Chapter 3
Random Variables and Distribution Functions

Abstract Section 3.1 introduces the formal definitions of random variable and its distribution, illustrated by several examples. The main properties of distribution functions, including a characterisation theorem for them, are presented in Sect. 3.2. This is followed by listing and briefly discussing the key univariate distributions. The second half of the section is devoted to considering the three types of distributions on the real line and the distributions of functions of random variables. In Sect. 3.3 multivariate random variables (random vectors) and their distributions are introduced and discussed in detail, including the two key special cases: the multinomial and the normal (Gaussian) distributions. After that, the concepts of independence of random variables and that of classes of events are considered in Sect. 3.4, establishing criteria for independence of random variables of different types. The theorem on independence of sigma-algebras generated by independent algebras of events is proved with the help of the probability approximation theorem. Then the relationships between the introduced notions are extensively discussed. In Sect. 3.5, the problem of existence of infinite sequences of random variables is solved with the help of Kolmogorov's theorem on families of consistent distributions, which is proved in Appendix 2. Section 3.6 is devoted to discussing the concept of integral in the context of Probability Theory (a formal introduction to Integration Theory is presented in Appendix 3). The integrals of functions of random vectors are discussed, including the derivation of the convolution formulae for sums of independent random variables.

3.1 Definitions and Examples

Let $\langle \Omega, \mathfrak{F}, \mathbf{P} \rangle$ be an arbitrary probability space.

Definition 3.1.1 A *random variable* ξ is a measurable function $\xi = \xi(\omega)$ mapping $\langle \Omega, \mathfrak{F} \rangle$ into $\langle \mathbb{R}, \mathfrak{B} \rangle$, where $\mathbb{R}$ is the set of real numbers and $\mathfrak{B}$ is the σ-algebra of all Borel sets, i.e. a function for which the inverse image $\xi^{(-1)}(B) = \{\omega : \xi(\omega) \in B\}$ of any Borel set $B \in \mathfrak{B}$ is a set from the σ-algebra $\mathfrak{F}$.

A.A. Borovkov, *Probability Theory*, Universitext,
DOI 10.1007/978-1-4471-5201-9_3,

For example, when tossing a coin once, Ω consists of two points: heads and tails. If we put 1 in correspondence to heads and 0 to tails, we will clearly obtain a random variable.

The number of points showed up on a die will also be a random variable.

The distance between the origin to a point chosen at random in the square $[0 \le x \le 1, 0 \le y \le 1]$ will also be a random variable, since the set $\{(x, y) : x^2 + y^2 < t\}$ is measurable. The reader might have already noticed that in these examples it is very difficult to come up with a non-measurable function of ω which would be related to any real problem. This is often the case, but not always. In Chap. 18, devoted to random processes, we will be interested in sets which, generally speaking, are not events and which require special modifications to be regarded as events.

As we have already mentioned above, it follows from the definition of a random variable that, for any set B from the σ-algebra $\mathfrak{B}$ of Borel sets on the real line,

$$\xi^{(-1)}(B) = \{\omega : \xi(\omega) \in B\} \in \mathfrak{F}.$$

Hence one can define a probability $\mathbf{F}_\xi(B) = \mathbf{P}(\xi \in B)$ on the measurable space $\langle \mathbb{R}, \mathfrak{B} \rangle$ which generates the probability space $\langle \mathbb{R}, \mathfrak{B}, \mathbf{F}_\xi \rangle$.

Definition 3.1.2 The probability $\mathbf{F}_\xi(B)$ is called the *distribution of the random variable* ξ.

Putting $B = (-\infty, x)$ one obtains the function

$$F_\xi(x) = \mathbf{F}_\xi(-\infty, x) = \mathbf{P}(\xi < x)$$

defined on the whole real line which is called the *distribution function*[1] *of the random variable* ξ.

We will see below that the distribution function of a random variable completely specifies its distribution and is often used to describe the latter.

Where it leads to no confusion, we will write just $\mathbf{F}$, $F(x)$ instead of $\mathbf{F}_\xi$, $F_\xi(x)$, respectively. More generally, in what follows, as a rule, we will be using boldface letters $\mathbf{F}$, $\mathbf{G}$, $\mathbf{I}$, $\mathbf{\Phi}$, $\mathbf{K}$, $\mathbf{\Pi}$, etc. to denote distributions, and the standard font letters F, G, I, Φ, ... to denote the respective distribution functions.

Since a random variable ξ is a mapping of Ω into $\mathbb{R}$, one has $\mathbf{P}(|\xi| < \infty) = 1$. Sometimes, it is also convenient to consider along with such random variables random variables which can assume the values $\pm\infty$ (they will be measurable mappings of Ω into $\mathbb{R} \cup \{\pm\infty\}$). If $\mathbf{P}(|\xi| = \infty) > 0$, we will call such random variables $\xi(\omega)$ *improper*. Each situation where such random variables appear will be explicitly noted.

Example 3.1.1 Consider the Bernoulli scheme with success probability p and sample size k (see Sect. 3.3). As we know, the set of elementary outcomes Ω in this case

[1] In the English language literature, the distribution function is conventionally defined as $F_\xi(x) = \mathbf{P}(\xi \le x)$. The only difference is that, with the latter definition, F will be right-continuous, cf. property F3 below.

is the set of all k-tuples of zeros and ones. Take the σ-algebra $\mathfrak{F}$ to be the system of all subsets of Ω. Define a random variable on Ω as follows: to each k-tuple of zeros and ones we relate the number of ones in this tuple.

The probability of r successes is, as we already know,

$$P(r,k) = \binom{k}{r} p^r (1-p)^{k-r}.$$

Therefore the distribution function $F(x)$ of our random variable will be defined as

$$F(x) = \sum_{r<x} P(r,k).$$

Here the summation is over all integers r which are less than x. If $x \le 0$ then $F(x) = 0$, and if $x > k$ then $F(x) = 1$.

Example 3.1.2 Suppose we choose a point at random from the segment $[a,b]$, i.e. the probability that the chosen point is in a subset of $[a,b]$ is taken to be proportional to the Lebesgue measure of this subset. Here, Ω is the segment $[a,b]$, the σ-algebra $\mathfrak{F}$ is the class of Borel subsets of $[a,b]$. Define a random variable ξ by

$$\xi(\omega) = \omega, \quad \omega \in [a,b],$$

i.e. the value of the random variable is equal to the number from $[a,b]$ we have chosen. It is a measurable function. If $x \le a$, then $F(x) = \mathbf{P}(\xi < x) = 0$. Let $x \in (a,b]$. Then $\{\xi < x\}$ means that the point is in the interval $[a,x)$. The probability of this event is proportional to the length of the interval, hence

$$F(x) = \mathbf{P}(\xi < x) = \frac{x-a}{b-a}.$$

If $x > b$, then clearly $F(x) = 1$. Finally, we find that

$$F(x) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \le x \le b, \\ 1, & x > b. \end{cases} \tag{3.1.1}$$

This distribution function defines the so-called *uniform distribution* on the interval $[a,b]$.

If $\mu(B)$ is the Lebesgue measure on $\langle \mathbb{R}, \mathfrak{B} \rangle$, then, as we will see in the next section, it is not hard to show that in this case $\mathbf{F}_\xi(B) = \mu(B \cap [a,b])/(b-a)$.

3.2 Properties of Distribution Functions. Examples

3.2.1 The Basic Properties of Distribution Functions

Let $F(x)$ be the distribution function of a random variable ξ. Then $F(x)$ has the following properties:

F1. *Monotonicity*: *if* $x_1 < x_2$, *then* $F(x_1) \le F(x_2)$.
F2. $\lim_{x\to-\infty} F(x) = 0$ *and* $\lim_{x\to\infty} F(x) = 1$.
F3. *Left-continuity*: $\lim_{x\uparrow x_0} F(x) = F(x_0)$.

Proof Since for $x_1 \le x_2$ one has $\{\xi < x_1\} \subseteq \{\xi < x_2\}$, F1 immediately follows from property 3 of probability (see Sect. 3.2.2).

To prove F2, consider two number sequences $\{x_n\}$ and $\{y_n\}$ such that $\{x_n\}$ is decreasing and $x_n \to -\infty$, while $\{y_n\}$ is increasing and $y_n \to \infty$. Put $A_n = \{\xi < x_n\}$ and $B_n = \{\xi < y_n\}$. Since x_n tends monotonically to $-\infty$, the sequence of sets A_n decreases monotonically to $\bigcap A_n = \varnothing$. By the continuity axiom (see Sect. 3.2.1), $\mathbf{P}(A_n) \to 0$ as $n \to \infty$ or, which is the same, $\lim_{n\to\infty} F(x_n) = 0$. This and the monotonicity of $F(x)$ imply that

$$\lim_{x\to-\infty} F(x) = 0.$$

Since the sequence $\{y_n\}$ tends monotonically to ∞, the sequence of sets B_n increases to $\bigcup B_n = \Omega$, and hence (see property 9 in Sect. 3.2.2) $\mathbf{P}(B_n) \to 1$. This implies, as above, that

$$\lim_{n\to\infty} F(y_n) = 1, \qquad \lim_{x\to\infty} F(x) = 1.$$

Property F3 is proved in a similar way. Let $\{x_n\}$ be an increasing sequence with $x_n \uparrow x_0$,

$$A = \{\xi < x_0\}, \qquad A_n = \{\xi < x_n\}.$$

The sequence of sets A_n also increases, and $\bigcup A_n = A$. Therefore, $\mathbf{P}(A_n) \to \mathbf{P}(A)$. This means that

$$\lim_{x\uparrow x_0} F(x) = F(x_0). \qquad \square$$

It is not hard to see that the function F would be *right-continuous* if we put $F(x) = \mathbf{P}(\xi \le x)$.

With our definition, the function F is generally speaking not right-continuous, since by the continuity axiom

$$\begin{aligned} F(x+0) - F(x) &= \lim_{n\to\infty}\left(F\left(x+\frac{1}{n}\right) - F(x)\right) \\ &= \lim_{n\to\infty} \mathbf{P}\left(x \le \xi < x+\frac{1}{n}\right) = \mathbf{P}\left(\bigcap_{n=1}^{\infty}\left\{\xi \in \left[x, x+\frac{1}{n}\right)\right\}\right) \\ &= \mathbf{P}(\xi = x). \end{aligned}$$

This means that $F(x)$ is continuous if and only if $\mathbf{P}(\xi = x) = 0$ for any x. Examples 3.1.1 and 3.1.2 show that both continuous and discontinuous $F(x)$ are quite common.

From the above relations it also follows that

$$\mathbf{P}(x \le \xi \le y) = \mathbf{F}_\xi\big([x, y]\big) = F(y+0) - F(x).$$

Theorem 3.2.1 *If a function $F(x)$ has properties* F1, F2 *and* F3, *then there exist a probability space $\langle \Omega, \mathfrak{F}, \mathbf{P}\rangle$ and a random variable ξ such that $F_\xi(x) = F(x)$.*

Proof First we construct a probability space $\langle \Omega, \mathfrak{F}, \mathbf{P}\rangle$. Take Ω to be the real line $\mathbb{R}$, $\mathfrak{F}$ the σ-algebra $\mathfrak{B}$ of Borel sets. As we already know (see Sect. 3.2.1), to construct a probability space $\langle \mathbb{R}, \mathfrak{B}, \mathbf{P}\rangle$ it suffices to define a probability on the algebra $\mathcal{A}$ generated, say, by the semi-intervals of the form $[\cdot,\cdot)$ (then $\sigma(\mathcal{A}) = \mathfrak{B}$). An arbitrary element of the algebra $\mathcal{A}$ has the form of a finite union of disjoint semi-intervals:

$$A = \bigcup_{i=1}^{n} [a_i, b_i), \quad a_i < b_i$$

(the values of a_i and b_i can be infinite). We define

$$\mathbf{P}(A) = \sum_{i=1}^{n} \big(F(b_i) - F(a_i)\big).$$

It is absolutely clear that axioms P1 and P2 are satisfied by virtue of F1 and F2. It remains to verify the countable additivity, or continuity, of $\mathbf{P}$ on the algebra $\mathcal{A}$. Let $B_n \in \mathcal{A}$, $B_{n+1} \subset B_n$, $\bigcap_{n=1}^{\infty} B_n = B \in \mathcal{A}$. One has to show that $\mathbf{P}(B_n) \to \mathbf{P}(B)$ as $n \to \infty$ or, which is the same, that $\mathbf{P}(B_n \overline{B}) \to 0$ ($B_n \overline{B} \in \mathcal{A}$). To this end, it suffices to prove that, for any fixed N, $\mathbf{P}(B_n \overline{B} C_N) \to 0$, where $C_N = [-N, N)$. Indeed, for any given $\varepsilon > 0$, by virtue of F2 we can choose an N such that $\mathbf{P}(\overline{C}_N) < \varepsilon$. Then $\mathbf{P}(B_n \overline{B}\, \overline{C}_N) \le \mathbf{P}(\overline{C}_N) < \varepsilon$ and

$$\limsup_{n\to\infty} \mathbf{P}(B_n \overline{B}) \le \limsup_{n\to\infty} \mathbf{P}(B_n \overline{B} C_N) + \varepsilon.$$

Since ε is arbitrary, the convergence $\mathbf{P}(B_n \overline{B} C_N) \to 0$ as $n \to \infty$ implies the required convergence $\mathbf{P}(B_n \overline{B}) \to 0$. It follows that we can assume that the sets B_n are bounded ($B_n \subset [-N, N)$ for some $N < \infty$). Moreover, we can assume without loss of generality that B is the empty set.

By the above remarks, B_n admits the representation

$$B_n = \bigcup_{i=1}^{k_n} \big[a_i^n, b_i^n\big), \quad k_n < \infty,$$

where a_i^n, b_i^n are finite. Further note that, for a given $\varepsilon > 0$ and any semi-interval $[a, b)$, one can always find an embedded interval $[a, b - \delta)$, $\delta > 0$, such that $\mathbf{P}([a, b - \delta)) \ge \mathbf{P}([a, b)) - \varepsilon$. This follows directly from property F3: $F(b - \delta) \to F(b)$ as $\delta \downarrow 0$. Hence, for a given $\varepsilon > 0$ and set B_n, there exist $\delta_i^n > 0$, $i = 1, \dots, k_n$, such that

$$\widetilde{B}_n = \bigcup_{i=1}^{k_n} \big[a_i^n, b_i^n - \delta_i^n\big) \subset B_n, \qquad \mathbf{P}(\widetilde{B}_n) > \mathbf{P}(B_n) - \varepsilon 2^{-n}.$$

Now add the right end points of the semi-intervals to the set $\widetilde{B}_n$ and consider the closed bounded set

$$K_n = \bigcup_{i=1}^{k_n} \left[a_i^n, b_i^n - \delta_i^n\right].$$

Clearly,

$$\widetilde{B}_n \subset K_n \subset B_n, \qquad K = \bigcap_{n=1}^{\infty} K_n = \varnothing,$$

$$\mathbf{P}(B_n - K_n) = \mathbf{P}(B_n \overline{K}_n) \le \varepsilon 2^{-n}.$$

It follows from the relation $K = \varnothing$ that $K_n = \varnothing$ for all sufficiently large n. Indeed, all the sets K_n belong to the closure $[C_N] = [N, -N]$ which is compact. The sets $\{\Delta_n = [C_N] - K_n\}_{n=1}^{\infty}$ form an *open covering* of $[C_N]$, since

$$\bigcup_n \Delta_n = [C_N]\left(\bigcup_n \overline{K}_n\right) = [C_N]\left(\overline{\bigcap_n K_n}\right) = [C_N].$$

Thus, by the Heine–Borel lemma there exists a *finite subcovering* $\{\Delta_n\}_{n=1}^{n_0}$, $n_0 < \infty$, such that $\bigcup_{n=1}^{n_0} \Delta_n = [C_N]$ or, which is the same, $\bigcap_{n=1}^{n_0} K_n = \varnothing$. Therefore

$$\mathbf{P}(B_{n_0}) = \mathbf{P}\left(B_{n_0}\left(\overline{\bigcap_{n=1}^{n_0} K_n}\right)\right) = \mathbf{P}\left(B_{n_0}\left(\bigcup_{n=1}^{n_0} \overline{K}_n\right)\right)$$

$$= \mathbf{P}\left(\bigcup_{n=1}^{n_0} B_{n_0}\overline{K}_n\right) \le \mathbf{P}\left(\bigcup_{n=1}^{n_0} B_n \overline{K}_n\right) \le \sum_{n=1}^{n_0} \varepsilon 2^{-n} < \varepsilon.$$

Thus, for a given $\varepsilon > 0$ we found an n_0 (depending on ε) such that $\mathbf{P}(B_{n_0}) < \varepsilon$. This means that $\mathbf{P}(B_n) \to 0$ as $n \to \infty$. We proved that axiom P3 holds.

So we have constructed a probability space. It remains to take ξ to be the identity mapping of $\mathbb{R}$ onto itself. Then

$$F_\xi(x) = \mathbf{P}(\xi < x) = \mathbf{P}(-\infty, x) = F(x). \qquad \square$$

The model of the *sample probability space* based on the assertion just proved is often used in studies of distribution functions.

Definition 3.2.1 A probability space $\langle \Omega, \mathfrak{F}, \mathbf{F}\rangle$ is called a *sample space* for a random variable $\xi(\omega)$ if Ω is a subset of the real line $\mathbb{R}$ and $\xi(\omega) \equiv \omega$.

The probability $\mathbf{F} = \mathbf{F}_\xi$ is called, in accordance with Definition 3.1.1 from Sect. 3.1, the *distribution* of ξ. We will write this as

$$\xi \Subset \mathbf{F}. \tag{3.2.1}$$

It is obvious that constructing a sample probability space is always possible. It suffices to put $\Omega = \mathbb{R}$, $\mathfrak{F} = \mathfrak{B}$, $\mathbf{F}(B) = \mathbf{P}(\xi \in B)$. For integer-valued variables

ξ the space $\langle \Omega, \mathfrak{F} \rangle$ can be chosen in a more "economical" way by taking $\Omega = \{\ldots, -1, 0, \ldots\}$.

Since by Theorem 3.2.1 the distribution function $F(x)$ of a random variable ξ uniquely specifies the distribution $\mathbf{F}$ of this random variable, along with (3.2.1) we will also write $\xi \Subset F$.

Now we will give examples of some of the most common distributions.

3.2.2 The Most Common Distributions

1. *The degenerate distribution* $\mathbf{I}_a$. The distribution $\mathbf{I}_a$ is defined by

$$\mathbf{I}_a(B) = \begin{cases} 0 & \text{if } a \in B, \\ 1 & \text{if } a \notin B. \end{cases}$$

This distribution is concentrated at the point a: if $\xi \Subset \mathbf{I}_a$, then $\mathbf{P}(\xi = a) = 1$. The distribution function of $\mathbf{I}_a$ has the form

$$F(x) = \begin{cases} 0 & \text{for } x \le a, \\ 1 & \text{for } x > a. \end{cases}$$

The next two distributions were described in Examples 3.1.1 and 3.1.2 of Sect. 3.1.

2. *The binomial distribution* $\mathbf{B}_p^n$. By the definition, $\xi \Subset \mathbf{B}_p^n$ ($n > 0$ is an integer, $p \in (0, 1)$) if $\mathbf{P}(\xi = k) = \binom{n}{k} p^k (1-p)^{n-k}$, $0 \le k \le n$. The distribution $\mathbf{B}_p^1$ will be denoted by $\mathbf{B}_p$.

3. *The uniform distribution* $\mathbf{U}_{a,b}$. If $\xi \Subset \mathbf{U}_{a,b}$, then

$$\mathbf{P}(\xi \in B) = \frac{\mu(B \cap [a, b])}{\mu([a, b])},$$

where μ is the Lebesgue measure. We saw that this distribution has distribution function (3.1.1).

The next distribution plays a special role in probability theory, and we will encounter it many times.

4. *The normal distribution* $\mathbf{\Phi}_{\alpha,\sigma^2}$ (the normal or Gaussian law). We will write $\xi \Subset \mathbf{\Phi}_{\alpha,\sigma^2}$ if

$$\mathbf{P}(\xi \in B) = \mathbf{\Phi}_{\alpha,\sigma^2}(B) = \frac{1}{\sigma\sqrt{2\pi}} \int_B e^{-(u-\alpha)^2/(2\sigma^2)}\, du. \tag{3.2.2}$$

The distribution $\mathbf{\Phi}_{\alpha,\sigma^2}$ depends on two parameters: α and $\sigma > 0$. If $\alpha = 0, \sigma = 1$, the normal distribution is called *standard*. The distribution function of $\mathbf{\Phi}_{0,1}$ is equal to

$$\Phi(x) = \mathbf{\Phi}_{0,1}\big((-\infty, x)\big) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du.$$

The distribution function of $\mathbf{\Phi}_{\alpha,\sigma^2}$ is obviously equal to $\Phi((x-\alpha)/\sigma)$, so that the parameters α and σ have the meaning of the "location" and "scale" of the distribution.

The fact that formula (3.2.2) defines a distribution follows from Theorem 3.2.1 and the observation that the function $\Phi(x)$ (or $\Phi((x-a)/\sigma)$) satisfies properties F1–F3, since $\Phi(-\infty)=0$, $\Phi(\infty)=1$, and $\Phi(x)$ is continuous and monotone. One could also directly use the fact that the integral in (3.2.2) is a countably additive set function (see Sect. 3.6 and Appendix 3).

5. *The exponential distribution* $\mathbf{\Gamma}_\alpha$. The relation $\xi \Subset \mathbf{\Gamma}_\alpha$ means that ξ is nonnegative and

$$\mathbf{P}(\xi \in B) = \mathbf{\Gamma}_\alpha(B) = \alpha \int_{B\cap(0,\infty)} e^{-\alpha u}\, du.$$

The distribution function of $\xi \Subset \mathbf{\Gamma}_\alpha$ clearly has the form

$$\mathbf{P}(\xi < x) = \begin{cases} 1 - e^{-\alpha x} & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases}$$

The exponential distribution is a special case of the gamma distribution $\mathbf{\Gamma}_{\alpha,\lambda}$, to be considered in more detail in Sect. 7.7.

6. A discrete analogue of the exponential distribution is called the *geometric distribution*. It has the form

$$\mathbf{P}(\xi = k) = (1-p)p^k, \quad p \in (0,1),\ k = 0, 1, \ldots$$

7. *The Cauchy distribution* $\mathbf{K}_{\alpha,\sigma}$. As was the case with the normal distribution, this distribution depends on two parameters α and σ which are also location and scale parameters. If $\xi \Subset \mathbf{K}_{\alpha,\sigma}$ then

$$\mathbf{P}(\xi \in B) = \frac{1}{\pi\sigma} \int_B \frac{du}{1 + ((u-a)/\sigma)^2}.$$

The distribution function $K(x)$ of $\mathbf{K}_{0,1}$ is

$$K(x) = \frac{1}{\pi} \int_{-\infty}^{x} \frac{du}{1+u^2}.$$

The distribution function of $\mathbf{K}_{\alpha,\sigma}$ is equal to $K((x-\alpha)\sigma)$. All the remarks made for the normal distribution continue to hold here.

Example 3.2.1 Suppose that there is a source of radiation at a point (α, σ), $\sigma > 0$, on the plane. The radiation is registered by a detector whose position coincides with the x-axis. An emitted particle moves in a random direction distributed uniformly over the circle. In other words, the angle η between this direction and the vector $(0,-1)$ has the uniform distribution $\mathbf{U}_{-\pi,\pi}$ on the interval $[-\pi,\pi]$. Observation results are the coordinates $\xi_1, \xi_2, \ldots$ of the points on the x-axis where the particles interacted with the detector. What is the distribution of the random variable $\xi = \xi_1$?

To find this distribution, consider a particle emitted at the point (α,σ) given that the particle hit the detector (i.e. given that $\eta \in [-\pi/2, \pi/2]$). It is clear that the conditional distribution of η given the last event (of which the probability is $\mathbf{P}(\eta \in [-\pi/2,\pi/2]) = 1/2$) coincides with $\mathbf{U}_{-\pi/2,\pi/2}$. Since $(\xi-\alpha)/\sigma = \tan\eta$, one obtains that

$$\mathbf{P}(\xi < x) = \mathbf{P}(\alpha + \sigma \tan\eta < x)$$
$$= \mathbf{P}\left(\frac{\eta}{\pi} < \frac{1}{\pi}\arctan\frac{x-\alpha}{\sigma}\right) = \frac{1}{2} + \frac{1}{\pi}\arctan\frac{x-\alpha}{\sigma}.$$

Recalling that $(\arctan u)' = 1/(1+u^2)$, we have

$$\arctan x = \int_0^x \frac{du}{1+u^2} = \int_{-\infty}^x \frac{du}{1+u^2} - \frac{\pi}{2},$$
$$\mathbf{P}(\xi < x) = \frac{1}{\pi}\int_{-\infty}^{(x-\alpha)/\sigma} \frac{du}{1+u^2} = K\left(\frac{x-\alpha}{\sigma}\right).$$

Thus the coordinates of the traces on the x-axis of the particles emitted from the point (α, σ) have the Cauchy distribution $\mathbf{K}_{\alpha,\sigma}$.

8. *The Poisson distribution* $\mathbf{\Pi}_\lambda$. We will write $\xi \in \mathbf{\Pi}_\lambda$ if ξ assumes nonnegative integer values with probabilities

$$\mathbf{P}(\xi = m) = \frac{\lambda^m}{m!}e^{-\lambda}, \quad \lambda > 0,\ m = 0, 1, 2, \ldots$$

The distribution function, as in Example 3.1.1, has the form of a sum:

$$F(x) = \begin{cases} \sum_{m<x} \frac{\lambda^m}{m!}\, e^{-\lambda} & \text{for } x > 0, \\ 0 & \text{for } x \le 0. \end{cases}$$

3.2.3 The Three Distribution Types

All the distributions considered in the above examples can be divided into two types.

I. Discrete Distributions

Definition 3.2.2 The distribution of a random variable ξ is called *discrete* if ξ can assume only finitely or countably many values $x_1, x_2, \ldots$ so that

$$p_k = \mathbf{P}(\xi = x_k) > 0, \qquad \sum p_k = 1.$$

A discrete distribution $\{p_k\}$ can obviously always be defined on a discrete probability space. It is often convenient to characterise such a distribution by a table:

Values	x_1	x_2	x_3	$\ldots$
Probabilities	p_1	p_2	p_3	$\ldots$

The distributions $\mathbf{I}_a$, $\mathbf{B}_p^n$, $\mathbf{\Pi}_\lambda$, and the geometric distribution are discrete. The derivative of the distribution function of such a distribution is equal to zero everywhere except at the points $x_1, x_2, \ldots$ where $F(x)$ is discontinuous, the jumps being

$$F(x_k + 0) - F(x_k) = p_k.$$

An important class of discrete distributions is formed by lattice distributions.

Definition 3.2.3 We say that random variable ξ has a *lattice distribution* with span h if there exist a and h such that

$$\sum_{k=-\infty}^{\infty} \mathbf{P}(\xi = a + kh) = 1. \tag{3.2.3}$$

If h is the greatest number satisfying (3.2.3) and the number a lies in the interval $[0, h)$ then these numbers are called the *span* and the *shift*, respectively, of the lattice.

If $a = 0$ and $h = 1$ then the distribution is called *arithmetic*. The same terms will be used for random variables.

Obviously the greatest common divisor (g.c.d.) of all possible values of an arithmetic random variable equals 1.

II. Absolutely Continuous Distributions

Definition 3.2.4 The distribution $\mathbf{F}$ of a random variable ξ is said to be *absolutely continuous*[2] if, for any Borel set B,

$$\mathbf{F}(B) = \mathbf{P}(\xi \in B) = \int_B f(x)\,dx, \tag{3.2.4}$$

where $f(x) \geq 0$, $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

The function $f(x)$ in (3.2.4) is called the *density* of the distribution.

It is not hard to derive from the proof of Theorem 3.2.1 (to be more precise, from the theorem on uniqueness of the extension of a measure) that the above definition of absolute continuity is equivalent to the representation

$$F_\xi(x) = \int_{-\infty}^{x} f(u)\,du$$

for all $x \in R$. Distribution functions with this property are also called absolutely continuous.

---

[2]The definition refers to absolute continuity with respect to the *Lebesgue measure*. Given a measure μ on $\langle \mathbb{R}, \mathfrak{B} \rangle$ (see Appendix 3), a distribution $\mathbf{F}$ is called *absolutely continuous with respect to* μ if, for any $B \in \mathfrak{B}$, one has

$$\mathbf{F}(B) = \int_B f(x)\mu(dx).$$

In this sense discrete distributions are also absolutely continuous, but with respect to the counting measure m. Indeed, if one puts $f(x_k) = p_k$, $m(B) = \{$*the number of points from the set* $(x_1, x_2, \ldots)$ *which are in* $B\}$, then

$$\mathbf{F}(B) = \sum_{x_k \in B} p_k = \sum_{x_k \in B} f(x_k) = \int_B f(x)m(dx)$$

(see Appendix 3).

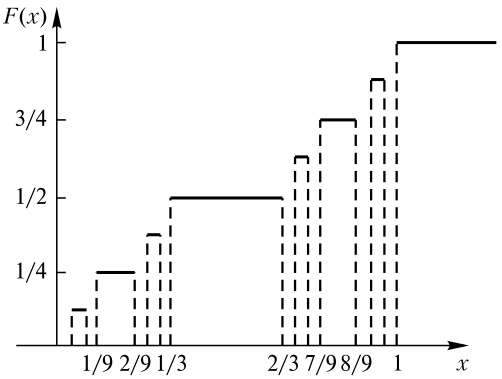

Fig. 3.1 The plot shows the result of the first three steps in the construction of the Cantor function

The function $f(x)$ is determined by the above equalities up to its values on a set of Lebesgue measure 0. For this function, the relation $f(x) = \frac{dF(x)}{dx}$ holds[3] almost everywhere (with respect to the Lebesgue measure).

The distributions $\mathbf{U}_{a,b}$, $\mathbf{\Phi}_{\alpha,\sigma^2}$, $\mathbf{K}_{\alpha,\sigma}$ and $\mathbf{\Gamma}_\alpha$ are absolutely continuous. The density of the normal distribution with parameters $\alpha$and σ is equal to

$$\phi_{\alpha,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-(x-\alpha)^2/(2\sigma^2)}.$$

From their definitions, one could easily derive the densities of the distributions $\mathbf{U}_{a,b}$, $\mathbf{K}_{\alpha,\sigma}$ and $\mathbf{\Gamma}_\alpha$ as well. The density of $\mathbf{K}_{\alpha,\sigma}$ has a shape resembling that of the normal density, but with "thicker tails" (it vanishes more slowly as $|x| \to \infty$).

We will say that a distribution $\mathbf{F}$ *has an atom* at point x_1 if $\mathbf{F}(\{x_1\}) > 0$. We saw that any discrete distribution consists of atoms but, for an absolutely continuous distribution, the probability of hitting a set of zero Lebesgue measure is zero. It turns out that there exists yet a *third* class of distributions which is characterised by the negation of both mentioned properties of discrete and absolutely continuous distributions.

III. Singular Distributions

Definition 3.2.5 A distribution $\mathbf{F}$ is said to be *singular* (with respect to Lebesgue measure) if it has no atoms and is concentrated on a set of zero Lebesgue measure.

Because a singular distribution has no atoms, its distribution function is continuous. An example of such a distribution function is given by the famous Cantor function of which the whole variation is concentrated on the interval $[0, 1]$: $F(x) = 0$ for $x \le 0$, $F(x) = 1$ for $x \ge 1$. It can be constructed as follows (the construction process is shown in Fig. 3.1).

[3]The assertion about the "almost everywhere" uniqueness of the function f follows from the Radon–Nikodym theorem (see Appendix 3).

Divide the segment [0, 1] into three equal parts [0, 1/3], [1/3, 2/3], and [2/3, 1]. On the inner segment put $F(x) = 1/2$. The remaining two segments are again divided into three equal parts each, and on the inner parts one sets $F(x)$ to be $1/4$ and $3/4$, respectively. Each of the remaining segments is divided in turn into three parts, and $F(x)$ is defined on the inner parts as the arithmetic mean of the two already defined neighbouring values of $F(x)$, and so on. At the points which do not belong to such inner segments $F(x)$ is defined by continuity. It is not hard to see that the total length of such "inner" segments on which $F(x)$ is constant is equal to

$$\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \cdots = \frac{1}{3} \sum_{k=0}^{\infty} \left(\frac{2}{3}\right)^k = \frac{1}{3} \frac{1}{1 - 2/3} = 1,$$

so that the function $F(x)$ grows on a set of measure zero but has no jumps.

From the construction of the Cantor distribution we see that $dF(x)/dx = 0$ almost everywhere.

It turns out that these three types of distribution exhaust all possibilities.

More precisely, there is a theorem belonging to Lebesgue[4] stating that any distribution function $F(x)$ can be represented in a unique way as a sum of three components: discrete, absolutely continuous, and singular. Hence an arbitrary distribution function cannot have more than a countable number of jumps (which can also be observed directly: we will count all the jumps if we first enumerate all the jumps which are greater than $1/2$, then the jumps greater than $1/3$, then greater than $1/4$, etc.). This means, in particular, that $F(x)$ is everywhere continuous except perhaps at a countable or finite set of points.

In conclusion of this section we will list several properties of distribution functions and densities that arise when forming new random variables.

3.2.4 Distributions of Functions of Random Variables

For a given function $g(x)$, to find the distribution of $g(\xi)$ we have to impose some measurability requirements on the function. The function $g(x)$ is called *Borel* if the inverse image

$$g^{-1}(B) = \{x : g(x) \in B\}$$

of any Borel set B is again a Borel set. For such a function g the distribution function of the random variable $\eta = g(\xi)$ equals

$$F_{g(\xi)}(x) = \mathbf{P}\big(g(\xi) < x\big) = \mathbf{P}\big(\xi \in g^{-1}(-\infty, x)\big).$$

If $g(x)$ is continuous and strictly increasing on an interval (a, b) then, on the interval $(g(a), g(b))$, the inverse function $y = g^{(-1)}(x)$ is defined as the solution to

[4] See Sect. 3.5 in Appendix 3.

the equation $g(y) = x$.[5] Since g is a monotone mapping we have

$$\{g(\xi) < x\} = \{\xi < g^{(-1)}(x)\} \quad \text{for } x \in (g(a), g(b)).$$

Thus we get the following representation for $F_{g(\xi)}$ in terms of F_ξ: for $x \in (g(a), g(b))$,

$$F_{g(\xi)}(x) = \mathbf{P}(\xi < g^{-1}(x)) = F_\xi(g^{-1}(x)). \tag{3.2.5}$$

Putting $g = F_\xi$ we obtain, in particular, that *if F_ξ is continuous and strictly increasing on (a, b) and $F(a) = 0$, $F(b) = 1$ ($-a$ and b may be ∞) then*

$$F_\xi(g^{(-1)}(x)) \equiv x$$

for $x \in [0, 1]$ and therefore the random variable $\eta = F_\xi(\xi)$ is uniformly distributed over $[0, 1]$.

Definition 3.2.6 *The quantile transform* $F^{(-1)}(f)$ of an arbitrary distribution $\mathbf{F}$ with the distribution function $F(x)$ is the "generalised" inverse of the function F

$$F^{(-1)}(y) := \sup\{x : F(x) < y\} \quad \text{for } y \in (0, 1];$$
$$F^{(-1)}(0) := \inf\{x : F(x) > 0\}.$$

In mathematical statistics, the number $F^{(-1)}(y)$ is called the *quantile* of order y of the distribution $\mathbf{F}$. The function $F^{(-1)}$ has a discontinuity of size $b - a$ at a point y if (a, b) is the interval on which F is constant and such that $F(x) = y \in [0, 1)$.

Roughly speaking, the plot of the function $F^{(-1)}$ can be obtained from that of the function $F(x)$ on the (x, y) plane in the following way: rotate the (x, y) plane in the counter clockwise direction by 90°, so that the x-axis becomes the ordinate axis, but the y-axis becomes the abscissa axis directed to the left. To switch to normal coordinates, we have to reverse the direction of the new x-axis.

Further, if x is a point of continuity and a point of growth of the function F (i.e., $F(x)$ is a point of continuity of $F^{(-1)}$) then $F^{(-1)}(y)$ is the unique solution of the equation $F(x) = y$ and the equality $F(F^{(-1)}(y)) = y$ holds.

In some cases the following statement proves to be useful.

Theorem 3.2.2 *Let $\eta \Subset \mathbf{U}_{0,1}$. Then, for any distribution* $\mathbf{F}$,

$$f^{(-1)}(\eta) \Subset \mathbf{F}.$$

Proof If $F(x) > y$ then $F^{(-1)}(y) = \sup\{v : F(v) < y\} < x$, and vice versa: if $F(x) < y$ then $F^{(-1)}(y) \geq x$ (recall that $F(x)$ is left-continuous). Therefore the following inclusions are valid for the sets in the (x, y) plane:

$$\{y < F(x)\} \subset \{F^{(-1)}(y) < x\} \subset \{y \leq F(x)\}.$$

[5] For an arbitrary non-decreasing function g, the inverse function $g^{(-1)}(x)$ is defined by the equation

$$g^{(-1)}(y) := \inf\{x : g(x) \geq y\} = \sup\{x : g(x) < y\}.$$

Substituting $\eta \in \mathbf{U}_{0,1}$ in place of y in these relations yields that, for any x, such inclusions hold for the respective events, and hence

$$\mathbf{P}\big(F^{(-1)}(\eta) < x\big) = \mathbf{P}\big(\eta < F(x)\big) = F(x).$$

The theorem is proved. □

Thus we have obtained an important method for constructing random variables with prescribed distributions from uniformly distributed random variables. For instance, if $\eta \in \mathbf{U}_{0,1}$ then $\xi = -(1/\alpha)\ln\eta \in \mathbf{\Gamma}_\alpha$.

In another special case, when $g(x) = a + bx$, $b > 0$, from (3.2.5) we get $F_{g(\xi)} = F_\xi((x-a)/b)$. We have already used this relation to some extent when considering the distributions $\mathbf{\Phi}_{\alpha,\sigma^2}$ and $\mathbf{K}_{\alpha,\sigma}$.

If a function g is strictly increasing and differentiable (the inverse function $g^{(-1)}$ is defined in this case), and ξ has a density $f(x)$, then there exists a density for $g(\xi)$ which is equal to

$$f_{g(\xi)}(y) = f\big(g^{(-1)}(y)\big)\big(g^{(-1)}(y)\big)' = f(x)\,\frac{dx}{dy},$$

where $x = g^{(-1)}(y)$, $y = g(x)$. A similar argument for decreasing g leads to the general formula

$$f_{g(\xi)}(y) = f(x)\left|\frac{dx}{dy}\right|.$$

For $g(x) = a + bx$, $b \neq 0$, one obtains

$$f_{a+b\xi}(y) = \frac{1}{|b|} f\left(\frac{y-a}{b}\right).$$

3.3 Multivariate Random Variables

Let $\xi_1, \xi_2, \ldots, \xi_n$ be random variables given on a common probability space $\langle \Omega, \mathfrak{F}, \mathbf{P}\rangle$. To each ω, these random variables put into correspondence an n-dimensional vector $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \ldots, \xi_n(\omega))$.

Definition 3.3.1 A mapping $\Omega \to \mathbb{R}^n$ given by random variables $\xi_1, \xi_2, \ldots, \xi_n$ is called a *random vector* or *multivariate random variable*.

Such a mapping $\Omega \to \mathbb{R}^n$ is a measurable mapping of the space $\langle \Omega, \mathbf{F}\rangle$ into the space $\langle \mathbb{R}^n, \mathfrak{B}^n\rangle$, where $\mathfrak{B}^n$ is the σ-algebra of Borel sets in $\mathbb{R}^n$. Therefore, for Borel sets B, the function $\mathbf{P}_\xi(B) = \mathbf{P}(\xi \in B)$ is defined.

Definition 3.3.2 The function $\mathbf{F}_\xi(B)$ is called the *distribution of the vector* ξ.

The function

$$F_{\xi_1\ldots\xi_n}(x_1, \ldots, x_n) = \mathbf{P}(\xi_1 < x_1, \ldots, \xi_n < x_n)$$

is called the *distribution function* of the random vector $(\xi_1, \ldots, \xi_n)$ or *joint distribution function* of the random variables $\xi_1, \ldots, \xi_n$.

The following properties of the distribution functions of random vectors, analogous to properties F1–F3 in Sect. 3.2, hold true.

FF1. *Monotonicity*: "Multiple" differences of the values of the function $F_{\xi_1 \ldots \xi_n}$, which correspond to probabilities of hitting arbitrary "open at the right" parallelepipeds, are nonnegative. For instance, in the two-dimensional case this means that, for any $x_1 < x_2$, $y_1 < y_2$ (the points (x_1, y_1) and (x_2, y_2) being the "extreme" vertices of the parallelepiped),

$$F_{\xi_1,\xi_2}(x_2, y_2) - F_{\xi_1,\xi_2}(x_2, y_1) - \big(F_{\xi_1,\xi_2}(x_1, y_2) - F_{\xi_1,\xi_2}(x_1, y_1)\big) \geq 0.$$

This double difference is nothing else but the probability of hitting the "semi-open" parallelepiped $[x_1, x_2) \times [y_1, y_2)$ by ξ.

In other words, the *differences*

$$F_{\xi_1,\xi_2}(t, y_2) - F_{\xi_1,\xi_2}(t, y_1) \quad \text{for } y_1 < y_2$$

must be monotone in t. (For this to hold, the monotonicity of the function $F_{\xi_1,\xi_2}(t, y_1)$ is not sufficient.)

FF2. The second property can be called *consistency*.

$$\lim_{x_n \to \infty} F_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n) = F_{\xi_1 \ldots \xi_{n-1}}(x_1, \ldots, x_{n-1}),$$
$$\lim_{x_n \to -\infty} F_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n) = 0.$$

FF3. *Left-continuity*:

$$\lim_{x_n' \uparrow \infty} F_{\xi_1 \ldots \xi_n}\big(x_1, \ldots, x_n'\big) = F_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n).$$

That the limits in properties FF2 and FF3 are taken in the last variable is inessential, for one can always renumber the components of the vectors.

One can prove these properties in the same way as in the one-dimensional case. As above, *any function $F(x_1, \ldots, x_n)$ possessing this collection of properties will be the distribution function of a (multivariate) random variable*.

As in the one-dimensional case, when considering random vectors $\xi = (\xi_1, \ldots, \xi_n)$, we can make use of the simplest sample model of the probability space $\langle \Omega, \mathbf{F}, \mathbf{P} \rangle$. Namely, let Ω coincide with $\mathbb{R}^n$ and $\mathbf{F} = \mathfrak{B}^n$ be the σ-algebra of Borel sets. We will complete the construction of the required probability space if we put $\mathbf{F}(B) = \mathbf{F}_\xi(B) = \mathbf{P}(\xi \in B)$ for any $B \in \mathfrak{B}^n$. It remains to define the random variable as the value of the elementary event itself, i.e. to put $\xi(\omega) = \omega$, where ω is a point in $\mathbb{R}^n$.

It is not hard to see that the distribution function $F_{\xi_1 \ldots \xi_n}$ uniquely determines the distribution $\mathbf{F}_\xi(B)$. Indeed, $F_{\xi_1 \ldots \xi_n}$ defines a probability on the σ-algebra $\mathcal{A}$ generated by rectangles $\{a_i \leq x_i < b_i;\ i = 1, \ldots, n\}$. For example, in the two-dimensional case

$$\begin{aligned}&\mathbf{P}(a_1 \le \xi_1 < b_1, a_2 \le \xi_2 < b_2)\\ &\quad = \mathbf{P}(\xi_1 < b_1, a_2 \le \xi_2 < b_2) - \mathbf{P}(\xi_1 < a_1, a_2 \le \xi_2 < b_2)\\ &\quad = \big[F_{\xi_1,\xi_2}(b_1, b_2) - F_{\xi_1,\xi_2}(b_1, a_2)\big] - \big[F_{\xi_1,\xi_2}(a_1, b_2) - F_{\xi_1,\xi_2}(a_1, a_2)\big].\end{aligned}$$

But $\mathfrak{B}^n = \sigma(\mathcal{A})$, and it remains to make use of the measure extension theorem (see Sect. 3.2.1).

Thus *from a distribution function $F_{\xi_1\dots\xi_n} = F$ one can always construct a sample probability space $\langle \mathbb{R}^n, \mathfrak{B}^n, \mathbf{F}_\xi \rangle$ and a random variable $\xi(\omega) \equiv \omega$ on it so that the latter will have the prescribed distribution $\mathbf{F}_\xi$.*

As in the one-dimensional case, we say that the distribution of a random vector is discrete if the random vector assumes at most a countable set of values.

The distribution of a random vector will be *absolutely continuous* if, for any Borel set $B \subset \mathbb{R}^n$,

$$\mathbf{F}_\xi(B) = \mathbf{P}(\xi \in B) = \int_B f(x)\,dx,$$

where clearly $f(x) \ge 0$ and $\int_\Omega f(x)\,dx = 1$.

This definition can be replaced with an equivalent one requiring that

$$F_{\xi_1\dots\xi_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, \dots, t_n)\,dt_1 \cdots dt_n. \tag{3.3.1}$$

Indeed, if (3.3.1) holds, we define a countably additive set function

$$\mathbf{Q}(B) = \int_B f(x)\,dx$$

(see properties of integrals in Appendix 3), which will coincide on rectangles with $\mathbf{F}_\xi$. Consequently, $\mathbf{F}_\xi(B) = \mathbf{Q}(B)$.

The function $f(x)$ is called the *density of the distribution* of ξ or *density of the joint distribution* of $\xi_1, \dots, \xi_n$. The equality

$$\frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{\xi_1\dots\xi_n}(x_1, \dots, x_n) = f(x_1, \dots, x_n)$$

holds for this function almost everywhere.

If a random vector ξ has density $f(x_1, \dots, x_n)$, then clearly any "subvector" $(\xi_{k_1} \dots \xi_{k_n})$, $k_i \le n$, also has a density equal (let for the sake of simplicity $k_i = i$, $i = 1, \dots, s$) to

$$f(x_1, \dots, x_s) = \int f(x_1, \dots, x_n)\,dx_{s+1} \cdots dx_n.$$

Let continuously differentiable functions $y_i = g_i(x_1, \dots, x_n)$ be given in a region $A \subset \mathbb{R}^n$. Suppose they are univalently resolvable for $x_1, \dots, x_n$: there exist functions $x_i = g_i^{(-1)}(y_1, \dots, y_n)$, and the Jacobian $J = |\partial x_i / \partial y_i| \ne 0$ in A. Denote by B the image of A in the range of $(y_1, \dots, y_n)$. Suppose further that a random vector $\xi = (\xi_1, \dots, \xi_n)$ has a density $f_\xi(x)$. Then $\eta_i = g_i(\xi_1, \dots, \xi_n)$ will be random variables with a joint density which, at a point $(y_1, \dots, y_n) \in B$, is equal to

$$f_n(y_1, \dots, y_n) = f_\xi(x_1, \dots, x_n)|J|; \tag{3.3.2}$$

moreover

$$\mathbf{P}(\xi \in A) = \int_A f_\xi(x_1, \dots, x_n)\, dx_1 \cdots dx_n = \int_B f_\xi(x_1, \dots, x_n)|J|\, dy_1 \cdots dy_n$$
$$= \int_B f_\eta(y_1, \dots, y_n)\, dy_1 \cdots dy_n = \mathbf{P}(\eta \in B). \tag{3.3.3}$$

This is clearly an extension to the multi-dimensional case of the property of densities discussed at the end of Sect. 3.2. Formula (3.3.3) for integrals is well-known in calculus as the change of variables formula and could serve as a proof of (3.3.2).

The distribution $\mathbf{F}_\xi$ of a random vector ξ is called *singular* if the distribution has no atoms ($\mathbf{F}_\xi(\{x\}) = 0$ for any $x \in \mathbb{R}^n$) and is concentrated on a set of zero Lebesgue measure.

Consider the following two important examples of multivariate distributions (we continue the list of the most common distribution from Sect. 3.2).

9. *The multinomial distribution* $\mathbf{B}_p^n$. We use here the same symbol $\mathbf{B}_p^n$ as we used for the binomial distribution. The only difference is that now by p we understand a vector $p = (p_1, \dots, p_r)$, $p_j \geq 0$, $\sum_{j=1}^r p_j = 1$, which could be interpreted as the collection of probabilities of disjoint events A_j, $\bigcup A_j = \Omega$. For an integer-valued random vector $\nu = (\nu_1, \dots, \nu_r)$, we will write $\nu \Subset B$ if for $k = (k_1, \dots, k_r)$, $k_j \geq 0$, $\sum_{j=1}^r k_j = n$ one has

$$\mathbf{P}(\nu = k) = \frac{n!}{k_1! \cdots k_r!}\, p_1^{k_1} \cdots p_r^{k_r}. \tag{3.3.4}$$

On the right-hand side we have a term from the expansion of the polynomial $(p_1 + \cdots + p_r)^n$ into powers of $p_1, \dots, p_r$. This explains the name of the distribution. If p is a number, then evidently $\mathbf{B}_p^n = \mathbf{B}_{(p,1-p)}^n$, so that the binomial distribution is a multinomial distribution with $r = 2$.

The numbers ν_j could be interpreted as the frequencies of the occurrence of events A_j in n independent trials, the probability of occurrence of A_j in a trial being p_j. Indeed, the probability of any fixed sequence of outcomes containing $k_1, \dots, k_r$ outcomes $A_1, \dots, A_r$, respectively, is equal to $p_1^{k_1} \cdots p_r^{k_r}$, and the number of different sequences of this kind is equal to $n!/k_1! \cdots k_r!$ (of $n!$ permutations we leave only those which differ by more than merely permutations of elements inside the groups of $k_1, \dots, k_r$ elements). The result will be the probability (3.3.4).

Example 3.3.1 The simplest model of a chess tournament with two players could be as follows. In each game, independently of the outcomes of the past games, the 1st player wins with probability p, loses with probability q, and makes a draw with probability $1 - p - q$. In that case the probability that, in n games, the 1st player wins i and loses j games ($i + j \leq n$), is

$$p(n; i, j) = \frac{n!}{i!\, j!\, (n - i - j)!}\, p^i q^j (1 - p - q)^{n-i-j}.$$

Suppose that the tournament goes on until one of the players wins N games (and thereby wins the tournament). If we denote by η the duration of the tournament (the

number of games played before its end) then

$$\mathbf{P}(\eta = n) = \sum_{i=0}^{N-1} p(n-1; N-1, i)p + \sum_{i=0}^{N-1} p(n-1; i, N-1)q.$$

10. *The multivariate normal* (or *Gaussian*) *distribution* $\mathbf{\Phi}_{\alpha,\sigma^2}$. Let $\alpha = (\alpha_1, \ldots, \alpha_r)$ be a vector and $\sigma^2 = \|\sigma_{ij}\|$, $i, j = 1, \ldots, r$, a symmetric positive definite matrix, and $A = \|a_{ij}\|$ the matrix inverse to $\sigma^2 = A^{-1}$. We will say that a vector $\xi = (\xi_1, \ldots, \xi_r)$ has the *normal distribution*: $\xi \Subset \mathbf{\Phi}_{\alpha,\sigma^2}$, if it has the density

$$\varphi_{\alpha,\sigma^2}(x) = \frac{\sqrt{|A|}}{(2\pi)^{r/2}} \exp\left\{-\frac{1}{2}(x-\alpha)A(x-\alpha)^T\right\}.$$

Here T denotes transposition:

$$xAx^T = \sum a_{ij}x_i x_j.$$

It is not hard to verify that

$$\int \varphi_{\alpha,\sigma^2}(x)\, dx_1 \cdots dx_r = 1$$

(see also Sect. 7.6).

3.4 Independence of Random Variables and Classes of Events

3.4.1 Independence of Random Vectors

Definition 3.4.1 Random variables $\xi_1, \ldots, \xi_n$ are said to be *independent* if

$$\mathbf{P}(\xi_1 \in B_1, \ldots, \xi_n \in B_n) = \mathbf{P}(\xi_1 \in B_1) \cdots \mathbf{P}(\xi_n \in B_n) \tag{3.4.1}$$

for any Borel sets $B_1, \ldots, B_n$ on the real line.

One can introduce the notion of a *sequence* of independent random variables. The random variables from the sequence $\{\xi_n\}_{n=1}^{\infty}$ given on a probability space $\langle \Omega, \mathfrak{F}, \mathbf{P}\rangle$, are independent if (3.4.1) holds for any integer n so that the independence of a sequence of random variables reduces to that of any finite collection of random variable from this sequence. As we will see below, for a sequence of independent random variables, any two events related to disjoint groups of random variables from the sequence are independent.

Another possible definition of independence of random variables follows from the assertion below.

Theorem 3.4.1 *Random variables $\xi_1, \ldots, \xi_n$ are independent if and only if*

$$F_{\xi_1 \ldots \xi_n}(x_1, \ldots, x_n) = F_{\xi_1}(x_1) \cdots F_{\xi_n}(x_n).$$

The proof of the theorem is given in the third part of the present section.

An important criterion of independence in the case when the distribution of $\xi = (\xi_1, \ldots, \xi_n)$ is absolutely continuous is given in the following theorem.

Theorem 3.4.2 *Let random variables $\xi_1, \dots, \xi_n$ have densities $f_1(x), \dots, f_n(x)$, respectively. Then for the independence of $\xi_1, \dots, \xi_n$ it is necessary and sufficient that the vector $\xi = (\xi_1, \dots, \xi_n)$ has a density $f(x_1, \dots, x_n)$ which is equal to*

$$f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n).$$

Thus, if it turns out that the density of ξ equals the product of densities of ξ_j, that will mean that the random variables ξ_j are independent.

We leave it to the reader to verify, using this theorem, that the components of a normal vector $(\xi_1, \dots, \xi_n)$ are independent if and only if $a_{ij} = 0$, $\sigma_{ij} = 0$ for $i \neq j$.

Proof of Theorem 3.4.2 If the distribution function of the random variable ξ_i is given by

$$F_{\xi_i}(x_i) = \int_{-\infty}^{x_i} f_i(t_i)\, dt_i$$

and ξ_i are independent, then the joint distribution function will be defined by the formula

$$\begin{aligned} F_{\xi_1 \dots \xi_n}(x_1, \dots, x_n) &= F_{\xi_1}(x_1) \cdots F_{\xi_n}(x_n) \\ &= \int_{-\infty}^{x_1} f_1(t_1)\, dt_1 \cdots \int_{-\infty}^{x_n} f_n(t_n)\, dt_n \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_1(t_1) \cdots f_n(t_n)\, dt_1 \cdots dt_n. \end{aligned}$$

Conversely, assuming that

$$F_{\xi_1 \dots \xi_n}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_1(t_1) \cdots f_n(t_n)\, dt_1 \cdots dt_n,$$

we come to the equality

$$F_{\xi_1 \dots \xi_n}(x_1, \dots, x_n) = F_{\xi_1}(x_1) \cdots F_{\xi_n}(x_n).$$

The theorem is proved. □

Now consider the discrete case. Assume for the sake of simplicity that the components of ξ may assume only integral values. Then for the independence of ξ_j it is necessary and sufficient that, for all $k_1, \dots, k_n$,

$$\mathbf{P}(\xi_1 = k_1, \dots, \xi_n = k_n) = \mathbf{P}(\xi_1 = k_1) \cdots \mathbf{P}(\xi_n = k_n).$$

Verifying this assertion causes no difficulties, and we leave it to the reader.

The notion of independence is very important for Probability Theory and will be used throughout the entire book. Assume that we are formalising a practical problem (constructing an appropriate probability model in which various random variables are to be present). How can one find out whether the random variables (or events) to appear in the model are independent? In such situations it is a *justified rule to consider events and random variables with no causal connection as independent*.

The detection of "probabilistic" independence in a mathematical model of a random phenomenon is often connected with a deep understanding of its physical essence.

Consider some simple examples. For instance, it is known that the probability of a new-born child to be a boy (event A) has a rather stable value $\mathbf{P}(A) = 22/43$. If B denotes the condition that the child is born on the day of the conjunction of Jupiter and Mars, then, under the assumption that the position of the planets does not determine individual fates of humans, the conditional probability $\mathbf{P}(A|B)$ will have the same value: $\mathbf{P}(A|B) = 22/43$. That is, the actual counting of the frequency of births of boys under these specific astrological conditions would give just the value $22/43$. Although such a counting might never have been carried out at a sufficiently large scale, we have no grounds to doubt its results.

Nevertheless, one should not treat the connection between "mathematical" and causal independence as an absolute one. For instance, by Newton's law of gravitation the flight of a missile undoubtedly influences the simultaneous flight of another missile. But it is evident that in practice one can ignore this influence. This example also shows that independence of events and variables in the concrete and relative meaning of this term does not contradict the principle of the universal interdependence of all events.

It is also interesting to note that the formal definition of independence of events or random variables is much wider than the notion of real independence in the sense of affiliation to causally unrelated phenomena. This follows from the fact that "mathematical" independence can take place in such cases when one has no reason for assuming no causal relation. We illustrate this statement by the following example. Let η be a random variable uniformly distributed over $[0, 1]$. Then in the expansion of η into a binary fraction

$$\eta = \frac{\xi_1}{2} + \frac{\xi_2}{4} + \frac{\xi_3}{8} + \cdots$$

the random variables ξ_k will be independent (see Example 11.3.1), although they all have a related origin.

One can see that this circumstance only enlarges the area of applicability of all the assertions we obtain below under the formal condition of independence.[6]

The notion of independence of random variables is closely connected with that of *independence of σ-algebras*.

3.4.2 Independence of Classes of Events

Let $\langle \Omega, \mathfrak{F}, \mathbf{P} \rangle$ be a probability space and $\mathcal{A}_1$ and $\mathcal{A}_2$ classes of events from the σ-algebra $\mathfrak{F}$.

[6] For a more detailed discussion of connections between causal and probabilistic independence, see [24], from where we borrowed the above examples.

Definition 3.4.2 The classes of events $\mathcal{A}_1$ and $\mathcal{A}_2$ are said to be *independent* if, for any events A_1 and A_2 such that $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$, one has

$$\mathbf{P}(A_1 A_2) = \mathbf{P}(A_1)\mathbf{P}(A_2).$$

The following definition introduces the notion of independence of a sequence of classes of events.

Definition 3.4.3 Classes of events $\{\mathcal{A}_n\}_{n=1}^{\infty}$ are *independent* if, for any collection of integers $n_1, \ldots, n_k$,

$$\mathbf{P}\left(\bigcap_{j=1}^{k} A_{n_j}\right) = \prod_{j=1}^{k} \mathbf{P}(A_{n_j})$$

for any $A_{n_j} \in \mathcal{A}_{n_j}$.

For instance, in a sequence of independent trials, the sub-σ-algebras of events related to different trials will be independent. The independence of a sequence of algebras of events also reduces to the independence of any finite collection of algebras from the sequence. It is clear that subalgebras of events of independent algebras are also independent.

Theorem 3.4.3 *σ-algebras $\mathfrak{A}_1$ and $\mathfrak{A}_2$ generated, respectively, by independent algebras of events $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent.*

Before proving this assertion we will obtain an approximation theorem which will be useful for the sequel. By virtue of the theorem, any event A from the σ-algebra $\mathfrak{A}$ generated by an algebra $\mathcal{A}$ can, in a sense, be approximated by events from $\mathcal{A}$. To be more precise, we introduce the "distance" between events defined by

$$d(A, B) = \mathbf{P}(A\overline{B} \cup \overline{A}B) = \mathbf{P}(A\overline{B}) + \mathbf{P}(\overline{A}B) = \mathbf{P}(A - B) + \mathbf{P}(B - A).$$

This distance possesses the following properties:

$$\begin{aligned} d(\overline{A}, \overline{B}) &= d(A, B), \\ d(A, C) &\le d(A, B) + d(B, C), \\ d(AB, CD) &\le d(A, C) + d(B, D), \\ \left|\mathbf{P}(A) - \mathbf{P}(B)\right| &\le d(A, B). \end{aligned} \tag{3.4.2}$$

The first relation is obvious. The triangle inequality follows from the fact that

$$\begin{aligned} d(A, C) &= \mathbf{P}(A\overline{C}) + \mathbf{P}(\overline{A}C) = \mathbf{P}(A\overline{C}B) + \mathbf{P}(A\overline{C}\overline{B}) + \mathbf{P}(\overline{A}CB) + \mathbf{P}(\overline{A}C\overline{B}) \\ &\le \mathbf{P}(\overline{C}B) + \mathbf{P}(A\overline{B}) + \mathbf{P}(\overline{A}B) + \mathbf{P}(C\overline{B}) = d(A, B) + d(B, C). \end{aligned}$$

The third relation in (3.4.2) can be obtained in a similar way by enlarging events under the probability sign. Finally, the last inequality in (3.4.2) is a consequence of the relations

$$\mathbf{P}(A) = \mathbf{P}(AB) + \mathbf{P}(A\overline{B}) = \mathbf{P}(B) - \mathbf{P}(B\overline{A}) + \mathbf{P}(A\overline{B}).$$

Theorem 3.4.4 (The approximation theorem) *Let $\langle \Omega, \mathfrak{F}, \mathbf{P}\rangle$ be a probability space and $\mathfrak{A}$ the σ-algebra generated by an algebra $\mathcal{A}$ of events from $\mathfrak{F}$. Then, for any $\mathcal{A} \in \mathfrak{A}$, there exists a sequence $A_n \in \mathcal{A}$ such that*

$$\lim_{n\to\infty} d(A, A_n) = 0. \tag{3.4.3}$$

By the last inequality from (3.4.2), the assertion of the theorem means that $\mathbf{P}(A) = \lim_{n\to\infty} \mathbf{P}(A_n)$ and that each event $A \in \mathfrak{A}$ can be represented, up to a set of zero probability, as a limit of a sequence of events from the generating algebra $\mathcal{A}$ (see also Appendix 1).

Proof[7] We will call an event $A \in \mathfrak{F}$ *approximable* if there exists a sequence $A_n \in \mathcal{A}$ possessing property (3.4.3), i.e. $d(A_n, A) \to 0$.

Since $d(A, A) = 0$, the class of approximable events $\mathfrak{A}^*$ contains $\mathcal{A}$. Therefore to prove the theorem it suffices to verify that $\mathfrak{A}^*$ is a σ-algebra.

The fact that $\mathfrak{A}^*$ is an algebra is obvious, for the relations $A \in \mathfrak{A}^*$ and $B \in \mathfrak{A}^*$ imply that $\overline{A}$, $A \cup B$, $A \cap B \in \mathfrak{A}$. (For instance, if $d(A, A_n) \to 0$ and $d(B, B_n) \to 0$, then by the third inequality in (3.4.2) one has $d(AB, A_nB_n) \le d(A, A_n) + d(B, B_n) \to 0$, so that $AB \in \mathfrak{A}^*$.)

Now let $C = \bigcap_{k=1}^{\infty} C_k$ where $C_k \in \mathfrak{A}^*$. Since $\mathfrak{A}^*$ is an algebra, we have $D_n = \bigcup_{k=1}^{n} C_k \in \mathfrak{A}^*$; moreover,

$$d(D_n, C) = \mathbf{P}(C - D_n) = \mathbf{P}(C) - \mathbf{P}(D_n) \to 0.$$

Therefore one can choose $A_n \in \mathcal{A}$ so that $d(D_n, A_n) < 1/n$, and consequently by virtue of (3.4.2) we have

$$d(C, A_n) \le d(C, D_n) + d(D_n, A_n) \to 0.$$

Thus $C \in \mathfrak{A}^*$ and hence $\mathfrak{A}^*$ forms a σ-algebra. The theorem is proved. □

Proof of Theorem 3.4.3 is now easy. If $A_1 \in \mathfrak{A}_1$ and $A_2 \in \mathfrak{A}_2$, then by Theorem 3.4.4 there exist sequences $A_{1n} \in \mathcal{A}_1$ and $A_{2n} \in \mathcal{A}_2$ such that $d(A_i, A_{in}) \to 0$ as $n \to \infty$, $i = 1, 2$. Putting $B = A_1A_2$ and $B_n = A_{1n}A_{2n}$, we obtain that

$$d(B, B_n) \le d(A_1, A_{1n}) + d(A_2, A_{2n}) \to 0$$

as $n \to \infty$ and

$$\mathbf{P}(A_1A_2) = \lim_{n\to\infty} \mathbf{P}(B_n) = \lim_{n\to\infty} \mathbf{P}(A_{1n})\mathbf{P}(A_{2n}) = \mathbf{P}(A_1)\mathbf{P}(A_2).$$ □

3.4.3 Relations Between the Introduced Notions

We will need one more definition. Let ξ be a random variable (or vector) given on a probability space $\langle \Omega, \mathfrak{F}, \mathbf{P}\rangle$.

[7] The theorem is also a direct consequence of the lemma from Appendix 1.

Definition 3.4.4 The class $\mathfrak{F}_\xi$ of events from $\mathfrak{F}$ of the form $A = \xi^{-1}(B) = \{\omega : \xi(\omega) \in B\}$, where B are Borel sets, is called the *σ-algebra generated by the random variable ξ*.

It is evident that $\mathfrak{F}_\xi$ is a σ-algebra since to each operation on sets A there corresponds the same operation on the sets $B = \xi(A)$ forming a σ-algebra.

The σ-algebra $\mathfrak{F}_\xi$ generated by the random variable ξ will also be denoted by $\sigma(\xi)$.

Consider, for instance, a probability space $\langle \Omega, \mathfrak{B}, \mathbf{P}\rangle$, where $\Omega = \mathbb{R}$ is the real line and $\mathfrak{B}$ is the σ-algebra of Borel sets. If

$$\xi = \xi(\omega) = \begin{cases} 0, & \omega < 0, \\ 1, & \omega \geq 0, \end{cases}$$

then $\mathfrak{F}_\xi$ clearly consists of four sets: $\mathbb{R}$, $\varnothing$, $\{\omega < 0\}$ and $\{\omega \geq 0\}$. Such a random variable ξ cannot distinguish "finer" sets from $\mathfrak{B}$. On the other hand, it is obvious that ξ will be measurable ($\{\xi \in B\} \in \mathfrak{B}_1$) with respect to any other "richer" sub-σ-algebra $\mathfrak{B}_1$, such that $\sigma(\xi) \subset \mathfrak{B}_1 \subset \mathfrak{B}$.

If $\xi = \xi(\omega) = \lfloor \omega \rfloor$ is the integral part of ω, then $\mathfrak{F}_\xi$ will be the σ-algebra of sets composed of the events $\{k \leq \omega < k+1\}$, $k = \ldots, -1, 0, 1, \ldots$

Finally, if $\xi(\omega) = \varphi(\omega)$ where φ is continuous and monotone, $\varphi(\infty) = \infty$ and $\varphi(-\infty) = -\infty$, then $\mathfrak{F}_\xi$ coincides with the σ-algebra of Borel sets $\mathfrak{B}$.

Lemma 3.4.1 *Let ξ and η be two random variables given on $\langle \Omega, \mathfrak{F}, \mathbf{P}\rangle$, the variable ξ being measurable with respect to $\sigma(\eta)$. Then ξ and η are functionally related, i.e. there exists a Borel function g such that $\xi = g(\eta)$.*

Proof By assumption,

$$A_{k,n} = \left\{ \xi \in \left[\frac{k}{2^n}, \frac{k+1}{2^n} \right) \right\} \in \sigma(\eta).$$

Denote by $B_{k,n} = \{\eta(\omega) : \omega \in A_{k,n}\}$ the images of the sets $A_{k,n}$ on the line $\mathbb{R}$ under the mapping $\eta(\omega)$ and put $g_n(x) = k/2^n$ for $x \in B_{k,n}$. Then $g_n(\eta) = [2^n \varepsilon]/2^n$ and because $A_{k,n} \in \sigma(\eta)$, $B_{k,n} \in \mathfrak{B}$ and g_n is a Borel function. Since $g_n(x) \uparrow$ for any x, the limit $\lim_{n\to\infty} g_n(x) = g(x)$ exists and is also a Borel function. It remains to observe that $\varepsilon = \lim_{n\to\infty} g_n(\eta) = g(\eta)$ by the very construction. □

Now we formulate an evident proposition relating independence of random variables and σ-algebras.

Random variables $\xi_1, \ldots, \xi_n$ are independent if and only if the σ-algebras $\sigma(\xi_1), \ldots, \sigma(\xi_n)$ are independent.

This is a direct consequence of the definitions of independence of random variables and σ-algebras.

Now we can prove Theorem 3.4.1. First note that finite unions of semi-intervals $[\cdot,\cdot)$ (perhaps with infinite end points) form a σ-algebra generating the Borel σ-algebra on the line: $\mathfrak{B} = \sigma(\mathcal{A})$.

Proof of Theorem 3.4.1 Since in one direction the assertion of the theorem is obvious, it suffices to verify that the equality $F(x_1,\dots,x_n) = F_{\xi_1}(x_1)\cdots F_{\xi_n}(x_n)$ for the joint distribution function implies the independence of $\sigma(\xi_1),\dots,\sigma(\xi_n)$. Put for simplicity $n=2$ and denote by Δ and Λ the semi-intervals $[x_1,x_2)$ and $[y_1,y_2)$, respectively. The following equalities hold:

$$\begin{aligned}\mathbf{P}(\xi_1\in\Delta,\xi_2\in\Lambda) &= \mathbf{P}\big(\xi_1\in[x_1,x_2),\xi_2\in[y_1,y_2)\big)\\ &= F(x_2,y_2)F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_1)\\ &= \big(F_{\xi_1}(x_2) - F_{\xi_1}(x_1)\big)\big(F_{\xi_2}(y_2) - F_{\xi_2}(y_1)\big)\\ &= \mathbf{P}\{\xi_1\in\Delta\}\mathbf{P}\{\xi_2\in\Lambda\}.\end{aligned}$$

Consequently, if Δ_i, $i=1,\dots,n$, and Λ_j, $j=1,\dots,m$, are two systems of disjoint semi-intervals, then

$$\begin{aligned}\mathbf{P}\left(\xi_1\in\bigcup_{i=1}^{n}\Delta_i,\xi_2\in\bigcup_{j=1}^{m}\Lambda_j\right) &= \sum_{i,j}\mathbf{P}(\xi_1\in\Delta_i,\xi_2\in\Lambda_j)\\ &= \sum_{i,j}\mathbf{P}(\xi_1\in\Delta_i)\mathbf{P}(\xi_2\in\Lambda_j)\\ &= \mathbf{P}\left(\xi_1\in\bigcup_{i=1}^{n}\Delta_i\right)\mathbf{P}\left(\xi_2\in\bigcup_{j=1}^{m}\lambda_j\right). \qquad (3.4.4)\end{aligned}$$

But the class of events $\{\omega:\xi(\omega)\in A\} = \xi^{-1}(A)$, where $A\in\mathcal{A}$, forms, along with $\mathcal{A}$, an algebra (we will denote it by $\alpha(\xi)$), and one has $\sigma(\alpha(\xi)) = \sigma(\xi)$. In (3.4.4) we proved that $\alpha(\xi_1)$ and $\alpha(\xi_2)$ are independent. Therefore by Theorem 3.4.3 the σ-algebras $\sigma(\xi_1) = \sigma(\alpha(\xi_1))$ and $\sigma(\xi_2) = \sigma(\alpha(\xi_1))$ are also independent. The theorem is proved. □

It is convenient to state the following fact as a theorem.

Theorem 3.4.5 *Let φ_1 and φ_2 be Borel functions and ξ_1 and ξ_2 be independent random variables. Then $\eta_1 = \varphi_1(\xi_1)$ and $\eta_2 = \varphi_2(\xi_2)$ are also independent random variables.*

Proof We have to verify that, for any Borel sets B_1 and B_2,

$$\mathbf{P}\big(\varphi_1(\xi_1)\in B_1,\varphi_2(\xi_2)\in B_2\big) = \mathbf{P}\big(\varphi_1(\xi_1)\in B_1\big)\mathbf{P}\big(\varphi_2(\xi_2)\in B_2\big). \qquad (3.4.5)$$

But the sets $\{x:\varphi_i(x)\in B_i\} = \varphi^{-1}(B_i) = B_i^*$, $i=1,2$, are again Borel sets. Therefore

$$\big\{\omega:\varphi_i(\xi_i)\in B_i\big\} = \big\{\omega:\xi_i\in B_i^*\big\},$$

and the required multiplicativity of probability (3.4.5) follows from the independence of ξ_i. The theorem is proved. □

Let $\{\xi_n\}_{n=1}^{\infty}$ be a sequence of independent random variables. Consider the random variables $\xi_k, \xi_{k+1}, \ldots, \xi_m$ where $k < m \leq \infty$. Denote by $\sigma(\xi_k, \ldots, \xi_m)$ (for $m = \infty$ we will write $\sigma(\xi_k, \xi_{k+1}, \ldots)$) the σ-algebra generated by the events $\bigcap_{i=k}^{m} A_i$, where $A_i \in \sigma(\xi_i)$.

Definition 3.4.5 The σ-algebra $\sigma(\xi_k, \ldots, \xi_m)$ is said to be *generated by the random variables* $\xi_k, \ldots, \xi_m$.

In the sequel we will need the following proposition.

Theorem 3.4.6 *For any* $k \geq 1$, *the* σ*-algebra* $\sigma(\xi_{n+k})$ *is independent of* $\sigma(\xi_1, \ldots, \xi_n)$.

Proof To prove the assertion, we make use of Theorem 3.4.3. To this end we have to verify that the algebra $\mathcal{A}$ generated by sets of the form $B = \bigcap_{i=1}^{n} A_i$, where $A_i \in \sigma(\xi_i)$, is independent of $\sigma(\xi_{n+k})$. Let $A \in \sigma(\xi_{n+k})$, then it follows from the independence of the σ-algebras $\sigma(\xi_1), \sigma(\xi_2), \ldots, \sigma(\xi_n), \sigma(\xi_{n+k})$ that

$$\mathbf{P}(AB) = \mathbf{P}(A)\mathbf{P}(A_1)\cdots\mathbf{P}(A_n) = \mathbf{P}(A)\cdot\mathbf{P}(B).$$

In a similar way we verify that

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i A\right) = \mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right)\mathbf{P}(A)$$

(one just has to represent $\bigcup_{i=1}^{n} A_i$ as a union of disjoint events from $\mathcal{A}$). Thus the algebra $\mathcal{A}$ is independent of $\sigma(\xi_{n+k})$. Hence $\sigma(\xi_1, \ldots, \xi_n)$ and $\sigma(\xi_{n+k})$ are independent. The theorem is proved. □

It is not hard to see that similar conclusions can be made about vector-valued random variables $\xi_1, \xi_2, \ldots$ defining their independence using the relation

$$\mathbf{P}(\xi_1 \in B_1, \ldots, \xi_n \in B_n) = \prod \mathbf{P}(\xi_j \in B_j),$$

where B_j are Borel sets in spaces of respective dimensions.

In conclusion of this section note that one can always construct a probability space $\langle \Omega, \mathfrak{F}, \mathbf{P}\rangle$ ($\langle \mathbb{R}^n, \mathfrak{B}^n, \mathbf{P}_\xi\rangle$) on which independent random variables $\xi_1, \ldots, \xi_n$ with prescribed distribution functions F_{ξ_j} are given whenever these distributions F_{ξ_j} are known. This follows immediately from Sect. 3.3, since in our case the joint distribution function $F_\xi(x_1, \ldots, x_n)$ of the vector $\xi = (\xi_1, \ldots, \xi_n)$ is uniquely determined by the distribution functions $F_{\xi_j}(x)$ of the variables ξ_j:

$$F_\xi(x_1, \ldots, x_n) = \prod_{1}^{n} F_{\xi_j}(x_j).$$

3.5 * On Infinite Sequences of Random Variables

We have already mentioned infinite sequences of random variables. Such sequences will repeatedly be objects of our studies below. However, there arises the question of whether one can define an infinite sequence on a probability space in such a way that its components possess certain prescribed properties (for instance, that they will be independent and identically distributed).

As we saw, one can always define a *finite* sequence of independent random variables by choosing for the "compound" random variable $(\xi_1, \dots, \xi_n)$ the sample space $\mathbb{R}_1 \times \mathbb{R}_2 \times \cdots \times \mathbb{R}_n = \mathbb{R}^n$ and σ-algebra $\mathfrak{B}_1 \times \mathfrak{B}_1 \times \cdots \times \mathfrak{B}_n = \mathfrak{B}^n$ generated by sets of the form $B_1 \times B_2 \times \cdots \times B_n \subset \mathbb{R}^n$, B_i being Borel sets. It suffices to define probability on the algebra of these sets. In the infinite-dimensional case, however, the situation is more complicated. Theorem 3.2.1 and its extensions to the multivariate case are insufficient here. One should define probability on an algebra of events from $\mathbb{R}^\infty = \prod_{k=1}^\infty \mathbb{R}_k$ so that its closure under countably many operations $\cup$ and $\cap$ form the σ-algebra $\mathfrak{B}^\infty$ generated by the products $\bigcap B_{j_k}$, $B_{j_k} \in \mathfrak{B}_{j_k}$.

Let N be a subset of integers. Denote by $\mathbb{R}^N = \prod_{k \in N} \mathbb{R}_k$ the direct product of the spaces $\mathbb{R}_k$ over $k \in N$, $\mathfrak{B}^N = \prod_{k \in N} \mathfrak{B}_k$. We say that distributions $\mathbf{P}_{N'}$ and $\mathbf{P}_{N''}$ on $\langle \mathbb{R}^{N'}, \mathfrak{B}^{N'} \rangle$ and $\langle \mathbb{R}^{N''}, \mathfrak{B}^{N''} \rangle$, respectively, are *consistent* if the measures induced by $\mathbf{P}_{N'}$ and $\mathbf{P}_{N''}$ on the intersection $\mathbb{R}^N = \mathbb{R}^{N'} \cap \mathbb{R}^{N''}$ (here $N = N' \cap N''$) coincide with each other. The measures on $\mathbb{R}^N$ are said to be the *projections* of $\mathbf{P}_{N'}$ and $\mathbf{P}_{N''}$, respectively, on $\mathbb{R}^N$. An answer to the above question about the existence of an infinite sequence of random variables is given by the following theorem (the proof of which is given in Appendix 2).

Theorem 3.5.1 (Kolmogorov) *Specifying a family of consistent distributions* $\mathbf{P}_N$ *on finite-dimensional spaces* $\mathbb{R}^N$ *defines a unique probability measure* $\mathbf{P}_\infty$ *on* $\langle \mathbb{R}^\infty, \mathfrak{B}^\infty \rangle$ *such that each probability* $\mathbf{P}_N$ *is the projection of* $\mathbf{P}_\infty$ *onto* $\mathbb{R}^N$.

It follows from this theorem, in particular, that one can always define on an appropriate space an infinite sequence of arbitrary independent random variables. Indeed, direct products of measures given on $\mathbb{R}_1, \mathbb{R}_2, \dots$ for different products $\mathbb{R}^{N'}$ and $\mathbb{R}^{N''}$ are always consistent.

3.6 Integrals

3.6.1 Integral with Respect to Measure

As we have already noted, defining a probability space includes specifying a finite countably additive measure. This enables one to consider integrals with respect to the measure,

$$\int g\big(\xi(\omega)\big)\mathbf{P}(d\omega) \tag{3.6.1}$$

over the set Ω for a Borel function g and any random variable ξ on $\langle \Omega, \mathfrak{F}, \mathbf{P} \rangle$ (recall that $g(x)$ is said to be Borel if, for any t, $\{x : g(x) < t\}$ is a Borel set on the real line).

The definition, construction and basic properties of the integral with respect to a measure are assumed to be familiar to the reader. If the reader feels his or her background is insufficient in this aspect, we recommend Appendix 3 which contains all the necessary information. However, the reader could skip this material if he/she is willing to restrict him/herself to considering only discrete or absolutely continuous distributions for which integrals with respect to a measure become sums or conventional Riemann integrals. It would also be useful for the sequel to know the Stieltjes integral; see the comments in the next subsection.

We already know that a random variable $\xi(\omega)$ induces a measure $\mathbf{F}_\xi$ on the real line which is specified by the equality

$$\mathbf{F}_\xi\big([x, y)\big) = \mathbf{P}(x \le \xi \le y) = F_\xi(y) - F_\xi(x).$$

Using this measure, one can write the integral (3.6.1) as

$$\int g\big(\xi(\omega)\big)\mathbf{P}(d\omega) = \int g(x)\mathbf{F}_\xi(dx).$$

This is just the result of the substitution $x = \xi(\omega)$. It can be proved simply by writing down the definitions of both integrals. The integral on the right hand side is called the *Lebesgue–Stieltjes integral* of the function $g(x)$ with respect to the measure $\mathbf{P}_\xi$ and can also be written as

$$\int g(x)\, dF_\xi(x). \tag{3.6.2}$$

3.6.2 The Stieltjes Integral

The integral (3.6.2) is often just called the Stieltjes integral, or the Riemann–Stieltjes integral which is defined in a somewhat different way and for a narrower class of functions.

If $g(x)$ is a continuous function, then the Lebesgue–Stieltjes integral coincides with the Riemann–Stieltjes integral which is equal by definition to

$$\int g(x)\, dF(x) = \lim_{\substack{b\to\infty \\ a\to-\infty}} \lim_{N\to\infty} \sum_{k=0}^{N} g(\widetilde{x}_k)\big[F(x_{k+1}) - F(x_k)\big], \tag{3.6.3}$$

where the limit on the right-hand side does not depend on the choice of partitions $x_0, x_1, \ldots, x_N$ of the semi-intervals $[a, b)$ and points $\widetilde{x}_k \in \Delta_k = [x_k, x_{k+1})$. Partitions $x_0, x_1, \ldots, x_N$ are different for different N's and have the property that $\max_k(x_{k+1} - x_k) \to 0$ as $N \to \infty$.

Indeed, as we know (see Appendix 3), the Lebesgue–Stieltjes integral is

$$\int g(x)\, dF(x) = \lim_{\substack{b\to\infty \\ a\to-\infty}} \lim_{N\to\infty} \int_a^b g_N(x)\mathbf{F}_\xi(dx), \tag{3.6.4}$$

where g_N is any sequence of simple functions (assuming finitely many values) converging monotonically to $g(x)$. We see from these definitions that it suffices to show that the integrals $\int_a^b$ with finite integration limits coincide. Since the Lebesgue–Stieltjes integral $\int_a^b g\,dF$ of a continuous function g always exists, we could obtain its value by taking the sequence g_N to be any of the two sequences of simple functions g_N^* and g_N^{**} which are constant on the semi-intervals Δ_k and equal on them to

$$g_N^*(x_k) = \sup_{x\in\Delta_k} g(x) \quad \text{and} \quad g_N^{**}(x_k) = \inf_{x\in\Delta_k} g(x),$$

respectively. Both sequences in (3.6.4) constructed from g_N^* and g_N^{**} will clearly converge monotonically from different sides to the same limit equal to the Lebesgue–Stieltjes integral

$$\int_a^b g(x)\,dF(x).$$

But for any $\widetilde{x}_k \in \Delta_k$, one has

$$g_N^{**}(x_k) \le g(\widetilde{x}_k) \le g_N^*(x_k),$$

and therefore the integral sum in (3.6.3) will be between the bounds

$$\int_a^b g_N^{**}\,dF(x) \le \sum_{k=0}^{N} g(\widetilde{X}_k)\big[F(x_{k+1}) - F(x_k)\big] \le \int_a^b g_N^*\,dF(x).$$

These inequalities prove the required assertion about the coincidence of the integrals.

It is not hard to verify that (3.6.3) and (3.6.4) will also coincide when $F(x)$ is continuous and $g(x)$ is a function of bounded variation. In that case,

$$\int_a^b g(x)\,dF(x) = g(x)F(x)\big|_a^b - \int_a^b F(x)\,dg(x).$$

Making use of this fact, we can extend the definition of the Riemann–Stieltjes integral to the case when $g(x)$ is a function of bounded variation and $F(x)$ is an arbitrary distribution function. Indeed, let $F(x) = F_c(x) + F_d(x)$ be a representation of $F(x)$ as a sum of its continuous and discrete components, and $y_1, y_2, \ldots$ be the jump points of $F_d(x)$:

$$p_k = F_d(y_k + 0) - F_d(y_k) > 0.$$

Then one has to put by definition

$$\int g(x)\,dF(x) = \sum p_k g(y_k) + \int g(x)\,dF_c(x),$$

where the Riemann–Stieltjes integral $\int g\,dF_c(x)$ can be understood, as we have already noted, in the sense of definition (3.6.3).

We will say, as is generally accepted, that $\int g\,dF$ exists if the integral $\int |g|\,dF$ is finite. It is easy to see from the definition of the Stieltjes integral that, for step

functions $F(x)$ (the distribution is discrete), the integral becomes the sum

$$\int g(x)\,dF(x) = \sum_k g(x_k)\big(F(x_k+0) - F(x_k)\big) = \sum_k g(x_k)\mathbf{P}(\xi = x_k),$$

where $x_1, x_2, \ldots$ are jump points of $F(x)$. If

$$F(x) = \int_{-\infty}^{x} p(x)\,dx$$

is absolutely continuous and $p(x)$ and $g(x)$ are Riemann integrable, then the Stieltjes integral

$$\int g(x)\,dF(x) = \int g(x)p(x)\,dx$$

becomes a conventional Riemann integral.

We again note that for a *reader who is not familiar with Stieltjes integral techniques and integration with respect to measures, it is possible to continue reading the book keeping in mind only the last two interpretations of the integral*. This would be quite sufficient for an understanding of the exposition. Moreover, most of the distributions which are important from the practical point of view are just of one of these types: either discrete or absolutely continuous.

We recall some other properties of the Stieltjes integral (following immediately from definitions (3.6.4) or (3.6.3) and (3.6.5)):

$$\int_a^b dF = F(b) - F(a);$$

$$\int_a^b g\,dF = \int_a^c g\,dF + \int_c^b g\,dF \quad \text{if } g \text{ or } F \text{ is continuous at the point } c;$$

$$\int (g_1 + g_2)\,dF = \int g_1\,dF + \int g_2\,dF;$$

$$\int cg\,dF = c\int g\,dF \quad \text{for } c = \text{const};$$

$$\int_a^b g\,dF = gF\big|_a^b - \int_a^b F\,dg$$

if g is a function of bounded variation.

3.6.3 *Integrals of Multivariate Random Variables. The Distribution of the Sum of Independent Random Variables*

Integrals with respect to measure (3.6.1) make sense for multivariate variables $\xi(\omega) = (\xi_1(\omega), \ldots, \xi_n(\omega))$ as well (one cannot say the same about Riemann–

Stieltjes integrals (3.6.3)). We mean here the integral

$$\int_{\Omega} g\big(\xi_1(\omega), \dots, \xi_n(\omega)\big)\mathbf{P}(d\omega), \tag{3.6.5}$$

where g is a measurable function mapping $\mathbb{R}^n$ into $\mathbb{R}$, so that $g(\xi_1(\omega), \dots, \xi_n(\omega))$ is a measurable mapping of Ω into $\mathbb{R}$.

If $\langle \mathbb{R}^n, \mathfrak{B}^n, \mathbf{F}_\xi \rangle$ is a sample probability space for ξ, then the integral (3.6.5) can be written as

$$\int_{\mathbb{R}^n} g(x)\mathbf{F}_\xi(dx), \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n.$$

Now turn to the case when the components $\xi_1, \dots, \xi_n$ of the vector ξ are independent and assume first that $n = 2$. For sets

$$B = B_1 \times B_2 = \big\{(x_1, x_2) : x_1 \in B_1,\ x_2 \in B_2\big\} \subset \mathbb{R}^2,$$

where B_1 and B_2 are measurable subsets of $\mathbb{R}$, one has the equality

$$\mathbf{P}(\xi \in B) = \mathbf{P}(\xi_1 \in B_1,\ \xi_2 \in B_2) = \mathbf{P}(\xi_1 \in B_1)\mathbf{P}(\xi_2 \in B_2). \tag{3.6.6}$$

In that case one says that the measure $\mathbf{F}_{\xi_1,\xi_2}(dx_1, dx_2) = \mathbf{P}(\xi_1 \in dx_1, \xi_2 \in dx_2)$ on $\mathbb{R}^2$, corresponding to (ξ_1, ξ_2), is a *direct product of the measures*

$$\mathbf{F}_{\xi_1}(dx_1) = \mathbf{P}(\xi_1 \in dx_1) \quad \text{and} \quad \mathbf{F}_{\xi_2}(dx_2) = \mathbf{P}(\xi_2 \in dx_2).$$

As we already know, equality (3.6.6) uniquely specifies a measure on $\langle \mathbb{R}^2, \mathfrak{B}^2 \rangle$ from the given distributions of ξ_1 and ξ_2 on $\langle \mathbb{R}, \mathfrak{B} \rangle$. It turns out that the integral

$$\int g(x_1, x_2)\mathbf{F}_{\xi_1\xi_2}(dx_1, dx_2) \tag{3.6.7}$$

with respect to the measure $\mathbf{F}_{\xi_1,\xi_2}$ can be expressed in terms of integrals with respect to the measures $\mathbf{F}_{\xi_1}$ and $\mathbf{F}_{\xi_2}$. Namely, Fubini's theorem holds true (for the proof see Appendix 3 or property 5A in Sect. 4.8).

Theorem 3.6.1 (Theorem on iterated integration) *For a Borel function $g(x, y) \geq 0$ and independent ξ_1 and ξ_2,*

$$\int g(x_1, x_2)\mathbf{F}_{\xi_1\xi_2}(dx_1, dx_2) = \int \left[\int g(x_1, x_2)\mathbf{F}_{\xi_2}(dx_2) \right] \mathbf{F}_{\xi_1}(dx_1). \tag{3.6.8}$$

If $g(x, y)$ can assume values of different signs, then the existence of the integral on the left-hand side of (3.6.8) *is required for the equality* (3.6.8). *The order of integration on the right-hand side of* (3.6.8) *may be changed.*

It is shown in Appendix 3 that the measurability of $g(x, y)$ implies that of the integrands on the right-hand side of (3.6.8).

Corollary 3.6.1 *Let $g(x_1, x_2) = g_1(x_1)g_2(x_2)$. Then, if at least one of the following three conditions is met*:

(1) $g_1 \geq 0$, $g_2 \geq 0$,

(2) $\int g_1(x_1)g_2(x_2)\mathbf{F}_{\xi_1\xi_2}(dx_1, dx_2)$ *exists*,
(3) $\int g_j(x_j)\mathbf{F}_{\xi_j}(dx_j)$, $j = 1, 2$, *exist*,

then

$$\int g_1(x_1)g_2(x_2)\mathbf{F}_{\xi_1\xi_2}(dx_1, dx_2) = \int g_1(x_1)\mathbf{F}_{\xi_1}(dx_1)\int g_2(x_2)\mathbf{F}_{\xi_2}(dx_2). \quad (3.6.9)$$

To avoid trivial complications, we assume that $\mathbf{P}(g_j(\xi_j) = 0) \neq 1$, $j = 1, 2$.

Proof Under any of the first two conditions, the assertion of the corollary follows immediately from Fubini's theorem. For arbitrary g_1, g_2, put $g_j = g_j^+ - g_j^-$, $g_j^\pm \geq 0$, $j = 1, 2$. If $\int g_j^\pm d\mathbf{F}_\xi < \infty$ (we will use here the abridged notation for integrals), then

$$\begin{aligned}
\int g_1 g_2\, d\mathbf{F}_{\xi_1}\, d\mathbf{F}_{\xi_2} &= \int g_1^+ g_2^+ d\mathbf{F}_{\xi_1}\, d\mathbf{F}_{\xi_2} - \int g_1^+ g_2^-\, d\mathbf{F}_{\xi_1}\, d\mathbf{F}_{\xi_2} \\
&\quad - \int g_1^- g_2^+\, d\mathbf{F}_{\xi_1}\, d\mathbf{F}_{\xi_2} + \int g_1^- g_2^-\, d\mathbf{F}_{\xi_1}\, d\mathbf{F}_{\xi_2} \\
&= \int g_1^+\, d\mathbf{F}_{\xi_1} \int g_2^+\, d\mathbf{F}_{\xi_2} - \int g_1^+\, d\mathbf{F}_{\xi_1} \int g_2^+\, d\mathbf{F}_{\xi_2} \\
&\quad - \int g_1^-\, d\mathbf{F}_{\xi_1} \int g_2^+\, d\mathbf{F}_{\xi_2} + \int g_1^-\, d\mathbf{F}_{\xi_1} \int g_2^-\, d\mathbf{F}_{\xi_2} \\
&= \int g_1\, d\mathbf{F}_{\xi_1} \int g_2\, d\mathbf{F}_{\xi_2}.
\end{aligned}$$

□

Corollary 3.6.2 *In the special case when $g(x_1, x_2) = I_B(x_1, x_2)$ is the indicator of a set $B \in \mathfrak{B}^2$, we obtain the formula for sequential computation of the measure of B*:

$$\mathbf{P}\big((\xi_1, \xi_2) \in B\big) = \int \mathbf{P}\big((x_1, \xi_2) \in B\big)\mathbf{F}_{\xi_1}(dx_1).$$

The probability of the event $\{(x_1, \xi_2) \in B\}$ could also be written as $\mathbf{P}(\xi_2 \in B_{x_1}) = \mathbf{P}_{\xi_2}(B_{x_1})$ where $B_{x_1} = \{x_2 : (x_1, x_2) \in B\}$ is the "section" of the set B at the point x_1.

If $B = \{(x_1, x_2) : x_1 + x_2 < x\}$, we get

$$\begin{aligned}
\mathbf{P}\big((\xi_1, \xi_2) \in B\big) &= \mathbf{P}(\xi_1 + \xi_2 < x) \equiv F_{\xi_1+\xi_2}(x) \\
&= \int \mathbf{P}(x_1 + \xi_2 < x)\mathbf{F}_{\xi_1}(dx_1) \\
&= \int F_{\xi_2}(x - x_1)\, dF_{\xi_1}(x_1). \qquad (3.6.10)
\end{aligned}$$

We have obtained a formula for the distribution function of the sum of independent random variables expressing $F_{\xi_1+\xi_2}$ in terms of F_{ξ_1} and F_{ξ_2}. The integral on the right-hand side of (3.6.10) is called the *convolution* of the distribution functions

$F_{\xi_1}(x)$ and $F_{\xi_2}(x)$ and is denoted by $F_{\xi_1} * F_{\xi_2}(x)$. In the same way one can obtain the equality

$$\mathbf{P}(\xi_1 + \xi_2 < x) = \int_{-\infty}^{\infty} F_{\xi_1}(x-t)\, dF_{\xi_2}(t).$$

Observe that the right-hand side here could also be considered as a result of integrating

$$\int d F_{\xi_1}(t) F_{\xi_2}(x-t)$$

by parts.

If at least one of the distribution functions has a density, the convolution also has a density. This follows immediately from the formulas for convolution. Let, for instance,

$$F_{\xi_2}(x) = \int_{-\infty}^{x} f_{\xi_2}(u)\, du.$$

Then

$$\begin{aligned} F_{\xi_1+\xi_2}(x) &= \int_{-\infty}^{\infty} \mathbf{F}_{\xi_1}(dt) \int_{-\infty}^{x} f_{\xi_2}(u-t)\, du \\ &= \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} \mathbf{F}_{\xi_1}(dt)\, f_{\xi_2}(u-t) \right) du, \end{aligned}$$

so that the density of the sum $\xi_1 + \xi_2$ equals

$$f_{\xi_1+\xi_2}(x) = \int_{-\infty}^{\infty} \mathbf{F}_{\xi_1}(dt)\, f_{\xi_2}(x-t) = \int_{-\infty}^{\infty} f_{\xi_2}(x-t)\, dF_{\xi_1}(t).$$

Example 3.6.1 Let $\xi_1, \xi_2, \dots$ be independent random variables uniformly distributed over $[0, 1]$, i.e. $\xi_1, \xi_2, \dots$ have the same distribution function with density

$$f(x) = \begin{cases} 1, & x \in [0,1], \\ 0, & x \notin [0,1]. \end{cases} \tag{3.6.11}$$

Then the density of the sum $\xi_1 + \xi_2$ is

$$f_{\xi_1+\xi_2}(x) = \int_0^1 f(x-t)\, dt = \begin{cases} 0, & x \notin [0,2], \\ x, & x \in [0,1], \\ 2-x, & x \in [1,2]. \end{cases} \tag{3.6.12}$$

The integral present here is clearly the length of the intersection of the segments $[0, 1]$ and $[x-1, x]$. The graph of the density of the sum $\xi_1 + \xi_2 + \xi_3$ will consist of three pieces of parabolas:

$$f_{\xi_1+\xi_2+\xi_3}(x) = \int_0^1 f_{\xi_1+\xi_2}(x-t)\, dt = \begin{cases} 0, & x \notin [0,3], \\ \frac{x^2}{2}, & x \in [0,1], \\ 1 - \frac{(2-x)^2}{2} - \frac{(x-1)^2}{2}, & x \in [1,2], \\ \frac{(3-x)^2}{2}, & x \in [2,3]. \end{cases}$$

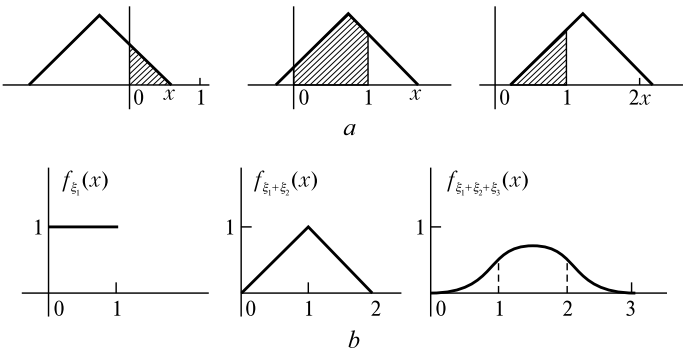

Fig. 3.2 Illustration to Example 3.6.1. The *upper row* visualizes the computation of the convolution integral for the density of $\xi_1 + \xi_2 + \xi_3$. The *lower row* displays the densities of ξ_1, $\xi_1 + \xi_2$, and $\xi_1 + \xi_2 + \xi_3$, respectively

The computation of this integral is visualised in Fig. 3.2, where the shaded areas correspond to the values of $f_{\xi_1+\xi_2+\xi_3}(x)$ for different x. The shape of the densities of ξ_1, $\xi_1 + \xi_2$ and $\xi_1 + \xi_2 + \xi_3$ is shown in Fig. 3.2b. The graph of the density of the sum $\xi_1 + \xi_2 + \xi_3 + \xi_4$ will consist of four pieces of cubic parabolas and so on. If we shift the origin to the point $n/2$, then, as n increases, the shape (up to a scaling transformation) of the density of the sum $\xi_1 + \cdots + \xi_n$ will be closer and closer to that of the function e^{-x^2}. We will see below that this is not due to chance.

In connection with this example we could note that if ξ and η are two independent random variables, ξ having the distribution function $F(x)$ and η being uniformly distributed over $[0, 1]$, then the density of the sum $\xi + \eta$ at the point x is equal to

$$f_{\xi+\eta}(x) = \int dF(t) f_\eta(x - t) = \int_{x-1}^{x} dF(t) = F(x) - F(x - 1).$$