the process $X(t)$ belongs to class $\mathfrak{M}_2$, then its correlation function $B(t+\tau,t)$ belongs, for any fixed τ, to the class of functions of t having a finite average over the positive half-axis (i.e. to the class of mean values of processes from class $\mathfrak{M}_1$). This last condition is evidently fulfilled if the function $B(t+\tau,t) = C(t,\tau)$ is periodic in τ (i.e. if $X(t)$ is a periodically correlated process). In this case, by virtue of (4.467), the average correlation function $B^{(0)}(\tau)$ coincides with $C_0(\tau)$ and, hence $F^{(0)}(\omega) = F_0(\omega)$, $f^{(0)}(\omega) = f_0(\omega)$. It is also clear that a *random*

process $X(t)$ belongs to class $\mathfrak{M}_2$ if $\lim\limits_{t\to\infty} B(t+\tau,\, t) = B^{(0)}(\tau)$ *exists*

for all $\tau \geqslant 0$ (in this case $X(t)$ may be called an *asymptotically stationary process*[175]). Moreover, *all harmonizable random processes* also *belong to class* $\mathfrak{M}_2$. In fact, it is not hard to show that if the correlation function $B(t,s)$ of $X(t)$ is given by (4.416), then for any finite $\tau > 0$

$$\lim_{T\to\infty} \frac{1}{T} \int_0^{T-\tau} B(t+\tau, t)dt = \lim_{T\to\infty} \frac{1}{T} \int_0^{T-\tau}\!\!\int_{-\infty}^{\infty}\!\!\int_{-\infty}^{\infty} e^{i\omega\tau} e^{i(\omega-\omega')t} F(d\omega, d\omega') dt$$

$$(4.501) \qquad = \lim_{T\to\infty} \int_{-\infty}^{\infty}\!\!\int_{-\infty}^{\infty} e^{i\omega\tau} \left\{ \frac{1}{T}\int_0^T e^{i(\omega-\omega')t} dt \right\} F(d\omega, d\omega')$$

$$= \int_{-\infty}^{\infty} e^{i\omega\tau} F(d\omega, d\omega).$$

Thus, $X(t)$ belongs to the class $\mathfrak{M}_2$, and its average spectral measure $F^{(0)}(\Delta)$ for $\Delta = [\omega_1, \omega_2]$ coincides with the contribution from the diagonal segment $\omega_1 \leqslant \omega = \omega' \leqslant \omega_2$ to the two-dimensional spectral measure $F(\Delta, \Delta')$.[176] There are also many nonharmonizable random processes of class $\mathfrak{M}_2$, and some of them will be considered below (see also Note 170).

Suppose now that a random process $X(t)$ belonging to class $\mathfrak{M}_2$ is applied to the input of a linear time-invariant system $\mathfrak{L}$ specified by an impulse response function $h(t)$ or by its Fourier transform — the transfer function $H(\omega)$. If $\mathfrak{L}\{X(t)\} = Y(t)$, then clearly

$$\lim_{T\to\infty} \frac{1}{T} \int_0^{T-\tau} B_y(t+\tau, t)\, dt$$

$$= \lim_{T\to\infty} \frac{1}{T} \int_0^{T-\tau}\!\!\int_{-\infty}^{\infty}\!\!\int_{-\infty}^{\infty} h(u)\overline{h(v)} B(t + \tau - u,\, t - v) du\, dv\, dt$$

$$= \lim_{T\to\infty} \int_{-\infty}^{\infty}\!\!\int_{-\infty}^{\infty} h(u)\overline{h(v)} \left\{ \frac{1}{T} \int_{-v}^{T-\tau-v} B(t_1 + \tau - u + v,\, t_1) dt_1 \right\} du\, dv$$

where $B_y(t,s) = \langle Y(t)\overline{Y(s)} \rangle$, $t_1 = t - v$. We now note that for any finite $\tau \geqslant 0$, u, and v we have

$$\lim_{T\to\infty} \frac{1}{T}\int_{-v}^{T-\tau-v} B(t_1 + \tau - u + v,\ t_1)dt_1$$

$$= \lim_{T\to\infty} \frac{1}{T}\int_0^{T-|T-u+v|} B(t_1 + \tau - u + v, t_1)\, dt_1 = B^{(0)}(|\tau - u + v|)$$

$$= \int_{-\infty}^{\infty} e^{i\omega(\tau-u+v)}\, dF^{(0)}(\omega).$$

Therefore

$$\lim_{T\to\infty} \frac{1}{T}\int_0^{T-\tau} B_y(t+\tau,t)dt\ t$$

(4.502)
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i\omega(\tau-u+v)}h(u)\overline{h(v)}\, du\, dv\, dF^{(0)}(\omega)$$

$$= \int_{-\infty}^{\infty} e^{i\omega\tau}|H(\omega)|^2\, dF^{(0)}(\omega).$$

We see that if

$$\int_{-\infty}^{\infty} |H(\omega)|^2 dF^{(0)}(\omega) < \infty,$$

then the application of the process $X(t)$ having the average spectral distribution function $F^{(0)}(\omega)$ to the input of a linear time-invariant system having the transfer function $H(\omega)$ yields an output process $Y(t)$ which has also an average spectral distribution function, and this function is

(4.503) $$F_y^{(0)}(\omega) = \int_{-\infty}^{\omega} |H(\omega')|^2 dF^{(0)}(\omega').$$

Thus, the main result relating to the spectral properties of linear time-invariant transformations of stationary random processes proves to hold also for nonstationary processes having an average spectral distribution.[177] Since precisely this result is applied most often in engineering sciences, it is no wonder that engineers often use one and the same term "power spectrum" for ordinary spectral densities of stationary processes and for more general average spectral densities of nonstationary processes.

Let us consider now the problem of determining average correlation and spectral characteristics of the process $X(t)$ of class $\mathfrak{M}_2$ from a single realization $x(t)$ of this process.   Note that $B(t+\tau, t)$ is the mean value of the random process $Y_\tau(t) = X(t+\tau)X(t)$ depending on parameter τ.   Let

$$b^{(4)}(t,s,\tau) = \langle X(t + \tau)X(t)X(s + \tau)X(s)\rangle$$

(4.504)

$$- B(t + \tau, t)B(s + \tau, s)$$

be the centered correlation function of the process $Y_T(t)$. Then it is possible to show, similarly to the derivation of condition (4.497) (and also Slutsky's condition (3.10a) and the related condition (3.28)) that

(4.505) $\quad B_T^{**}(\tau) = \dfrac{1}{T} \displaystyle\int_0^{T-\tau} X(t + \tau)X(t)dt$

is a consistent estimator of the average correlation function $B^{(0)}(\tau)$ (i.e. $B_T^{**}(\tau) \to B^{(0)}(\tau)$ for $T \to \infty$) if, and only if,

(4.506) $\quad \displaystyle\lim_{T\to\infty} \dfrac{1}{T} \int_0^T b^{(4)}(T,s,\tau)ds = 0$

for every $\tau \geqslant 0$.[178]

If we now have to determine the average spectral density $f^{(0)}(\omega)$ or the spectral distribution function $F^{(0)}(\omega)$ from a single realization $x(t)$ of length T, then it is reasonable to begin by considering a sample periodogram $i_T(\omega)$ and corresponding random periodogram $I_T(\omega)$; see (3.36) and (3.37). In this case (3.35) will also be valid, i.e.,

(4.507) $\quad I_T(\omega) = \dfrac{1}{2\pi} \displaystyle\int_{-T}^{T} e^{-i\omega\tau} B_T^{**}(\tau)d\tau, \quad B_T^{**}(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau}I_T(\omega)d\omega$

where $B_T^{**}(\tau)$ is the estimator (4.505) of $B^{(0)}(\tau)$. Since $\langle B_T^{**}(\tau)\rangle \to B^{(0)}(\tau)$ for $T \to \infty$ by definition of $B^{(0)}(\tau)$ (i.e. $B_T^{**}(\tau)$ is always an asymptotically unbiased estimator of $B^{(0)}(\tau)$), it follows from (4.507) that if the average spectral density $f^{(0)}(\omega)$ exists, then

(4.508) $\quad \langle I_T(\omega)\rangle \to f^{(0)}(\omega) \quad$ for $T \to \infty$.

Hence the periodogram $I_T(\omega)$ is always an asymptotically unbiased estimator of $f^{(0)}(\omega)$.

Naturally, asymptotic unbiasedness does not mean that $I_T(\omega)$ is an acceptable estimator of $f^{(0)}(\omega)$. In fact we know that even in the simplest case where $X(t)$ is a stationary process the periodogram $I_T(\omega)$ does not tend to the spectral density as $T \to \infty$, but is very erratic at large values of T; see Sec. 18. To obtain a consistent estimator, one must always first smooth out the periodogram in some way. Let $A(\omega)$ be a fixed spectral window

(i.e. some bounded function which is everywhere continuous except, perhaps, for a finite number of discontinuities at some continuity points of $F^{(0)}(\omega)$), and let

$$(4.509) \quad \Phi(A) = \int_{-\infty}^{\infty} A(\omega')dF^{(0)}(\omega')$$

be a spectral average corresponding to $A(\omega)$. Then, similar to the reasoning in Sec. 18, one may show that, under wide conditions (guaranteeing the consistency of the estimator $B_T^{**}(\tau)$ of $B^{(0)}(\tau)$),

$$(4.510) \quad \Phi_T(A) = \int_{-\infty}^{\infty} A(\omega')I_T(\omega')d\omega'$$

will be a consistent estimator of $\Phi(A)$.[179] It follows from this, in particular, that under wide conditions

$$(4.511) \quad F_T^*(\omega) = \int_{-\infty}^{\omega} I_T(\omega')d\omega'$$

is a consistent estimator of the average spectral distribution function $F^{(0)}(\omega)$. If, however, the determination of the average spectral density $f^{(0)}(\omega)$ is of primary interest, then the spectral estimation techniques described in Secs. 18 and 19 must be applied: Instead of $A(\omega')$ one should use the spectral window $A_T(\omega - \omega')$, which has a narrow (not too narrow, however, but having a width greatly exceeding $1/T$) peak at $\omega' = \omega$. Indeed, it may be shown that most of the results of Secs. 18 and 19 relating to spectral density estimators for stationary processes can be generalized to the case where the average spectral density of a process of class $\mathfrak{m}_2$ must be estimated. In particular, one can show that passing the realization $x(t)$, where $0 \leqslant t \leqslant T$, of a process of class $\mathfrak{m}_2$ (given in the form of a fluctuating current) through a narrow band-pass filter and then squaring and averaging the output signal, we obtain a reasonable estimate of $f^{(0)}(\omega)$.[180] Thus, most of the conventional engineering methods for determining the spectral density of stationary processes can be applied to the determination of the average spectral density $f^{(0)}(\omega)$.

We now assume that the mean value $\langle X(t) \rangle = m(t)$ of a process $X(t)$ of class $\mathfrak{m}_2$ differs from zero and that the time average (4.491) of the mean value function exists, so that the process $X(t)$ also belongs to class $\mathfrak{m}_1$. Let $F^{(0)}(\omega)$ be the average spectral distribution function of a centered process $\overset{\circ}{X}(t) = X(t) - m(t)$. Then it may readily be shown that

(4.512) $\lim\limits_{T\to\infty} \langle |M_T^* - M^{(0)}|^2 \rangle = F^{(0)}(+0) - F^{(0)}(-0).$

Therefore, $M_T^* = T^{-1}\int_0^T X(t)dt$ *is a consistent estimator of* $\mathrm{M}^{(0)}$ *if,*

and only if, the average distribution function $\mathrm{F}^{(0)}(\omega)$ *is continuous at the point* $\omega = 0$.[181] In the particular case where the process $X(t)$ is stationary or, at least, harmonizable, this result coincides with the one given on pp. 220 and 463. However, the class $\mathfrak{M}_2$ of random processes is much wider than the class of harmonizable processes, and therefore (4.512) is a substantial generalization of the preceding results.

In conclusion we consider a few simple examples of calculation of the average correlation functions and the average spectral distribution or density functions.

Example 1. *Modulated Stationary Process.* Consider a random process of the form (4.425), where $X_0(t)$ is a stationary random process having the correlation function $B_0(\tau)$ and the spectral distribution function $F_0(\omega)$, while $A(t)$ is a nonrandom real bounded function. Then evidently

$$B(t + \tau, t) = A(t + \tau)A(t)\langle X_0(t + \tau)X_0(t)\rangle = A(t + \tau)A(t)B_0(\tau)$$

and therefore

$$\lim\limits_{T\to\infty} \frac{1}{T} \int_0^{T-\tau} B(t + \tau, t)dt = B_0(\tau) \lim\limits_{T\to\infty} \frac{1}{T} \int_0^{T-\tau} A(t + \tau)A(t)dt.$$

It follows from this that the random process (4.425) belongs to the class $\mathfrak{M}_2$ if, and only if,

(4.513) $\lim\limits_{T\to\infty}\dfrac{1}{T}\int_0^{T-\tau} A(t + \tau)A(t)dt = \lim\limits_{T\to\infty}\dfrac{1}{T}\int_0^{T} A(t + \tau)A(t)dt = B^{(A)}(\tau)$

exists for all $\tau > 0$ and is continuous in τ (i.e. if $A(t)$ is a function possessing Wiener's generalized harmonic analysis; see p. 95 and Note 15 to Chap. 2). In this case

(4.514) $B^{(0)}(\tau) = B^{(A)}(\tau)B_0(\tau),$

which implies that the· average spectral distribution function of $X(t)$ is a convolution of $F_0(\omega)$ and of Wiener's "spectrum of the function $A(t)$". In particular, if

(4.515) $A(t) = \sum_{j=1}^{n} c_j \cos\omega_j t,$

then it is easy to see that

(4.516) $B^{(A)}(\tau) = \frac{1}{2} \sum_{j=1}^{n} |c_j|^2 \cos\omega_j \tau$

and Wiener's "spectral distribution" of $A(t)$ consists of a finite number of point masses ("spectral lines") at $\omega = \omega_1, ..., \omega_n, -\omega_1, ..., -\omega_n$. In this case clearly

(4.517) $F^{(0)}(\omega) = \frac{1}{4} \sum_{j=1}^{n} |c_j|^2 \{F_0(\omega - \omega_j) + F(\omega_0 + \omega_j)\}.$

If $n = 1$ and $c_1 = 1$, then $X(t)$ is the amplitude modulated harmonic oscillation (4.439), and a comparison of (4.517) with (4.442) confirms that in the considered case of a periodically correlated process the average spectral distribution $dF^{(0)}(\omega)$ coincides with the distribution of the two-dimensional spectral measure $F(d\omega, d\omega')$ along the diagonal $\omega = \omega'$.[182]

Note that the average correlation function (4.514) does not depend on the behavior of $A(t)$ on the negative half-axis $t < 0$, so that this behavior may be quite arbitrary. Moreover, even if the limit (4.513) can be replaced by a two-sided limit (2.56) (i.e., if the interval $[0,T]$ may be replaced by $[-T,T]$), then the existence of the limiting function $B^{(A)}(\tau)$ does not imply that the function $A(t)$ is representable as a Fourier–Stieltjes integral of the form (4.426). Therefore a process of the form (4.425) having an average correlation function can very well be unharmonizable.

Example 2.   Let $X(t)$ be a *process with equidistant discontinuities and adjoined random variables* (see p. 470). Let us first assume that $\langle X_n \rangle = 0$ and $T_0 = 1$. Then, if $\tau \geqslant 0$, $t = k + t'$, $t + \tau = k_1 + t''$, where k and k_1 are integers and $0 \leqslant t$, $t' < 1$,

(4.518) $B(t + \tau, \tau) = \begin{cases} \sigma_0^2 & \text{for } k = k_1, \\ 0 & \text{for } k \neq k_1 \end{cases}$

(i.e. $B(t + \tau, \tau) = \sigma_0^2 \neq 0$ if, and only if, the integral parts of t and $t + \tau$ are the same). Note that the correlation function (4.518) is discontinuous, whence it follows at once that the random process $X(t)$ cannot be harmonizable. However, it is easy to see that

$$(4.519) \ B^{(0)}(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_0^{T-\tau} B(t+\tau, t)dt = \begin{cases} \sigma_0^2(1 - \tau) \text{ for } 0 \leqslant \tau < 1, \\ \\ 0 \qquad \text{for } \tau \geqslant 1. \end{cases}$$

Hence $X(t)$ belongs to class $\mathfrak{M}_2$.[183] The corresponding average spectral density is given by (2.124) with $C = \sigma_0^2$, $T = 1$.

In the more general case, where $\langle X_n \rangle = m_0 \neq 0$ and the distance between adjacent discontinuities of $X(t)$ is equal to T_0, we similarly obtain

$$(4.520) \qquad B^{(0)}(\tau) = \begin{cases} \sigma_0^2 \left(1 - \dfrac{|\tau|}{T_0} \right) + m_0^2 & \text{for } |\tau| < T_0, \\ \\ m_0^2 & \text{for } |\tau| \geqslant T_0. \end{cases}$$

The corresponding average spectral distribution function $F^{(0)}(\omega)$ has a single discontinuity at zero, where it increases by m_0^2, and an absolutely continuous part corresponding to the spectral density (2.124) with $C = \sigma_0^2/T_0$, $T = T_0$. Thus,

$$(4.521) \qquad f^{(0)}(\omega) = \frac{2\sigma_0^2}{\pi T_0} \frac{\sin^2(\omega T_0/2)}{\omega^2} + m_0^2 \delta(\omega).$$

Example 3. *Pulse-Amplitude-Modulated Signal.* A process with equidistant discontinuities and adjoined random variables may be interpreted as a pulse-amplitude-modulated signal consisting of adjacent rectangular pulses of width T_0 having random amplitudes X_n. We now consider a more general pulse-amplitude-modulated signal of the form (4.449), where $\Gamma(t)$ is a given function specifying the pulse shape, and the amplitudes X_n, $n = 0, \pm 1, \pm 2, ...$, are independent identically distributed random variables. We suppose, for simplicity, that $\Gamma(t) = 0$ for $t < 0$ and $t > T_0$ and that $\langle X_n \rangle = 0, \langle X_n^2 \rangle = \sigma_0^2$. In this case one may easily compute the average spectral density $f^{(0)}(\omega)$ of the process $X(t)$ by evaluating the corresponding periodogram $I_T(\omega)$ and then exploiting the general relation $f^{(0)}(\omega) = \lim_{T \to \infty} \langle I_T(\omega) \rangle$.

In fact, if $T = NT_0$ where N is an integer, then

$$\int_0^{NT_0} e^{-i\omega t} X(t)dt = \sum_{n=0}^{n-1} X_n \int_0^{T_0} e^{-i\omega t} \Gamma(t - nT_0)dt$$

$$(4.522)$$

$$= D(\omega) \sum_{n=0}^{N-1} X_n e^{-in\omega T_0},$$

where

(4.523) $D(\omega) = \int_0^{T_0} e^{-i\omega t} \Gamma(t) dt = \int_{-\infty}^{\infty} e^{-i\omega t} \Gamma(t) dt.$

Using the relations $\langle X_n^2 \rangle = \sigma_0^2$, $\langle X_n X_m \rangle = 0$ for $n \neq m$, we obtain

(4.524)

$$f^{(0)}(\omega) = \langle I_{NT_0}(\omega) \rangle = \frac{1}{2\pi NT_0} \langle \left| \int_0^{NT_0} e^{-i\omega t} X(t) dt \right|^2 \rangle$$

$$= \frac{|D(\omega)|^2}{2\pi NT_0} \sum_{n=0}^{N-1} \langle X_n^2 \rangle = \frac{\sigma_0^2}{2\pi T_0} |D(\omega)|^2.$$

(It is unnecessary to assume that $N \to \infty$, since we see that $\langle I_{NT_0}(\omega) \rangle$ is independent of N in this case.) In particular, if

$\Gamma(t) = 1$ for $0 \leq t < T_0$, then $D(\omega) = 2\exp(i\omega T_0/2)\omega^{-1}\sin\omega T_0/2$ and (4.524) coincides with (4.521), where $m_0 = 0$.

The case where $\langle X_n \rangle = m_0 \neq 0$, $\langle X_n^2 \rangle = \sigma_0^2 + m_0^2$ is more complicated. Here, direct evaluation of the periodogram $I_{NT_0}(\omega)$ cannot be used since in this case $I_{NT_0}(\omega)$ includes a

divergent series. This suggests that the corresponding spectral distribution function $F(\omega)$ is not absolutely continuous. Therefore it is reasonable to examine first of all the process $\mathring{X}(t) = X(t) - \langle X(t) \rangle$, which has an average spectral density of the form (4.524). Then one may evaluate the contribution from the non-zero mean value $\langle X(t) \rangle = m(t) = m_0 \sum_{n=-\infty}^{\infty} \Gamma(t - nT_0)$ to the average correlation function $B^{(0)}(\tau)$, which is given by

(4.525) $B^{(m)}(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_0^{T-\tau} m(t + \tau) m(t) dt.$

It is easy to see that $B^{(m)}(\tau)$ is a periodic positive definite function with period T_0. Therefore $B^{(m)}(\tau)$ is representable as a Fourier series in $\exp(ik\omega_0)$, $\omega_0 = 2\pi/T_0$, $k = 0, \pm1, \pm2, \ldots$, with nonnegative coefficients. Hence it follows that if $\langle X_n \rangle = m_0 \neq 0$, then the average spectral distribution of $X(t)$ consists in general, of a discrete (i.e. line) spectrum at frequencies $k\omega_0$, $k = 0, \pm1, \ldots$, and of a continuous spectrum given by a spectral density of the form (4.524).[184]

The case of a pulse-phase-modulated (i.e., pulse-position-modulated) signal

$$X(t) = \sum_{n=-\infty}^{\infty} \Gamma(t - nT_0 - \epsilon_n)$$

where ϵ_n are random variables, can be studied similarly. Here, too, as a rule, the average spectral distribution $F^{(0)}(\omega)$ includes a line spectrum concentrated at discrete frequencies $\omega = k\omega_0$, $k = 0, \pm1, \ldots$, and a continuous spectrum specified by the spectral density $f^{(0)}(\omega)$.[185]

As a final example we consider a simple model of a *phase-modulated sine oscillation*.

Example 4. Let

$$(4.526) \quad X(t) = C\cos\{\omega_0 t + X_0(t)\}$$

where $X_0(t)$ is a random process with stationary increments (in particular, it may also be stationary). In radioengineering applications it is usually reasonable to assume that fluctuations of the phase $X_0(t)$ are normally distributed; we shall also use this assumption. The principal attention will be given below to a special case of great practical importance, where $X_0(t)$ is Wiener's process, i.e. a normal random process such that

$$\langle X_0(t) \rangle = 0, \quad \langle X_0(t)X_0(s) \rangle = \sigma^2 \min(t,s).$$

Here

$$\begin{aligned}
B(t + \tau, t) &= C^2 \langle \cos\{\omega_0(t+\tau) + X_0(t+\tau)\}\cos\{\omega_0 t + X_0(t)\} \rangle \\
&= \frac{C^2}{4} \langle \left[e^{i\{\omega_0(t+\tau)+X_0(t+\tau)\}} \right. \\
&\quad + \left. e^{-i\omega_0(t+\tau)+X_0(t+\tau)\}} \right] \left[e^{i\{\omega_0 t+X_0(t)\}} + e^{-i\{\omega_0 t+X_0(t)\}} \right] \rangle \\
(4.527) \quad &= \frac{C^2}{4} \left[e^{i\omega_0(2t+\tau)} \langle e^{i\{X_0(t+\tau)+X_0(t)\}} \rangle \right. \\
&\quad + e^{i\omega_0\tau} \langle e^{i\{X_0(t+\tau)-X_0(t)\}} \rangle + e^{-i\omega_0\tau} \langle e^{-i\{X_0(t+\tau)-X_0(t)\}} \rangle \\
&\quad + \left. e^{-i\omega_0(2t+\tau)} \langle e^{-i\{X_0(t+\tau)+X_0(t)\}} \rangle \right].
\end{aligned}$$

We now take into account that if Y is a normal (Gaussian) random variable with mean value zero and variance σ_y^2, then

$$\langle e^{iY} \rangle = e^{-\sigma_y^2/2}$$

(cf. equation (0.11c') in Note 6 to the Introduction). Therefore, for $t \geqslant 0$, $\tau \geqslant 0$

$$(4.528) \quad B(t + \tau, t) = \frac{C^2}{4} \left\{ e^{i\omega_0(2t+\tau)-\sigma^2(4t+\tau)/2} + e^{i\omega_0\tau-\sigma^2\tau/2} \right.$$
$$\left. + e^{-i\omega_0\tau-\sigma^2\tau/2} + e^{-i\omega_0(2t+\tau^2)-\sigma^2(4t+\tau)/2} \right\}.$$

(4.528) implies that

$$\lim_{t\to\infty} B(t + \tau, t) = B^{(0)}(\tau)$$

exists for all $\tau \geqslant 0$ and

$$(4.529) \quad B^{(0)}(\tau) = \frac{1}{2}\, C^2 e^{-\sigma^2\tau/2}\, \cos\omega_0\tau.$$

Clearly, in this case

$$\lim_{T\to\infty} \frac{1}{T}\int_0^{T-\tau} B(t + \tau, t)dt = B^{(0)}(\tau)$$

i.e., $B^{(0)}(\tau)$ is the average correlation function of the phase-modulated oscillation (4.526). According to (2.102) the corresponding average spectral density is

$$(4.530) \quad f^{(0)}(\omega) = \frac{C^2\sigma^2}{8\pi} \left\{ \frac{1}{(\omega - \omega_0)^2 + \sigma^4/4} + \frac{1}{(\omega + \omega_0)^2 + \sigma^4/4} \right\}.$$

We see that the phase fluctuations produce a spreading of discrete spectral lines at $\omega = \omega_0$ and $\omega = -\omega_0$, which correspond to the random sine oscillation of a frequency ω_0, over a continuous frequency range.[186]

The case where the phase fluctuations $X_0(t)$ are described by an arbitrary normal process with stationary increments having the given structure function

$$(4.531) \quad \langle |X_0(t + \tau) - X_0(t)|^2 \rangle = D\{\tau\}$$

can be analyzed in quite a similar manner. Such an analysis shows[187] that, under wide conditions, which are nearly always met in practical applications, the average correlation function of the considered phase-modulated oscillation (4.526) has the form

(4.532) $B^{(0)}(\tau) = \dfrac{C^2}{2} e^{-D(\tau)/2} \cos\omega_0\tau.$

The average spectral density $f^{(0)}(\omega)$ corresponding to this correlation function is given by

(4.533) $f^{(0)}(\omega) = \dfrac{C^2}{4} \left\{ G(\omega - \omega_0) + G(\omega + \omega_0) \right\}$

where

(4.534) $G(\omega) = \dfrac{2}{\pi} \displaystyle\int_0^\infty e^{-D(\tau)/2} \cos\omega\tau\, d\tau.$

Formulae (4.532) − (4.534) clearly generalize (4.529) and (4.530).

A number of additional examples of the calculation of average correlation functions, spectral distribution functions, and spectral densities of nonstationary processes of class $\mathfrak{M}_2$ may be found in the available literature.[188] There are also many other topics concerning the correlation and spectral theory of random functions which are not considered at all in this book. However, it is impossible to cover all such topics, even briefly, in a single book; therefore, we conclude the book with the examples given above.

BIBLIOGRAPHY

The following list includes some textbooks and monographs on cognate subjects, a few reference books and mathematical texts containing the material used above, a number of classical papers of historical interest, and the books and papers directly referred to in the present book. A much more comprehensive bibliography can be found in Volume II.

A. Mathematical Reference Books and Books on Fourier Analysis

Abramowitz, M., and I. A. Stegun (1964). *Handbook of Mathematical Functions.* Nat. Bureau Stand., Washington, D.C.

Bochner, S. (1959). *Lectures on Fourier Integrals.* Princeton University Press, Princeton, N.J.

Erdélyi, A., W. Magnus, F. Oberhettinger, and F. G. Tricomi (1953). *Higher Transcendental Functions,* Vols. 1–3. McGraw–Hill, New York.

Goldberg, R. R. (1961). *Fourier Transforms.* Cambridge University Press, London.

Gradshteyn, I. S., and I. M. Ryzhik (1980). *Tables of Integrals, Series, and Products: Corrected and Enlarged Edition.* Academic Press, New York.

Jeffreys, H., Sir, and Lady Jeffreys (B. Swirles) (1950). *Methods of Mathematical Physics.* Cambridge University Press, Cambridge.

Oberhettinger, F. (1973). *Fourier Transforms of Distribution Functions and Their Inverses. A Collection of Tables.* Academic Press, New York.

Paley, R. E. A. C., and N. Wiener (1934). *Fourier Transforms in the Complex Domain.* Amer. Math. Soc., Providence, Rhode Island.

Watson, G. N. (1958). *A Treatise on the Theory of Bessel Functions.* Cambridge University Press, Cambridge.

Wiener, N. (1933). *The Fourier Integral and Certain of Its Applications.* Cambridge University Press, Cambridge, and Dover, New York.

B. Textbooks and Monographs on Probability Theory and Statistics

Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press, Princeton, N.J.

Cramér, H. (1962). *Random Variables and Probability Distributions.* 2nd ed. Cambridge University Press, Cambridge.

Feller, W. (1968,71). *An Introduction to Probability Theory and Its Applications.* Vols. 1 (3rd ed.) and 2 (2nd ed.). Wiley, New York.

Gnedenko, B.V. (1962). *The Theory of Probability.* Chelsea, New York.

Loève, M. (1977,78). *Probability Theory.* Vols. I & II. 4th ed. Springer-Verlag, New York.

Parsen, E. (1960). *Modern Probability Theory and Its Applications.* Wiley, New York.

Silvey, S. D. (1970). *Statistical Inference.* Chapman and Hall, London.

Wilks, S. S. (1962). *Mathematical Statistics.* Wiley, New York.

Yule, G. U., and M. G. Kendall (1950). *An Introduction to the Theory of Statistics.* 14th ed. Griffin, London.

C. Theory of Random Processes

Blanc–Lapierre, A., and R. Fortet (1953). *Théorie des fonctions aléatoires.* Masson, Paris. English translation: *Theory of Random Functions, Vols. 1 and 2,* Gordon and Breach, New York, 1965 and 1968.

Doob, J. L. (1953). *Stochastic Processes.* Wiley, New York.

Gikhman, I. I., and A. V. Skorokhod (1969). *Introduction to the Theory of Random Processes.* Saunders, Philadelphia.

Lamperti, J. (1977). *Stochastic Processes. A Survey of the Mathematical Theory.* Springer-Verlag, New York.

Parzen, E. (1962). *Stochastic Processes.* Holden–Day, San Francisco.

Wentzell, A. D. (1981). *A Course in the Theory of Stochastic Processes.* McGraw–Hill, New York.

Wong, E. (1971). *Stochastic Processes in Information and Dynamical Systems.* McGraw–Hill, New York.

D. Theory of Stationary Random Processes and Homogeneous Random Fields

Cramér, H., and M. R. Leadbetter (1967). *Stationary and Related Stochastic Processes.* Wiley, New York.

Lambert, P. J., and D. S. Poskitt (1983). *Stationary Processes in Time Series Analysis: The Mathematical Foundations.* Vandenhoeck and Ruprecht, Göttingen.

Rosenblatt, M. (1985). *Stationary Sequences and Random Fields.* Birkhäuser, Boston.

Rozanov, Yu. A. (1967). *Stationary Random Processes.* Holden-Day, San Francisco.

Vanmarcke, E. (1984). *Random Fields: Analysis and Synthesis.* MIT Press, Cambridge (Mass.).

Yadrenko, M. I. (1983). *Spectral Theory of Random Fields.* Optimization Software Inc., Publication Division, New York.

Yaglom, A. M. (1962). *An Introduction to the Theory of Stationary Random Functions.* Prentice–Hall, Englewood Cliffs, N.J.; also Dover, New York.

Yaglom, A. M. (1981). *Correlation Theory of Stationary Random Function with Examples from Meteorology* (in Russian). Leningrad, Gidrometeoizdat.

E. Engineering, Physical, and Economical Applications of Stationary and Related Random Functions

Bath, M. (1974). *Spectral Analysis in Geophysics.* Elsevier, Amsterdam.

Bendat, J. S. (1958). *Principles and Applications of Random Noise Theory.* Wiley, New York.

Box, G. E. P., and G. M. Jenkins (1970). *Time Series Analysis, Forecasting and Control.* Holden–Day, San Francisco.

Davenport, W. B., Jr., and W. L. Root (1958). *An Introduction to the Theory of Random Signals and Noise.* McGraw–Hill, New York.

Einstein, A. (1956). *Investigations on the Theory of the Brownian Movement,* ed. by R. Fürth. Dover, New York.

Granger, C. W. J., and M. Hatanaka (1964). *Spectral Analysis of Economic Time Series.* Princeton University Press, Princeton, N.J.

Khinchin, A. Ya. (1949). *Mathematical Foundations of Statistical Mechanics.* Dover, New York.

Laning, J. H., Jr., and R. H. Battin (1956). *Random Processes in Automatic Control.* McGraw–Hill, New York.

Malinvaud, E. (1969). *Méthodes statistiques de l'econométrie.* Dunod, Paris.

Moddleton, D. (1960). *Introduction to Statistical Communication Theory.* McGraw–Hill, New York.

Monin, A. S., and A. M. Yaglom (1971, 1975). *Statistical Fluid Mechanics, Vols. 1 and 2.* MIT Press, Cambridge (Mass.).

Pugachev, V. S. (1965). *Theory of Random Functions and Its Application to Control Problems.* Pergamon, Oxford.

Rytov, S. M. (1976). *Introduction to Statistical Radiophysics. Part 1: Random Processes.* Nauka, Moscow. (English translation is to be published by Springer-Verlag, New York.)

Rytov, S. M., Yu. A. Kravtsov, and V. I. Tatarskii (1978). *Introduction to Statistical Radiophysics. Part 2: Random Fields.* Nauka, Moscow. (English translation is to be published by Springer-Verlag, New York.)

Smoluchowski, M., von (1923). *Abhandlungen über die Brownsche Bewegung und verwandte Erscheinungen.* Ostwalds Kalssiker der exakten Wissenschaften Nr. 207, Leipzig.

Wiener, N. (1949). *Extrapolation, Interpolation and Smoothing of Stationary Time Series.* MIT Press, Cambridge (Mass.).

Wax, N. (ed.) (1954). *Selected Papers on Noise and Stochastic Processes.* Dover, New York.

F. Time Series Analysis: Statistical Inference for Stationary Random Functions

Anderson, T. W. (1970). *The Statistical Analysis of Time Series.* Wiley, New York.

Bendat, J. S., and A. G. Piersol (1986). *Random Data: Analysis and Measurement Procedures.* 2nd ed. Wiley–Interscience, New York.

Blackman, R. B., and J. W. Tukey (1959). *The Measurement of Power Spectra from the Point of View of Communications Engineering.* Dover, New York.

Brillinger, D. R. (1975). *Time Series. Data Analysis and Theory.* Holt, Rinehart, and Winston, New York.

Childers, D. G. (ed.) (1978). *Modern Spectrum Analysis.* IEEE Press, New York.

Elliott, D. F., and K. R. Rao (1982). *Fast Transforms. Algorithms, Analysis, Applications.* Academic Press, New York.

Grenander, U., and M. Rosenblatt (1956). *Statistical Analysis of Stationary Time Series.* Almqvist and Wiksell, Stockholm; also Wiley, New York, 1957 and (2nd ed.) Chelsea, New York, 1984.

Hannan, E. J. (1960). *Time Series Analysis.* Methuen, London.

Hannan, E. J. (1970). *Multiple Time Series.* Wiley, New York.

Haykin, S. (ed.) (1983). *Nonlinear Methods of Spectral Analysis.* 2nd ed. Springer–Verlag, Berlin.

Haykin, S. and J. A. Cadzow (eds.) (1982). *Spectral Estimation.* The special issue of Proc. IEEE, 70, No. 9.

Jenkins, G. M., and D. G. Watts (1968). *Spectral Analysis and Its Applications.* Holden–Day, San Francisco.

Koopmans, L. H. (1974). *The Spectral Analysis of Time Series.* Academic Press, New York.

Priestley, M. (1981). *Spectral Analysis and Time Series. Vols. 1 and 2.* Academic Press, London.

Ripley, B. D. (1981). *Spatial Statistics.* Wiley, New York.

Zhurbenko, I. G. (1982). *Spectral Analysis of Time Series* (in Russian). Izdat. MGU (Moscow State University Press), Moscow. English translation published by North–Holland, Amsterdam, 1986.)

G. Some Classical Papers of Historical Interest

Cramér, H. (1942). *On harmonic analysis in certain functional spaces.* Arkiv Mat. Astr. Fyzik, 28, No. 12, 1–7.

Einstein, A. (1914). *Méthode pour la détermination de valeurs statistiques d'observations concernant des grandeurs soumises à des fluctuations irrégulières.* Arch. Sci. Phys. et Natur., 37, Ser. 4, 254–256.

Karhunen, K. (1947). *Über lineare Methoden in der Wahrscheinlichkeitsrechnung.* Ann. Acad. Sci. Fennicae, Ser. A, I, No. 37, 3–79.

Khinchin, A. (1934). *Korrelationstheorie der stationären stochastischen Prozesse.* Math. Ann., 109, 604–615.

Kolmogorov, A. N. (1940). *Curves in Hilbert space which are invariant with respect to a one-parametric group of motions.* Dokl. Akad. Nauk SSSR, 26, No. 1, 6–9.

Kolmogorov, A. N. (1941). *Stationary sequences in Hilbert space.* Bull. MGU, ser. matem. (Bull. Moscow State Univ., Ser. math.), 2, No. 6, 1–40. English translation in *Linear Least-Square Estimation*, ed. by T. Kailath (Vol. 17 of "Benchmark Papers in Electrical Engineering and Computer Science"), pp. 66–89, Dowden, Hutchinson, and Ross, Stroudsburg, PA, 1977.

H. Books and Papers Referred to in the Book and Missing in the Previous Sections of Bibliography

Alisov, B. P., O. A. Drozdov, and E. S. Rubinstein (1952). *A Course in Climatology* (in Russian). Gidrometeoizdat, Leningrad.

Beveridge, W. H. (1921). *Weather and harvest cycles.* Econ. J., 31, 429–452.

Brown, T. J., and J. C. Hardin (1973). *A note on Kendall's autoregressive series.* J. Appl. Prob., 10, No. 2, 475–478.

Favre, A., J. Gaviglio, and R. Dumas (1954). *Quelques fonctions d'autocorrélation et derépartitions spectrales d'energie, pour la turbulence en aval des diverses grilles.* C.r. Acad. Sci., 238, No. 15, 1561–1563.

GARP Publication Series (1975). *The Physical Basis of Climate and Climate Modelling.* Report of the Intern. Study Confer. in Stockholm, 29 July – 10 August 1974. Global Atmosph. Res. Program Publ. Ser. No. 16, World Meteor. Organization, Geneva, Switzerland.

James, H. M., N. B. Nichols, and R. S. Phillips (eds.) (1947). *Theory of Servomechanisms.* McGraw–Hill, New York; also Dover, New York.

Jones, R. H. (1974). *Identification and autoregressive spectrum estimation.* IEEE Trans. Automat. Contr., 19, No. 6, 894–898.

Kendall, M. G. (1946). *Contributions to the Study of Oscillatory Time Series.* Cambridge University Press, Cambridge.

Kozhevnikov, V. A., and R. M. Meshcherskii (1963). *Modern Methods of Electroencephalogram Analysis* (in Russian). Medgiz, Moscow.

Otnes, R. K., and L. Enochson (1972). *Digital Time Series Analysis.* Wiley, New York.

Perrin, J. (1916). *Atoms.* Van Nostrand, Princeton, N.J.

Polyak, I. I. (1975). *Numerical Methods for Analysis of Observations* (in Russian). Gidrometeoizdat, Leningrad.

Privalsky, V. E. (1973). *Linear extrapolation of sea level variations by a parametric model.* Izv. Akad. Nauk SSSR, Fiz. Atmosf. i Okeana (Bull. Acad. Sci. USSR, Atmosph. Ocean. Phys.), 9, No. 3, 973–981.

Sheppard, P. A. (1951). *Atmospheric turbulence.* Weather, 6, No. 2, 42–49.

Taira, K., A. Takeda, and K. Ishikawa (1971). *A shipborne wave-recording system with digital data processing.* J. Oceanographical Soc. of Japan, 27, No. 4, 175–186.

Taylor, G. I. (1938). *The spectrum of turbulence.* Proc. Roy. Soc., Ser. A, 164, No. 919, 476–490.

Vinnik, L. P. (1968). *The Structure of Microseisms and Grouping Methods in Seismology* (in Russian). Nauka, Moscow.

Waldmeir, M. (1961). *The Sunspot Activity in the Years 1610–1960.* Schulthess, Zurich.

Wold, H. (1965). *A graphic introduction to stochastic processes.* In *Bibliography on Time Series and Stochastic Processes* (ed. by H. Wold), pp. 7–76, Oliver and Boyd, Edinburgh and London.

Yaglom, A. M. (1955). *Correlation Theory of Continuous Processes and Fields with Applications to Statistical Extrapolation of Time Series and to Turbulence Theory.* Doctor of Sci. Dissertation, Institute of Atmosph. Phys., Acad. Sci. of the USSR, Moscow.

INDEX

In the cases where an author is referred to in the text only implicitly (as et al.), the corresponding page is also indicated in the index after the author's name.