

Chapter 2

Basis of Probability Theory

Abstract We discuss the concept of a ‘random event’. The classical and statistical approaches used to formalize the notion of probability are described, along with the basic concepts of set theory and measure theory. The Kolmogorov approach for axiomatizing probability theory is presented. The probability space is introduced. The axioms of probability theory are presented, together with the addition and multiplication theorems. The notion of a scalar random variable is formalized. We present ways to describe a random variable in terms of the distribution function, probability density function, and moments, including in particular, the expectation and variance. Examples of scalar random variables with different distribution laws are presented. Methods for describing a scalar random variable are generalized to a vector random variable. The transformation of random variables and arithmetic operations on them are briefly examined.

2.1 The Concept of Random Phenomena

There are many different interpretations of the concept of a random phenomenon. Currently, there is no consensus on this issue, even among scientists. In everyday language the term *random phenomenon* usually refers to a *mass phenomenon* for which the results of observation cannot be predicted. In this context, one distinguishes *unpredictable events, values, processes, and fields*.

A more exact interpretation of this concept leads to probability theory. A key role is played by the concept of an *event*, that is, an already occurred or possible result (outcome) of an experiment or test. Note that one sometimes sees a distinction between an experiment and a test. An *experiment* is any observation, while a *test* is an observation under *controlled* (or *partly controlled*) conditions (situations). However, in the following we shall use these words synonymously.

A *random event* is any event observed under partly controlled conditions and characterized by a probability. A random event is an event that *has not happened yet*, but which *can occur with a certain probability*. An event that *has already*

This chapter is based on material from the books (Gorban 2003, 2016)

occurred is not considered as random. This remark also applies to any other random phenomenon.

In probability theory, a *random phenomenon* is any *mass phenomenon* (event, variable, process, or field) observed under partially controlled conditions and having a *probability measure*. Mass phenomena that *do not have a probability measure* are not considered as random.

2.2 Options for the Definition of Probability

Probability is a concept that is not as trivial as it seems at first glance. There are many options for its definition, and even several mathematical options. The best known variants are the classical, statistical, and axiomatic definitions.

2.2.1 Classical Approach

Apparently, the classical approach is historically the first. It is based on the idea of the incompatibility and equal possibility of a *finite number* of elementary events. We need to explain what is meant by ‘elementary’, ‘incompatible’, and ‘equally possible’ events.

Among events there are events that make up other events. They are called the *elementary events*. We denote the set of all elementary events by Ω . An event A occurs when any elementary event ω belonging to some set Ω_A occurs. Concerning elementary events ω from the set Ω_A , one says that these events *favor* (*foster*) the event A .

Let us consider a simple example. In a standard pack of playing cards there are thirty-six cards of four suits (hearts, diamonds, clubs, and spades). Each suit contains nine cards of different denominations (six, seven, eight, nine, ten, jack, queen, king, and ace). The result of choosing one of the thirty-six cards can be considered as an elementary event ω . In this case, the number of elementary events equals thirty-six. These elementary events form the set Ω .

A set of random events is not confined only to these elementary events. A random *non-elementary event*, for example, would be the result of choosing cards of a certain suit from the pack, for example, hearts. The nine hearts cards favor this event A , forming a set Ω_A . The event A occurs when one chooses any card from this set.

Incompatible events are ones that cannot occur together in the experiment. An example of incompatible events would be to choose a card of two different suits or

two different denominations. *Compatible (joint) events* are ones that can occur simultaneously in the experiment. An example of compatible events would be to choose a card of a definite suit and a definite denomination. These events can occur simultaneously in the experiment.

Equal possibility (likelihood) of elementary events means that none has priority over the others. So, in a well-shuffled pack of cards, the possibility of choosing any particular card is the same.

Another example of *incompatible elementary events* would be two numbers coming out together when tossing a dice with numbers from one to six on its sides. In the case of a symmetric (balanced) dice, the events consisting of particular numbers coming up are *equally possible elementary events*, whereas in an asymmetric dice, they are *not equally possible events*. An even (or odd) number in the roll of the dice is an example of a *non-elementary event*.

The *classical probability* $P(A)$ of an event A is the ratio of the number L_A of *incompatible equally possible elementary events* ω favorable to the event A to the total number L of elementary events: $P(A) = L_A/L$. In other words, in this case, the probability $P(A)$ of an event A is the ratio of the number L_A of elementary events belonging to the set Ω_A to the total number L of elementary events which belong to the set Ω .

In the example with cards, the total number of elementary events is $L = 36$, the number of elementary events favorable to choosing hearts (event A) is $L_A = 9$, and the probability of choosing hearts is $P(A) = 9/36 = 0.25$.

In the example with a perfectly balanced dice, the total number of elementary events is $L = 6$, the number of elementary events favorable to an even number is $L_A = 3$, and the probability of getting an even number is $P(A) = 3/6 = 0.5$.

The disadvantage with the classical definition of probability is the difficulty in selecting equally possible elementary events in asymmetric cases and the impossibility of extending this approach to the case of an infinite number of events.

2.2.2 Statistical Approach

A statistical approach avoids the difficulty in selecting equally possible elementary events. In this approach, the probability is determined on the basis of observation results. Such a probability is called statistical probability. The *statistical probability* $P(A)$ of a mass event A is a limit of the relative frequency $p_n(A)$ of the event when the number of tests n tends to infinity.

The statistical approach is of a physical nature and has a simple clear interpretation, and is thus widely used among physicists, engineers, and specialists in other applied fields. A zealous supporter of the statistical approach was *R. von Mises*, who proposed, on this basis, a variant axiomatization of probability theory (Mises 1919, 1964). Among mathematicians, the statistical approach has not found wide acceptance.

A significant disadvantage of this approach is that, in practice, the number of tests n is always limited, so it is in principle impossible to calculate the statistical probability simply by analyzing experimental results. One can only obtain a more or less accurate assessment.

When testing is carried out under the same statistical conditions, the accuracy of the assessment increases with the number of tests. However, this occurs only when the statistical experimental conditions remain unchanged. In real life, conditions do change, so it is impossible to achieve high accuracy when estimating the statistical probability.

It is possible to avoid these difficulties by defining the concept of probability without using statistical data. From a mathematical point of view, the most correct axiomatic definition of the notion of probability was suggested by A. N. Kolmogorov (1929, 1956, 1974). This approach is based on *set theory* and *measure theory*. Let us briefly consider the main concepts of these theories.

2.2.3 Main Concepts of Set Theory

A *set* is a collection of objects of an arbitrary nature. These objects are called *set members* or *set elements*. In mathematics, the concept of the set is the initial, *strictly undefinable* concept, *introduced axiomatically*.

A set can be assigned by listing its elements, for instance a, b, \dots, x . In this case one writes $\{a, b, \dots, x\}$. Often a set is assigned using some rule or contingency $R(x)$. Then one writes $\{x | R(x)\}$.

A *subset* A of the set Ω is a set all elements of which are elements of the set Ω . In this case, using *concatenation signs* \subset, \supset , we write $A \subset \Omega$ or $\Omega \supset A$. Note that the set Ω is itself a subset of the set Ω . If the set member x is contained in the set A , then using a *membership (adhesion) sign* \in , we write $x \in A$. By *definition*, the *empty set* \emptyset is a subset of *any set*. When counting the number of elements of a set, the empty set is *not taken into account*.

In the example of cards, the thirty-six cards in a pack form the set Ω . The subsets of it include for example the set Ω , six cards of a given suit, four cards of a given denomination, and so on.

Two sets A and B are said to be *equal* if they consist of the same elements (i.e., $A \subset B$ and $B \subset A$). For example, the set A of red suit cards and the set B of non-black suit cards are equal. Note that the sets $\{a, b, c\}$ and $\{c, a, b, c\}$ are considered equal because both consist of the same elements a, b, c .

If the number of set members is finite, then the set is said to be *finite*, while if the number is infinite, the set is said to be *infinite*. In the latter case, if the number of set members can be counted, then the set is said to be *countable*, and if the number of set members cannot be counted, the set is said to be *uncountable*. A *discrete set* is a set consisting of a finite or countable number of elements.

To characterize the number of elements of a set, the concept of *cardinal number* is used. *Two sets* have the same *cardinal number* if a *one-to-one correspondence*,

can be established among all their members, i.e., to every element of one set we can associate a unique element of the other set, and vice versa.

The cardinal number of a finite set is equal to the number of its elements. The *cardinal numbers of all countable sets* are the same (denoted by \aleph_0). The *cardinal number of all real numbers* is called the *continuum* (denoted by \aleph). Note that there are uncountable sets whose cardinal numbers are greater than the continuum.

Here are a few examples. An example of a finite set is a set whose members are the numeral 1, the number π , some geometric figure (for instance, a triangle), and the sign $*$. This set consists of four elements, so its cardinal number equals four. The sets of all integer numbers (from minus infinity to plus infinity) and positive integer numbers (from one to plus infinity) are countable sets. From the standpoint of set theory, the numbers of elements in these sets are identical. The cardinal numbers of these sets are \aleph_0 .

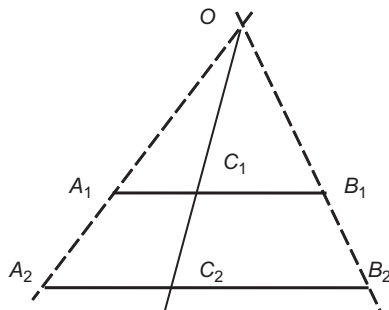
The cardinal numbers of points on the segments of any finite length, on an infinite line, and on the plane are the same. The cardinal numbers of all these sets equal to the continuum. These strange and, at first glance, unexpected statements are easy to prove. Let us show, for example, the statement that on two segments A_1B_1 and A_2B_2 of *different lengths* the numbers of points are *equal*.

To do this, we arrange these segments in parallel one above the other (Fig. 2.1) and draw the lines A_1A_2 and B_1B_2 . Their point of intersection is denoted by the letter O .

Then an arbitrary line OC_1 passing through the point O and intersecting the segment A_1B_1 intersects the segment A_2B_2 , and conversely, an arbitrary line OC_2 passing through the point O and intersecting the segment A_2B_2 intersects the segment A_1B_1 . Thus, we find a one-to-one correspondence between the points C_1 and C_2 of the segments A_1B_1 and A_2B_2 . If there is such a correspondence, the number of points on the segments are equal. Other parts of the above claim are proven similarly.

We now introduce the *union* (sum (logical addition)) \cup , *intersection* (logical multiplication) \cap , and *complement* \ operations. The union $A_1 \cup A_2$ of sets A_1 and A_2 is a set A that consists of all elements included in at least one of the sets A_1 and A_2 (Fig. 2.2). The intersection $A_1 \cap A_2$ of sets A_1 and A_2 is a set B containing precisely those elements included simultaneously in both A_1 and A_2 . The complement ΩA of

Fig. 2.1 One-to-one correspondence between the points C_1 and C_2 of the segments A_1B_1 and A_2B_2



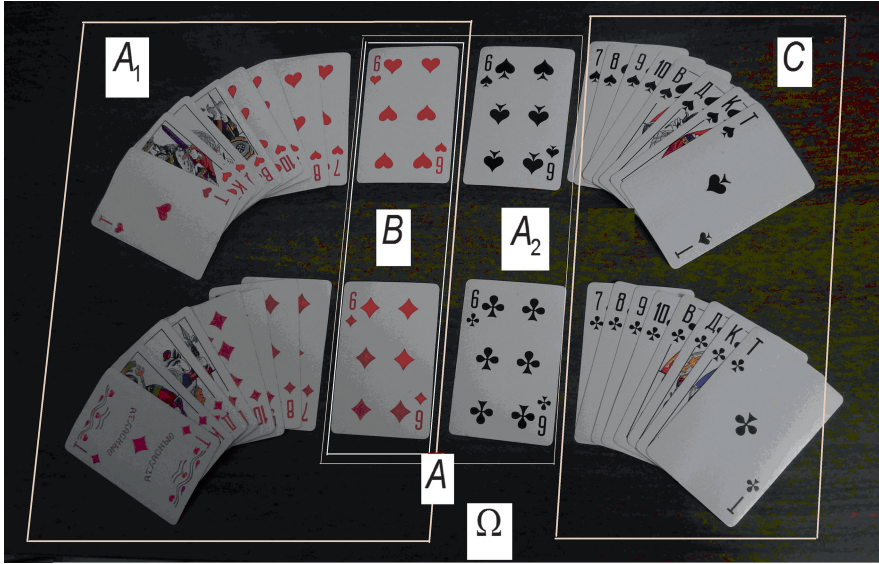


Fig. 2.2 Illustration of the union, intersection, and complement operations in the card example

the set $A \subset \Omega$ is the set $C = \bar{A}$ containing precisely those elements of the set Ω that do not belong to the set A .

For example, the union $A_1 \cup A_2$ of eighteen cards of red suit A_1 and four sixes A_2 (Fig. 2.2) forms the set A of twenty cards, namely, eighteen cards of red suit and two sixes of black suit. The intersection $A_1 \cap A_2$ of eighteen cards of red suit A_1 and four sixes A_2 forms a set B that consists of two sixes of red suit. The complement $\Omega \setminus A = C$ is a set of sixteen cards containing all cards of black suit, except two sixes.

Sets A_1 and A_2 are *disjoint* (*mutually exclusive*) if their intersection is the empty set: $A_1 \cap A_2 = \emptyset$. An *algebra of sets* is a class of subsets \mathfrak{F} of the set Ω with the operations of *union*, *intersection*, and *complement* such that the following three conditions are satisfied:

- (1) The set Ω belongs to the set \mathfrak{F} , i.e., $\Omega \in \mathfrak{F}$;
- (2) If a set $A \in \mathfrak{F}$, then its complement also belongs to the set \mathfrak{F} , i.e., $\bar{A} \in \mathfrak{F}$;
- (3) If sets A_1 and A_2 belong to the set \mathfrak{F} ($A_1 \in \mathfrak{F}$ и $A_2 \in \mathfrak{F}$), their union and intersection also belong to the set \mathfrak{F} , i.e., $A_1 \cup A_2 \in \mathfrak{F}$ and $A_1 \cap A_2 \in \mathfrak{F}$.

Thus, in an algebra of sets, the operations of union, intersection, and complement *do not take one outside the set \mathfrak{F}* , i.e., the set \mathfrak{F} is *closed* under these operations.

In the card example the set \mathfrak{F} can be formed from any combinations of cards.

A set Ω may be either finite (as is the case of the cards) or countable. If Ω is countable, the set \mathfrak{I} is also countable.

In addition to properties (1–3), a *countable set* \mathfrak{I} may have the following property:

- (4) If sets $A_n \in \mathfrak{I}$ for all $n = 1, 2, \dots$, then the union of a *countable number* of these sets also belongs to \mathfrak{I} , i.e., $\bigcup_{n=1}^{\infty} A_n \in \mathfrak{I}$.

Then the algebra is called a σ -algebra (*sigma-algebra*). Note that the fourth property, which plays an important role in measure theory and probability theory, is *not a consequence* of the third property.

2.2.4 Main Concepts of Measure Theory

A measure is a mathematical concept that generalizes the notion of length, area, volume, mass, energy, and other quantities that characterize a subset of a set *quantitatively*. A *measure* $\mu(A)$ is a function of the subset A of a non-empty set Ω which has the following properties:

- (1) The measure $\mu(A)$ has a non-negative value for all sets A on which it is defined, i.e., $\mu(A) \geq 0$. For the empty set \emptyset , the measure equals zero, i.e., $\mu(\emptyset) = 0$;
- (2) The measure of the set formed by union of a *countable number of disjoint subsets* A_n ($n = 1, 2, \dots$) is equal to the sum of the measures of the subsets A_n , i.e., $\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n)$.

The latter condition is called the *axiom of countable additivity*.

For any measure μ , we require that the measure be specified *on all subsets of a σ -algebra* (i.e., for all elements of the set \mathfrak{I}). If this condition is satisfied, we say that the triple $(\Omega, \mathfrak{I}, \mu)$ forms a *space with measure*.

The measure is said to be *normalized* if the measure of the set Ω is equal to unity (i.e., $\mu(\Omega) = 1$). For a normalized measure, $\mu(A) \leq 1$ for all sets $A \in \mathfrak{I}$.

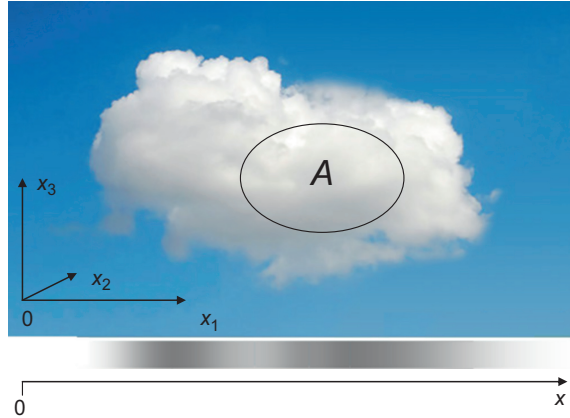
2.2.5 Axiomatic Definition of Probability

In the *axiomatic definition of probability*

- (1) There is a space Ω of elementary events $\omega \in \Omega$;
- (2) There is a σ -algebra \mathfrak{I}^1 of subsets, called *events*;
- (3) For any *event* A , there is a normalized measure $P(A)$, called a *probability measure*, or simply a *probability*.

¹More correctly, the smallest σ -algebra, called the *Borel σ -algebra*.

Fig. 2.3 Figurative representation of a probability space by a cloud



A space with a probability measure is called a *probability space*. Like any measure space, the probability space is given by a triple $(\Omega, \mathfrak{F}, P)$. The probability space $(\Omega, \mathfrak{F}, P)$ can be represented figuratively by a cloud consisting of a countable set Ω of not necessarily identical water drops (Fig. 2.3).

Every elementary event ω is represented by a water drop, the set Ω by the cloud, and the σ -algebra \mathfrak{F} by the set of all subsets formed from cloud droplets. The probability $P(A)$ of the subset A corresponds to the mass of water in the relevant part of the cloud, divided by the total mass of water in the cloud.

A *persistent (certain) event* I is an event taking place for all the elementary events $\omega \in \Omega$. An *impossible event* is an event that does not occur for any $\omega \in \Omega$.

Events A_1 and A_2 are *disjoint (mutually exclusive)* if their intersection is the empty set, i.e., $A_1 \cap A_2 = \emptyset$. Events A_1, A_2, \dots, A_I are *pairwise disjoint* if the intersection of any two different events is the empty set, i.e., if $A_i \cap A_j = \emptyset$ for any $i \neq j$ ($i, j = \overline{1, I}$).

The *probability* $P(A)$ is defined as a normalized measure by the following three *axioms*:

- (1) The probability of any event A is a non-negative number, i.e., $P(A) \geq 0$;
- (2) For pairwise disjoint events A_1, A_2, \dots (both finite and *countable*), the probability of their union is the sum of the probabilities of the events, i.e.,
$$P\left(\bigcup_n A_n\right) = \sum_n P(A_n);$$
- (3) The probability of the event Ω is equal to unity (i.e., $P(\Omega) = 1$).

2.2.6 Random Events

Events that satisfy the axioms of Sect. 2.2.5 are called *random events*. It follows from these axioms that the probability of any random event lies in the range

from zero to unity (i.e., $0 \leq P(A) \leq 1$), and the probability of an empty event is equal to zero (i.e., $P(\emptyset) = 0$). Note that if the probability of an event A is unity ($P(A) = 1$), it *does not generally follow* that this event is persistent, and if the probability of an event B is zero ($P(B) = 0$), it *does not follow* that it is an impossible event.

For general joint events A_1 and A_2 , the probability $P(A_1 \cap A_2)$ of the intersection $A_1 \cap A_2$ is equal to the product of the probability $P(A_1)$ of the event A_1 and the conditional probability $P(A_2/A_1)$ of the event A_2 under the condition that event A_1 has occurred²:

$$P(A_1 \cap A_2) = P(A_1)P(A_2/A_1). \quad (2.1)$$

This statement is known as the *multiplication theorem*.

Let us return to the card example (Fig. 2.2). Assume that the probability of choosing a certain card from the pack does not depend on its suit and denomination. Then this probability is $1/36$. Let event A_1 be the choice of any card of red suit and event A_2 the choice of a six. Then the event A_2/A_1 is the choice of a six from a half pack containing only red suit cards, and the event $A_1 \cap A_2 = B$ is the choice of a six of red suit. In this case $P(A_1) = 18 \times (1/36) = 1/2$, $P(A_2) = 4 \times (1/36) = 1/9$, and $P(A_2/A_1) = 2 \times (1/18) = 1/9$. According to (2.1), the probability of choosing a six of red suit is $P(B) = P(A_1 \cap A_2) = (1/2) \times (1/9) = 1/18$.

Random events A_1 and A_2 are said to be *independent* (*independent in probability*) if the probability of their intersection equals the product of their probabilities, i.e., $P(A_1 \cap A_2) = P(A_1)P(A_2)$. It follows from this definition that events A_1 and A_2 are independent if the appearance of one of them does not cause changes in the probability of occurrence of the other.

Let A_1 and A_2 be possibly *dependent events*. Then the probability of the sum of these events is the sum of the probabilities of the events A_1 and A_2 , minus the probability of their intersection:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2). \quad (2.2)$$

This easily proved formula is known as the *summing theorem*. Continuing with the card example, the probability of choosing a red card or any six $P(A_1 \cup A_2)$ can be calculated by (2.2): $P(A_1 \cup A_2) = 1/2 + 1/9 - 1/18 = 5/9$.

Using the simple rules above, one can solve not only trivial tasks such as those described above, but also some very difficult combinatorial problems.

²It is assumed that $P(A_1) \neq 0$. Otherwise, the probability $P(A_2/A_1)$ is not determined.

2.3 Random Variables

2.3.1 Basic Definitions

A *random variable* X is a measurable function defined on the space Ω of elementary events ω . The *value of a random variable* X can be represented by the function $x = \psi(\omega)$, where $\omega \in \Omega$. The set of values of the random variable forms the *value space of the random variable*. Before a test, it is *not known* what kind of elementary event will occur and therefore it is *not known* what value x a random variable X will take.

A random variable is given not only by the space of values of the random variable, but also characteristics which characterize the probabilities of these values. If a random variable is defined as the space of random values and the probabilities of these values, then we say that the random variable is *probabilistically defined*.

Random variables may be *scalar* or *vector*. Returning to the ‘cloud’ representation of the probability space (Fig. 2.3), a three-dimensional random variable $\vec{X} = (X_1, X_2, X_3)$ can be thought of as a random three-dimensional vector whose specific components x_1, x_2, x_3 describe the location of a particular water droplet in space, while a scalar random variable X is thought of as a scalar magnitude, describing the location of a particular water droplet on the axis x .

In the following, random variables will be denoted by *capital letters*, and their values by *small letters*. Random variables are described by different characteristics. The most complete description provides probability characteristics and a less complete description numerical characteristics (parameters).

2.3.2 Probabilistic Characteristics of a Scalar Random Variable

A comprehensive description of a scalar random variable X gives the *distribution function (cumulative distribution function)*³ $F_x(x)$ representing the probability that a random variable is less than x , i.e.,

$$F_x(x) = P\{X < x\}. \quad (2.3)$$

Examples of distribution functions of a *discrete random variable* X (taking discrete values x_i , $i = \overline{1, I}$) and continuous random variables are shown in Figs. 2.4a and 2.5a.

³If it is clear from the text which random variable the distribution function concerns, the subscript on the symbol is often omitted.

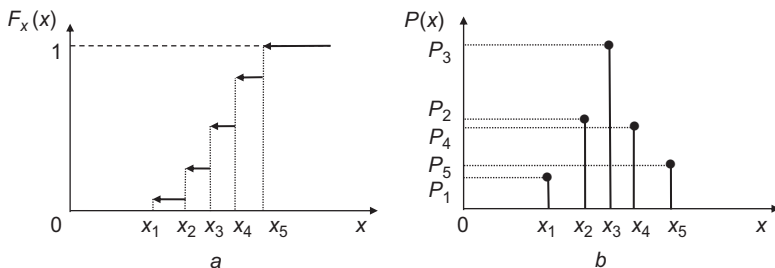


Fig. 2.4 The distribution function (a) and the probability mass function (b) of a discrete random variable

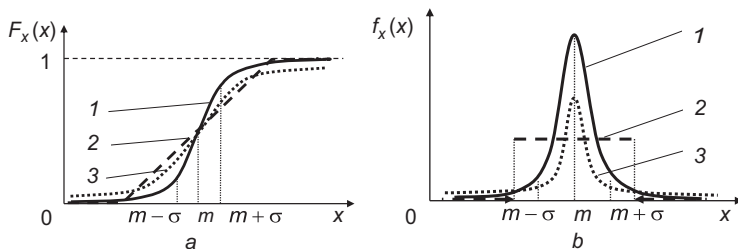


Fig. 2.5 Distribution functions (a) and probability densities (b) of continuous random variables: Gaussian distributed (1), uniformly distributed (2), and Cauchy distributed (3)

Each random variable is *uniquely described* by its distribution function $F_x(x)$. However, to a fixed distribution function there correspond in general a number of random variables.

The distribution function $F_x(x)$ takes *non-negative values*, lying in the range from zero to unity. This function is non-decreasing and continuous on the left (see Figs. 2.4a and 2.5a).

A random variable is *continuous* if its distribution function is continuous and has piecewise continuous derivative. Comprehensive description of such random variable also gives the *probability density function (probability distribution)*⁴ $f_x(x)$ (Fig. 2.5b). The probability distribution $f_x(x)$ and the distribution function $F_x(x)$ are *uniquely related* by

$$f_x(x) = \frac{dF_x(x)}{dx}, \quad (2.4)$$

$$F_x(x) = \int_{-\infty}^x f_x(x) dx. \quad (2.5)$$

⁴If it is clear from the text which random variable the distribution function concerns, the subscript on the symbol is often omitted.

Like the distribution function, the probability distribution takes only non-negative values. The area under the entire curve $f_x(x)$ equals unity (*normalization to unity*). The probability that a random variable takes a value in the interval $[x_1, x_2]$ is equals to the area under the curve $f_x(x)$ corresponding to this interval. This property is called the *additive property of the probability distribution*.

In the ‘cloud’ interpretation (Fig. 2.3), the probability distribution of a random variable X can be understood as the mass density distribution of water along the axis x . This distribution is imaged in the bottom part of the figure by a strip of showing the different densities on a grey scale.

Note that not all mathematical functions can be distribution functions or probability density functions. Those functions that have the properties listed above can be distribution functions and probability density functions of random variables, and only those.

2.3.3 Probabilistic Characteristics of a Discrete Random Variable

Since the distribution function of a discrete random variable is discontinuous, strictly speaking, a discrete random variable does not have a probability density function in the classical framework. The role of the probability density function in this case is played by the *probability mass function* (probability distribution) P_i , whose domain is all the x_i , $i = \overline{1, I}$ (Fig. 2.4b). Note that $\sum_{i=1}^I P_i = 1$. Using the generalized *Dirac delta function* for a discrete random variable, we can introduce the concept of probability density function.

The *Dirac delta function* (delta function) $\delta(x - x_0)$ is a function that possesses an infinite value at the point $x = x_0$ and zero at all other points on the real axis (Fig. 2.6a). Formally, it is a derivative of the unit step function at the point $x = x_0$ (Fig. 2.6b):

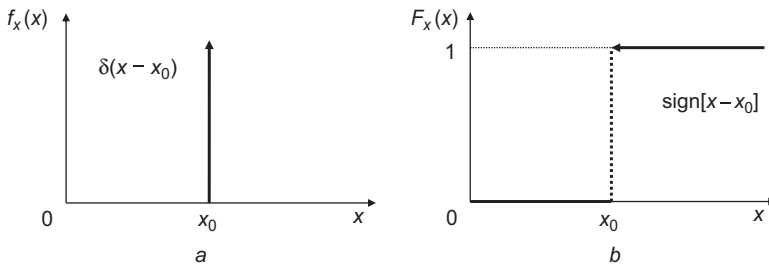


Fig. 2.6 Dirac delta function (a) and unit step function (b)

$$\text{sign}[x - x_0] = \begin{cases} 0 & \text{if } x \leq x_0, \\ 1 & \text{if } x > x_0. \end{cases} \quad (2.6)$$

An important property of the δ -function is that, for any function $\psi(x)$ whose domain is the set of all real numbers, we have

$$\int_{-\infty}^{\infty} \psi(x) \delta(x - x_0) dx = \psi(x_0).$$

The probability distribution of a discrete random variable taking the values x_1, \dots, x_l can be represented by the expression

$$f_x(x) = \{P_i \delta(x - x_i) | i = \overline{1, l}\}.$$

The use of the δ -function not only opens up the possibility of a unified description of discrete and continuous random variables, but also provides a way to describe determinate and random variables from a common position, namely viewing a determinate value x_0 *approximately* as a random variable X whose probability distribution $f_x(x)$ is a δ -function at the point x_0 (Fig. 2.6a) and whose distribution function $F_x(x)$ is the unit step function at this point (Fig. 2.6b).

2.3.4 Examples of Random Variables

Hundreds of different types of distributions have been investigated (Gubarev 1981, 1992; Muller et al. 1979), but in practice, as a rule, only a few of them are actually used. The most frequently used is the *Gaussian (normal) distribution*, whose probability density function $f_x(x)$ is described by a symmetric bell-shaped curve (continuous curve 1 in Fig. 2.5b):

$$f_x(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp \left[-\frac{(x - m_x)^2}{2\sigma_x^2} \right], \quad \sigma_x > 0. \quad (2.7)$$

This is defined by two parameters m_x and σ_x . The parameter m_x characterizes the position of the peak of the curve on the axis, and the parameter σ_x the width of the bell and its maximum value: the greater the value of σ_x , the wider the bell and the

smaller its maximum. Note that the parameter m_x is the *expectation* and the parameter σ_x the *standard deviation*.⁵

The Gaussian distribution function $F_x(x)$ cannot be expressed in terms of elementary functions. However, it can be calculated using the tabulated function called *Laplace function* or *probability integral*:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-z^2/2) dz, \quad (2.8)$$

which describes the distribution when $m_x = 0$ and $\sigma_x = 1$.

The expectation m_x characterizes the position of the curve $F_x(x)$ on the x axis and the standard deviation σ_x its slope: the smaller the standard deviation, the steeper the curve of the distribution function.

A *uniformly distributed* probability distribution $f_x(x)$ is often used. This is constant on an interval $[a, b]$ and zero outside this interval (dashed line 2 in Fig. 2.5b):

$$f_x(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b], \\ 0 & \text{if } x \notin [a, b]. \end{cases}$$

The distribution function $F_x(x)$ of such a random variable is a linear function within the interval $[a, b]$, increasing from zero to unity (dotted line 2 in Fig. 2.5a). The slope of this line segment is determined by the *range of the random variable*, that is, the difference between the upper b and lower a boundaries of its distribution. The smaller the range, the steeper the line.

The *Cauchy distribution* (*Student's⁶ distribution of the first order*) has interesting properties. We shall have more to say about these properties. But for now we note only that the distribution function $F_x(x)$ (curve 3 in Fig. 2.5a) and probability distribution $f_x(x)$ (curve 3 in Fig. 2.5b) of such a random variable are defined by two parameters, viz., the shift parameter x_0 and the scale parameter $\gamma > 0$:

$$F_x(x) = \frac{1}{\pi} \operatorname{arctg}\left(\frac{x - x_0}{\gamma}\right) + \frac{1}{2}, \quad (2.9)$$

$$f_x(x) = C[x_0, \gamma] = \frac{1}{\pi} \left[\frac{\gamma}{(x - x_0)^2 + \gamma^2} \right]. \quad (2.10)$$

These parameters in the Cauchy distribution play the same role as the expectation m_x and the standard deviation σ_x in the Gaussian distribution.

⁵These concepts are defined in the next subsection.

⁶*Student* is an alias of *W. S. Gosset*.

2.3.5 Numerical Parameters of Scalar Random Variables

In addition to the probabilistic characteristics, different *numerical characteristics* (*parameters*) can be used to provide an image of the distribution and, if necessary, to approximate it by other distributions. The best known among these are the *moments*, in particular the *expectation* m_x (the *moment of the first order*), the *variance* D_x (the *central moment of the second order*), and other parameters of the random variable.

The moments are defined by the concept of *mathematical expectation of a function*. By the expectation $E[\varphi(X)]$ of the function $\varphi(X)$ of a random variable X we mean the average (taking into account the probability distribution $f_x(x)$) of the function $\varphi(X)$:

$$E[\varphi(X)] = \int_{-\infty}^{\infty} \varphi(x)f_x(x)dx. \quad (2.11)$$

Note that the expectation of a function is a *functional*, the value of which is *determinate*, and $E[\cdot]$ in (2.11) is the *expectation operator* that acts on the function $\varphi(X)$.

Let us recall the concepts of *function*, *functional*, and *operator*. A rule assigning to each object x of a class A an object y of a class B is called a transformation (mapping) of class A to class B . The class A is the *applicable domain* and the class B the *actual range*.

If classes A and B are sets of numbers, then y is called the *function* of argument x , and if class A is a set of functions and class B is a set of numbers, then y is called a *functional*. Finally, if both classes A and B represent a set of functions, the transformation is called an *operator*. The most general concept is an operator. The concepts of function and functional are special cases of operators. The functional can be interpreted as a ‘function of a function’.

The *expectation* m_x of a random variable X is the expectation of the function $\varphi(X) = X$. In this way, $m_x = E[X]$. The *variance* D_x of a random variable X is the mathematical expectation of the function $\varphi(X) = (X - m_x)^2$, centered on the expectation m_x , i.e., $D_x = E[(X - m_x)^2] = \text{Var}[X]$, where $\text{Var}[X]$ is the variance operator. The *standard deviation* (SD) σ_x of a random variable X is the square root of the variance D_x , i.e., $\sigma_x = \sqrt{D_x}$.

The expectation m_x of a random variable X with the probability density $f_x(x)$ characterizes the *average value* of the variable, just as the variance D_x and standard deviation σ_x characterize the *scattering* of the random variable around the expectation m_x .

Sometimes, instead of the expectation and standard deviation one uses so-called *robust parameters*, such as the median and median absolute deviation. The *median* $e_x = \text{med}[X]$ of the random variable X is that value of the variable that divides the

area under the probability density function $f_x(x)$ in half, i.e., it is a solution of the equation

$$F_x(x) = 0.5.$$

If the distribution function is strictly increasing, this equation has a unique solution; otherwise it has a set of solutions. If the distribution is discrete, then one usually takes the *average of the two middle values* as the median.

The *median absolute deviation* is the median of the modulus of the random variable X relative to the median e_x :

$$s_x = \text{med}[|X - e_x|].$$

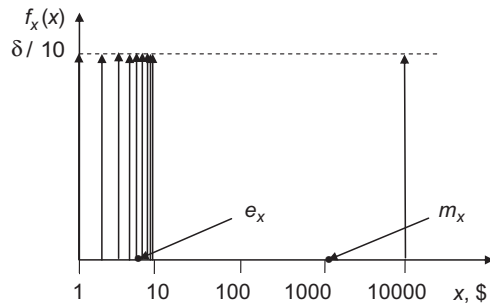
Note that there are many other useful parameters for the description of random variables, such as the *crude* and *central moments of order ν* defined by $m_\nu = E[X^\nu]$ and $\mu_\nu = E[(X - m_x)^\nu]$, respectively, the *coefficients of asymmetry, excess, variations*, etc., *quantiles, cumulants*, and others. They will not be used in the following, so we shall not dwell further on them.

2.3.6 Numerical Parameters of Various Random Variables

It follows from (2.7) that the Gaussian distribution is entirely determined by the expectation m_x and standard deviation σ_x (or variance D_x). The *odd central moments of this distribution are equal to zero, and the even ones are expressed in terms of the variance D_x* . The expectation of the *uniform distribution* is $m_x = (a + b)/2$, and the variance is $D_x = (b - a)^2/12$. A random variable that approximates the *determinate value x_0* has expectation m_x equal to x_0 and variance D_x equal to zero.

When the probability density function has *one maximum (unimodal)* the expectation and standard deviation are *particularly informative*. Let us consider an example. Gaussian random models often give an adequate description of actual measurement results of different physical quantities. The Gaussian distribution is unimodal. In this case, the estimate $m_x^* = y_n$ of the expectation m_x calculated by

Fig. 2.7 Probability density function of the bottle prices



averaging measurement results x_1, \dots, x_n [see (1.1)] is close to the measurand and the estimate $\sigma_x^* = z_n$ of the standard deviation σ_x described by (1.2) adequately characterizes the dispersion of the measurement results.

However, if the probability density function is not unimodal, these parameters can be *spurious*. Let us consider an example which, although not typical for probability theory, will allow us to illustrate the possibility of applying the theory to solve, not only probabilistic, but also determinate tasks, and also to demonstrate the usefulness of the robust parameters.

Imagine a grocer's that sells nine sorts of ordinary wine, with prices of \$1, \dots, \$9 a bottle and also a collection wine, priced at \$10,000 a bottle. Let us ask what is the average price of a bottle. Note that the price is a *determinate value* and the concept of average price is *not specified*, so the average can be regarded, for example, as the arithmetic mean, geometric mean, median, etc.

Solving the task by the methods of probability theory, the prices of the bottles can be described by a probability density function, represented by the ten δ -functions normalized by ten (Fig. 2.7). Calculation of the parameters gives the following results: the expectation is $m_x = \$1004.5$, the standard deviation $\sigma_x = \$3160.7$, the median $e_x = \$5.5$, and the median absolute deviation $s_x = \$2.5$.

Obviously, in this example, the expectation m_x and standard deviation σ_x are completely uninformative. At the same time, the robust parameters e_x and s_x have a clear interpretation: the median e_x characterizes the average price of the bottles which are in the middle of the series of prices (the middle of \$5 and \$6), and the median absolute deviation s_x characterizes the variation in prices in the region of this middle value. Note that, if the price of the collection wine bottle exceeds \$9, the median value and median absolute deviation are independent of this price.

Note also that *not all distributions have moments*. For instance, the Cauchy distribution does not have any.⁷ If the distribution has moments, the moments of the first two orders, i.e., the expectation and variance (or standard deviation), are often used instead of the distribution function and probability density function. If the distribution has no moments or, as in the latter example, they are not informative, the robust parameters are used.

2.4 Vector Random Variables

The statements of the previous section can be generalized to *vector random variables* (system of random variables). For simplicity, we begin our considerations with a two-dimensional variable represented by a column vector $\vec{X} = (X_1, X_2)^T$ with scalar random components X_1 and X_2 (system $\{X_1, X_2\}$ of scalar random variables), where T is the *transpose operator*.

⁷More correctly, for the Cauchy distribution, there is the principal value integral, which describes the first moment (first order moment). The value of this integral is x_0 .

2.4.1 Probabilistic Characteristics of a System of Two Random Variables

To describe a two-dimensional random variable $\vec{X} = (X_1, X_2)^T$, we use a *two-dimensional distribution function (cumulative distribution function)* $F_{\vec{x}}(x_1, x_2)$ representing the probability of the inequalities $\{X_1 < x_1, X_2 < x_2\}$:

$$F_{\vec{x}}(x_1, x_2) = P\{X_1 < x_1, X_2 < x_2\}, \quad (2.12)$$

and associated with it the *two-dimensional probability density function* $f_{\vec{x}}(x_1, x_2)$:

$$\frac{f_{\vec{x}}(x_1, x_2)}{\partial x_1 \partial x_2} = \frac{\partial^2 F_{\vec{x}}(x_1, x_2)}{\partial x_1 \partial x_2} \quad (2.13)$$

When solving practical tasks one often applies the conditional probability. The *conditional (marginal) distribution function of a random variable X_2 with the proviso $X_1 = x_1$* is a *one-dimensional distribution function* $F_{x_2/x_1}(x)$ defined with the proviso that the random variable X_1 has adopted a specific value x_1 .

According to (2.1), the probability distribution $f_{\vec{x}}(x_1, x_2)$ of a two-dimensional random variable $\vec{X} = (X_1, X_2)^T$ is equal to the product of the probability distribution $f_{x_1}(x_1)$ of the random variable X_1 and the conditional probability distribution $f_{x_2/x_1}(x_2)$ of the random variable X_2 :

$$f_{\vec{x}}(x_1, x_2) = f_{x_1}(x_1)f_{x_2/x_1}(x_2). \quad (2.14)$$

Random variables X_1 and X_2 are said to be independent if their joint probability distribution $f_{\vec{x}}(x_1, x_2)$ is equal to the product of the one-dimensional probability distributions $f_{x_1}(x_1)$ and $f_{x_2}(x_2)$ of the variables X_1 and X_2 , respectively:

$$f_{\vec{x}}(x_1, x_2) = f_{x_1}(x_1)f_{x_2}(x_2), \quad (2.15)$$

i.e., the conditional probability distribution $f_{x_2/x_1}(x_2)$ is equal to the unconditional probability distribution $f_{x_2}(x_2)$.

For independent random variables X_1 and X_2 , and only for these, the two-dimensional distribution function $F_{\vec{x}}(x_1, x_2)$ equals the product of the one-dimensional distribution functions $F_{x_1}(x_1)$ and $F_{x_2}(x_2)$:

$$F_{\vec{x}}(x_1, x_2) = F_{x_1}(x_1)F_{x_2}(x_2). \quad (2.16)$$

The independence of random variables X_1 and X_2 does not mean that these variables are in no way related. They may be linked, but on the probability level, their relationship does not become apparent.

2.4.2 Numerical Parameters of a System of Two Random Variables

The expectation $E[\varphi(X_1, X_2)]$ of a function $\varphi(X_1, X_2)$ of random variables X_1 and X_2 that have joint probability distribution $f_{\vec{x}}(x_1, x_2)$ is the average of the function $\varphi(X_1, X_2)$:

$$E[\varphi(X_1, X_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x_1, x_2) f_{\vec{x}}(x_1, x_2) dx_1 dx_2. \quad (2.17)$$

The mathematical expectation $\vec{m}_{\vec{x}} = (m_{x_1}, m_{x_2})^T$ of a two-dimensional random variable $\vec{X} = (X_1, X_2)^T$ is a vector, the components m_{x_1} and m_{x_2} of which are the expectations of the random variables X_1 and X_2 , i.e., $m_{x_1} = E[X_1]$ and $m_{x_2} = E[X_2]$.

The variance $\vec{D}_{\vec{x}} = (D_{x_1}, D_{x_2})^T = D[\vec{X}]$ of a two-dimensional random variable $\vec{X} = (X_1, X_2)^T$ is a vector whose components D_{x_1} and D_{x_2} are equal to the variances of the random variables X_1 and X_2 .

The standard deviation $\vec{\sigma}_{\vec{x}} = (\sigma_{x_1}, \sigma_{x_2})^T$ of a two-dimensional random variable $\vec{X} = (X_1, X_2)^T$ is a vector whose components σ_{x_1} and σ_{x_2} are equal to the square root of the corresponding variances D_{x_1} and D_{x_2} of the random variables X_1 and X_2 .

The expectation vector \vec{m}_x characterizes the average value of the vector \vec{X} . The variance vector \vec{D}_x and the standard deviation vector $\vec{\sigma}_{\vec{x}}$ characterize the scattering of the components of the vector \vec{X} along the x_1 and x_2 axes relative to the corresponding components of the vector \vec{m}_x .

The relationship between the random variables X_1 and X_2 is characterized by the product correlation (crude second order) moment

$$K_{x_1 x_2} = E[X_1 X_2]$$

and the product covariance (central second order) moment

$$R_{x_1 x_2} = \text{Cov}[X_1, X_2] = E[(X_1 - m_{x_1})(X_2 - m_{x_2})].$$

The product covariance moment $R_{x_1 x_2}$ normalized by the standard deviations σ_{x_1} and σ_{x_2} is called the correlation coefficient:

$$r_{x_1 x_2} = \frac{R_{x_1 x_2}}{\sigma_{x_1} \sigma_{x_2}}. \quad (2.18)$$

The product covariance moment $R_{x_1 x_2}$ is associated with the product correlation moment and the expectations m_{x_1} and m_{x_2} by the simple relation

$$R_{x_1 x_2} = K_{x_1 x_2} - m_{x_1} m_{x_2}. \quad (2.19)$$

The product covariance moment $R_{x_1x_2}$ characterizes the linear and only the linear relationship between the given random variables. If there is no linear connection, the product covariance moment $R_{x_1x_2}$ and the correlation coefficient $r_{x_1x_2}$ are both equal to zero. In this case, the random variables X_1 and X_2 are said to be *uncorrelated* (linearly independent).

Note that the noncorrelatedness and independence of random variables are *different concepts*. The independence of random variables *implies* their noncorrelatedness. However, in general, noncorrelatedness *does not imply* independence. These concepts differ when there is a *nonlinear relationship* between the random variables. The concepts of independence and noncorrelatedness coincide only in special cases, such as when the random variables X_1 and X_2 are jointly Gaussian.

The random variables X_1 and X_2 are said to be *orthogonal* if the product correlation moment $K_{x_1x_2}$ equals zero. The concepts of noncorrelatedness and orthogonality are *different notions*. But if at least one of the expectations m_{x_1} , m_{x_2} equals zero, then the orthogonality of the random variables implies their noncorrelatedness and noncorrelatedness implies their orthogonality [see (2.19)].

2.4.3 System of Two Jointly Gaussian Random Variables

As an example of a system of two random variables we consider a *system of two jointly Gaussian random variables* described by a *two-dimensional Gaussian (normal) distribution law*. The probability density function of such a system of jointly Gaussian random variables X_1 and X_2 has a bell-shaped form (Fig. 2.8) and is defined by five parameters, viz., the mathematical expectations m_{x_1} , m_{x_2} of the

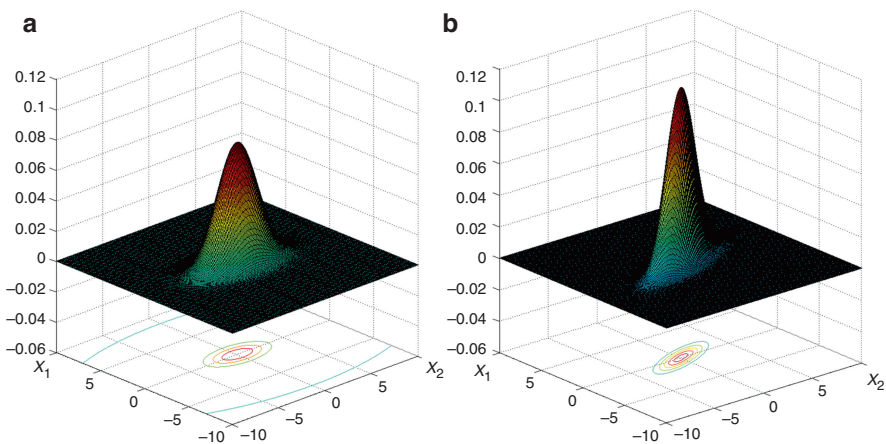


Fig. 2.8 Probability density function of a system of jointly Gaussian random variables X_1 and X_2 when the correlation coefficient $r_{x_1x_2} = 0$ (a) and $r_{x_1x_2} = 0.7$ (b) ($\sigma_{x_1} = 1$, $\sigma_{x_2} = 2$)

variables X_1 and X_2 , their standard deviations σ_{x_1} , σ_{x_2} , and their correlation coefficient $r_{x_1x_2}$:

$$f_{\vec{x}}(x_1, x_2) = \frac{1}{2\pi\sqrt{|R_{\vec{x}}|}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{m}_{\vec{x}})^T R_{\vec{x}}^{-1} (\vec{x} - \vec{m}_{\vec{x}})\right), \quad (2.20)$$

where $\vec{x} = (x_1, x_2)^T$ is a column vector whose components are the values of the random variables X_1 and X_2 ; $\vec{m}_{\vec{x}} = (m_{x_1}, m_{x_2})^T$ is a column vector of whose components are the expectations of the variables X_1 and X_2 ; $R_{\vec{x}}$ is a square covariance matrix

$$R_{\vec{x}} = \begin{vmatrix} \sigma_{x_1}^2 & r_{x_1x_2} \sigma_{x_1} \sigma_{x_2} \\ r_{x_1x_2} \sigma_{x_1} \sigma_{x_2} & \sigma_{x_2}^2 \end{vmatrix};$$

$R_{\vec{x}}^{-1}$ is the inverse of the matrix $R_{\vec{x}}$; and $|R_{\vec{x}}|$ is the determinant of the matrix $R_{\vec{x}}$.

The cut set of the probability density function on any fixed level C is the ellipse (*scattering ellipse*). Its center is at the point (m_{x_1}, m_{x_2}) and its sizes are determined by the standard deviations σ_{x_1} and σ_{x_2} , the correlation coefficient $r_{x_1x_2}$, and the value of the parameter C . Varying the parameter C , one can obtain scattering ellipses of different sizes, although the center and orientation of their axes coincide (Fig. 2.8).

If $m_{x_1} = m_{x_2} = 0$, the centers of the scattering ellipses are located at the coordinate origin. If the random variables are uncorrelated ($r_{x_1x_2} = 0$), the axes of the ellipses are oriented along the coordinate axes. If, in addition, the standard deviations are equal ($\sigma_{x_1} = \sigma_{x_2} = \sigma$), the ellipses degenerate into a circle of radius σC .

By a linear transformation of the coordinate axes, a system of correlated jointly Gaussian random variables X_1, X_2 can be reduced to a system of uncorrelated (independent) jointly Gaussian random variables Y_1, Y_2 with zero expectations. Note that a jointly Gaussian system of random variables *remains jointly Gaussian* under any *linear transformation* of the coordinates.

2.4.4 Characteristics and Parameters of a System of more than Two Random Variables

The above relations for a system of two random variables can be generalized to a system of $N > 2$ random variables. The *N-dimensional distribution function (cumulative distribution function) of an N-dimensional vector random variable (system of random variables)* is described by the expression

$$F_{\vec{x}}(\vec{x}) = F_{\vec{x}}(x_1, \dots, x_N) = P\{X_1 < x_1, \dots, X_N < x_N\}$$

and the corresponding *multidimensional probability density function (probability distribution)* by the expression

$$f_{\vec{x}}(\vec{x}) = f_{\vec{x}}(x_1, \dots, x_N) = \frac{\partial^N F_{\vec{x}}(x_1, \dots, x_N)}{\partial x_1 \dots \partial x_N}.$$

The dependence of the probability distribution $f_{\vec{x}}(\vec{x})$ on the value $\vec{x} = (x_1, x_2, x_3)^T$ of the three-dimensional random vector $\vec{X} = (X_1, X_2, X_3)^T$ can be represented figuratively by a cloud (see Fig. 2.3).

A *special case* of a multidimensional vector random variable \vec{X} is the *multidimensional Gaussian random variable* whose components have a *jointly Gaussian distribution*. In the N -dimensional case the probability distribution is

$$f_{\vec{x}}(\vec{x}) = \frac{1}{(2\pi)^{N/2} |R_{\vec{x}}|^{1/2}} \exp \left[-\frac{1}{2} (\vec{x} - \vec{m}_{\vec{x}})^T R_{\vec{x}}^{-1} (\vec{x} - \vec{m}_{\vec{x}}) \right], \quad (2.21)$$

where $\vec{x} = (x_1, \dots, x_N)^T$ is a column vector that describes the values of the random vector $\vec{X} = (X_1, \dots, X_N)^T$; $\vec{m}_{\vec{x}} = (m_{x_1}, \dots, m_{x_N})^T$ is an expectation column vector of the vector \vec{X} ; $R_{\vec{x}}$ is an $N \times N$ square *covariance matrix*:

$$R_{\vec{x}} = \begin{bmatrix} R_{x_1 x_1} & \dots & R_{x_1 x_N} \\ \dots & \dots & \dots \\ R_{x_N x_1} & \dots & R_{x_N x_N} \end{bmatrix},$$

$R_{x_n x_m}$ is the covariance moment of the variables X_n and X_m ($n, m = \overline{1, N}$).

Note that the diagonal elements of the matrix $R_{\vec{x}}$ are the variances of the random variables X_1, \dots, X_N ($R_{x_n x_n} = D_{x_n}, n = \overline{1, N}$).

2.5 Operations on Random Variables

The distribution function is an exhaustive characteristic of the random variable. On this basis, one often assumes that the *random variables* X_1, X_2 *are equal*, if their distribution functions $F_{x_1}(x), F_{x_2}(x)$ *coincide*, viz. $F_{x_1}(x) = F_{x_2}(x)$.

Note that from a mathematical point of view, such a definition of the equality of random variables is *not correct*. This is so because the random variable is uniquely described by the distribution function, but in general, the distribution function does not uniquely describe the random variable. As an example, *two different random variables* differing in sign have the same probability density function if it is symmetrical with respect to zero. In practice, one is not usually interested in random variables themselves, but in their *probabilistic and numerical characteristics*. Therefore, this mathematical incorrectness *can be ignored* when solving practical tasks.

Transformation of random variables leads to new random variables. Their probabilistic and numerical characteristics will in general differ from those of the original variables. From knowledge of the transformation function and the characteristics of the original random variables, one can calculate the characteristics of random variables obtained as a result of the conversion. The formulas are generally quite bulky, so we shall not present them here, but only give the formulas describing arithmetic operations. Let the initial operands be the random variables X_1 and X_2 described by a two-dimensional probability density function $f_{\vec{x}}(x_1, x_2)$ and the result of the operation be the random variable Y described by the probability density function $f_y(y)$. For *summation* of the variables, the probability density function is described by the expression

$$f_y(y) = \int_{-\infty}^{\infty} f_{\vec{x}}(y - x_2, x_2) dx_2, \quad (2.22)$$

for their *subtraction*,

$$f_y(y) = \int_{-\infty}^{\infty} f_{\vec{x}}(y + x_2, x_2) dx_2, \quad (2.23)$$

for their *multiplication*,

$$f_y(y) = \int_{-\infty}^{\infty} f_{\vec{x}}\left(\frac{y}{x_2}, x_2\right) \frac{dx_2}{|x_2|}, \quad (2.24)$$

and for their *division*,

$$f_y(y) = \int_{-\infty}^{\infty} |x_2| f_{\vec{x}}(yx_2, x_2) dx_2. \quad (2.25)$$

The expectation m_y and variance D_y of the *sum* of random variables X_1 and X_2 with expectations m_{x_1} , m_{x_2} , variances D_{x_1} , D_{x_2} , and covariance moment $R_{x_1x_2}$ are given by the expressions

$$m_y = m_{x_1} + m_{x_2}, \quad D_y = D_{x_1} + D_{x_2} + 2R_{x_1x_2},$$

while the expectation and variance of the *difference* of these random variables are given by the expressions

$$m_y = m_{x_1} - m_{x_2}, \quad D_y = D_{x_1} + D_{x_2} - 2R_{x_1x_2}.$$

References

- Gorban, I.I.: *Teoriya Ymovirnostey i Matematychna Statystika dla Naukovykh Pratsivnykiv ta Inzheneriv* (Probability Theory and Mathematical Statistics for Scientists and Engineers). IMMSP, NAS of Ukraine, Kiev (2003)
- Gorban, I.I.: *Sluchaynost i gipersluchaynost* (Randomness and Hyper-randomness). Naukova Dumka, Kiev (2016)
- Gubarev, V.V.: *Tablitci Kharakteristik Sluchainykh Velichin I Vektorov* (Tables of Characteristics of Random Variables and Vectors). Novosibirskiy elektrotekhnicheskyy institut, Rukopis deponirovana v VINITI, 3146-81, Novosibirsk (1981)
- Gubarev, V.V.: *Veroytnostnye modeli. Chast 1, 2.* (Probability models. Parts 1, 2). Novosibirskiy elektrotekhnicheskyy institut, Novosibirsk (1992)
- Kolmogorov, A.N.: *Obschaya teoriya mery i ischislenie veroyatnostey* (General measure theory and calculation of probability). In: *Proceedings of Communist Academy. Mathematics*, pp. 8–21 (1929)
- Kolmogorov, A.N.: *Foundations of the Theory of Probability*. Chelsea Publishing, New York (1956)
- Kolmogorov, A.N.: *Osnovnye Ponyatiya Teorii Veroyatnostey* (Fundamentals of Probability Theory). ONTI, Moscow (1974)
- Mises, R.: *Grundlagen der Wahrscheinlichkeitsrechnung*. Math. Z. **5**, 52–99 (1919)
- Mises, R.: *Mathematical Theory of Probability and Statistics*. Academic, New York (1964)
- Muller, P.H., Neumann, P., Storm, R.: *Tafeln der Mathematischen Statistic*. VEB Fachbuchverlag, Leipzig (1979)

Randomness and Hyper-randomness

Gorban, I.

2018, XXXII, 216 p. 30 illus., 23 illus. in color., Hardcover

ISBN: 978-3-319-60779-5