3

Random Quantities and Random Vectors

By definition, values of random signals at a given sampling time are random quantities which can be distributed over a certain range of values. The tools for the precise, quantitative description of those distributions are provided by the classical *probability theory*. However natural, its development has to be handled with care since the overly heuristic approach can easily lead to apparent paradoxes.[10] But the basic intuitive idea that for independently repeated experiments, probabilities of their particular outcomes correspond to their relative frequencies of appearance, is correct. Although the concept of probability is more elementary than the concept of cumulative probability distribution functions, we assume that the reader is familiar with the former at the high school level, and start our exposition with the latter, which not only applies universally to all types of data, both discrete and continuous, but also gives us a tool to immediately introduce the probability calculus ideas, including the physically appealing probability density function.

Think here about an electrical engineer whose responsibility is to monitor the voltage on the electrical outlets in the university's circuits laboratory. The record of a month's worth of daily readings on a very sensitive voltmeter may look as follows:

```
109.779, 109.37,  110.733, 109.762, 110.364, 110.73,
109.906, 110.378, 109.132, 111.137, 109.365, 108.968,
111.275, 110.806, 110.99,  111.522, 110.728, 109.689,
111.163, 107.22,  109.661, 108.933, 111.057, 111.055,
112.392, 109.55,  111.042, 110.679, 111.431, 112.06.
```

Surprisingly, the voltage varies from day to day and this variability is visualized in Figure 3.0.1.

In the presence of such uncertainty he may want to get a better idea of how the voltage values are distributed within its range and he is

[10] See, e.g., Problem 3.7.25.

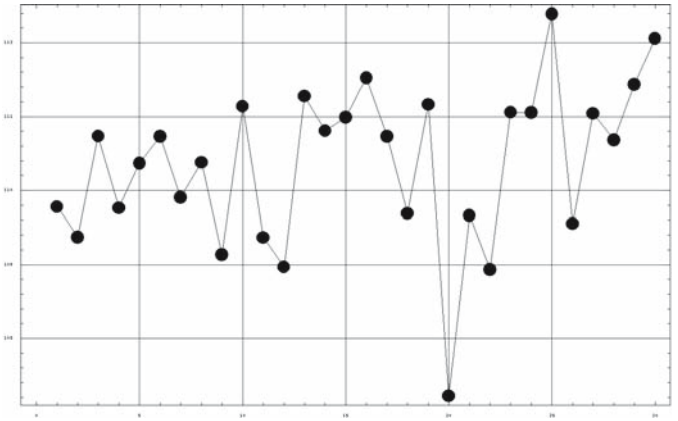

Fig. 3.0.1. Variability of daily voltage readings on an electrical outlet.

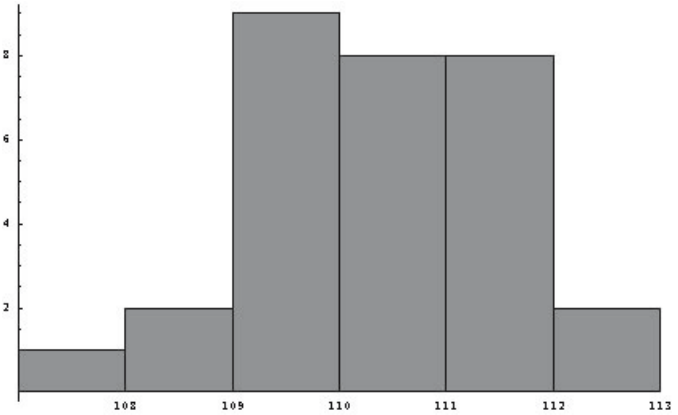

Fig. 3.0.2. The histogram of daily voltage readings on an electrical outlet.

likely to visualize this information in the form of a histogram shown in Figure 3.0.2.

In this chapter, we will discuss analytical tools for the study of such random quantities. The discrete and continuous random quantities are introduced, but we also show that, in the presence of fractal phenomena, the above classification is not exhaustive.

3.1 Discrete, continuous, and singular random quantities

For the purposes of this book, *random quantities* (also called *random variables* in the literature), denoted by capital letters X, Y, etc., will symbolize measurements of experiments with uncertain outcomes. A random quantity X will be fully characterized by its *probability distribution* $\mathbf{P}_X$, which, for any numbers $a < b$, assigns the probability

$$\mathbf{P}_X(a,b] = \mathbf{P}(a < X \le b) = \mathbf{P}(X \in (a,b])$$

that X takes values in the interval $(a,b]$. It is customary to assume that the probability measure $\mathbf{P}_X$ is *normalized*, that is,

$$\mathbf{P}_X(-\infty,+\infty) = \mathbf{P}(-\infty < X < +\infty) = 1, \tag{3.1.1}$$

and it is natural to demand that, if $a < b < c$, than

$$\mathbf{P}(a < X \le c) = \mathbf{P}(a < X \le b) + \mathbf{P}(b < X \le b). \tag{3.1.2}$$

This fundamental property of probabilities, called *additivity*, can be extended from disjoint intervals to more general disjoint[11] sets A and B, yielding the formula

$$\mathbf{P}(X \in A \cup B) = \mathbf{P}(X \in A) + \mathbf{P}(X \in B).$$

In other words, probability measure behaves like the area measure of planar sets.

Equivalently, one can completely characterize the probability distribution $\mathbf{P}$ of X by its *cumulative distribution function* (*c.d.f.*)

$$F_X(x) := \mathbf{P}(X \le x),$$

which gives the probability that the outcomes of experiment X do not exceed number x. Note that, in a sense, c.d.f. $F_X(x)$, which depends only on one variable x, is a simpler object than the probability distribution $\mathbf{P}_X(a,b]$, which depends on two. Properties (3.1.1)–(3.1.2) of $\mathbf{P}$ immediately imply the *normalization* and *monotonicity* of F_X,

$$F_X(-\infty) = 0, \quad x < y \Rightarrow F_X(x) \le F_X(y), \quad F_X(+\infty) = 1, \tag{3.1.3}$$

and the formula recovering $\mathbf{P}$ from F_X:

$$\mathbf{P}(a < X \le b) = F_X(b) - F_X(a). \tag{3.1.4}$$

Discrete probability distributions. A random quantity X with a discrete probability distribution takes on only (finitely or infinitely many) discrete values, say, $x_1, x_2, \ldots$, so that

$$\mathbf{P}(X = x_i) = p_i, \quad i = 1,2,\ldots, \quad 0 < p_i < 1, \qquad \sum p_i = 1. \tag{3.1.5}$$

In the discrete case, the c.d.f.

$$F_X(x) = \sum_{i=1}^{\infty} p_i u(x - x_i), \tag{3.1.6}$$

where $u(x)$ is the unit step function. In other words, the c.d.f. has jumps of size p_i at locations x_i, and is constant at other points of the real line.

[11] Recall that sets A and B are called *disjoint* if their intersection is the empty set, i.e., $A \cap B = \varnothing$.

Example 3.1.1 (Bernoulli distribution). In this case the values of X, that is the outcomes of the experiment, are assumed to be either 1 or 0 (think about it as a model of an experiment in which "success" or "failure" are the only possible outcomes), with $\mathbf{P}(X = 1) = p > 0$, $\mathbf{P}(X = 0) = q > 0$, with p, q satisfying condition $p + q = 1$. The c.d.f. of the Bernoulli random quantity is

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0; \\ q = 1 - p & \text{for } 0 \le x < 1; \\ 1 & \text{for } 1 \le x. \end{cases}$$

The Bernoulli family of distributions has one parameter p which must be a number between 0 and 1. Then $q = 1 - p$.

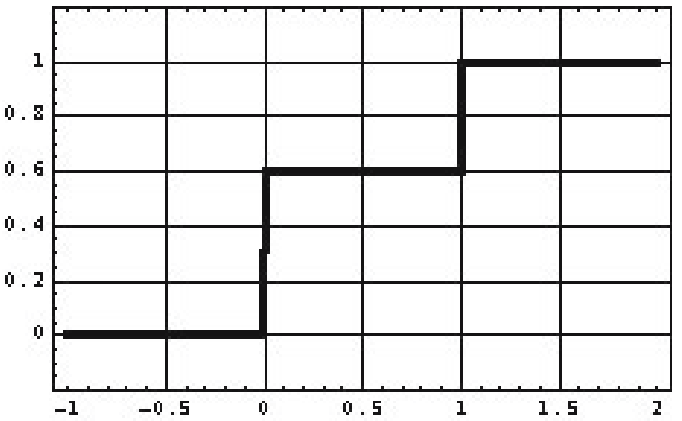

Fig. 3.1.1. Cumulative distribution function $F_X(x)$ of a Bernoulli random quantity X with parameter $p = 0.4$ has a jump of size $q = 1 - 0.4 = 0.6$ at $x = 0$ and a jump of size $p = 0.4$ at $x = 1$.

Example 3.1.2 (binomial distribution). The binomial random quantity X can take values $0, 1, \ldots, n$, with corresponding probabilities

$$p_k = \mathbf{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, 2, \ldots, n,$$

where the binomial coefficient is defined by

$$\binom{n}{k} = \frac{n!}{k!(n - k)!}.$$

Recall, that the name "binomial coefficient" comes from the elementary *binomial formula*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k},$$

familiar in the special cases:

$$(a+b)^2 = a^2 + 2ab + b^2,$$
$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3,$$

and so on.

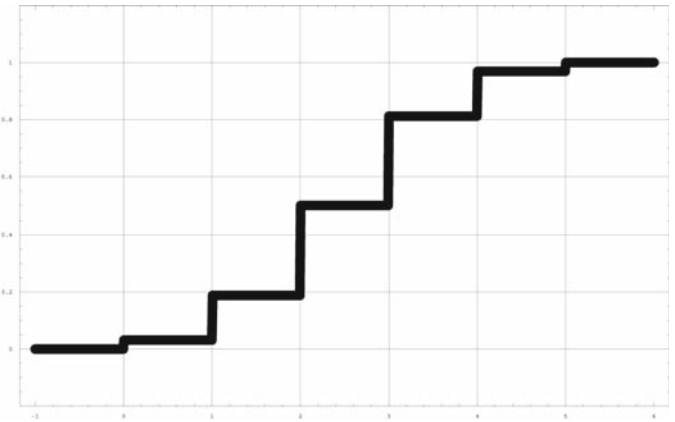

Fig. 3.1.2. Cumulative distribution function $F_X(x)$ of a binomial random quantity X with parameters $p = 0.5$ and $n = 5$.

Probabilities $p_k = p_k(n,p)$ in the binomial probability distribution are probabilities that exactly k "successes" occur in n independent[12] Bernoulli experiments in each of which the probability of "success" is p.

The normalization condition $\sum_k p_k = 1$ (3.1.5) is here satisfied because, in view of the above-mentioned binomial formula,

$$1 = (p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}.$$

The binomial family of distributions has two parameters: p, which must be between 0 and 1, and n, which can be an arbitrary positive integer.

Example 3.1.3 (Poisson distribution). The values of a Poisson random quantity X can be arbitrary nonnegative integers $0, 1, 2, \ldots$, and their probabilities are defined by the formula

[12] A rigorous definition of the concept of independence of random quantities will be discussed later on in this chapter.

$$p_k = \mathbf{P}(X = k) = e^{-\mu}\frac{\mu^k}{k!}, \quad k = 0, 1, 2, \ldots.$$

The normalization condition $\sum_k p_k = 1$ is satisfied in this case because of the power series expansion for the exponential function:

$$\sum_{k=0}^{\infty} e^{-\mu}\frac{\mu^k}{k!} = e^{-\mu}\sum_{k=0}^{\infty}\frac{\mu^k}{k!} = e^{-\mu}e^{\mu} = 1.$$

The family of Poisson distributions has one parameter $\mu > 0$. Poisson random quantities are often used as models of numbers of arrivals of "customers" in queuing systems (an Internet website, a line at the checkout counter, etc.) within a given time interval.

Continuous distributions. A random quantity X is said to have a continuous probability distribution[13] if its c.d.f. $F_X(x)$ can be written as an integral of a certain nonnegative function $f_X(x)$ which traditionally is called the *probability density function* (*p.d.f.*) of X, that is,

$$F_X(x) = \mathbf{P}(X \le x) = \int_{-\infty}^{x} f_X(z)dz. \tag{3.1.7}$$

Then, of course, the probability of the random quantity to assume values between a and b is just the integral of the p.d.f. over the interval $[a, b]$; see Figure 3.1.3, where $f_X(x)$ was selected to be $\frac{3}{5\sqrt{\pi}}e^{-x^2} + \frac{2}{5\sqrt{\pi}}e^{-(x-2)^2}$. Note that in the continuous case it does not matter whether the interval between a and b is open or closed. Thus we have

$$\mathbf{P}(a < X \le b) = F_X(b) - F_X(a) = \int_a^b f_X(z)dz. \tag{3.1.8}$$

Also, necessarily, we have the normalization condition

$$\int_{-\infty}^{\infty} f_X(x)dx = 1, \tag{3.1.9}$$

and, in view of (3.1.7), and the fundamental theorem of calculus, we can obtain the p.d.f. $f_X(x)$ by differentiation of the c.d.f. $F_X(x)$:

[13] Strictly speaking, c.d.f.s that admit the integral representation (3.1.7), that is, have densities, are called *absolutely continuous distributions* as there exist continuous c.d.f.s which do not admit this integral representation; see an example of a singular c.d.f. later in this section and, e.g., M. Denker and W. A. Woyczyński, *Introductory Statistics and Random Phenomena: Uncertainty, Complexity, and Chaotic Behavior in Engineering and Science*, Birkhäuser Boston, Cambridge, MA, 1998.

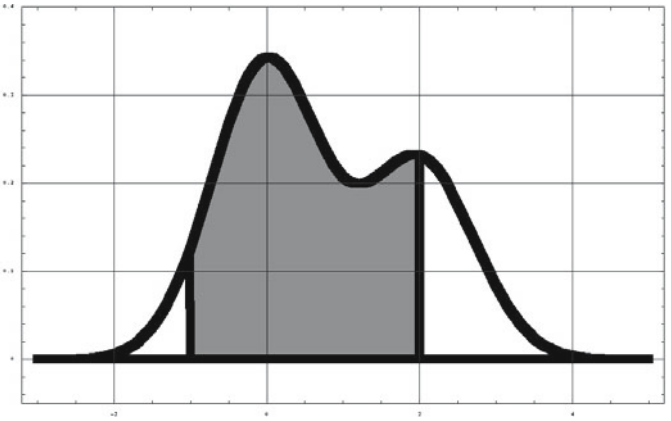

Fig. 3.1.3. The shaded area under $f_X(x)$, and above the interval $[-1,2]$ is equal to the probability that a random quantity X with p.d.f. $f_X(x)$ takes values in the interval $[-1,2]$.

$$\frac{d}{dx}F_X(x) = f_X(x).$$

Example 3.1.4 (uniform distribution). The density of a uniformly distributed random quantity X is defined to be a positive constant within a certain interval, say $[c,d]$, and zero outside this interval. Thus, because of the normalization condition (3.1.9),

$$f_X(x) = \begin{cases} (d-c)^{-1} & \text{for } c \le x \le d; \\ 0 & \text{elsewhere.} \end{cases}$$

The family of uniform densities is parametrized by two parameters c and d, with $c < d$.

The c.d.f. of a uniform random quantity is

$$F_X(x) = \begin{cases} 0 & \text{for } x < c; \\ \frac{x-c}{d-c} & \text{for } c \le x \le d; \\ 1 & \text{for } d \le x. \end{cases}$$

Example 3.1.5 (exponential distribution). An exponentially distributed random quantity X has the density of the form

$$f_X(x) = \begin{cases} 0 & \text{for } x < 0; \\ \frac{e^{-x/\mu}}{\mu} & \text{for } x \ge 0. \end{cases}$$

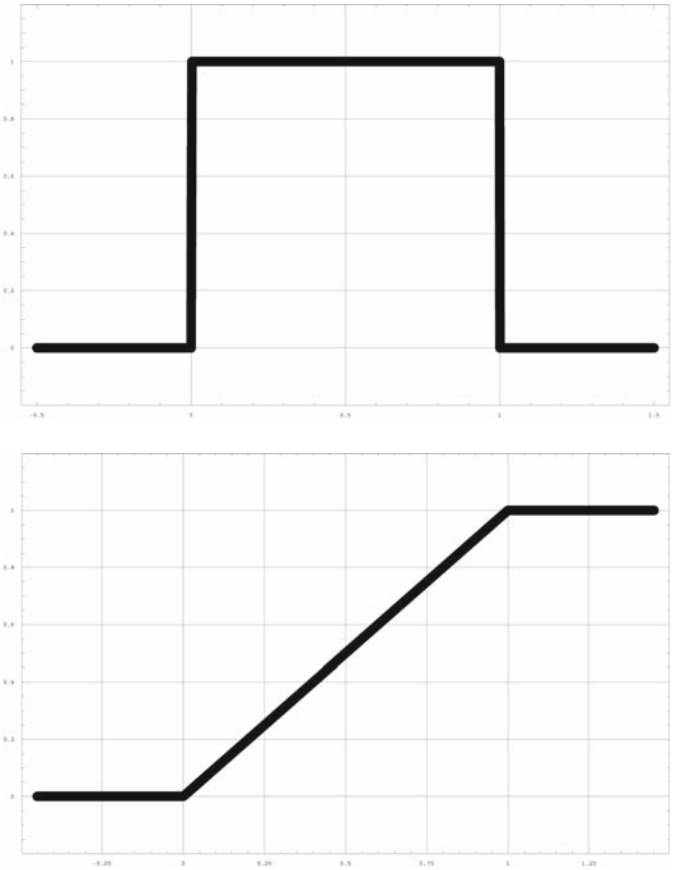

Fig. 3.1.4. *Top*: Probability density function (p.d.f) $f_X(x)$ for a random quantity with values uniformly distributed over the interval $[0, 1]$. *Bottom*: C.d.f. $F_X(x)$ for the same random quantity.

There is one parameter, $\mu > 0$. The c.d.f. in this case is easily computable:

$$F_X(x) = \begin{cases} 0 & \text{for } x < 0; \\ 1 - e^{-x/\mu} & \text{for } x \geq 0. \end{cases}$$

An exponential p.d.f. and the corresponding c.d.f. are pictured in Figure 3.1.3.

Exponential p.d.f.s often appear in applications as probability distributions of random waiting times between Poisson events discussed earlier in this section. For example, under certain simplifying assumptions, it can be proven that the time intervals between consecutive hits of a website have an exponential probability distribution. For this rea-

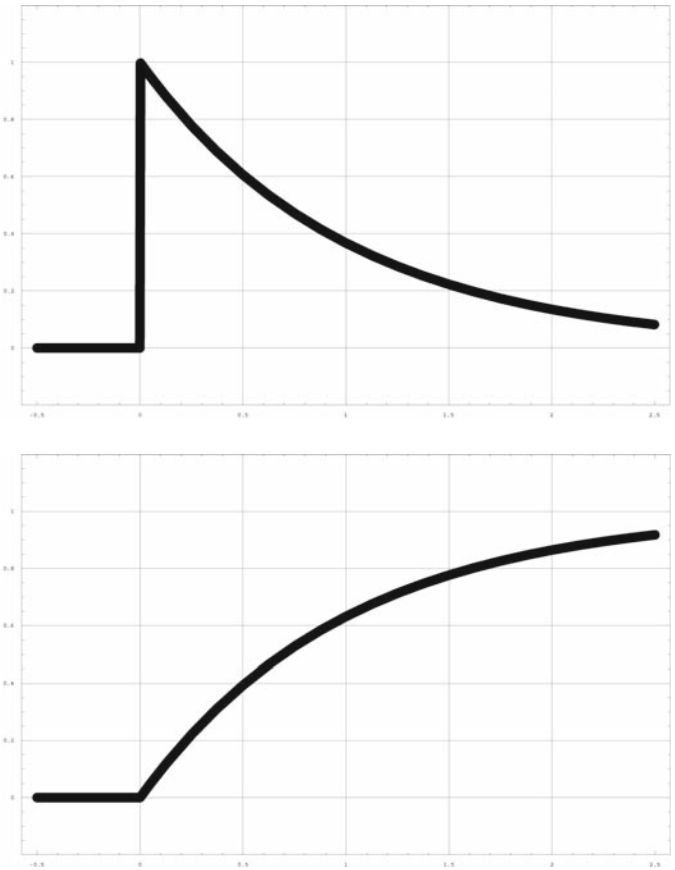

Fig. 3.1.5. *Top*: Probability density function (p.d.f.) $f_X(x)$ of an exponentially distributed random quantity with parameter $\mu = 1$. *Bottom*: Cumulative distribution function (c.d.f.) $F_X(x)$ for the same random quantity.

son, exponential p.d.f.s plays a crucial role in the analysis of Internet traffic and other queuing networks.

Example 3.1.6 (Gaussian (normal) distribution). The density of a Gaussian (also called normal) random quantity X is defined by the formula

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}.$$

There are two parameters—μ, which is a real number, and $\sigma > 0$—and this distribution is often called the $N(\mu, \sigma^2)$ p.d.f. (N for "normal"). The Gaussian c.d.f. is of the form (see Figure 3.1.4)

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-(z-\mu)^2/2\sigma^2} dz,$$

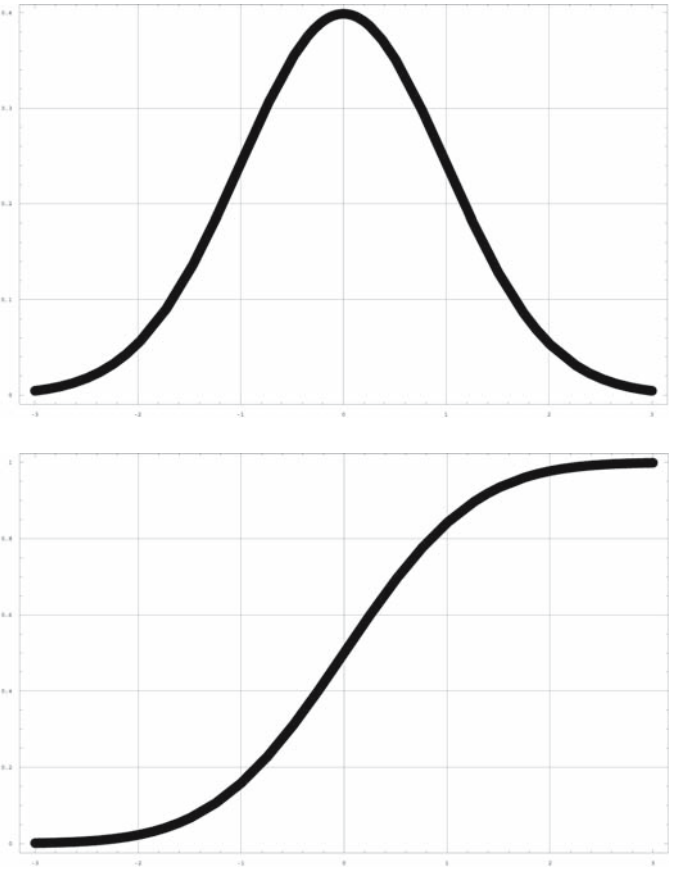

Fig. 3.1.6. *Top*: Probability density function (p.d.f.) $f_X(x)$ for a Gaussian random quantity with parameters $\mu = 0$, $\sigma = 1$. *Bottom*: Cumulative distribution function (c.d.f.) $F_X(x)$ for the same random quantity.

but, unfortunately, the integral cannot be expressed in terms of the elementary functions of the variable x. Thus the values of this c.d.f., and the probabilities of a Gaussian random quantity taking values within a given interval, have to be evaluated numerically, using tables (provided at the end of this chapter), or mathematical software such as MATLAB, MAPLE, or *Mathematica*; see the continuation of Example 3.1.6 below.

However, the normalization condition for the Gaussian p.d.f. can be verified directly analytically by a clever trick that replaces the square of the integral by a double integral which is then evaluated in polar coordinates r, θ. We carry out this calculation in the special case $\mu = 0$, $\sigma^2 = 1$:

$$\left(\int_{-\infty}^{\infty} f_X(x)dx \right)^2 = \int_{-\infty}^{\infty} f_X(x)dx \cdot \int_{-\infty}^{\infty} f_X(y)dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_X(x)\cdot f_X(y)dxdy$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-x^2-y^2/2}dxdy$$
$$= \frac{1}{2\pi}\int_{0}^{2\pi}\int_{0}^{\infty} e^{-r^2/2}rdrd\theta = 1.$$

Example 3.1.7 (calculations of $N(0,1)$ probabilities). The values of the Gaussian $N(0,1)$ cumulative distribution, traditionally denoted $\Phi(x)$, are tabulated at the end of this chapter. They are listed only for positive values of variable x, because, in view of the symmetry of the $N(0,1)$ density, we have

$$\Phi(-x) = 1 - \Phi(x).$$

Thus

$$\mathbf{P}(-1.53 < X < 2.11) = \Phi(2.11) - \Phi(-1.53) = \Phi(2.11) - (1 - \Phi(1.53))$$
$$\approx 0.9826 - (1 - 0.9370) = 0.9196.$$

Remark. The fundamental importance of the Gaussian probability distribution stems from the *central limit theorem* (see Section 3.5), which asserts that for a large number of independent repetitions of experiments with random outcomes, the fluctuations (errors) of the outcomes around their mean value have, approximately, a Gaussian p.d.f.

Mixed and singular distributions. A random quantity is said to have a c.d.f. of *mixed type* if it has both discrete and continuous components. The c.d.f. thus has both discrete jumps, perhaps infinitely many, as well as points of continuous increase where its derivative is well defined. For example, the c.d.f.

$$F_X(x) = \begin{cases} 0 & \text{for } x < -1; \\ \frac{x}{6} + \frac{2}{6} & \text{for } -1 \le x < 0; \\ \frac{x}{6} + \frac{4}{6} & \text{for } 0 \le x < 1; \\ 1 & \text{for } 1 \le x, \end{cases} \tag{3.1.10}$$

represents a random quantitiy X which is uniformly distributed on the interval $[-1,1]$ with probability $\frac{1}{3}$, but also takes the discrete values $-1, 0, 1$, with positive probabilites equal to the jump sizes of the c.d.f at those points. Thus, for example,

$$\mathbf{P}\left(-\frac{1}{2} < X \le \frac{1}{2}\right) = F_X\left(\frac{1}{2}\right) - F_X\left(-\frac{1}{2}\right) = \left(\frac{1}{12} + \frac{4}{6}\right) - \left(-\frac{1}{12} + \frac{2}{6}\right) = \frac{1}{2},$$

and

$$\mathbf{P}(X = 0) = \lim_{\epsilon\to 0}\mathbf{P}(-\epsilon < X \le \epsilon) = \lim_{\epsilon\to 0}(F_X(\epsilon) - F_X(-\epsilon))$$

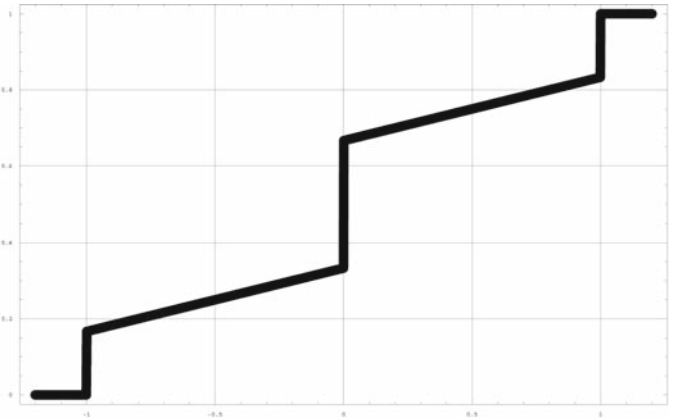

Fig. 3.1.7. Cumulative distribution function (c.d.f.) $F_X(x)$ of mixed type described by formula (3.1.10). This distribution has both discrete and continuous components.

$$= \lim_{\epsilon \to 0} \left[\left(\frac{\epsilon}{6} + \frac{4}{6} \right) - \left(-\frac{\epsilon}{6} + \frac{2}{6} \right) \right] = \frac{1}{3}.$$

Similarly,

$$\mathbf{P}(X = -1) = \frac{1}{6}, \qquad \mathbf{P}(X = 0) = \frac{2}{6}, \qquad \mathbf{P}(X = 1) = \frac{1}{6}.$$

Remark. The reader will notice that the example of a p.d.f. which appeared in Figure 3.1.3 is a mixture of two Gaussian p.d.f.s.

It is tempting to venture a guess that all c.d.f.s have to be either discrete, continuous, or of mixed type. This, however, is not the case.

The limit of the so-called "devil's staircase" c.d.f.s shown in Figure 3.1.8 is an example of a c.d.f. which, although continuous, does not have a p.d.f.

Observe that inside the interval $[0, 1]$ its derivative is 0 on the union of the infinite family of disjoint intervals whose lengths add up to 1. Indeed, as is clear from the construction displayed in Figure 3.1.8, this set has the linear measure

$$\lim_{n \to \infty} \left(\frac{1}{3} + 2 \cdot \frac{1}{3^2} + \cdots + 2^2 \cdot \frac{1}{3^n} \right) = \frac{1}{3} \sum_{i=0}^{\infty} \left(\frac{2}{3} \right)^i = \frac{1}{3} \cdot \frac{1}{1 - \frac{2}{3}} = 1,$$

in view of the formula for the sum of a geometric series. Thus integration of this derivative cannot possibly give a c.d.f. that grows from 0 to 1. Distributions of this type are called *singular* and they arise in studies of fractal phenomena. One can prove that the set of points of

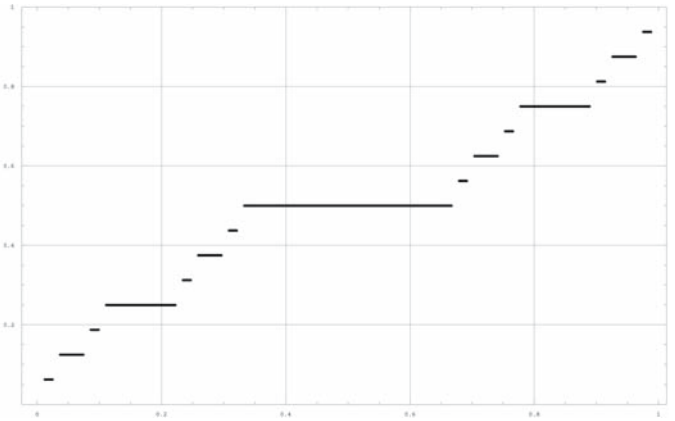

Fig. 3.1.8. The construction of the singular "devil's staircase" c.d.f. $F_X(x)$. It continuously grows from 0 at $x = 0$ to 1 at $x = 1$, and yet it has no density; its derivative is equal to 0 on disjoint intervals whose lengths add up to 1.

increase of the limit "devil's staircase," i.e., the set of points on which the probability is concentrated, has a fractional dimension equal to $\frac{\ln 2}{\ln 3} = 0.6309\ldots$.[14]

Distributions of functions of random quantities. One often measures random quantities through devices that distort the original quantity X to produce a new random quantity, say, $Y = g(X)$, and the natural question is how the c.d.f. $F_X(x)$ of X is affected by such a transformation. In other words, the question is: Can $F_Y(y)$ be expressed in terms of g and $F_X(x)$? In the case when the transforming function $g(x)$ is *monotonically increasing* the answer is simple:

$$F_{g(X)}(y) = \mathbf{P}(g(X) \le y) = \mathbf{P}(X \le g^{-1}(y)) = F_X(g^{-1}(y)), \quad (3.1.11)$$

where $g^{-1}(y)$ is the inverse function of $g(x)$, that is $g^{-1}(g(x)) = x$, or, equivalently, if $y = g(x)$ then $x = g^{-1}(y)$.

Remembering the chain rule of elementary calculus, and the formula for the derivative of the inverse function $g^{-1}(y)$, we also immediately obtain, in the case of monotonically increasing $g(x)$, the *expression of the p.d.f. of $Y = g(X)$ in terms of the p.d.f. of X itself*:

[14] See, for example, M. Denker and W. A. Woyczyński, *Introductory Statistics and Random Phenomena: Uncertainty, Complexity, and Chaotic Behavior in Engineering and Science*, Birkhäuser Boston, Cambridge, MA, 1998.

$$f_{g(X)}(y) = \frac{d}{dy} F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \cdot \frac{1}{g'(g^{-1}(y))}. \qquad (3.1.12)$$

Example 3.1.8 (linear transformation of a standard Gaussian random quantity). A Gaussian random quantity X is called *standard* (or $N(0,1)$) if its p.d.f. is of the form

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

It is a special case of the general Gaussian p.d.f. introduced in Example 3.1.6, with parameters μ and σ specified to be 0 and 1, respectively. Consider now a new random quantity Y obtained from X by a linear transformation

$$Y = aX + b, \quad a > 0.$$

Think about this transformation as representing the change in units of measurement and the choice of the origin (like changing the temperature measurements from degrees Celsius to Fahrenheit: if X represents temperature measurements in degrees Celsius, then $Y = 1.8 \cdot X + 32$ gives the same measurements in degrees Fahrenheit).

The transforming function in this case, $y = g(x) = ax + b$, is monotonically increasing, and

$$g'(x) = a \quad \text{and} \quad g^{-1}(y) = \frac{y - b}{a}.$$

Formula (3.1.12) now gives the following expression for the p.d.f. of Y:

$$f_Y(y) = 1\sqrt{2\pi} e^{-((y-b)/a)^2/2} \cdot \frac{1}{a} = \frac{1}{\sqrt{2\pi a^2}} e^{-(y-b)^2/2a^2}.$$

The conclusion is that the transformed random quantity Y also has a Gaussian p.d.f., but with parameters $\mu = b$ and $\sigma^2 = a^2$; in other words, Y is $N(b, a^2)$-distributed (in short, $Y \sim N(b, a^2)$).

Example 3.1.9 (calculation of general $N(\mu, \sigma^2)$ probabilities). The relationship established in Example 3.1.8 permits utilization of tables of the $N(0,1)$ distributions supplied at the end of this chapter to calculate $N(\mu, \sigma^2)$ probabilities for arbitrary values of parameter μ and $\sigma > 0$. Indeed, if a random quantity Y has the $N(\mu, \sigma^2)$ distribution, then it is of the form

$$Y = \sigma X + \mu,$$

where X has the $N(0,1)$ distribution, so that

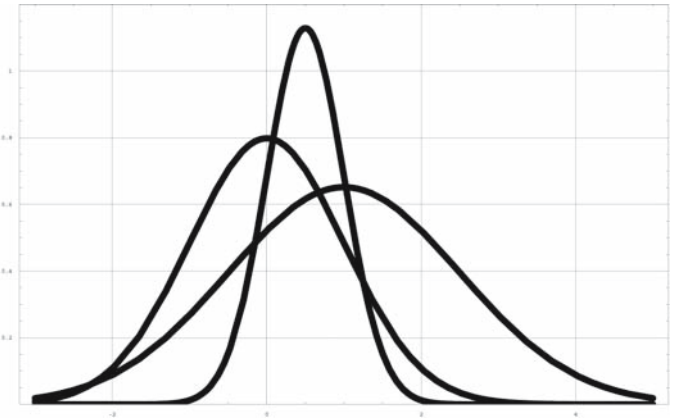

Fig. 3.1.9. Probability density functions of $N(0,1)$, $N(0.5,0.25)$, and $N(1,2.25)$ random quantities (from left to right).

$$F_Y(y) = \mathbf{P}(Y \le y) = \mathbf{P}(\sigma X + \mu \le y) = \mathbf{P}\left(X \le \frac{y-\mu}{\sigma}\right) = \Phi\left(\frac{y-\mu}{\sigma}\right), \tag{3.1.13}$$

and the values of the latter can be taken from the tables. For example, if Y is Gaussian with parameters $\sigma = 1.8$ and $\mu = 32$, then

$$\begin{aligned}\mathbf{P}(30 < Y < 36) &= \Phi\left(\frac{36-32}{1.8}\right) - \Phi\left(\frac{30-32}{1.8}\right) \\ &= \Phi(2.22) - (1 - \Phi(-1.11)) \\ &\approx 0.9868 - (1 - 0.8665) = 0.8533.\end{aligned}$$

In the next two examples we will consider the quadratic transformation $Y = \frac{X^2}{2}$ corresponding to calculation of the (random) kinetic energy[15] Y of an object of unit mass $m = 1$, traveling with random velocity X.

Example 3.1.10 (kinetic energy of a unit mass traveling with random, exponentially distributed velocity). Suppose that the random quantity X has an exponential c.d.f. and p.d.f. given in Example 3.1.5 with parameter $\mu = 1$. It is transformed by a quadratic "device" $g(x) = \frac{x^2}{2}$ into the random quantity $Y = \frac{X^2}{2}$. Note that the exponential p.d.f. is concentrated on the positive half-line and that the transforming function $g(x)$ is monotonically increasing in that domain. Then the c.d.f. $F_Y(y) = 0$ for $y \le 0$, and for $y > 0$ we can repeat the argument from formula (3.1.11) to obtain

[15] Recall that an object of mass m traveling with velocity v has kinetic energy $E = \frac{mv^2}{2}$.

$$F_Y(y) = \mathbf{P}(Y \le y) = \mathbf{P}\left(\frac{X^2}{2} \le y\right)$$
$$= \mathbf{P}\left(X \le \sqrt{2y}\right) = F_X\left(\sqrt{2y}\right) = 1 - e^{-\sqrt{2y}}.$$

Similarly, using (3.1.12), one gets the p.d.f. of $\frac{X^2}{2}$:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \begin{cases} 0 & \text{for } y \le 0; \\ \frac{e^{-\sqrt{2y}}}{\sqrt{2y}} & \text{for } y > 0. \end{cases}$$

Note that this p.d.f. has a singularity at the origin; indeed, $f_Y(y) \uparrow +\infty$ as $y \downarrow 0+$. Observe, however, that the singularity does not affect the p.d.f. normalization condition $\int_{-\infty}^{\infty} f_Y(y)dy = 1$.

If the transforming function $y = g(x)$ is *not monotonically increasing* (or decreasing; see Problem 3.7.26 and Sections 8.1–8.2) over the range of the random quantity X (as, for example, $g(x) = x^2$ in the case when X takes both positive and negative values), then a more subtle analysis is required to find the p.d.f. of the random quantity $Y = g(X)$.

Example 3.1.11 (square of a standard Gaussian random quantity). Assume that X has the standard $N(0,1)$ Gaussian p.d.f. and that the transforming function is quadratic: $y = g(x) = x^2$. The quadratic function is monotonically increasing only over the positive half-line; it is monotonically decreasing over the negative half-line. So, we have to proceed with caution, and start with the analysis of the c.d.f. of $Y = X^2$ by taking advantage of the symmetry of the Gaussian p.d.f.:

$$F_Y(y) = \mathbf{P}(Y \le y) = \mathbf{P}(X^2 \le y)$$
$$= 2\mathbf{P}(0 \le X \le \sqrt{y}) = 2\left(F_X(\sqrt{y}) - \frac{1}{2}\right).$$

The above formula, obviously, is valid only for $y > 0$; on the negative half-line the c.d.f. vanishes. Thus the p.d.f. of $Y = X^2$ is

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \begin{cases} 0 & \text{for } y \le 0; \\ \frac{e^{-y/2}}{\sqrt{2\pi y}} & \text{for } y > 0. \end{cases}$$

This p.d.f. is traditionally called the *chi-square* probability density function. We'll see its importance in Section 3.6, where it plays the central role in the statistical parameter estimation problems.

3.2 Expectations and moments of random quantities

The *expected value*, or, in brief, the *expectation* of a random quantity X is its mean value (or, for a physics-minded reader, the center of the

probability mass) with different values of X given weights equal to their probabilities. The expectation of X will be denoted $\mathbf{E}X$, or $\mathbf{E}(X)$, whichever is more convenient. So for a discrete random quantity X with $\mathbf{P}(X = x_i) = p_i$, $\sum_i p_i = 1$, we have

$$\mathbf{E}X = \sum_i x_i p_i, \tag{3.2.1}$$

and for an (absolutely) continuous random quantity with probability density $f_X(x)$

$$\mathbf{E}X = \int_{-\infty}^{\infty} x f_X(x)\,dx. \tag{3.2.2}$$

More generally, one can consider the expectation of a function $g(X)$ of a random quantity X, which is defined by the formulas

$$\mathbf{E}[g(X)] = \begin{cases} \sum_i g(x_i)p_i & \text{in the discrete case;} \\ \int_{-\infty}^{\infty} g(x) f_X(x)\,dx & \text{in the continuous case.} \end{cases} \tag{3.2.3}$$

In particular, if $g(x) = x^k$, $k = 1, 2, \ldots$, then the numbers

$$\mu_k(X) = \mathbf{E}g(X) = \mathbf{E}X^k = \begin{cases} \sum_i x_i^k p_i & \text{in the discrete case;} \\ \int_{-\infty}^{\infty} x^k f_X(x)\,dx & \text{in the continuous case} \end{cases} \tag{3.2.4}$$

are called *kth moments* of X. The first moment $\mu_1 = \mu_1(X)$ is just the expectation of $\mathbf{E}X$ of the random quantity X.

If $g(x) = |x|^\alpha$, $-\infty < \alpha < \infty$, then

$$m_k(X) = \mathbf{E}|X|^\alpha$$

are called *αth absolute moments*, and for $g(x) = |x - \mu_1|^\alpha$, the numbers

$$\mathbf{E}|X - \mu_1|^\alpha = \mathbf{E}|X - \mathbf{E}X|^\alpha$$

are called αth central moments of X. The latter measure the mean value of the αth power of the deviation of the random quantity X from its expectation $\mathbf{E}X$. In other words, they provide a family of parameters which measure how the values of the random quantity are spread around its "center of mass." In the special case $\alpha = 2$, the second central moment

$$\mathbf{E}(X - \mathbf{E}X)^2 = \begin{cases} \sum_i (x_i - \mu_1)^2 p_i & \text{in discrete case;} \\ \int_{-\infty}^{\infty} (x - \mu_1)^2 f_X(x)\,dx & \text{in continuous case} \end{cases} \tag{3.2.5}$$

is called the *variance* of the random quantity X and denoted $\mathrm{Var}(X)$. Again, for a physics-minded reader, it is worth noticing that the variance is just the moment of inertia of the probability mass distribution. A simple calculation gives the formula

$$\mathrm{Var}(X) = \mathbf{E}X^2 - (\mathbf{E}X)^2, \tag{3.2.6}$$

which is sometimes simpler computationally than (3.2.5); the variance is thus the difference between the second moment (sometimes also called the *mean square* of a random quantity) and the square of the first moment. This rule is then often phrased as follows: *Variance is equal to the mean square minus the squared mean.*

Example 3.2.1 (moments of the Bernoulli distribution). For the Bernoulli random quantity X, with distribution given in Example 3.1.1, all the moments are

$$\mu_k(X) = 1^k \cdot p + 0^k \cdot (1-p) = p,$$

and the variance is

$$\mathrm{Var}(X) = (1-p)^2 p + (0-p)^2(1-p) = p(1-p).$$

Example 3.2.2 (mean and variance of the uniform distribution). A uniformly distributed random quantity X (see Example 3.1.4) has expectation

$$\mathbf{E}X = \int_c^d x \frac{1}{d-c} dx = \frac{d+c}{2}.$$

Its variance is

$$\mathrm{Var}(X) = \int_c^d \left(x - \frac{d+c}{2}\right)^2 \frac{1}{d-c} dx = \frac{(d-c)^2}{12}.$$

Notice that the *expectation* or *expected value* $\mathbf{E}X$ *of a random quantity* X *scales linearly*, that is,

$$\mathbf{E}(\alpha X) = \alpha \mathbf{E}(X), \quad -\infty < \alpha < \infty, \tag{3.2.7}$$

so that the change of scale of the measurements affects the expectations proportionally: if, for example, X is measured in meters, then $\mathbf{E}X$ is also measured in meters. Indeed, in the continuous case,

$$\mathbf{E}(\alpha X) = \int_{-\infty}^{\infty} (\alpha x) f_X(x) dx = \alpha \int_{-\infty}^{\infty} x f_X(x) dx = \alpha \mathbf{E}(X),$$

and the discrete case can be verified in an analogous fashion.

On the other hand, the *variance* $\mathrm{Var}(X)$ *has a quadratic scaling*

$$\mathrm{Var}(\alpha X) = \alpha^2 \mathrm{Var}(X). \tag{3.2.8}$$

This follows immediately from the linearity of the expectations (3.2.7) and formula (3.2.6). Thus the mean-square deviation has a somewhat unpleasant nonlinear property which implies that if X is measured, say, in meters, then its variance is measured in meters squared.

For this reason, one often considers the *standard deviation* $\mathrm{Std}(X)$ of a random quantity X which is defined as the square root of the variance:

$$\mathrm{Std}(X) = \sqrt{\mathrm{Var}(X)}. \tag{3.2.9}$$

The standard deviation scales linearly, at least for positive α, since

$$\mathrm{Std}(\alpha X) = |\alpha| \, \mathrm{Std}(X), \quad -\infty < \alpha < \infty. \tag{3.2.10}$$

This means that changing the measurement units affects the standard deviation proportionately as well. If a random quantity is measured in meters, then its standard deviation is also measured in meters.

Additionally, observe that the *expectation is additive with respect to constants*, that is, for any constant β, $-\infty < \beta < \infty$,

$$\mathbf{E}(X + \beta) = \mathbf{E}(X) + \beta. \tag{3.2.11}$$

The verification is again immediate and follows from the additivity property of the integrals (or, in the discrete case, sums):

$$\begin{aligned} \mathbf{E}(X + \beta) &= \int_{-\infty}^{\infty} (x + \beta) f_X(x) dx \\ &= \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} \beta f_X(x) dx = \mathbf{E}(X) + \beta \end{aligned}$$

because $\int_{-\infty}^{\infty} f_X(x) dx = 1$.

Finally, the *variance is invariant under translations*, that is, for any constant β, $-\infty < \beta < \infty$,

$$\mathrm{Var}(X + \beta) = \mathrm{Var}(X). \tag{3.2.12}$$

Indeed,

$$\mathrm{Var}(X+\beta) = \mathbf{E}((X+\beta) - \mathbf{E}(X+\beta))^2 = \mathbf{E}(X+\beta-\mathbf{E}(X)-\beta)^2 = \mathrm{Var}(X).$$

The above properties indicate that any random quantity X can be *standardized* by first centering it and then rescaling it so that the standardized random quantity has expectation 0 and variance 1. Indeed, if

$$Z = \frac{X - \mathbf{E}X}{\mathrm{Std}(X)}, \tag{3.2.13}$$

then it immediately follows from (3.2.10)–(3.2.11) that $\mathbf{E}Z = 0$ and $\sigma^2(Z) = 1$.

Example 3.2.3 (mean and variance of the Gaussian distribution). Let us begin with a random quantity X with the standard $N(0,1)$ p.d.f. Its expectation is

$$\mathbf{E}(X) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0$$

because the integrand is an odd function and is integrated over the interval $(-\infty, \infty)$ which is symmetric about the origin. Its variance is thus just the second moment (mean square) of X, which can be evaluated easily by integration by parts[16]

$$\begin{aligned}\mathrm{Var}(X) &= \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \cdot (x e^{-x^2/2}) dx. \\ &= \frac{1}{\sqrt{2\pi}} \left(-x \cdot e^{-x^2/2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-x^2/2} dx \right) = 1,\end{aligned}$$

because $\lim_{x \to \pm\infty} x \cdot e^{-x^2/2} = 0$ and $(\frac{1}{\sqrt{2\pi}}) \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1$.

Now let us consider a general Gaussian random quantity Y with $N(\mu, \sigma^2)$ p.d.f. In view of Example 3.1.8,

$$Y = \sigma X + \mu.$$

The above properties of the expectation and the variance ((3.2.7)–(3.2.8) and (3.2.11)–(3.2.12)) immediately give

$$\mathbf{E}(Y) = \mathbf{E}(\sigma X + \mu) = \sigma \mathbf{E}(X) + \mu = \mu$$

and

$$\mathrm{Var}(Y) = \mathrm{Var}(\sigma X + \mu) = \mathrm{Var}(\sigma X) = \sigma^2 \mathrm{Var}(X) = \sigma^2.$$

Thus the parameters μ and σ^2 in the Gaussian $N(\mu, \sigma^2)$ p.d.f. are, simply, its expectation and variance.

[16] Recall the integration-by-parts formula: $\int f(x) g'(x) dx = f(x)g(x) - \int f'(x) g(x) dx$.

3.3 Random vectors, conditional probabilities, statistical independence, and correlations

A *random vector* $\boldsymbol{X}$ has components $X_1, X_2, \ldots, X_d$, which are scalar random quantities, that is,

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_d),$$

where d is the dimension of the random vector and its statistical properties are completely determined by its *joint c.d.f.*

$$F_{(X_1,\ldots,X_d)}(x_1, \ldots, x_d) = \mathbf{P}(X_1 \le x_1, \ldots, X_d \le x_d).$$

For the sake of simplicity of notation, we shall consider first the case of dimension $d = 2$, and we shall write $\boldsymbol{X} = (X, Y)$. In the discrete case, for a random vector $\boldsymbol{X}$ taking discrete values $\boldsymbol{x} = (x, y)$, the joint probability distribution is

$$\mathbf{P}(\boldsymbol{X} = \boldsymbol{x}) = \mathbf{P}(X = x, Y = y) = p_{\boldsymbol{X}}(x, y), \tag{3.3.1}$$

and

$$\sum_{(x,y)} p_{\boldsymbol{X}}(x, y) = 1. \tag{3.3.2}$$

Example 3.3.1 (a Bernoulli random vector). The random vector (X, Y) takes values $(0, 0), (0, 1), (1, 0), (1, 1)$, with the following joint probabilities:

$$p_{(X,Y)}(0,0) = (1-p)^2, \qquad p_{(X,Y)}(0,1) = p(1-p),$$
$$p_{(X,Y)}(0,1) = (1-p)p, \qquad p_{(X,Y)}(1,1) = p^2.$$

It is easy to check that

$$\sum_{x=0}^{1} \sum_{y=0}^{1} p_{(X,Y)}(x, y) = 1.$$

In the special case $p = \frac{1}{2}$ all four possible values of this random vector are taken with the same probability equal to $\frac{1}{4}$.

A continuous random vector is characterized by its *joint p.d.f.* $f_{(X,Y)}(x, y)$, which is a nonnegative function of two variables x, y, such that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{(X,Y)}(x, y)\, dx\, dy = 1. \tag{3.3.3}$$

In this case, the probability that the random vector (X, Y) takes values in a certain domain A of the 2D space is calculated by evaluating the double integral of the joint p.d.f. over the domain A:

$$\mathbf{P}((X,Y) \in A) = \int\int_A f_{(X,Y)}(x,y)dxdy. \tag{3.3.4}$$

For example, if the domain A is a rectangle $[a,b] \times [c,d] = \{(x,y) : a \le x \le b, c \le y \le d\}$, then

$$\mathbf{P}((X,Y) \in A) = \mathbf{P}(a \le X \le b, c \le Y \le d) = \int_a^b \int_c^d f_{(X,Y)}(x,y)dydx. \tag{3.3.5}$$

If the domain $B = \{(x,y) : x^2 + y^2 \le R^2\}$ is a centered disk of radius R, then

$$\mathbf{P}((X,Y) \in B) = \mathbf{P}(X^2 + Y^2 \le R^2) = \int_{-R}^{R} \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} f_{(X,Y)}(x,y)dydx. \tag{3.3.6}$$

The graph of a 2D joint p.d.f. is a surface over the (x,y)-plane such that the volume underneath it is equal to 1; see (3.3.3).

Example 3.3.2 (a 2D Gaussian random vector). An example of the 2D Gaussian joint p.d.f. is given by the formula

$$f_{(X,Y)}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2}\right]. \tag{3.3.7}$$

where $\sigma_x, \sigma_y > 0$, and μ_x, μ_y are arbitrary real numbers. Figure 3.3.1 shows the plot of the surface representing a 2D Gaussian joint p.d.f. in the case $\sigma_x, \sigma_y = 1$ and $\mu_x, \mu_y = 0$.

Calculation of the probabilities $\mathbf{P}(a \le X \le b, c \le Y \le d)$ is here reduced to calculation of one-dimensional Gaussian probabilities since the joint 2D density in this case is a product of two 1D Gaussian densities, one depending only on x and the other on y,[17] and the double integral splits into a product of two single integrals. To obtain numerical values, tables of (or software for) 1D $N(0,1)$ c.d.f. have to be used; see Section 3.5.

In the special case of equal variances $\sigma_x^2 = \sigma_y^2 = \sigma^2$, the probability that the above Gaussian random vector takes values in a disk of radius R centered at (μ_x, μ_y) can, however, be carried out explicitly by calculation of the integral in polar coordinates (θ, r):

$$\begin{aligned}&\mathbf{P}((X-\mu_x)^2 + (Y-\mu_y)^2 \le R^2)\\ &\quad = \int_{-R}^{R} \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \frac{1}{2\pi\sigma^2} \exp\left[-\frac{x^2+y^2}{2\sigma^2}\right] dydx\end{aligned}$$

[17] We will have more to say about joint p.d.f.s of this type in the next few pages. The multiplicative property is equivalent to the concept of statistical independence of components of a random vector.

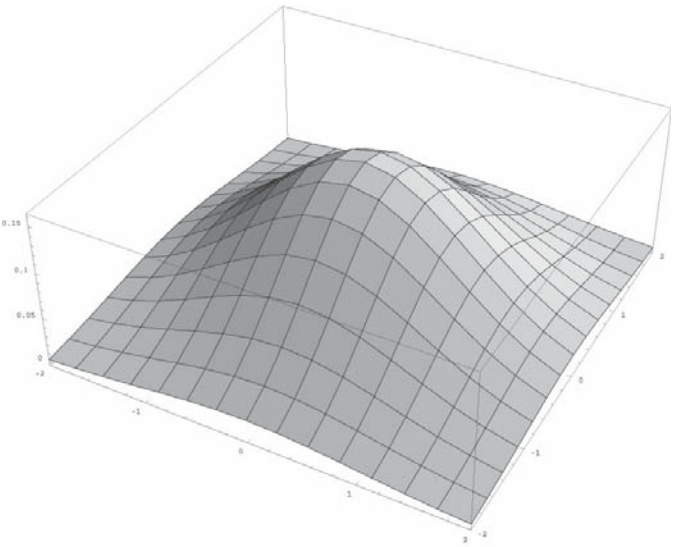

Fig. 3.3.1. Plot of the surface representing a 2D Gaussian joint p.d.f. (3.3.7) in the case $\sigma_x, \sigma_y = 1$ and $\mu_x, \mu_y = 0$.

$$= \frac{1}{2\pi\sigma^2} \int_0^{2\pi} \int_0^R \exp\left[-\frac{r^2}{2\sigma^2}\right] r\,dr\,d\theta$$
$$= \frac{1}{\sigma^2} \left[-\sigma^2 \exp\left[-\frac{r^2}{2\sigma^2}\right]\right]_0^R = 1 - e^{-R^2/2\sigma^2}.$$

Because the joint p.d.f. gives complete information about the random vector (X, Y), it also yields complete information about the probability distributions of each of the component random quantities. These distributions are called *marginal distributions* of the random vector. In particular, for a discrete random vector, the marginal distribution of the component X is

$$p_X(x) = \sum_y p_{(X,Y)}(x, y). \tag{3.3.8}$$

To find the probability of X taking a particular value x_0 we simply need to sum, over all possible ys, the probabilities of (X, Y) taking values (x_0, y). For a continuous random vector the marginal p.d.f. of the component X is

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x, y)\,dy. \tag{3.3.9}$$

It is important to observe that the marginal distributions of components of a random vector do not determine its joint distribution. Indeed, the example provided below shows that it is quite possible for

random vectors to have the same marginal probability distributions of their components while their joint probability distributions are different.

Example 3.3.3 (different random vectors with the same marginal probability distributions). A random vector (X, Y) has components X and Y that take values $1, 2, 3$ and $1, 2$, respectively. The joint probability distribution of this random vector is given in Table 3.3.1.

Table 3.3.1.

$Y \backslash X$	1	2	3	Y
1	$\frac{5}{24}$	$\frac{4}{24}$	$\frac{3}{24}$	$\frac{6}{12}$
2	$\frac{5}{24}$	$\frac{4}{24}$	$\frac{3}{24}$	$\frac{6}{12}$
X	$\frac{5}{12}$	$\frac{4}{12}$	$\frac{3}{12}$	$\sum = 1$

Thus, for example, $\mathbf{P}((X, Y) = (3, 2)) = \frac{3}{24}$. The last row in the above table gives the marginal probability distribution for the component X, and the last column, the marginal probability distribution for the component Y.

Now consider another random vector (W, Z) with components W and Z which also take values $1, 2, 3$ and $1, 2$, respectively. The joint distrbution of this random vector is given by Table 3.3.2.

Table 3.3.2.

$Z \backslash W$	1	2	3	Z
1	$\frac{1}{12}$	$\frac{2}{12}$	$\frac{3}{12}$	$\frac{6}{12}$
2	$\frac{4}{12}$	$\frac{2}{12}$	0	$\frac{6}{12}$
W	$\frac{5}{12}$	$\frac{4}{12}$	$\frac{3}{12}$	$\sum = 1$

This time, $\mathbf{P}((X, Y) = (3, 2)) = 0$. The last row in the above table gives the marginal probability distribution for the component W, and the last column, the marginal probability distribution for the component Z. The marginal probability distributions for vectors (X, Y) and (W, Z) are the same, while their joint distributions are different.

Conditional probabilities. Knowledge of the joint p.d.f. permits us also to introduce the concept of the *conditional probability* (in the discrete case) and the *conditional density* (in the continuous case). Thus, the conditional probability of the component X taking value x, given that the second component Y took value y, is given by the formula

$$p_{X|Y}(x|y) \equiv \mathbf{P}(X = x|Y = y) = \frac{\mathbf{P}(X = x, Y = y)}{\mathbf{P}(Y = y)} = \frac{p_{(X,Y)}(x, y)}{p_Y(y)}, \tag{3.3.10}$$

and the conditional probability density function of X given $Y = y$ is given by the formula

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)}. \tag{3.3.11}$$

In other words, conditional probability distributions are distributions of values of one component of a random vector calculated under the assumption that the value of the other component has already been determined.

Conditional probabilities are bona fide probabilities, as they satisfy the normalization property. Indeed, say, in the continuous case, for each fixed y,

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y)dx = \frac{\int_{-\infty}^{\infty} f_{(X,Y)}(x, y)dx}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} = 1,$$

in view of formula (3.3.9), which calculates the marginal density from the joint density.

If the component X of random vector (X, Y) takes on distinct values $x_1, x_2, \ldots, x_n$, then the additive property of probabilities immediately gives the following *total probability formula*:

$$\mathbf{P}(Y = y) = \sum_{i=1}^{n} \mathbf{P}(Y = y|X = x_i) \cdot \mathbf{P}(X = x_i).$$

Remark. Heuristically, one can think about conditional probabilities as probabiilities obtained under additional constraints. Think here about the probability of your running into a bear during a hike. If you are hiking in the city park, the probability of the event may be only 0.0001; in Yellowstone the similar conditional probability may be as high as 0.75. Now assume you participate, with 51 of your classmates, in a raffle and the prize is a trip to Yellowstone; the consolation prize is a group hike in the city park. The total probability of your running into a bear would then be $0.0001 \cdot \frac{51}{52} + 0.75 \cdot \frac{1}{52} \approx 0.015$.

One of the corollaries of the total probability formula is the celebrated *Bayes formula for reverse conditional probabilities* which, loosely speaking, computes the conditional probability of X, given Y, in terms of the conditional probabilities of Y, given X:

$$\mathbf{P}(X = x_i|Y = y) = \frac{\mathbf{P}(Y = y|X = x_i) \cdot \mathbf{P}(X = x_i)}{\sum_{i=1}^{n} \mathbf{P}(Y = y|X = x_i) \cdot \mathbf{P}(X = x_i)}.$$

Indeed,

$$\begin{aligned}\mathbf{P}(X = x_i|Y = y) &= \frac{\mathbf{P}(X = x_i, Y = y)}{\mathbf{P}(Y = y)} \cdot \frac{\mathbf{P}(X = x_i)}{\mathbf{P}(X = x_i)} \\ &= \frac{\mathbf{P}(Y = y|X = x_i) \cdot \mathbf{P}(X = x_i)}{\mathbf{P}(Y = y)},\end{aligned}$$

and an application of the total probability formula immediately gives the final result.

Example 3.3.4 (transmission of a binary signal in the presence of random errors). A channel transmits binary symbols 0 and 1 with random errors. The probability that the symbols 0 and 1 appear at the input of the channel are, respectively, 0.45 amd 0.55. Because of transmission errors, if the symbol 0 appears at the input, then the probability of it being received as 0 at the output is 0.95. The analogous conditional probability is 0.9, for the symbol 1 to be received, given that it was transmitted. Our task is to find the reverse conditional probability that the symbol 1 was transmitted given that 1 was received.

The random vector here is (X, Y), where X is the input signal and Y is the output signal. The problem's description contains the following information:

$$\mathbf{P}(X = 0) = 0.45, \qquad \mathbf{P}(X = 1) = 0.55,$$

and

$$\mathbf{P}(Y = 0|X = 0) = 0.95, \qquad \mathbf{P}(Y = 1|X = 1) = 0.9,$$

so that

$$\mathbf{P}(Y = 1|X = 0) = 0.05, \qquad \mathbf{P}(Y = 0|X = 1) = 0.1.$$

We are seeking $\mathbf{P}(X = 1|Y = 1)$ and the Bayes formula gives the answer:

$$\begin{aligned}&\mathbf{P}(X = 1|Y = 1) \\ &\quad = \frac{\mathbf{P}(Y = 1|X = 1) \cdot \mathbf{P}(X = 1)}{\mathbf{P}(Y = 1|X = 0) \cdot \mathbf{P}(X = 0) + \mathbf{P}(Y = 1|X = 1) \cdot \mathbf{P}(X = 1)} \\ &\quad = \frac{0.9 \cdot 0.55}{0.05 \cdot 0.45 + 0.9 \cdot 0.55} \approx 0.9565.\end{aligned}$$

Statistical independence. Components X and Y of a random vector $\boldsymbol{X} = (X, Y)$ are said to be *statistically independent* if the conditional probabilities of X given Y are independent of Y and vice versa. In the discrete case, this means that, for all x and y,

$$\mathbf{P}(X = x | Y = y) = \mathbf{P}(X = x),$$

which is equivalent to the statement that the joint p.d.f. is the product of the marginal p.d.f.s. Indeed, the above independence assumption and the formula defining the conditional probabilities yield

$$\begin{aligned} \mathbf{P}(X = x, Y = y) &= \mathbf{P}_{(X,Y)}(x, y) \\ &= \mathbf{P}_X(x) \cdot \mathbf{P}_Y(y) = \mathbf{P}(X = x) \cdot \mathbf{P}(Y = y). \end{aligned} \tag{3.3.12}$$

In the continuous case, the analogous definition of independence of X and Y can be stated via the multiplicative formula for the joint p.d.f.:

$$f_{(X,Y)}(x, y) = f_X(x) \cdot f_Y(y). \tag{3.3.13}$$

Note that both the 2D Bernoulli distribution of Example 3.3.1 and the 2D Gaussian distribution of Example 3.3.2 have statistically independent components X and Y. Also, components of the random vector (X, Y) in Example 3.3.3 are independent, as the table was actually obtained by multiplying the marginal probabilities in the corresponding rows and columns. However, the components W and Z of random vector (W, Z) in Example 3.3.3 are not statistically independent. To see this, it is sufficient to observe that

$$\mathbf{P}(W = 3, Z = 2) = 0,$$

but

$$\mathbf{P}(W = 3) \cdot \mathbf{P}(Z = 2) = \frac{3}{12} \cdot \frac{6}{12} = \frac{3}{24} \neq 0.$$

Moments of random vectors and correlations. If a random quantity Z is a function of a random vector (X, Y), say,

$$Z = g(X, Y),$$

then as in Section 3.2, we can calculate the expectation of Z using the joint p.d.f. Indeed,

$$\mathbf{E}Z = \sum_x \sum_y g(x, y) p_{(X,Y)}(x, y) \tag{3.3.14}$$

in the discrete case, and

$$\mathbf{E}Z = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{(X,Y)}(x,y)\,dx\,dy, \tag{3.3.15}$$

in the continuous case.

A mixed second-order moment corresponding to function $g(x,y) = xy$ will play a pivotal role in the analysis of random signals. The number

$$\varphi_{X,Y} = \mathbf{E}(X \cdot Y) \tag{3.3.16}$$

is called the *correlation* of random quantities X and Y. The related parameter corresponding to $g(x,y) = (x-\mu_x)(y-\mu_y)$,

$$\mathrm{Cov}(X,Y) = \mathbf{E}[(X-\mu_X)(Y-\mu_Y)] = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y), \tag{3.3.17}$$

is called the *covariance* of X and Y. Obviously, the covariance of X and X is just the variance of X:

$$\mathrm{Cov}(X,X) = \mathbf{E}[(X-\mu_X)(X-\mu_Y)] = \mathrm{Var}(X). \tag{3.3.18}$$

By the Cauchy–Schwartz inequality,[18]

$$|\,\mathrm{Cov}(X,Y)| \le \mathrm{Std}(X) \cdot \mathrm{Std}(Y). \tag{3.3.19}$$

This suggests the introduction of yet another parameter for a 2D random vector which is called the *correlation coefficient* of X and Y:

$$\mathrm{Cor}(X,Y) \equiv \rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Std}(X) \cdot \mathrm{Std}(Y)}. \tag{3.3.20}$$

In view of (3.3.19) the correlation coefficient is always contained between -1 and $+1$:

$$-1 \le \rho_{X,Y} \le 1, \tag{3.3.21}$$

and, in view of (3.3.18), if random components X and Y are linearly dependent, that is, $Y = \alpha X$, then the correlation coefficient takes its extreme values

$$\rho_{X,\alpha X} = \pm 1, \tag{3.3.22}$$

[18] Recall that if $\mathbf{a} = (a_1,\ldots,a_d)$ and $\mathbf{b} = (b_1,\ldots,b_d)$ are two d-dimensional vectors, then the Cauchy–Schwartz inequality says that the absolute value of their scalar (dot) product is not larger than the product of their norms (magnitudes), i.e., $|\langle \mathbf{a},\mathbf{b}\rangle| \le \|\mathbf{a}\| \cdot \|\mathbf{b}\|$, where $\langle \mathbf{a},\mathbf{b}\rangle = a_1b_1 + \cdots + a_db_d$, and $\|\mathbf{a}\|^2 = a_1^2 + \cdots + a_d^2$; see Section 3.7.

depending on whether α is positive or negative. In those cases we say that the random quantities X and Y are perfectly (positively or negatively) correlated.

The opposite case is that of statistically independent random quantities X and Y. Then, because of the multiplicative property $f_{(X,Y)}(x,y) = f_X(x)f_Y(y)$ (3.3.12)–(3.3.13) of the joint p.d.f., we always have

$$\mathbf{E}(XY) = \int\int xy f_X(x) f_Y(y) dx dy = \mathbf{E}X \cdot \mathbf{E}Y, \tag{3.3.23}$$

so that

$$\mathrm{Cov}(X,Y) = \mathbf{E}(X-\mu_X)(Y-\mu_Y) = \mathbf{E}(X-\mu_X)\cdot\mathbf{E}(Y-\mu_Y) = 0 \tag{3.3.24}$$

and the correlation coefficient $\rho_{X,Y} = 0$. In other words, statistically independent random quantities are always uncorrelated.[19] The correlation coefficient $\rho_{X,Y}$ is often considered as a measure of "independence" of random quantities X and Y; more appropriately, it should be interpreted as a measure of the "linear association" of random quantities X and Y.

Example 3.3.5 (a discrete 2D distribution with nontrivial correlation). Consider the random vector (W,Z) from Example 3.3.3. The expectations of the components are

$$\mathbf{E}W = 1\left(\frac{5}{12}\right) + 2\left(\frac{4}{12}\right) + 3\left(\frac{3}{12}\right) = \frac{11}{6},$$
$$\mathbf{E}Z = 1\left(\frac{6}{12}\right) + 2\left(\frac{6}{12}\right) = \frac{3}{2}.$$

The variances are

$$\mathrm{Var}(W) = \left(1-\frac{11}{6}\right)^2\left(\frac{5}{12}\right) + \left(2-\frac{11}{6}\right)^2\left(\frac{4}{12}\right) + \left(3-\frac{11}{6}\right)^2\left(\frac{3}{12}\right) = \frac{23}{36},$$
$$\mathrm{Var}(Z) = \left(1-\frac{3}{2}\right)^2\left(\frac{6}{12}\right) + \left(2-\frac{3}{2}\right)^2\left(\frac{6}{12}\right) = \frac{1}{4}.$$

The expectation of the product is

$$\mathbf{E}(WZ) = (1\cdot1)\left(\frac{1}{12}\right) + (2\cdot1)\left(\frac{2}{12}\right) + (3\cdot1)\left(\frac{3}{12}\right) + (1\cdot2)\left(\frac{4}{12}\right) + (2\cdot2)\left(\frac{2}{12}\right) + (3\cdot2)0 = \frac{5}{2}.$$

Thus the covariance is

$$\mathrm{Cov}(W,Z) = \mathbf{E}(WZ) - \mathbf{E}(W)\mathbf{E}(Z) = \frac{5}{2} - \left(\frac{11}{6}\right)\left(\frac{3}{2}\right) = -\frac{1}{4},$$

[19] The opposite statement is, in general, not true; see Problem 3.7.28.

and, finally, the correlation coefficient of W and Z is

$$\mathrm{Cor}(W,Z) = \frac{\mathrm{Cov}(W,Z)}{\mathrm{Std}(W)\cdot\mathrm{Std}(Z)} = -\frac{\frac{1}{4}}{\sqrt{\frac{23}{36}}\cdot\sqrt{\frac{1}{4}}} = -\sqrt{\frac{3}{23}} \approx -0.361.$$

Example 3.3.6 (a continuous 2D distribution with nontrivial correlation). A random vector (X,Y) has a continuous joint p.d.f. of the form

$$f_{(X,Y)}(x,y) = \begin{cases} C(1-(x+y)) & \text{for } x,y \geq 0,\ x+y \leq 1; \\ 0 & \text{elsewhere.} \end{cases}$$

The constant C can be determined from the normalization condition,

$$\int_0^1 \int_0^{1-x} C(1-(x+y))\,dy\,dx = 1,$$

which gives $C = 6$. The plot of the surface representing this density is given in Figure 3.3.2.

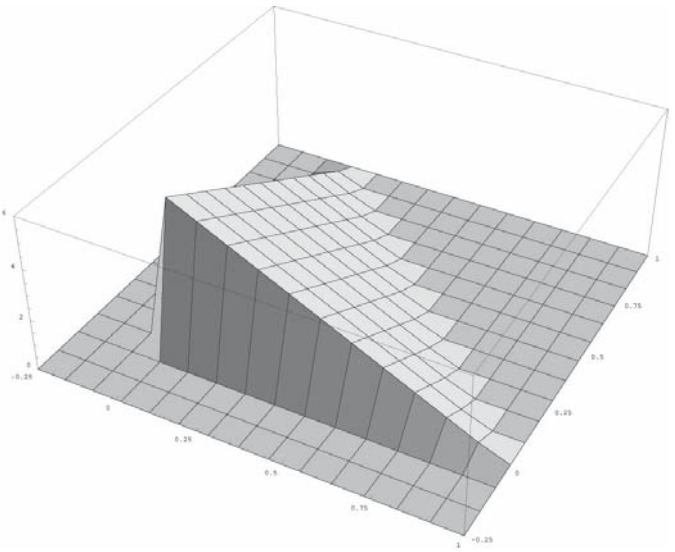

Fig. 3.3.2. The plot of the surface representing the joint p.d.f. from Example 3.3.6.

The marginal density of the component X,

$$f_X(x) = \int_0^{1-x} 6(1-(x+y))\,dy = 3(1-x)^2,$$

for $0 < x < 1$. It is equal to 0 elsewhere, and its plot is pictured in Figure 3.3.3.

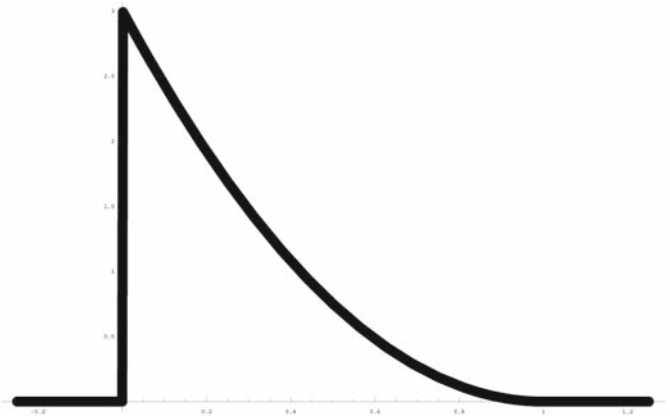

Fig. 3.3.3. The marginal density $F_X(x)$ of the X component of the random vector from Example 3.3.6.

The expectations of X and Y are easily evaluated using the marginal p.d.f.:

$$\mathbf{E}X = \mathbf{E}Y = \int_0^1 x \cdot 3(1-x)^2 dx = \frac{1}{4}.$$

Similarly, the variances are

$$\sigma^2(X) = \sigma^2(Y) = \int_0^1 \left(x - \frac{1}{4}\right)^2 \cdot 3(1-x)^2 dx = \frac{3}{80}.$$

Finally, the covariance is

$$\mathrm{Cov}(X,Y) = \int_0^1 \int_0^{1-x} \left(x - \frac{1}{4}\right)\left(y - \frac{1}{4}\right) \cdot 6(1-(x+y)) dy dx = -\frac{1}{80}.$$

So the random components X and Y are not independent; they are negatively correlated. The correlation coefficient itself is now easily evaluated to be

$$\rho_{X,Y} = \frac{-\frac{1}{80}}{\frac{3}{80}} = -\frac{1}{3}.$$

3.4 The least-squares fit, regression line

The roles of the covariance and the correlation coefficient will become better understood in the context of the following *least-squares regression* problem.

Consider a sample,

$$(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N),$$

of N 2D vectors. Its representation in the (x, y)-plane is called the *scatterplot* of the sample; see, for example, Figure 3.4.1. Our goal is to find a line

$$y = ax + b$$

that would provide the best approximation to the scatterplot in the sense of minimizing the sum of the squares of the errors of the approximation measured in the vertical direction. To be more precise, the error of the approximation for the ith sample point is expressed by the formula

$$\epsilon_i(a, b) = |y_i - (ax_i - b)|, \qquad i = 1, 2, \ldots, N,$$

and the sum of the squares of the errors,

$$\sum_{i=1}^{N} \epsilon_i^2(a, b) = \sum_{i=1}^{N} (y_i - (ax_i - b))^2,$$

is a nice, differentiable function of two variables a and b. We can find its minimum by taking partial derivatives with respect to a and b and equating them to 0:[20]

$$\frac{\partial}{\partial a} \sum_{i=1}^{N} \epsilon_i^2(a, b) = -2 \sum_{i=1}^{N} (y_i - (ax_i + b))x_i = 0,$$
$$\frac{\partial}{\partial b} \sum_{i=1}^{N} \epsilon_i^2(a, b) = -2 \sum_{i=1}^{N} (y_i - (ax_i + b)) = 0.$$

These two equations, sometimes called the *normal equations*, are linear in a and b and can be easily solved by the substitution method. To make the next step more transparent, we will introduce the following simplified notation for different sample means (think here about the means of random quantities with N possible values with each value assigned probability $\frac{1}{N}$):

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{M} x_i, \qquad \overline{y} = \frac{1}{N} \sum_{i=1}^{M} y_i,$$
$$\overline{x^2} = \frac{1}{N} \sum_{i=1}^{M} x_i^2, \qquad \overline{y^2} = \frac{1}{N} \sum_{i=1}^{M} y_i^2,$$
$$\overline{xy} = \frac{1}{N} \sum_{i=1}^{M} x_i y_i.$$

[20] This explains why we consider quadratic errors rather than the straight absolute errors; in the latter case the calculus tools would not work so well.

Now the normal equations for a and b can be written in the form

$$a\overline{x} + b - \overline{y} = 0 \quad \text{and} \quad a\overline{x}^2 + b\overline{x} - \overline{xy} = 0,$$

which can be immediately solved to give

$$b = \overline{y} - a\overline{x}, \quad a = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x}^2 - (\overline{x})^2}.$$

The first of the above two equations indicates that the point with coordinates formed by the sample means $\overline{x}$ and $\overline{y}$ is located on the regression line. To better see the meaning of the second equation, observe that

$$\overline{xy} - \overline{x} \cdot \overline{y} = \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})(y_i - \overline{y}) = \mathrm{Cov}(x, y)$$

is just the sample covariance of the x- and y-coordinates, and that

$$\overline{x}^2 - (\overline{x})^2 = \mathrm{Var}(x), \qquad \overline{y}^2 - (\overline{y})^2 = \mathrm{Var}(y).$$

Thus the equation $y = ax + b$ of the regression line can now be rewritten in the elegant form

$$\frac{y - \overline{y}}{\sqrt{\mathrm{Var}(y)}} = \rho_{x,y} \frac{x - \overline{x}}{\sqrt{\mathrm{Var}(x)}}, \tag{3.4.1}$$

where

$$\rho_{x,y} = \frac{\mathrm{Cov}(x, y)}{\sqrt{\mathrm{Var}(x)}\sqrt{\mathrm{Var}(y)}} = \frac{\mathrm{Cov}(x, y)}{\sigma_x \sigma_y},$$

is the sample correlation coefficient. Its significance is now clear: $\rho_{x,y}$ is the slope of the regression line but only after the x- and y-coordinates were standardized (see (3.2.11)), that is, they were centered by the means $\overline{x}$ and $\overline{y}$, and rescaled by the standard deviations σ_x and σ_y, respectively.

Example 3.4.1. Consider a 2D vector sample of size 10 (see Table 3.4.1).

Table 3.4.1.

x	1.06	2.08	3.28	4.13	5.28	6.39	7.12	8.04	9.23	10.38
y	1.10	3.37	3.23	6.92	7.66	6.78	8.12	9.94	9.55	10.87

The coefficients are $a = 0.9934$ and $b = 1.0925$, so that the equation of the regression line is

$$y = 0.9934 \cdot x + 1.0925$$

and the correlation coefficient is

$$\rho_{X,Y} = 0.1063.$$

The scatterplot of these data as well as the plot of the regression line are shown in Figure 3.4.1.

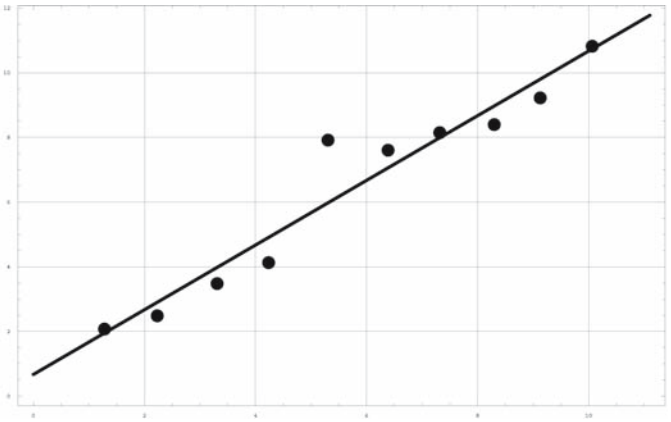

Fig. 3.4.1. The scatterplot and the least-squares fit regression line for data from Example 3.4.1.

3.5 The law of large numbers and the stability of fluctuations law

One of the fundamental theorems of statistics, called *the law of large numbers* (*LLN*), says that if $X_1, X_2, \ldots, X_n$ are independent random quantities with identical probability distributions (i.i.d.) and finite identical expectations $\mathbf{E}X_i = \mu_X$, then as $n \to \infty$, the averages converge to that expectation, i.e.,

$$\overline{X}_n \equiv \frac{X_1 + X_2 + \cdots + X_n}{n} \longrightarrow \mu_X \quad \text{as } n \to \infty. \tag{3.5.1}$$

Of course, the immediate issue is what do we mean by the convergence of random variables $\overline{X}_n$. For the purposes of these lectures the convergence of $\overline{X}_n$ to μ_X will mean that the standard deviation of the fluctuations of the averages $\overline{X}_n$ around the mean μ_X, that is, the differences $\overline{X}_n - \mu_X$, converge to zero as $n \to \infty$. More formally,

$$\lim_{n\to\infty} \mathrm{Std}(\overline{X}_n - \mu_X) = 0. \tag{3.5.2}$$

The statement (3.5.2) can be easily verified if we observe first that, for independent random quantities X and Y with finite variances, the variance

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y), \tag{3.5.3}$$

which follows immediately from the multiplicative property (3.3.23) of the expectations of independent random variables; see Section 3.3. Indeed, if X and Y are independent, then $X - \mu_X$ and $Y - \mu_Y$ are also independent, so that

$$\begin{aligned}\mathrm{Var}(X+Y) &= \mathbf{E}((X-\mu_X)+(Y-\mu_Y))^2 \\ &= \mathbf{E}(X-\mu_X)^2 + 2\mathbf{E}(X-\mu_X)\mathbf{E}(Y-\mu_Y) + \mathbf{E}(Y-\mu_Y)^2 \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y),\end{aligned}$$

because $\mathbf{E}(X-\mu_X) = \mathbf{E}(Y-\mu_Y) = 0$. Hence

$$\mathrm{Var}(\overline{X}_n - \mu_X) = \mathrm{Var}\left(\frac{X_1-\mu_X}{n} + \cdots + \frac{X_n-\mu_X}{n}\right) = \frac{\mathrm{Var}(X)}{n}, \tag{3.5.4}$$

which obviously approaches 0 as $n \to \infty$. Thus the *law of large numbers* (3.5.1), also often called the law of averages, is verified, at least in the situation when random quantities X_i have well-defined finite variances.[21]

A more subtle insight about the averages is provided by the following *stability of fluctuations law*, usually called the *central limit theorem* (*CLT*) in the mathematical and statistical literature. It states that as the averages $\overline{X}_n$ fluctuate around the expectation μ_X, the fluctuations, if viewed under a "magnifying glass," turn out to follow, asymptotically as $n \to \infty$, a Gaussian or normal probability distribution. More precisely, the c.d.f. of the standardized (see (3.2.13) and (3.5.4)) random fluctuations of the averages $\overline{X}_n$ around the mean μ_X,

$$Z_n = \frac{\sqrt{n}}{\mathrm{Std}(X)} \cdot (\overline{X}_n - \mu_X), \tag{3.5.5}$$

converges to the standard $N(0,1)$ Gaussian c.d.f., that is,

$$\lim_{n\to\infty} \mathbf{P}(Z_n \le z) = \Phi(z) \equiv \int_{-\infty}^{z} \phi(x)\,dx, \tag{3.5.6}$$

where the density is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}. \tag{3.5.7}$$

The important assumption of the central limit theorem is that the common variance of X_is is finite. It can be immediately verified that all of Z_ns have mean zero and variance one; see (3.2.13) and (3.5.4), but the

[21] Note that not all random quantities have well-defined, finite variances; see Section 3.7.

proof of the convergence to a Gaussian limit is more delicate; for its sketch, see Section 3.7.

So the central limit theorem can be loosely rephrased as follows:

Standardized random fluctuations of averages of independent and identically distributed random quantities around their common expected value have a limiting standard Gaussian p.d.f.

3.6 Estimators of parameters and their accuracy; confidence intervals

The law of large numbers can be reinterpreted as follows: If $X_1, X_2, \ldots, X_n$ is an i.i.d. sample from a certain probability distribution $F_X(x)$, then as n increases, the sample means $\overline{X}_n$, $n = 1, 2, \ldots$, become better and better estimators for the expectation of that distribution. In the statistical terminology the law of large numbers (3.5.1) says that $\overline{X}_n$ is a *consistent estimator* for parameter μ_X.

The central limit theorem (3.5.5)–(3.5.7) permits us to say what is the error of approximation of the theoretical mean μ_X by the sample mean $\overline{X}_n$, or, in other words, to establish the accuracy of the above estimation. Indeed, for a given sample of size n, the CLT says that the difference between the parameter μ_X and its estimator, the sample mean $\overline{X}_n$, is, after normalization by $\frac{\sqrt{n}}{\mathrm{Std}(X)}$, approximately $N(0,1)$ distributed so that, for large n,

$$\begin{aligned} \mathbf{P}\left(-\epsilon\frac{\mathrm{Std}(X)}{\sqrt{n}} \le \overline{X}_n - \mu_X \le \epsilon\frac{\mathrm{Std}(X)}{\sqrt{n}}\right) &\approx \Phi(\epsilon) - \Phi(-\epsilon) \\ &= 2\Phi(\epsilon) - 1 = C, \end{aligned} \tag{3.6.1}$$

where $C = C(\epsilon)$ is a nonrandom constant, depending on the choice of ϵ only.

If X itself has a Gaussian p.d.f., then the above approximate equality becomes exact for all n. This follows from the fact that the sum of two independent Gaussian random quantities is again a Gaussian random quantity, obviously with the mean and variance being the sums of means and variances, respectively, of the corresponding random summands; see Section 3.7.

The above statement can be reformulated as follows: *the parameter μ_X is contained in the random interval*

$$\left(\overline{X}_n - \epsilon\frac{\mathrm{Std}(X)}{\sqrt{n}}, \overline{X}_n + \epsilon\frac{\mathrm{Std}(X)}{\sqrt{n}}\right)$$

with probability C. This statement is sometimes abbreviated by writing

$$\mu_X = \overline{X}_n \pm \epsilon \frac{\mathrm{Std}(X)}{\sqrt{n}}$$

at the confidence level C. Note that it is the center of the above random interval that is random; its length is not random unless $\mathrm{Std}(X) = \sigma_X$ itself has to be estimated from the sample.

Example 3.6.1 (a 95% confidence interval for μ_X with known σ_X). One hundred independently repeated measurements of a random quantity X were conducted, resulting in $\overline{X}_{100} = 7.1$. Suppose that we know that $\mathrm{Std}(X) = 0.5$. To find the 95% confidence interval for μ_X using (3.5.1), we need to find ϵ such that $2\Phi(\epsilon) - 1 = 0.95$. From the table of the Gaussian $N(0,1)$ c.d.f. we have $z = 1.96$. Thus at the 95% confidence level,

$$7.1 - 1.96\frac{0.5}{\sqrt{100}} \le \mu_X \le 7.1 + 1.96\frac{0.5}{\sqrt{100}},$$

that is, $\mu_X = 7.1 \pm 0.098$ at the 95% confidence level. The above approximate confidence interval is exact if X has a Gaussian distribution.

Remark (*error of the Gaussian approximation in the CLT*). To be honest, we left open the essential, but delicate question of how good is the approximate equality in the basic formula (3.5.1) or, equivalently, the question of precise estimation of the error in the central limit theorem (3.5.6), which just says that the difference

$$\mathbf{P}(Z_n \le z) - \Phi(z) \to 0 \quad \text{as } n \to \infty,$$

where

$$Z_n = \frac{(X_1 + \cdots + X_n) - n\mu_X}{\sqrt{n} \cdot \mathrm{Std}(X)}$$

are standardized sums $X_1 + \cdots + X_n$. It turns out that the accuracy in CLT is actually pretty good if the X_is have finite higher absolute moments. In particular, if the third central moment $M_3 = \mathbf{E}|X - \mu_X|^3 < \infty$, then, for all $-\infty < x < \infty$ and $n = 1, 2, \ldots$,

$$|\mathbf{P}(Z_n \le z) - \Phi(z)| \le \frac{kM_3}{\sqrt{n}\sigma_X^3},$$

where k is a universal (independent of n and X) constant contained in the interval $(0.4097, 0.7975)$. Its exact value is not known.[22]

Of course, the above procedure used in Example 3.4.1 requires advance knowledge of the standard deviation $\mathrm{Std}(X)$. If that parameter is

[22] This error estimate in the CLT is known as the Berry–Esseen theorem and its proof can be found, for example, in V. V. Petrov's monograph *Sums of Independent Random Variables*, Springer-Verlag, Berlin, 1975.

unknown, then the obvious step is to try to estimate it from the sample $X_1, X_2, \ldots, X_n$ itself using the sample variance estimator

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2. \tag{3.6.2}$$

But in this case, even in the case of Gaussian X_i, the random quantity

$$T = \frac{\sqrt{n}}{S_n}(\overline{X} - \mu_X) \tag{3.6.3}$$

is no longer $N(0,1)$ distributed, so a simple construction of the confidence interval for μ_X is no longer possible.

However, in the case of a Gaussian random sample $X_1, X_2, \ldots, X_n$, it is known[23] that the random quantity T has the p.d.f.

$$f_T(x; n-1) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n-1}{2}\right)}\left(1 + \frac{x^2}{n-1}\right)^{-n/2}, \tag{3.6.4}$$

which traditionally is called the *Student-T* p.d.f. with $(n-1)$ degrees of freedom. The Gamma function $\Gamma(\alpha)$ appearing in the definition of f_T is defined by the formula

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx, \quad \alpha > 0. \tag{3.6.5}$$

It is worth noting that

$$\alpha\Gamma(\alpha) = \Gamma(\alpha+1) \quad \text{and} \quad \Gamma(n) = (n-1)!, \tag{3.6.6}$$

if n is a positive integer. Thus the Gamma function is an interesting extension of the concept of the factorial to noninteger numbers.

Therefore, in this case, the $C = (2F_T(\epsilon) - 1)$ confidence interval for μ_X is of the form

$$\left(\overline{X}_n - \epsilon\frac{S_n}{\sqrt{n}}, \overline{X}_n + \epsilon\frac{S_n}{\sqrt{n}}\right). \tag{3.6.7}$$

It is convenient to tabulate the quantiles $Q_T(\alpha; n)$ defined by the condition

$$F_T(Q_T(\alpha; n)) = \alpha, \tag{3.6.8}$$

[23] See, for example, M. Denker and W. A. Woyczyński, *Introductory Statistics and Random Phenomena: Uncertainty, Complexity, and Chaotic Behavior in Engineering and Science*, Birkhäuser Boston, Cambridge, MA, 1998, for more details on the statistical issues discussed in this section.

rather than the c.d.f. itself. Note that the quantile is just the function inverse to c.d.f. The tables of selected quantiles $Q_T(\alpha; n)$ are provided at the end of the chapter. Using the T-quantiles allows the C confidence level interval for μ_X to be simply written in the form

$$\left(\overline{X}_n - Q_T\,(1 + C2; n-1)\,\frac{S_n}{\sqrt{n}}, \overline{X}_n + Q_T\left(\frac{1+C}{2}; n-1\right)\frac{S_n}{\sqrt{n}}\right). \quad (3.6.9)$$

For large n, say, $n > 20$, the Student-T p.d.f. with n degress of freedom becomes almost indistinguishable from the $N(0,1)$ p.d.f. (see Exercise 3.7.17), and the latter can be used in the construction of confidence intervals even in the case of unknown variance.

Example 3.6.2 (a 90% confidence interval for μ_X with unknown Std(X)). Nine independent measurements of a Gaussian random quantity X resulted in $\overline{X}_9 = 2.56$ and $S_9 = 0.12$. With the desired confidence level $C = 0.9$, the table yields the quantile

$$Q_T\left(\frac{1+0.9}{2}; 8\right) = Q_T(0.95; 8) = 1.86.$$

Hence the 90% confidence interval for the expectation μ_X is of the form

$$\left(2.56 - 1.86 \cdot \frac{0.12}{\sqrt{9}}, 2.56 + 1.86 \cdot \frac{0.12}{\sqrt{9}}\right) = (2.56 - 0.07, 2.56 + 0.07)$$

or, in other words, $\mu_X = 2.56 \pm 0.07$ at the 90% confidence level.

The final question in this section is, how good is the sample variance estimator S_n^2 introduced in (3.6.2)? Here again, the answer is difficult for a general c.d.f. F_X. However, in the case of a Gaussian sample one can prove that the nonnegative random quantity

$$\chi^2 = \frac{1}{\sigma_X^2}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \quad (3.6.10)$$

has a p.d.f. of the form

$$f_{\chi^2}(x; n-1) = \frac{1}{2^{(n-1)/2}\Gamma\left(\frac{n-1}{2}\right)} x^{(n-3)/2} e^{-x/2}, \quad x \geq 0, \quad (3.6.11)$$

which traditionally is called the chi-square p.d.f. with $(n-1)$ degrees of freedom.[24] Thus the C confidence level interval for σ_X^2 is of the form

$$\left(\frac{(n-1)S_X^2}{Q_{\chi^2}\left(\frac{1+C}{2}; n-1\right)}, \frac{(n-1)S_X^2}{Q_{\chi^2}\left(\frac{1-C}{2}; n-1\right)}\right) \quad (3.6.12)$$

Selected quantiles $Q_{\chi^2}(\alpha; n)$ for the chi-square distributions are given in the tables at the end of this chapter.

[24] Compare this definition with the calculation of the p.d.f. of the square of the $N(0,1)$-distributed random quantity in Example 3.1.11.

Example 3.6.3 (a 99% confidence interval for Var(X)). Twenty-six independent measurements of a Gaussian random quantity X resulted in the estimate $S_{26}^2 = 1.37$ for the variance $\mathrm{Var}(X)$. With $C = 0.99$, the tables yield

$$Q_{\chi^2}\left(\frac{1+0.99}{2}; 25\right) = Q_{\chi^2}(0.995; 25) = 46.928$$

and

$$Q_{\chi^2}\left(\frac{1-0.99}{2}; 25\right) = Q_{\chi^2}(0.005; 25) = 10.520.$$

Thus the 99% confidence level interval for the variance σ_X^2 is

$$\left(\frac{25 \cdot 1.37}{46.928}, \frac{25 \cdot 1.37}{10.520}\right) = (0.723, 3.255).$$

The interval is relatively large because the confidence level demanded is very high. Note that it is not symmetric about the estimated value $S_{26}^2 = 1.37$.

3.7 Problems, exercises, and tables

Use *Mathematica*, MAPLE, or MATLAB as needed throughout this and other problem sections.

3.7.1. Plot the c.d.f.s of binomial quantities X with $p = 0.21$ and $n = 5, 13, 25$. Calculate the probabilities that X takes values between 1.3 and 3.7. Repeat the same exercise for $p = 0.5$ and $p = 0.9$.

3.7.2. Calculate the probability that a random quantity uniformly distributed over the interval $[0, 3]$ takes values between 1 and 3. Do the same calculation for the exponentially distributed random quantity with parameter $\mu = 1.5$, and the Gaussian random quantity with parameters $\mu = 1.5$, $\sigma^2 = 1$.

3.7.3. Prove that $\alpha\Gamma(\alpha) = \Gamma(\alpha + 1)$, and that $\Gamma(n) = (n-1)!$. Use the integration-by-parts formula. Verify analytically that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Use the idea employed in Example 3.1.6 to prove that the standard Gaussian density is normalized.

3.7.4. The p.d.f. of a random variable X is expressed by the quadratic function $f_X(x) = ax(1-x)$, for $0 < x < 1$, and is zero outside the unit interval. Find a from the normalization condition and then calculate $F_X(x)$, $\mathbf{E}X$, $\mathrm{Var}(X)$, $\mathrm{Std}(X)$, the nth central moment, and $\mathbf{P}(0.4 < X < 0.9)$. Graph $f_X(x)$ and $F_X(x)$.

3.7.5. Find the c.d.f and p.d.f. of the random quantity $Y = X^3$, where X is uniformly distributed on the interval $[1, 3]$.

3.7.6. Find the c.d.f and p.d.f. of the random quantity $Y = \tan X$, where X is uniformly distributed over the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$. Find a physical (geometric) interpretation of this result.

3.7.7. Verify that $\mathrm{Var}(X) = \mathbf{E}X^2 - (\mathbf{E}X)^2$; see formula (3.2.6).

3.7.8. Calculate the expectation and the variance of the binomial distribution from Example 3.1.2.

3.7.9. Calculate the expectation and the variance of the Poisson distribution from Example 3.1.3.

3.7.10. Calculate the expectation, the variance, and the nth moment of the exponential distribution from Example 3.1.5.

3.7.11. Calculate the nth central moment of the Gaussian distribution from Example 3.1.6.

3.7.12. Derive the formula for the binomial distribution from Example 3.1.2, relying on the observation that it is the distribution of the sum of n independent and identically distributed Bernoulli random qunatities. Show that if $p = \frac{\mu}{n}$ and $n \to \infty$, then the binomial probabilities converge to the Poisson probabilities.

3.7.13. A random quantity X has an even p.d.f. $f_X(x)$ of the triangular shape shown in Figure 3.7.1.

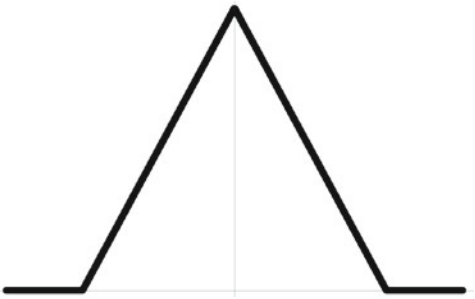

Fig. 3.7.1.

(a) How many parameters do you need to describe this p.d.f.? Find an explicit analytic formula for p.d.f. $f_X(x)$ and c.d.f. $F_X(x)$. Graph both of them.
(b) Find the expectation and variance of X.
(c) Let $Y = X^3$. Find the p.d.f. $f_Y(y)$ and graph it.

3.7.14. A discrete 2D random vector (X, Y) has the following joint p.d.f.:

$$\mathbf{P}(X = 1, Y = 1) = \frac{2}{12}, \qquad \mathbf{P}(X = 2, Y = 1) = \frac{1}{12},$$
$$\mathbf{P}(X = 3, Y = 1) = \frac{1}{12}, \qquad \mathbf{P}(X = 1, Y = 3) = \frac{2}{12},$$
$$\mathbf{P}(X = 2, Y = 3) = \frac{4}{12}, \qquad \mathbf{P}(X = 3, Y = 2) = \frac{2}{12}.$$

Find the marginal distributions of X and Y, their expectations and variances, as well as the covariance and the correlation coefficient of X and Y. Are X and Y independent?

3.7.15. Verify the Cauchy–Schwartz inequality (3.3.19). *Hint*: Take $Z = \frac{X-\mathbf{E}X}{\sigma(X)}$ and $W = \frac{Y-\mathbf{E}Y}{\sigma(Y)}$, and consider the discriminant of the expression $\mathbf{E}(Z + xW)^2$. The latter is quadratic in variable x and necessarily always nonnegative, so it can have at most one root.

3.7.16. The following sample of random vector (X, Y) was obtained: $(1, 1.7)$, $(2, 2)$, $(5, 4.3)$, $(7, 5.9)$, $(9, 8)$, $(9, 8.7)$. Produce the scatterplot of the sample and the corresponding least-squares regression line.

3.7.17. Using the table of $N(0, 1)$ c.d.f. provided at the end of this chapter calculate $\mathbf{P}(-1 \le Y \le 2)$ if $Y \sim N(0.7, 4)$.

3.7.18. Produce graphs of the Student-T p.d.f. $f_T(x, n)$, for $n = 2, 5, 12, 20$, and compare them with the standard normal p.d.f.

3.7.19. Produce graphs of the chi-square p.d.f. $f_{\chi^2}(x, n)$ for $n = 2, 5, 12, 20$.

3.7.20. Find a constant $c > 0$ such that the function

$$f_X(x) = \begin{cases} c(1+x)^{-3} & \text{for } x > 0; \\ 0 & \text{for } x \le 0. \end{cases}$$

is a valid p.d.f. Find $\mathbf{P}(\frac{1}{5} < X < 5)$, $\mathbf{E}(X)$, and the p.d.f. $f_Y(y)$ of $Y = X^{1/5}$. Show that $\mathrm{Var}(x) = \infty$.

3.7.22. Measurements of voltage V and current I on a resistor yielded the following $n = 5$ paired data: $(1.0, 2.3)$, $(2.0, 4.1)$, $(3.0, 6.4)$, $(4.0, 8.5)$, $(5.0, 10.5)$. Draw the scatterplot and find the regression line providing the least-squares fit for the data.

3.7.23. Independent measurements of the leakage current I on a capacitor yielded the following data: 2.71, 2.66, 2.78, 2.67, 2.71, 2.69, 2.70, 2.73 mA. Assuming that the distribution of the random quantity I is Gaussian, find the 95% confidence intervals for the expectation $\mathbf{E}I$ and the variance σ_X^2.

3.7.24. Complete the following sketch of the proof of the central limit theorem.

(a) Define $L_X(u)$ as the Laplace transform of c.d.f. $F_X(x)$:

$$L_X(u) = \int_{\infty}^{\infty} e^{ux} dF_X(x).$$

Find $L'_X(0)$ and $L''_X(0)$.

(b) Calculate $L_X(u)$ for the Gaussian $N(0, 1)$ random quantity.

(c) Prove that, for independent random quantities X and Y,

$$L_{X+Y}(u) = L_X(u) \cdot L_Y(u).$$

(d) Utilizing (c), calculate

$$L_{\sqrt{n}(\overline{X}-\mu_X)/\,\mathrm{Std}(X)}(u),$$

(it is easier to work here with the logarithm of the Laplace transform) and find its limit as $n \to \infty$. Compare it with the Laplace transform of the Gaussian $N(0,1)$ random quantity.

3.7.25. Use the introduced above Laplace transform technique to prove that the sum of two independent Gaussian random quantities is again a Gaussian random quantity.

3.7.26. What is the probability P that a randomly selected chord is shorter than the side S of an equilateral triangle inscribed in the circle? Here are two seemingly reasonable solutions:[25]

(a) A chord is determined by its two endpoints. Fix one of them to be A. For the chord to be shorter than the side S, the other endpoint must be chosen on either the arc AB or on the arc CA, and each of them is subtended by an engle of 120°. Thus $P = \frac{2}{3}$.

(b) A chord is completely determined by its center. For the chord to be shorter than the side S, the center must lie outside the circle of radius equal to the half of the radius of the original circle and the same center. Hence, the probability P equals the ratio of the annular area between two circles and the area of the original circle, which is $\frac{3}{4}$.

These two solutions are different. How is that possible?

3.7.27. Derive formulas for the c.d.f. $F_Y(y)$, and the p.d.f. $f_Y(y)$, of a transformation $Y = g(X)$ of a random quantity X, in terms of its c.d.f. $F_X(x)$, and p.d.f. $f_X(x)$, in the case when the transforming function $y = g(x)$ is *monotonically decreasing*. Follow the line of reasoning used to derive the analogous formulas (3.1.11)–(3.1.12) for *monotonically increasing* transformations. How would you extend these formulas to transformations that are monotonically increasing on some intervals and decreasing on their complement?

3.7.28. Verify that the components X, Y of the random vector with probability distribution $P((X,Y) = (1,0)) = P((X,Y) = (0,1)) = P((X,Y) = (-1,0)) = P((X,Y) = (0,-1)) = \frac{1}{4}$ are uncorrelated but not statistically independent. Calculate probability distribution of a random vector (W,Z) with statistically independent components and the same marginal distribution as (X,Y).

---

[25] For more information, see M. Denker and W. A. Woyczyński, *Introductory Statistics and Random Phenomena: Uncertainty, Complexity, and Chaotic Behavior in Engineering and Science*, Birkhäuser Boston, Cambridge, MA, 1998, Example 5.1.1.

Table 3.7.1. Gaussian $N(0,1)$ c.d.f. $\Phi(z) = (2\pi)^{-1/2} \int_{-\infty}^{z} e^{-x^2/2} dx$.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5395	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6296	.6331	.6366	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6884	.6879
0.5	.6915	.6956	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7857	.7389	.7422	.7454	.7486	7517	.7549
0.7	.7580	.7611	.7642	7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8075	.8106	.8133
0.9	.8195	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8503	.8531	.8554	.8577	.8599	.8621
1.1	.8613	.8665	.8686	.8708	.8729	.8749	.8770	.8796	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8977	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9359	.9370	.9382	.9309	.9404	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9606	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	9678.	9666.	9693.	9699.	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9773	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850.	9854	.9857
2.2	.9891	.9861	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.4	.9918	.9820	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9983	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.8876	.8876	.8877	.8877	.8878	.8878	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992.	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Table 3.7.2. Student-T distribution quantiles $Q_T(\alpha; n)$.

$n \backslash \alpha$	0.1000	0.0500	0.0250	0.0100	0.0050	0.0010	0.0005
1	3.078	6.314	12.706	31.821	63.657	318.317	636.61
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.451	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	8.610
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.500	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.813	2.228	2.764	3.169	4.144	4.587
11	1.364	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.141
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.584	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.553	2.879	3.610	3.992
19	1.328	1.729	2.093	2.540	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.849
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.320	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.059	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.312	1.701	2.049	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.311	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Table 3.7.3. Chi-square distribution quantiles $Q_{\chi^2}(\alpha; n)$.

$n \backslash \alpha$	0.9950	0.9900	0.9750	0.9500	0.9000	0.1000	0.0500	0.0250	0.0100	0.0050
1	0.000	0.000	0.001	0.004	0.016	2.706	3.843	5.025	6.637	7.882
2	0.010	0.020	0.051	0.103	0.211	4.605	5.992	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.344	12.937
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.160	9.236	11.070	12.832	15.085	16.748
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.440	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.17	14.067	16.012	18.474	20.276
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.534	20.090	21.954
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.022	21.665	23.587
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.724	26.755
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.041	19.812	22.362	24.735	27.687	29.817
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.600	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.577	32.799
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.407	7.564	8.682	10.085	24.769	27.587	30.190	33.408	35.716
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.843	7.632	8.906	10.117	11.651	27.203	30.143	32.852	36.190	38.580
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.033	8.897	10.283	11.591	13.240	29.615	32.670	35.479	38.930	41.399
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.195	11.688	13.090	14.848	32.007	35.172	38.075	41.637	44.179
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	10.519	11.523	13.120	14.611	16.473	34.381	37.652	40.646	44.313	46.925
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.807	12.878	14.573	16.151	18.114	36.741	40.113	43.194	46.962	49.642
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.120	14.256	16.147	17.708	19.768	39.087	42.557	45.772	49.586	52.333
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
31	14.457	15.655	17.538	19.280	21.433	41.422	44.985	48.231	52.190	55.000
32	15.134	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486	56.328
33	15.814	17.073	19.046	20.866	23.110	43.745	47.400	50.724	54.774	57.646
34	16.501	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061	58.964
35	17.191	18.508	20.569	22.465	24.796	46.059	49.802	53.203	57.340	60.272
36	17.887	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619	61.581
37	18.584	19.960	22.105	24.075	26.492	48.363	52.192	55.667	59.891	62.880
38	19.289	20.691	22.878	24.884	27.343	49.513	53.384	56.896	61.162	64.181
39	19.994	21.425	23.654	25.695	28.196	50.660	54.572	58.119	62.462	65.473
40	20.706	22.164	24.433	26.509	29.050	51.805	55.758	59.342	63.691	66.766