5. Some Important Distributions and Theorems

We shall now discuss in detail some specific distributions. This chapter could therefore be regarded as a collection of examples. These distributions, however, are of great practical importance and are often encountered in many applications. Moreover, their study will lead us to a number of important theorems.

5.1 The Binomial and Multinomial Distributions

Consider an experiment having only two possible outcomes. The sample space can therefore be expressed as

$$E = A + \bar{A} \tag{5.1.1}$$

with the probabilities

$$P(A) = p \quad , \qquad P(\bar{A}) = 1 - p = q \quad . \tag{5.1.2}$$

One now performs n independent trials of the experiment defined by (5.1.1). One wishes to find the probability distribution for the quantity $\mathsf{x} = \sum_{i=1}^{n} \mathsf{x}_i$, where one has $\mathsf{x}_i = 1$ (or 0) when A (or $\bar{A}$) occurs as the result of the ith experiment.

The probability that the first k trials result in A and all of the rest in $\bar{A}$ is, using Eq. (2.3.8),

$$p^k q^{n-k} \quad .$$

S. Brandt, *Data Analysis: Statistical and Computational Methods for Scientists and Engineers*,
DOI 10.1007/978-3-319-03762-2_5,

Using the rules of combinatorics, the event "outcome A k times in n trials" occurs in $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ different ways, according to the order of the occurrences of A and $\bar{A}$ (see Appendix B). The probability of this event is therefore

$$P(k) = W_k^n = \binom{n}{k} p^k q^{n-k} \quad . \tag{5.1.3}$$

We are interested in the mean value and variance of x. We first find these quantities for the variable x_i of an individual event. According to (3.3.2) one has

$$E(\mathsf{x}_i) = 1 \cdot p + 0 \cdot q \tag{5.1.4}$$

and

$$\begin{aligned} \sigma^2(\mathsf{x}_i) &= E\{(x_i - p)^2\} = (1-p)^2 p + (0-p)^2 q \quad , \\ \sigma^2(\mathsf{x}_i) &= pq \quad . \end{aligned} \tag{5.1.5}$$

From the generalization of (3.5.3) for $\mathsf{x} = \sum \mathsf{x}_i$ it follows that

$$E(\mathsf{x}) = \sum_{i=1}^{n} p = np \quad , \tag{5.1.6}$$

and from (3.5.10), since all of the covariances vanish because the x_i are independent, one has

$$\sigma^2(\mathsf{x}) = npq \quad . \tag{5.1.7}$$

Figure 5.1 shows the distribution W_k^n for various n and for fixed p, and Fig. 5.2 shows it for fixed n and various values of p. Finally in Fig. 5.3 n and p are both varied but the product np is held constant. The figures will help us to see relationships between the *binomial distribution* (5.1.3) and other distributions.

A logical extension of the binomial distribution deals with experiments where more than two different outcomes are possible. Equation (5.1.1) is then replaced by

$$E = A_1 + A_2 + \cdots + A_\ell \quad . \tag{5.1.8}$$

Let the probability for the outcome A_j be

$$P(A_j) = p_j \quad , \qquad \sum_{j=1}^{\ell} p_j = 1 \quad . \tag{5.1.9}$$

We consider again n trials and ask for the probability that the outcome A_j occurs k_j times. This is given by

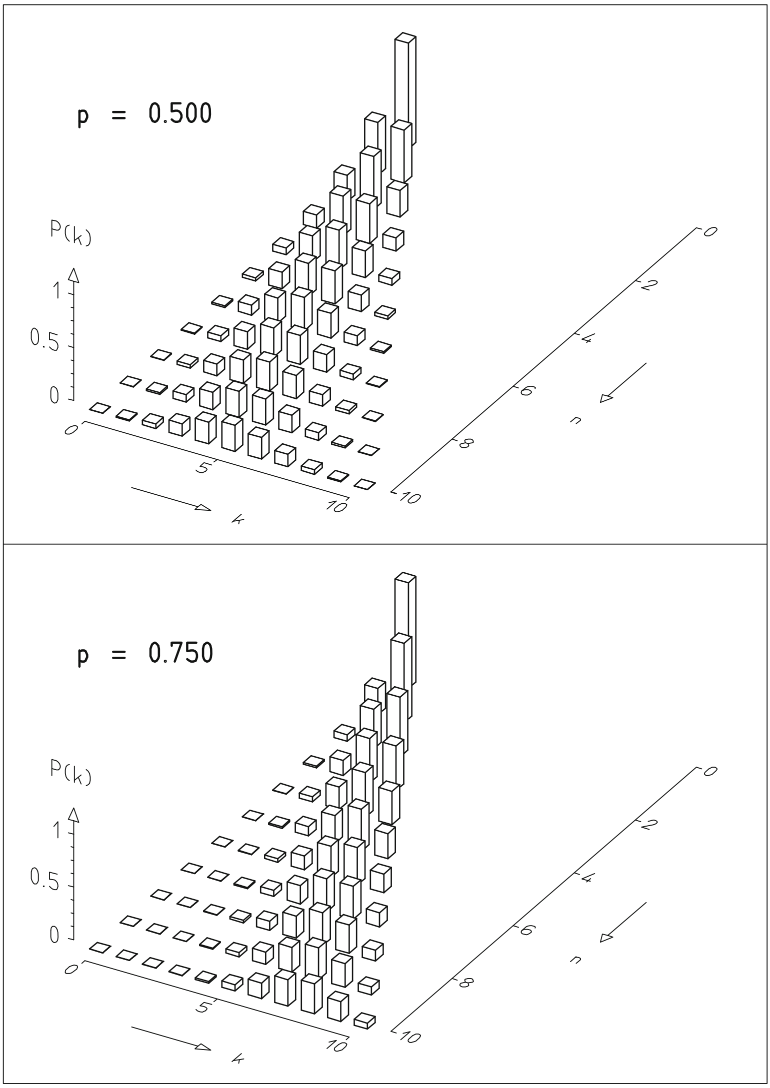

Fig. 5.1: Binomial distribution for various values of n, with p fixed.

$$W^n_{(k_1,k_2,\dots,k_\ell)} = \frac{n!}{\prod_{j=1}^{\ell} k_j!} \prod_{j=1}^{\ell} p_j^{kj} \quad , \qquad \sum_{j=1}^{\ell} k_j = n \quad . \tag{5.1.10}$$

The proof is left to the reader. The probability distribution (5.1.10) is called the *multinomial distribution*.

We can define a random variable x_{ij} that takes on the value 1 when the ith trial leads to the outcome A_j, and is zero otherwise. In addition define $\mathsf{x}_j = \sum_{i=1}^{n} \mathsf{x}_{ij}$. The expectation value of x_j is then

$$E(\mathsf{x}_j) = \widehat{x}_j = np_j \quad . \tag{5.1.11}$$

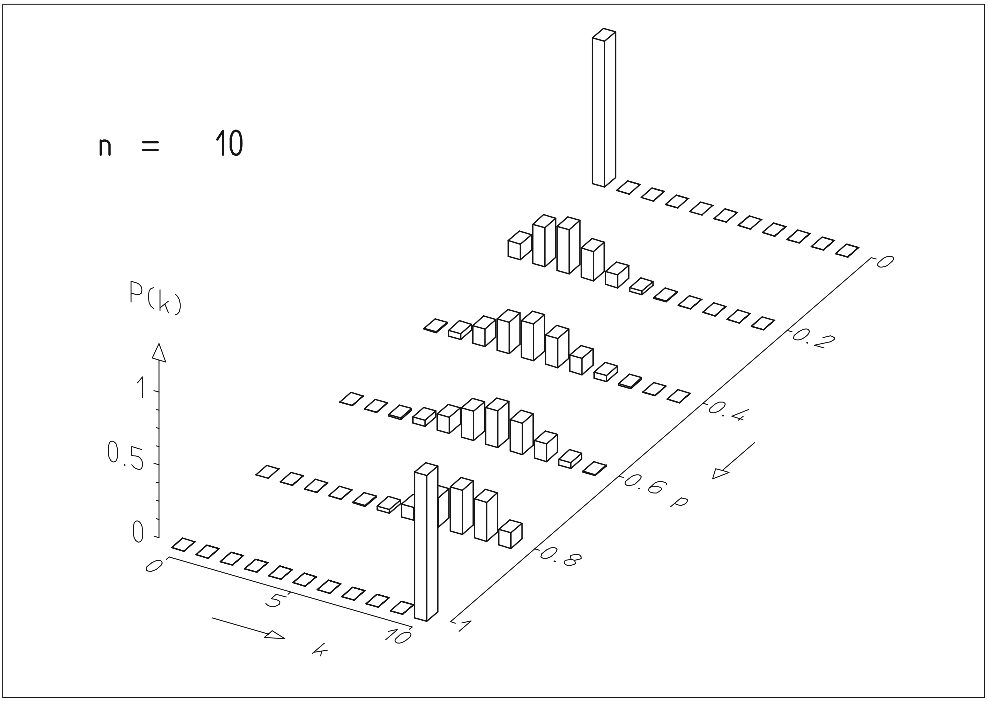

Fig. 5.2: Binomial distribution for various values of p, with n fixed.

The elements of the covariance matrix of the x_j are

$$c_{ij} = np_i(\delta_{ij} - p_j) \quad . \tag{5.1.12}$$

The off-diagonal elements are clearly not zero. This was to be expected, since from Eq. (5.1.9) the variables x_j are not independent.

5.2 Frequency: The Law of Large Numbers

Usually the probabilities for the different types of events, e.g., p_j in the case of the multinomial distribution, are not known but have to be obtained from experiment. One first measures the *frequency* of the events in n experiments,

$$\mathsf{h}_j = \frac{1}{n}\sum_{i=1}^{n} \mathsf{x}_{ij} = \frac{1}{n}\mathsf{x}_j \quad . \tag{5.2.1}$$

Unlike the probability, the frequency is a random quantity, since it depends on the outcomes of the n individual experiments. By use of (5.1.11), (5.1.12), and (3.3.15) we obtain

$$E(\mathsf{h}_j) = \widehat{h}_j = E\left(\frac{\mathsf{x}_j}{n}\right) = p_j \tag{5.2.2}$$

and

$$\sigma^2(\mathsf{h}_j) = \sigma^2\left(\frac{\mathsf{x}_j}{n}\right) = \frac{1}{n^2}\sigma^2(\mathsf{x}_j) = \frac{1}{n}p_j(1-p_j) \quad . \tag{5.2.3}$$

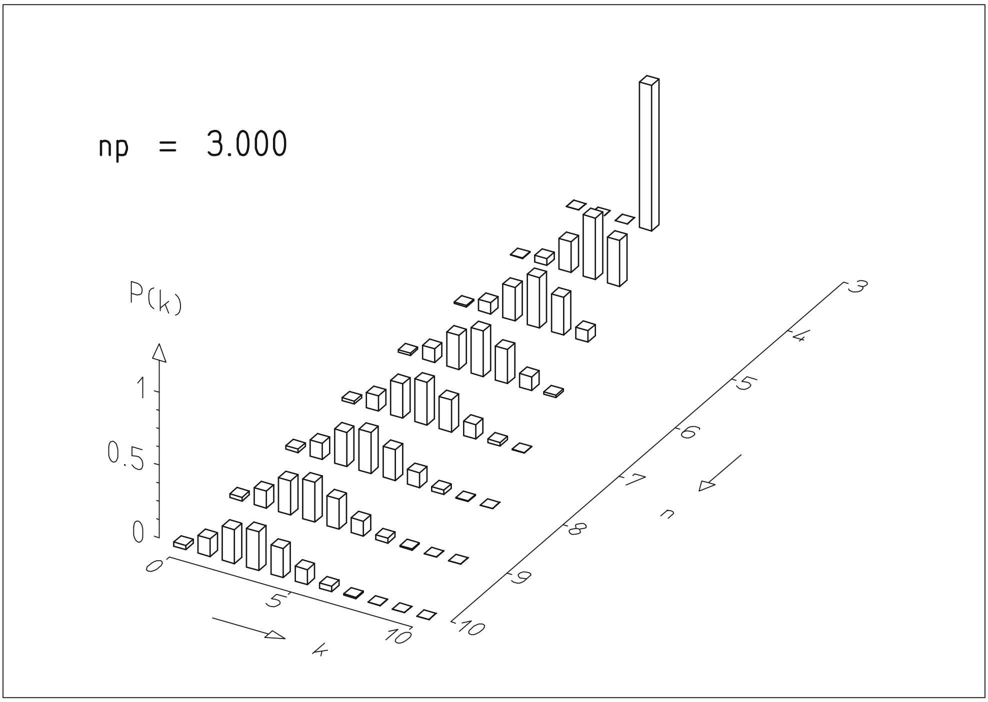

Fig. 5.3: Binomial distribution for various values of n but fixed product np. For higher values of n the distribution changes very little.

The product $p_j(1-p_j)$ in Eq. (5.2.3) is at most $1/4$. One sees that the expectation value of the frequency of an event is exactly equal to the probability that the event will occur, and that the variance of frequency about this expectation value can be made arbitrarily small as the number of trials increases. Since pq is at most $1/4$, one can always say that the standard deviation of h_j is at most $1/\sqrt{n}$. This property of the frequency is known as the *law of large numbers*. It is clearly the reason for the frequency definition of probability given by Eq. (2.2.1).

Frequently the purpose of an experimental investigation is to determine the probability for the occurrence of a certain type of event. According to (5.2.2) we can use the frequency as an approximation of the probability. The square of the error of this approximation is then inversely proportional to the number of individual experiments. This kind of error, which originates from the fact that only a finite number of experiments can be performed, is called the *statistical error*. It is of prime importance for applications that are concerned with the counting of individual events, e.g., nuclear particles passing through a counter, animals with certain traits in heredity experiments, defective items in quality control, and so forth.

Example 5.1: Statistical error

Suppose it is known from earlier experiments that a fraction $R \approx 1/200$ of a sample of fruit flies *(Drosophila)* develop a certain property A if exposed to a given dose of X-rays. An experiment is planned to determine the fraction R with an accuracy of 1%. How large must the original sample be in order to achieve this accuracy?

We use Eq. (5.2.3) and find $p_j = 0.005, (1 - p_j) \approx 1$. We must now choose n such that $\sigma(h_j)/h_j = 200\sigma(h_j) = 0.01$. This gives $\sigma(h_j) = 0.000\,05$ and $\sigma^2(h_j) = 0.25 \times 10^{-8}$. Equation (5.2.3) gives

$$0.25 \times 10^{-8} = \frac{1}{n} \times 0.005$$

and therefore

$$n = 2 \times 10^6 \quad .$$

A total of two million fruit flies would have to be used. This is practically impossible. To determine the fraction R with an accuracy of 10% would require 20 000 flies. ■

5.3 The Hypergeometric Distribution

Although we shall rigorously introduce the concept of random sampling at a later point, we will now discuss a typical problem of sampling. We consider a container – we shall not break with the habit of mathematicians of calling such a container an urn – with K white and $L = N - K$ black balls. We want to determine the probability that in drawing n balls (without replacing them) we will find exactly k white and $l = n - k$ black ones. The problem is rendered difficult by the fact that the drawing of a ball of a particular color changes the ratio of white and black balls and therefore influences the outcome of the next draw. One clearly has $\binom{N}{n}$ equally likely ways to choose n out of N balls. The probability that one of these possibilities will occur is therefore $1/\binom{N}{n}$. There are $\binom{K}{k}$ ways to choose k of the K white balls, and $\binom{L}{\ell}$ ways to choose ℓ of the L black ones. The required probability is therefore

$$W_k = \frac{\binom{K}{k}\binom{L}{\ell}}{\binom{N}{n}} \quad . \tag{5.3.1}$$

As in Sect. 5.1 we define the random variable $\mathsf{x} = \sum_{i=1}^{n} \mathsf{x}_i$ with $\mathsf{x}_i = 1$ when the ith draw results in a black ball, and $\mathsf{x}_i = 0$ otherwise. (In other words, we define k as the random variable x.)

To compute the expectation values of x we cannot simply add the expectation values of the x_i, since these are no longer independent. Instead we must return to the definition (3.3.2),

$$\begin{aligned} E(\mathsf{x}) &= \frac{1}{\binom{N}{n}} \sum_{i=1}^{n} i \binom{K}{i} \binom{N-K}{n-i} \\ &= \frac{(N-n)!n!}{N!} \sum_{i=1}^{n} \frac{i\,K!(N-K)!}{i!(K-i)!(n-i)!(N-K-n+i)!} \\ &= \frac{n(n-1)!(N-n)!}{N(N-1)!} \sum_{i=1}^{n} \frac{K!}{(i-1)!(K-1-(i-1))!} \\ &\quad \times \frac{(N-K)!}{(n-1-(i-1))!(N-K-(n-1)+(i-1))!} \quad . \end{aligned}$$

If we substitute $i-1=j$, this gives

$$\begin{aligned} E(\mathsf{x}) &= n\frac{K}{N}\frac{(n-1)!(N-n)!}{(N-1)!} \\ &\quad \times \sum_{j=0}^{n-1} \frac{(K-1)!(N-K)!}{j!(K-1-j)!(n-1-j)!(N-K-(n-1)+j)!} \\ &= n\frac{K}{N}\frac{1}{\binom{N-1}{n-1}} \sum_{j=0}^{n-1} \binom{K-1}{j}\binom{N-K}{n-1-j} \quad . \end{aligned}$$

With Eq. (B.5) we obtain

$$E(\mathsf{x}) = n\frac{K}{N} \quad . \tag{5.3.2}$$

The calculation of the variance follows along the same lines but is rather lengthy. The result is

$$\sigma^2(\mathsf{x}) = \frac{n\,K(N-K)(N-n)}{N^2(N-1)} \quad . \tag{5.3.3}$$

Figures 5.4 and 5.5 depict several examples of the distribution. If $n \ll N$, then drawing a white ball has little influence on the probabilities for the next draw. We therefore expect that in this case W_k behaves in a manner similar to a binomial distribution with $p = \frac{K}{N}$ and $q = \frac{N-K}{N}$. This is also made clear by the similarity of Figs. 5.5 and 5.1. One obtains in fact the same expectation value,

$$E(\mathsf{x}) = n\frac{K}{N} = np \quad ,$$

as for the binomial distribution. The variance is then

$$\sigma^2(\mathrm{x}) = \frac{npq(N-n)}{N-1} \quad ,$$

which for the case $n \ll N$ becomes

$$\sigma^2 = npq \quad .$$

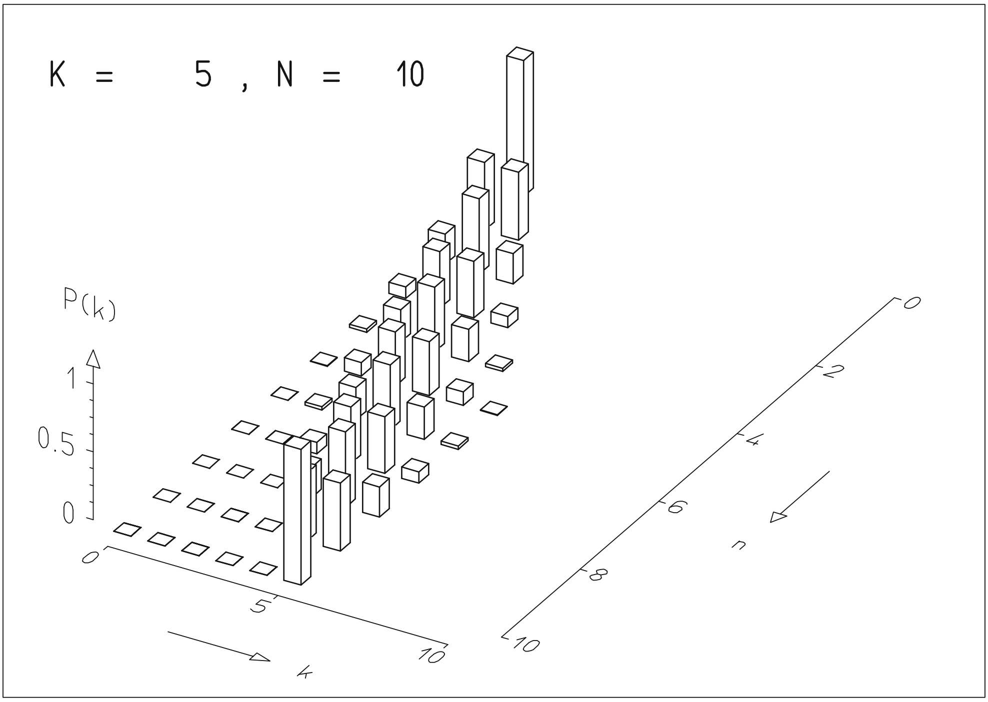

Fig. 5.4: Hypergeometric distribution for various values of n and small values of K and N.

There are many applications of the hypergeometric distribution. Opinion polls, quality controls, and so forth are all based on the experimental scheme of taking (polling) an object without replacement back into the original sample or *population*. The distribution can be generalized in two ways. First we can of course consider more properties instead of just two (white and black balls). This leads us to a similar transition as the one from the binomial to the multinomial distribution. The original sample (population) contains N elements each of which possesses one of l properties,

$$N = N_1 + N_2 + \cdots + N_\ell \quad .$$

The probability that n draws (without replacement) will be composed as

$$n = n_1 + n_2 + \cdots + n_\ell$$

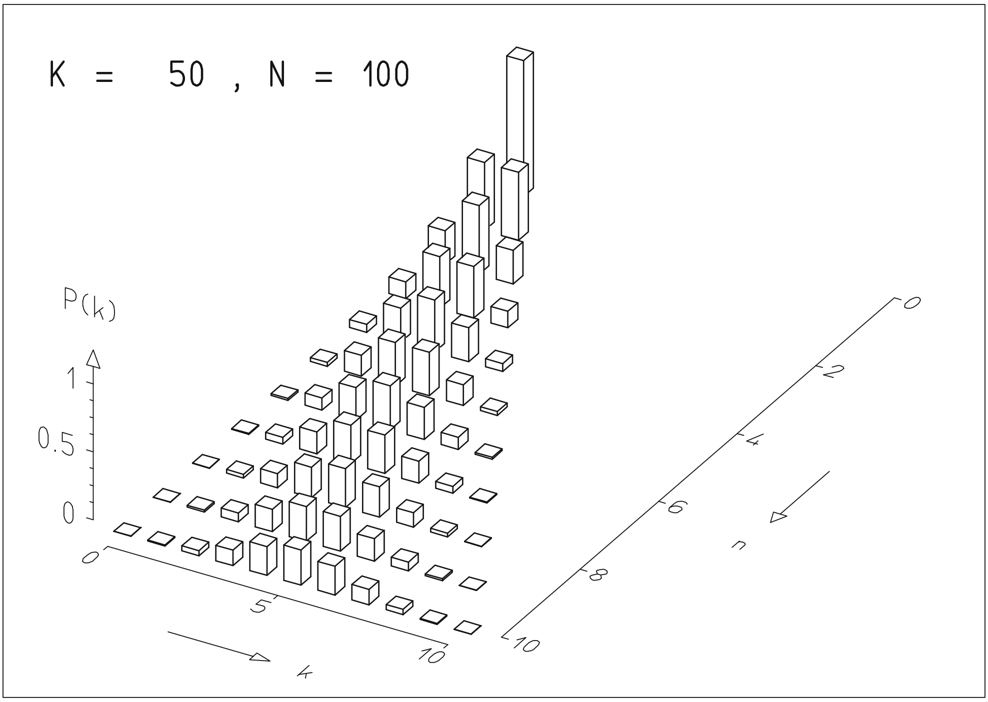

Fig. 5.5: Hypergeometric distribution for various values of n and for large values of K and N.

is, in analogy to Eq. (5.3.1),

$$W_{n_1,n_2,\ldots,n_\ell} = \frac{\binom{N_1}{n_1}\binom{N_2}{n_2}\cdots\binom{N_\ell}{n_\ell}}{\binom{N}{n}} \quad . \tag{5.3.4}$$

Another extension of the hypergeometric distribution is obtained in the following way. We saw earlier that consecutive drawings ceased to be independent because the balls were not replaced. If now each time we draw a ball of one type we place more balls of that type back in the urn, this dependence can be enhanced. One then obtains the *Polya distribution*. It is of importance in the study of epidemic diseases, where the appearance of a case of the disease enhances the probability of future cases.

Example 5.2: Application of the hypergeometric distribution for determination of zoological populations

From a pond K fish are taken and marked. They are then returned to the pond. After a short while n fish are caught, k of which are found to be marked. Before the second time that the fish are taken, the pond contains a total of N fish, of which K are marked. The probability of finding k marked out of n removed fish is given by Eq. (5.3.1). We will return to this problem in Example 7.3. ∎

5.4 The Poisson Distribution

Looking at Fig. 5.3 it appears that if n tends to infinity, but at the same time $np = \lambda$ is kept constant, the binomial distribution approaches a certain fixed distribution. We rewrite Eq. (5.1.3) as

$$
\begin{aligned}
W_k^n &= \binom{n}{k} p^k q^{n-k} \\
&= \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k} \\
&= \frac{\lambda^k}{k!}\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}\frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k} \\
&= \frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)}{\left(1-\frac{\lambda}{n}\right)^k} \quad .
\end{aligned}
$$

In the limiting case all of the many individual factors of the term on the right approach unity. In addition one has

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = \mathrm{e}^{-\lambda} \quad ,$$

so that in the limit one has

$$\lim_{n\to\infty} W_k^n = f(k) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda} \quad . \tag{5.4.1}$$

The quantity $f(k)$ is the probability of the *Poisson distribution*. It is plotted in Fig. 5.6 for various values of λ. As is the case for the other distributions we have encountered so far, the Poisson distribution is only defined for integer values of k.

The distribution satisfies the requirement that the total probability is equal to unity,

$$
\begin{aligned}
\sum_{k=0}^{\infty} f(k) &= \sum_{k=0}^{\infty} \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!} \\
&= \mathrm{e}^{-\lambda}\left(1+\lambda+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\cdots\right) \\
&= \mathrm{e}^{-\lambda}\mathrm{e}^{\lambda} \quad ,
\end{aligned}
$$

$$\sum_{k=0}^{\infty} f(k) = 1 \quad . \tag{5.4.2}$$

The expression in parentheses is in fact the Taylor expansion of e^{λ}.

We now want to determine the mean, variance, and skewness of the Poisson distribution. The definition (3.3.2) gives

$$\begin{aligned} E(\mathsf{k}) &= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \sum_{k=1}^{\infty} \frac{\lambda \lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda} \end{aligned}$$

and using this with (5.4.2),

$$E(\mathsf{k}) = \lambda \quad . \tag{5.4.3}$$

We would now like to find $E(\mathsf{k}^2)$. One obtains in a corresponding way

$$\begin{aligned} E(\mathsf{k}^2) &= \sum_{k=1}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \\ &= \lambda \sum_{j=0}^{\infty} (j+1) \frac{\lambda^j}{j!} e^{-\lambda} = \lambda \left(\sum_{j=0}^{\infty} j \frac{\lambda^j}{j!} e^{-\lambda} + 1 \right) \end{aligned}$$

and therefore

$$E(\mathsf{k}^2) = \lambda(\lambda+1) \quad . \tag{5.4.4}$$

We will use Eqs. (5.4.3) and (5.4.4) to compute the variance. According to Eq. (3.3.16) one has

$$\sigma^2(\mathsf{k}) = E(\mathsf{k}^2) - \{E(\mathsf{k})\}^2 = \lambda(\lambda+1) - \lambda^2 \tag{5.4.5}$$

or

$$\sigma^2(\mathsf{k}) = \lambda \quad . \tag{5.4.6}$$

We now consider the skewness (3.3.13) of the Poisson distribution. Following Sect. 3.3 we easily find that

$$\mu_3 = E\{(\mathsf{k} - \widehat{k})^3\} = \lambda \quad .$$

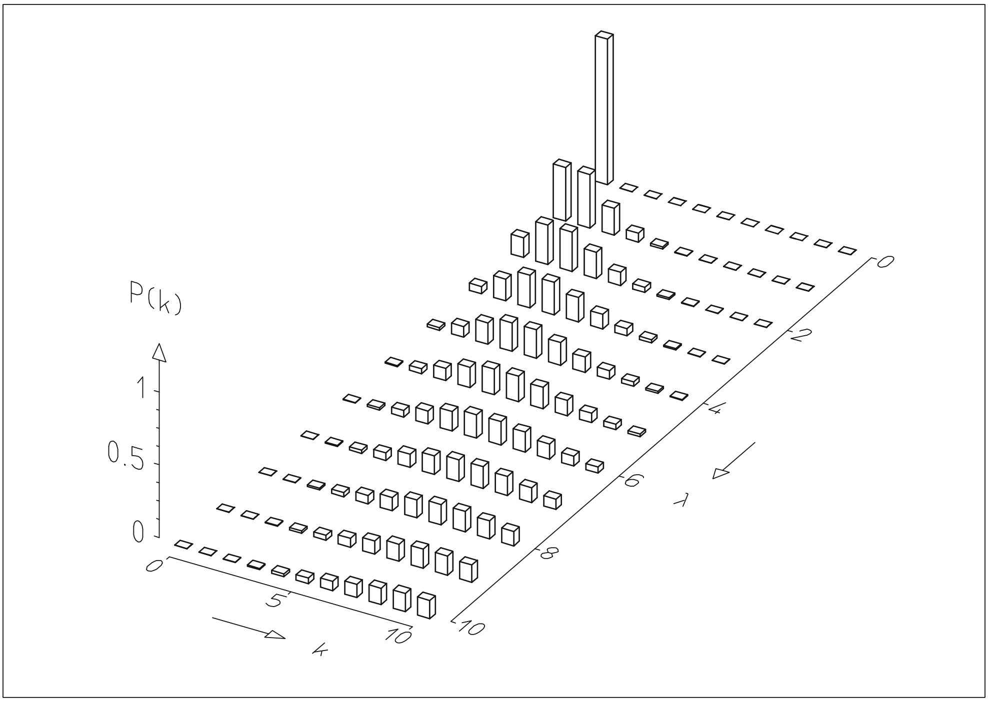

Fig. 5.6: Poisson distribution for various values of λ.

The skewness (3.3.13) is then

$$\gamma = \frac{\mu_3}{\sigma^3} = \frac{\lambda}{\lambda^{\frac{3}{2}}} = \lambda^{-\frac{1}{2}} \quad , \tag{5.4.7}$$

that is, the Poisson distribution becomes increasingly symmetric as λ increases. Figure 5.6 shows the distribution for various values of λ. In particular the distribution with $\lambda = 3$ should be compared with Fig. 5.3.

We have obtained the Poisson distribution from the binomial distribution with large n but constant $\lambda = np$, i.e., small p. We therefore expect it to apply to processes in which a large number of events occur but of which only very few have a certain property of interest to us (i.e., a large number of "trials" but few "successes").

Example 5.3: Poisson distribution and independence of radioactive decays

We consider a radioactive nucleus with mean lifetime τ and observe it for a time $T \ll \tau$. The probability that it decays within this time interval is $W \ll 1$. We break the observation time T into n smaller time intervals of length t, so that $T = nt$. The probability for the nucleus to decay in a particular time interval is $p \approx W/n$. We now observe a radioactive source containing N nuclei which decay independently from each other for a total time T, and detect a_1

decays in time interval 1, a_2 decays in interval 2, etc. Let $h(k)$ be the frequency of decays observed in the interval k, with $(k = 0, 1, \ldots)$. That is, if n_k is the number of intervals with k decays, then $h(k) = n_k/n$. In the limit $N \to \infty$ and for large n the frequency distribution $h(k)$ becomes the probability distribution (5.4.1). The statistical nature of radioactive decay was established in this way in a famous experiment by RUTHERFORD and GEIGER. ■

Similarly, the frequency of finding k stars per element of the celestial sphere or k raisins per volume element of a fruit cake is distributed according to the Poisson law, but not, however, the frequency of finding k animals of a given species per element of area, at least if these animals live in herds, since in this case the assumption of independence is not fulfilled.

As a quantitative example of the Poisson distribution many textbooks discuss the number of Prussian cavalrymen killed during a period of 20 years by horse kicks, an example originally due to VON BORTKIEWICZ [6]. We prefer to turn our attention to a somewhat less macabre example taken from a lecture of DE SOLLA PRICE [7].

Example 5.4: Poisson distribution and the independence of scientific discoveries

The author first constructs the model of an apple tree with 1000 apples and 1000 pickers with blindfolded eyes who each try at the same time to pick an apple. Since we are dealing with a model, they do not hinder each other but it can happen that two or several of them will attempt to pick the same apple at the same time. The number of apples grabbed simultaneously by k people $(k = 0, 1, 2, \ldots)$ follows a Poisson distribution. It was determined by DE SOLLA PRICE that the number of scientific discoveries made independently twice, three times, etc. is also distributed according to the Poisson law, in a way similar to the principle of the blindfolded apple pickers (Table 5.1). One gets the impression that scientists are not concerned with the activities of their colleagues. DE SOLLA PRICE believes that this can be explained by the assumption that scientists have a strong urge write papers, but feel only a relatively mild need to read them. ■

5.5 The Characteristic Function of a Distribution

So far we have only considered real random variables. In fact, in Sect. 3.1 we have introduced the concept of a random quantity as a real number associated with an event. Without changing this concept we can formally construct a *complex random variable* from two real ones by writing

$$\mathsf{z} = \mathsf{x} + i\mathsf{y} \quad . \tag{5.5.1}$$

Table 5.1: Simultaneous discovery and the Poisson distribution.

Number of simultaneous discoveries	Cases of simultaneous discovery	Prediction of Poisson distribution
0	Not defined	368
1	Not known	368
2	179	184
3	51	61
4	17	15
5	6	3
≥6	8	1

As its expectation value we define

$$E(\mathsf{z}) = E(\mathsf{x}) + \mathrm{i}\,E(\mathsf{y}) \quad . \tag{5.5.2}$$

By analogy with real variables, complex random variables are independent if the real and imaginary parts are independent among themselves.

If x is a real random variable with distribution function $F(x) = P(\mathsf{x} < x)$ and probability density $f(x)$, we define its characteristic function to be the expectation value of the quantity $\exp(\mathrm{i}t\mathsf{x})$:

$$\varphi(t) = E\{\exp(\mathrm{i}t\mathsf{x})\} \quad . \tag{5.5.3}$$

That is, in the case of a continuous variable the characteristic function is a Fourier integral with its known transformation properties:

$$\varphi(t) = \int_{-\infty}^{\infty} \exp(\mathrm{i}tx) f(x)\,\mathrm{d}x \quad . \tag{5.5.4}$$

For a discrete variable we obtain instead from (3.3.2)

$$\varphi(t) = \sum_i \exp(\mathrm{i}tx_i) P(\mathsf{x} = x_i) \quad . \tag{5.5.5}$$

We now consider the moments of x about the origin,

$$\lambda_n = E(\mathsf{x}^n) = \int_{-\infty}^{\infty} x^n f(x)\,\mathrm{d}x \quad , \tag{5.5.6}$$

and find that λ_n can be obtained simply by differentiating the characteristic function n times at the point $t = 0$:

$$\varphi^{(n)}(t) = \frac{\mathrm{d}^n \varphi(t)}{\mathrm{d}t^n} = \mathrm{i}^n \int_{-\infty}^{\infty} x^n \exp(\mathrm{i}tx) f(x)\,\mathrm{d}x$$

and therefore

$$\varphi^{(n)}(0) = \mathrm{i}^n \lambda_n \quad . \tag{5.5.7}$$

If we now introduce the simple coordinate translation

$$\mathsf{y} = \mathsf{x} - \widehat{x} \tag{5.5.8}$$

and construct the characteristic function

$$\varphi_y(t) = \int_{-\infty}^{\infty} \exp\{\mathrm{i}\,t(x-\widehat{x})\} f(x)\,\mathrm{d}x = \varphi(t)\exp(-\mathrm{i}\,t\widehat{x}) \quad , \tag{5.5.9}$$

then its nth derivative is (up to a power of i) equal to the nth moment of x about the expectation value [cf. (3.3.8)]:

$$\varphi_y^{(n)}(0) = \mathrm{i}^n \mu_n = \mathrm{i}^n E\{(\mathsf{x}-\widehat{x})^n\} \quad , \tag{5.5.10}$$

and in particular

$$\sigma^2(x) = -\varphi_y''(0) \quad . \tag{5.5.11}$$

Inverting the Fourier transform (5.5.4) we see that it is possible to obtain the probability density from the characteristic function,

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-\mathrm{i}tx)\varphi(t)\,\mathrm{d}t \quad . \tag{5.5.12}$$

It is possible to show that a distribution is determined *uniquely* by its characteristic function. This is the case even for discrete variables where one has

$$F(b) - F(a) = \frac{\mathrm{i}}{2\pi} \int_{-\infty}^{\infty} \frac{\exp(\mathrm{i}tb) - \exp(\mathrm{i}ta)}{t} \varphi(t)\,\mathrm{d}t \quad , \tag{5.5.13}$$

since in this case the probability density is not defined. Often it is more convenient to use the characteristic function rather than the original distribution. Because of the unique relation between the two it is possible to switch back and forth at any place in the course of a calculation.

We now consider the sum of two independent random variables

$$\mathsf{w} = \mathsf{x} + \mathsf{y} \quad .$$

Its characteristic function is

$$\varphi_w(t) = E[\exp\{\mathrm{i}t(\mathsf{x}+\mathsf{y})\}] = E\{\exp(\mathrm{i}t\mathsf{x})\exp(\mathrm{i}t\mathsf{y})\} \quad .$$

Generalizing relation (3.5.13) to complex variables we obtain

$$\varphi_w(t) = E\{\exp(\mathrm{i}tx)\}E\{\exp(\mathrm{i}ty)\} = \varphi_x(t)\varphi_y(t) \quad , \tag{5.5.14}$$

i.e., the characteristic function of a sum of independent random variables is equal to the product of their respective characteristic functions.

Example 5.5: Addition of two Poisson distributed variables with use of the characteristic function

From Eqs. (5.5.5) and (5.4.1) one obtains for the characteristic function of the Poisson distribution

$$\begin{aligned} \varphi(t) &= \sum_{k=0}^{\infty} \exp(\mathrm{i}tk)\frac{\lambda^k}{k!}\exp(-\lambda) = \exp(-\lambda)\sum_{k=0}^{\infty}\frac{(\lambda\exp(\mathrm{i}t))^k}{k!} \\ &= \exp(-\lambda)\exp(\lambda\,\mathrm{e}^{\mathrm{i}t}) = \exp\{\lambda(\mathrm{e}^{\mathrm{i}t}-1)\} \quad . \end{aligned} \tag{5.5.15}$$

We now form the characteristic function of the sum of two independent Poisson distributed variables with mean values λ_1 and λ_2,

$$\begin{aligned} \varphi_{\mathrm{sum}}(t) &= \exp\{\lambda_1(\mathrm{e}^{\mathrm{i}t}-1)\}\exp\{\lambda_2(\mathrm{e}^{\mathrm{i}t}-1)\} \\ &= \exp\{(\lambda_1+\lambda_2)(\mathrm{e}^{\mathrm{i}t}-1)\} \quad . \end{aligned} \tag{5.5.16}$$

This is again of the form of Eq. (5.5.15). Therefore the distribution of the sum of two independent Poisson distributed variables is itself a Poisson variable. Its mean is the sum of the means of the individual distributions. ■

5.6 The Standard Normal Distribution

The probability density of the *standard normal distribution* is defined as

$$f(x) = \phi_0(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2} \quad . \tag{5.6.1}$$

This function is depicted in Fig. 5.7a. It has a bell shape with the maximum at $x = 0$. From Appendix D.1 we have

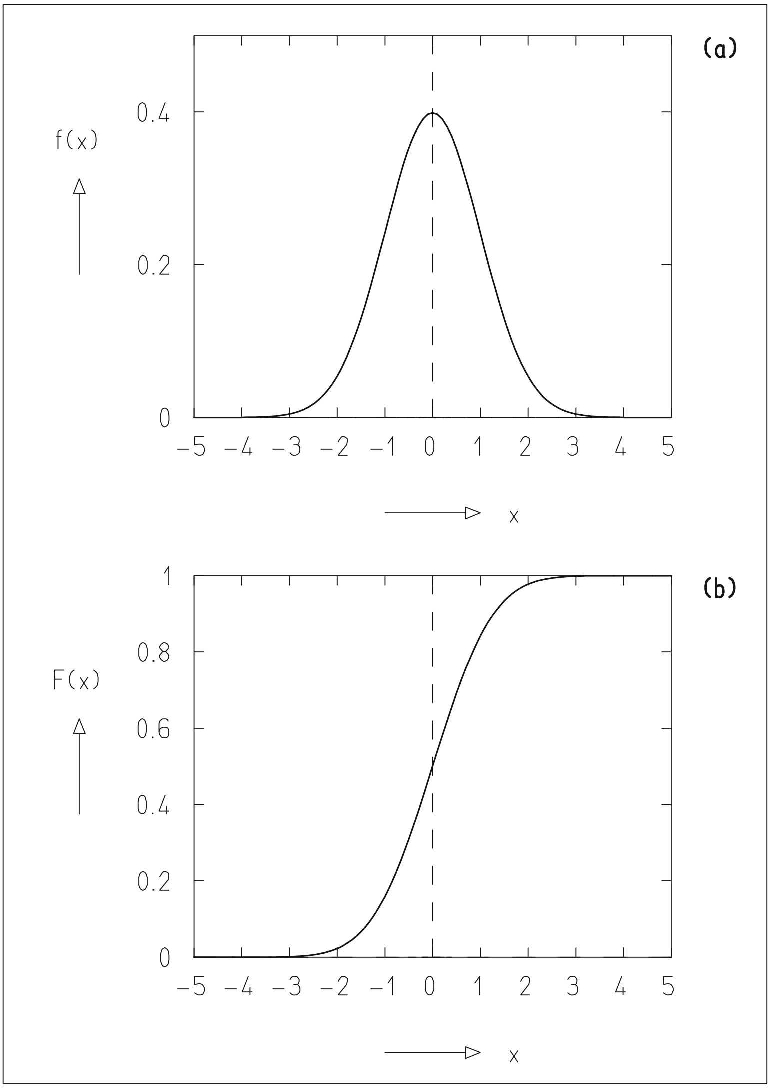

Fig. 5.7: Probability density (**a**) and distribution function (**b**) of the standard normal distribution.

$$\int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2}\,\mathrm{d}x = \sqrt{2\pi} \quad , \tag{5.6.2}$$

so that $\phi_0(x)$ is normalized to one as required. Using the symmetry of Fig. 5.7a, or alternatively, using the antisymmetry of the integrand we conclude that the expectation value is

$$\widehat{x} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x\,\mathrm{e}^{-x^2/2}\,\mathrm{d}x = 0 \quad . \tag{5.6.3}$$

By integrating by parts we can compute the variance to be

$$\begin{aligned}\sigma^2 &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2\,\mathrm{e}^{-x^2/2}\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}}\left\{\left[-x\,\mathrm{e}^{-x^2/2}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty}\mathrm{e}^{-x^2/2}\,\mathrm{d}x\right\} = 1 \quad , \qquad (5.6.4)\end{aligned}$$

since the expression in the square brackets vanishes at the integral's boundaries and the integral in curly brackets is given by Eq. (5.6.2).

The distribution function of the standard normal distribution

$$F(x) = \psi_0(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-t^2/2}\,\mathrm{d}t \qquad (5.6.5)$$

is shown in Fig. 5.7b. It cannot be expressed in analytic form. It is tabulated numerically in Appendix C.4.

5.7 The Normal or Gaussian Distribution

The standardized distribution of the last section had the properties $\widehat{x} = E(\mathsf{x}) = 0$, $\sigma^2(\mathsf{x}) = 1$, i.e., the variable x had the properties of the standardized variable u in Eq. (3.3.17). If we now replace x by $(\mathsf{x} - a)/b$ in (5.6.1), we obtain the probability density of the *normal* or *Gaussian distribution*,

$$f(x) = \phi(x) = \frac{1}{\sqrt{2\pi}\,b}\exp\left\{-\frac{(x-a)^2}{2b^2}\right\} \quad , \qquad (5.7.1)$$

with

$$\widehat{x} = a \quad , \qquad \sigma^2(\mathsf{x}) = b^2 \quad . \qquad (5.7.2)$$

The characteristic function of the normal distribution (5.7.1) is, using Eq. (5.5.4),

$$\varphi(t) = \frac{1}{\sqrt{2\pi}\,b}\int_{-\infty}^{\infty}\exp(\mathrm{i}tx)\exp\left(-\frac{(x-a)^2}{2b^2}\right)\mathrm{d}x \quad . \qquad (5.7.3)$$

With $u = (x-a)/b$ one obtains

$$\begin{aligned}\varphi(t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\{-\frac{1}{2}u^2 + \mathrm{i}t(bu+a)\}\,\mathrm{d}u \\ &= \frac{1}{\sqrt{2\pi}}\exp(\mathrm{i}ta)\int_{-\infty}^{\infty}\exp\{-\frac{1}{2}u^2 + \mathrm{i}tbu\}\,\mathrm{d}u \quad . \qquad (5.7.4)\end{aligned}$$

By completing the square the integral can be rewritten as

$$\begin{aligned}\int_{-\infty}^{\infty} \exp\{-\frac{1}{2}u^2 + \mathrm{i}tbu\}\,\mathrm{d}u \\ &= \int_{-\infty}^{\infty} \exp\{-\frac{1}{2}(u-\mathrm{i}tb)^2 - \frac{1}{2}t^2b^2\}\,\mathrm{d}u \\ &= \exp\{-\frac{1}{2}t^2b^2\}\int_{-\infty}^{\infty} \exp\{-\frac{1}{2}(u-\mathrm{i}tb)^2\}\,\mathrm{d}u \quad . \qquad (5.7.5)\end{aligned}$$

With $r = u - \mathrm{i}tb$ the last integral takes on the form

$$\int_{-\infty-\mathrm{i}tb}^{\infty-\mathrm{i}tb} \exp\{-\frac{1}{2}r^2\}\,\mathrm{d}r \quad .$$

The integrand does not have any singularities in the complex r plane. According to the residue theorem, therefore, the contour integral around any closed path vanishes. Consider a path that runs along the real axis from $r = -L$ to $r = L$, and then parallel to the imaginary axis from $r = L$ to $r = L - \mathrm{i}tb$ and from there antiparallel to the real axis to $r = -L - \mathrm{i}tb$, and finally back to the starting point $r = L$. In the limit $L \to \infty$ the integrand vanishes along the parts of the path that run parallel to the imaginary axis. One then has

$$\int_{-\infty-\mathrm{i}tb}^{\infty-\mathrm{i}tb} \exp\{-\frac{1}{2}r^2\}\,\mathrm{d}r = \int_{-\infty}^{\infty} \exp\{-\frac{1}{2}r^2\}\,\mathrm{d}r \quad ,$$

i.e., we can extend the integral to cover the entire real axis. The integral is computed in Appendix D.1 and has the value

$$\int_{-\infty}^{\infty} \exp\{-\frac{1}{2}r^2\}\,\mathrm{d}r = \sqrt{2\pi} \quad . \qquad (5.7.6)$$

Substituting this into Eqs. (5.7.5) and (5.7.4) we obtain finally the characteristic function of the normal distribution

$$\varphi(t) = \exp(\mathrm{i}ta)\exp(-\frac{1}{2}b^2t^2) \quad . \qquad (5.7.7)$$

For the case $a = 0$ one obtains from this the following interesting **theorem**:

> A normal distribution with mean value zero has a characteristic function that has itself (up to normalization) the form of a normal distribution. The product of the variances of both functions is one.

If we now consider the sum of two independent normal distributions, then by applying Eq. (5.5.14) one immediately sees that the characteristic function of the sum is again of the form of Eq. (5.7.7). The sum of independent normally distributed quantities is therefore itself normally distributed. The Poisson distribution behaves in a similar way (cf. Example 5.5).

5.8 Quantitative Properties of the Normal Distribution

Figure 5.7a shows the probability density of the standard Gaussian distribution $\phi_0(x)$ and the corresponding distribution function. By simple computation one can determine that the points of inflection of (5.6.1) are at $x = \pm 1$. [In the case of a general Gaussian distribution (5.7.1) they are at $x = a \pm b$.] The distribution function $\psi_0(x)$ gives the probability for the random variable to take on a value smaller than x:

$$\psi_0(x) = P(\mathsf{x} < x) \quad . \tag{5.8.1}$$

By symmetry one has

$$P(|\mathsf{x}| > x) = 2\psi_0(-|x|) = 2\{1 - \psi_0(|x|)\} \tag{5.8.2}$$

or conversely, the probability to obtain a random value within an interval of width $2x$ about zero (the expectation value) is

$$P(|\mathsf{x}| \le x) = 2\psi_0(|x|) - 1 \quad . \tag{5.8.3}$$

Since the integral (5.6.5) is not easy to evaluate, one typically finds the values of (5.8.1) and (5.8.3) from statistical tables, e.g., in Tables I.2 and I.3 of the appendix.

One can now extend this relation to the general Gaussian distribution given by Eq. (5.7.1). Its distribution function is

$$\psi(x) = \psi_0\left(\frac{x-a}{b}\right) \quad . \tag{5.8.4}$$

We are interested in finding the probability to obtain a random value inside (or outside) of a given multiple of $\sigma = b$ about the mean value:

$$P(|\mathsf{x} - a| \le n\sigma) = 2\psi_0\left(\frac{nb}{b}\right) - 1 = 2\psi_0(n) - 1 \quad . \tag{5.8.5}$$

From Table I.3 we find

$$\begin{aligned} P(|\mathsf{x}-a| \le \sigma) &= 68.3\% \;, & P(|\mathsf{x}-a| > \sigma) &= 31.7\% \;, \\ P(|\mathsf{x}-a| \le 2\sigma) &= 95.4\% \;, & P(|\mathsf{x}-a| > 2\sigma) &= 4.6\% \;, \\ P(|\mathsf{x}-a| \le 3\sigma) &= 99.8\% \;, & P(|\mathsf{x}-a| > 3\sigma) &= 0.2\% \;. \end{aligned} \tag{5.8.6}$$

As we will see later in more detail, one can often assume that the measurement errors of a quantity are distributed according to a Gaussian distribution about zero. This means that the probability to obtain a value between x and $x + \mathrm{d}x$ is given by

$$P(x \leq \mathsf{x} < x + \mathrm{d}x) = \phi(x)\,\mathrm{d}x \quad .$$

The dispersion σ of the distribution $\phi(x)$ is called the *standard deviation* or *standard error*. If the standard error of an instrument is known and one carries out a single measurement, then Eq. (5.8.6) tells us that the probability that the true value is within an interval given by plus or minus the standard error about the measured value is 68.3%. It is therefore a common practice to multiply the standard error with a more or less arbitrary factor in order to improve this percentage. (One obtains around 99.8% for the factor 3.) This procedure is, however, misleading and often harmful. If this factor is not explicitly stated, a comparison of different measurements of the same quantity and especially the calculation of a weighted average (cf. Example 9.1) is rendered impossible or is liable to be erroneous.

The quantiles [see Eq. (3.3.25)] of the standard normal distribution are of considerable interest. For the distribution function (5.6.5) one obtains by definition

$$P(x_p) = P(\mathsf{x} < x_p) = \psi_0(x_p) \quad . \tag{5.8.7}$$

The quantile x_p is therefore given by the inverse function

$$x_p = \Omega(P) \tag{5.8.8}$$

of the distribution function $\psi_0(x_p)$. This is computed numerically in Appendix C.4 and is given in Table I.4. Figure 5.8 shows a graphical representation.

We now consider the probability

$$P'(x) = P(|\mathsf{x}| < x) \quad , \qquad x > 0 \quad , \tag{5.8.9}$$

for a quantity distributed according to the standard normal distribution to differ from zero in absolute value by less than x. Since

$$\begin{aligned} P(x) &= P(\mathsf{x} < x) = \psi_0(x) = \int_{-\infty}^{x} \phi_0(x)\,\mathrm{d}x \\ &= \frac{1}{2} + \int_0^x \phi_0(x)\,\mathrm{d}x = \frac{1}{2} + \frac{1}{2}P'(x) = \frac{1}{2}(P'(x)+1) \quad , \end{aligned}$$

it is the inverse function and with it the quantiles of the distribution function $P'(x)$ are obtained by substituting $(P'(x)+1)/2$ for the argument of the inverse function of P:

$$x_p = \Omega'(P') = \Omega((P'+1)/2) \quad . \tag{5.8.10}$$

This function is tabulated in Table I.5.

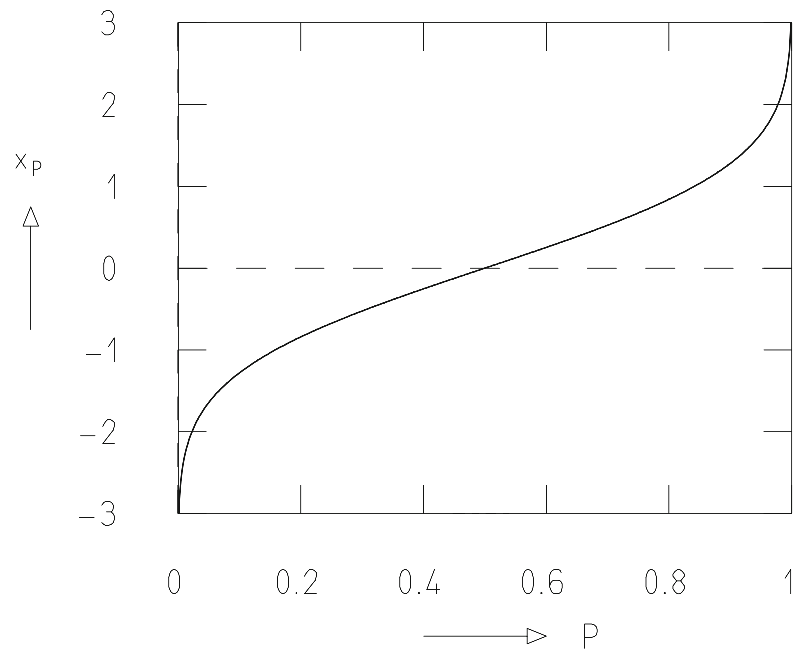

Fig. 5.8: Quantile of the standard normal distribution.

5.9 The Central Limit Theorem

We will now prove the following important theorem. If x_i are independent random variables with mean values a and variances b^2, then the variable

$$\mathsf{x} = \lim_{n\to\infty} \sum_{i=1}^{n} \mathsf{x}_i \qquad (5.9.1)$$

is normally distributed with

$$E(\mathsf{x}) = na \quad , \qquad \sigma^2(\mathsf{x}) = nb^2 \quad . \qquad (5.9.2)$$

From Eq. (3.3.15) one then has that the variable

$$\xi = \frac{1}{n}\mathsf{x} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathsf{x}_i \qquad (5.9.3)$$

is normally distributed with

$$E(\xi) = a \quad , \qquad \sigma^2(\xi) = b^2/n \quad . \qquad (5.9.4)$$

To prove this we assume for simplicity that all of the x_i have the same distribution. If we denote the characteristic function of the x_i by $\varphi(t)$, then the sum of n variables has the characteristic function $\{\varphi(t)\}^n$. We now assume that $a = 0$. (The general case can be related to this by a simple coordinate translation $\mathsf{x}'_i = \mathsf{x}_i - a$.) From (5.5.10) we have the first two derivatives of $\varphi(t)$ at $t = 0$,

$$\varphi'(0) = 0 \quad , \qquad \varphi''(0) = -\sigma^2 \quad .$$

We can therefore expand,

$$\varphi_{x'}(t) = 1 - \frac{1}{2}\sigma^2 t^2 + \cdots \quad .$$

Instead of x_i let us now choose

$$\mathsf{u}_i = \frac{\mathsf{x}'_i}{b\sqrt{n}} = \frac{\mathsf{x}_i - a}{b\sqrt{n}}$$

as the variable. If we consider n to be fixed for the moment, then this implies a simple translation and a change of scale. The corresponding characteristic function is

$$\varphi_{u_i}(t) = E\{\exp(\mathrm{i}tu_i)\} = E\left\{\exp\left(\mathrm{i}t\frac{x_i - a}{b\sqrt{n}}\right)\right\} = \varphi_{x'_i}\left(\frac{t}{b\sqrt{n}}\right)$$

or

$$\varphi_{u_i}(t) = 1 - \frac{t^2}{2n} + \cdots \quad .$$

The higher-order terms are at most of the order n^{-2}. If we now consider the limiting case and use

$$\mathsf{u} = \lim_{n\to\infty}\sum_{i=1}^{n}\mathsf{u}_i = \lim_{n\to\infty}\sum_{i=1}^{n}\frac{\mathsf{x}_i - a}{b\sqrt{n}} = \lim_{n\to\infty}\frac{(\mathsf{x} - na)}{b\sqrt{n}} \tag{5.9.5}$$

we obtain

$$\varphi_u(t) = \lim_{n\to\infty}\{\varphi_{u_i}(t)\}^n = \lim_{n\to\infty}\left(1 - \frac{t^2}{2n} + \cdots\right)^n$$

or

$$\varphi_u(t) = \exp\left(-\frac{1}{2}t^2\right) \quad . \tag{5.9.6}$$

This, however, is exactly the characteristic function of the standard normal distribution $\phi_0(u)$. One therefore has $E(\mathsf{u}) = 0$, $\sigma^2(\mathsf{u}) = 1$. Using Eqs. (5.9.5) and (3.3.15) leads directly to the theorem.

Example 5.6: Normal distribution as the limiting case of the binomial distribution

Suppose that the individual variables x_i in (5.9.1) are described by the simple distribution given by (5.1.1) and (5.1.2), i.e., they can only take on the value 1 (with probability p) or 0 (with probability $1-p$). One then has $E(\mathsf{x}_i) = p$, $\sigma^2(\mathsf{x}_i) = p(1-p)$. The variable

$$\mathsf{x}^{(n)} = \sum_{i=1}^{n} \mathsf{x}_i \tag{5.9.7}$$

then follows the binomial distribution, $P(\mathsf{x}^{(n)} = k) = W_k^n$ [see Eqs. (5.1.3), (5.1.6), (5.1.7)]. As done for (5.9.5) let us consider the distribution of

$$\mathsf{u}^{(n)} = \sum_{i=1}^{n} \frac{\mathsf{x}_i - p}{\sqrt{np(1-p)}} = \frac{1}{\sqrt{np(1-p)}} \left(\sum_{i=1}^{n} \mathsf{x}_i - np \right) \quad . \tag{5.9.8}$$

One clearly has $P(\mathsf{x} = k) = P\left(\mathsf{u}^{(n)} = (k-np)/\sqrt{np(1-p)}\right) = W_k^n$. These values, however, lie increasingly closer to each other on the $\mathsf{u}^{(n)}$ axis as n increases. Let us denote the distance between two neighboring values of $\mathsf{u}^{(n)}$ by $\Delta\mathsf{u}^{(n)}$. Then the distribution of a discrete variable $P(\mathsf{u}^{(n)})/\Delta\mathsf{u}^{(n)}$ finally becomes the probability density of a continuous variable. According to the Central Limit Theorem this must be a standard normal distribution. This is illustrated in Fig. 5.9, where $P(\mathsf{u}^{(n)})/\Delta\mathsf{u}^{(n)}$ is shown for various possible values of $\mathsf{u}^{(n)}$. ■

Example 5.7: Error model of Laplace

In 1783 LAPLACE made the following remarks concerning the origin of errors of an observation. Suppose the true value of a quantity to be measured is m_0. Now let the measurement be disturbed by a large number n of independent causes, each resulting in a disturbance of magnitude ε. For each disturbance there exists an equal probability for a variation of the measured value in either direction, i.e., $+\varepsilon$ or $-\varepsilon$. The measurement error is then composed of the sum of the individual disturbances. It is clear that in this model the probability distribution of measurement errors will be given by the binomial distribution. It is interesting nevertheless to follow the model somewhat further, since it leads directly to the famous Pascal triangle.

Figure 5.10 shows how the probability distribution is derived from the model. The starting point is with no disturbance where the probability of measuring m_0 is equal to one. With one disturbance this probability is split equally between the neighboring possibilities $m_0+\varepsilon$ and $m_0-\varepsilon$. The same happens with every further disturbance. Of course the individual probabilities leading to the same measured value must be added.

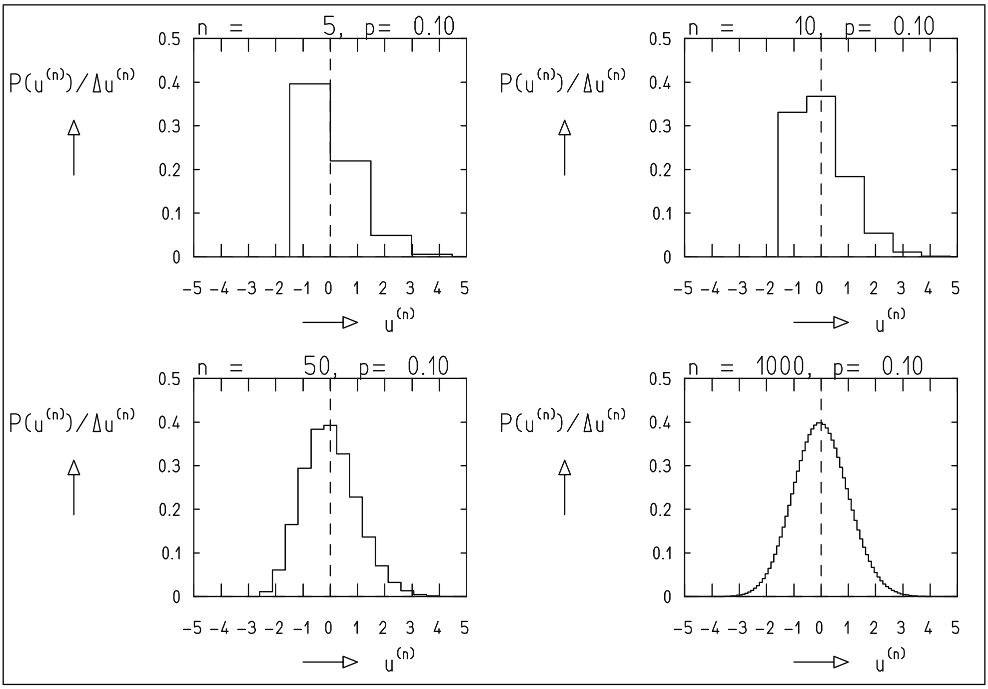

Fig. 5.9: The quantity $P(u^{(n)})/\Delta u^{(n)}$ for various values of the discrete variable $\mathsf{u}^{(n)}$ for increasing n.

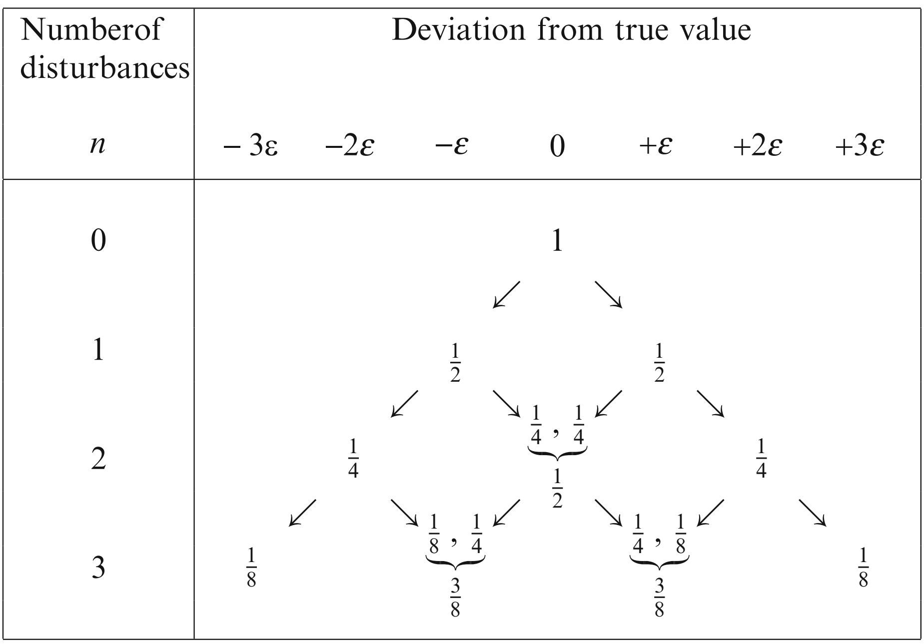

Fig. 5.10: Connection between the Laplacian error model and the binomial distribution.

Each line of the resulting triangle contains the distribution $W_k^n(k = 0, 1, \ldots, n)$ of Eq. (5.1.3) for the case $p = q = 1/2$. Multiplied by $1/(p^k q^{n-k}) = 2^n$ it becomes a line of binomial coefficients of Pascal's triangle (cf. Appendix B).

It is easy to relate this to Example 5.6 by extending Eq. (5.9.8) and substituting $p = 1/2$. For $n \to \infty$ the quantity

$$\mathsf{u}^{(n)} = \frac{2\left(\sum_{i=1}^{n} \varepsilon \mathsf{x}_i - n\varepsilon/2\right)}{\sqrt{n}\varepsilon}$$

follows a normal distribution with expectation value zero and standard deviation $\sqrt{n}\varepsilon/2$. Thus Gaussian measurement errors can result from a large number of small independent disturbances. ■

The identification of the measurement error distribution as Gaussian is of great significance in many computations, particularly for the method of least squares. The normal distribution for measurement errors is, however, not a law of nature. The causes of experimental errors can be individually very complicated. One cannot, therefore, find a distribution function that describes the behavior of measurement errors in all possible experiments. In particular, it is not always possible to guaranty symmetry and independence. One must ask in each individual case whether the measurement errors can be modeled by a Gaussian distribution. This can be done, for example, by means of a χ^2-test, applied to the distribution of a measured quantity (see Sect. 8.7). It is always necessary to check the distribution of experimental errors before more lengthy computations can be used whose results are only meaningful for the case of a Gaussian error distribution.

5.10 The Multivariate Normal Distribution

Consider a vector $\mathbf{x}$ of n variables,

$$\mathbf{x} = (\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_n) \quad .$$

We define the probability density of the joint normal distribution of the x_i to be

$$\phi(\mathbf{x}) = k \exp\{-\frac{1}{2}(\mathbf{x}-\mathbf{a})^{\mathrm{T}} B(\mathbf{x}-\mathbf{a})\} = k \exp\{-\frac{1}{2} g(\mathbf{x})\} \tag{5.10.1}$$

with

$$g(\mathbf{x}) = (\mathbf{x}-\mathbf{a})^{\mathrm{T}} B(\mathbf{x}-\mathbf{a}) \quad . \tag{5.10.2}$$

Here $\mathbf{a}$ is an n-component vector and B is an $n \times n$ matrix, which is symmetric and positive definite. Since $\phi(\mathbf{x})$ is clearly symmetric about the point $\mathbf{x} = \mathbf{a}$, one has

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} (\mathbf{x} - \mathbf{a})\phi(\mathbf{x})\, \mathrm{d}x_1\, \mathrm{d}x_2 \ldots \mathrm{d}x_n = 0 \quad , \tag{5.10.3}$$

that is,

$$E(\mathbf{x} - \mathbf{a}) = 0$$

or

$$E(\mathbf{x}) = \mathbf{a} \quad . \tag{5.10.4}$$

The vector of expectation values is therefore given directly by $\mathbf{a}$.

We now differentiate Eq. (5.10.3) with respect to $\mathbf{a}$,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [I - (\mathbf{x} - \mathbf{a})(\mathbf{x} - \mathbf{a})^{\mathrm{T}} B]\phi(\mathbf{x})\, \mathrm{d}x_1\, \mathrm{d}x_2 \ldots \mathrm{d}x_n = 0 \quad .$$

This means that the expectation value of the quantity in square brackets vanishes,

$$E\{(\mathbf{x} - \mathbf{a})(\mathbf{x} - \mathbf{a})^{\mathrm{T}}\}\, B = I$$

or

$$C = E\{(\mathbf{x} - \mathbf{a})(\mathbf{x} - \mathbf{a})^{\mathrm{T}}\} = B^{-1} \quad . \tag{5.10.5}$$

Comparing with Eq. (3.6.19) one sees that C is the covariance matrix of the variables $\mathbf{x} = (\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n)$.

Because of the practical importance of the normal distribution, we would like to investigate the case of two variables in somewhat more detail. In particular we are interested in the correlation of the variables. One has

$$C = B^{-1} = \begin{pmatrix} \sigma_1^2 & \mathrm{cov}(\mathrm{x}_1, \mathrm{x}_2) \\ \mathrm{cov}(\mathrm{x}_1, \mathrm{x}_2) & \sigma_2^2 \end{pmatrix} \quad . \tag{5.10.6}$$

By inversion one obtains for B

$$B = \frac{1}{\sigma_1^2 \sigma_2^2 - \mathrm{cov}(\mathrm{x}_1, \mathrm{x}_2)^2} \begin{pmatrix} \sigma_2^2 & -\mathrm{cov}(\mathrm{x}_1, \mathrm{x}_2) \\ -\mathrm{cov}(\mathrm{x}_1, \mathrm{x}_2) & \sigma_1^2 \end{pmatrix} \quad . \tag{5.10.7}$$

One sees that B is a diagonal matrix if the covariances vanish. One then has

$$B_0 = \begin{pmatrix} 1/\sigma_1^2 & 0 \\ 0 & 1/\sigma_2^2 \end{pmatrix} \quad . \tag{5.10.8}$$

If we substitute B_0 into Eq. (5.10.1), we obtain – as expected – the joint probability density of two *independently* normally distributed variables as the product of two normal distributions:

$$\phi = k \exp\left(-\frac{1}{2}\frac{(x_1 - a_1)^2}{\sigma_1^2}\right) \exp\left(-\frac{1}{2}\frac{(x_2 - a_2)^2}{\sigma_2^2}\right) \quad . \tag{5.10.9}$$

In this simple case the constant k takes on the value

$$k_0 = \frac{1}{2\pi\sigma_1\sigma_2} \quad ,$$

as can be determined by integration of (5.10.9) or simply by comparison with Eq. (5.7.1). In the general case of n variables with non-vanishing covariances, one has

$$k = \left(\frac{\det B}{(2\pi)^n}\right)^{\frac{1}{2}} \quad . \tag{5.10.10}$$

Here $\det B$ is the determinant of the matrix B. If the variables are not independent, i.e., if the covariance does not vanish, then the expression for the normal distribution of two variables is somewhat more complicated.

Let us consider the reduced variables

$$\mathsf{u}_i = \frac{\mathsf{x}_i - a_i}{\sigma_i} \quad , \qquad i = 1, 2 \quad ,$$

and make use of the correlation coefficient

$$\rho = \frac{\mathrm{cov}(\mathsf{x}_1, \mathsf{x}_2)}{\sigma_1\sigma_2} = \mathrm{cov}(\mathsf{u}_1, \mathsf{u}_2) \quad .$$

Equation (5.10.1) then takes on the simple form

$$\phi(u_1, u_2) = k \exp(-\frac{1}{2}\mathbf{u}^{\mathrm{T}} B \mathbf{u}) = k \exp\left(-\frac{1}{2}g(\mathbf{u})\right) \tag{5.10.11}$$

with

$$B = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \quad . \tag{5.10.12}$$

Contours of equal probability density are characterized by a constant exponent in (5.10.11):

$$-\frac{1}{2} \cdot \frac{1}{(1-\rho^2)}(u_1^2 + u_2^2 - 2u_1u_2\rho) = -\frac{1}{2}g(\mathbf{u}) = \mathrm{const} \quad . \tag{5.10.13}$$

Let us take for the moment $g(\mathbf{u}) = 1$.

In the original variables Eq. (5.10.13) becomes

$$\frac{(x_1 - a_1)^2}{\sigma_1^2} - 2\rho\frac{x_1 - a_1}{\sigma_1}\frac{x_2 - a_2}{\sigma_2} + \frac{(x_2 - a_2)^2}{\sigma_2^2} = 1 - \rho^2 \quad . \tag{5.10.14}$$

This is the equation of an ellipse centered around the point (a_1, a_2). The principal axes of the ellipse make an angle α with respect to the axes x_1 and x_2. This angle and the half-diameters p_1 and p_2 can be determined from Eq. (5.10.14) by using the known properties of conic sections:

$$\tan 2\alpha = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2} \quad , \tag{5.10.15}$$

$$p_1^2 = \frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_2^2\cos^2\alpha - 2\rho\sigma_1\sigma_2\sin\alpha\cos\alpha + \sigma_1^2\sin^2\alpha} \quad , \tag{5.10.16}$$

$$p_2^2 = \frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_2^2\sin^2\alpha + 2\rho\sigma_1\sigma_2\sin\alpha\cos\alpha + \sigma_1^2\cos^2\alpha} \quad . \tag{5.10.17}$$

The ellipse with these properties is called the *covariance ellipse* of the bivariate normal distribution. Several such ellipses are depicted in Fig. 5.11. The covariance ellipse always lies inside a rectangle determined by the point (a_1, a_2) and the standard deviations σ_1 and σ_2. It touches the rectangle at four points. For the extreme cases $\rho = \pm 1$ the ellipse becomes one of the two diagonals of this rectangle.

From (5.10.14) it is clear that other lines of constant probability (for $g \neq 1$) are also ellipses, concentric and similar to the covariance ellipse and situated inside (outside) of it for larger (smaller) probability. The bivariate

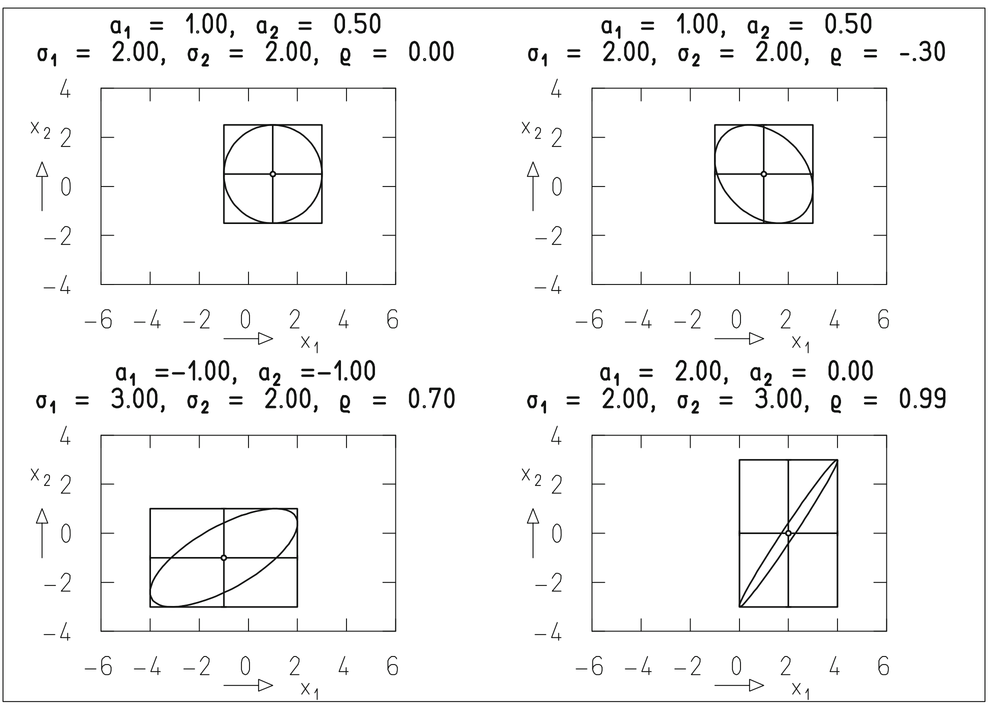

Fig. 5.11: Covariance ellipses.

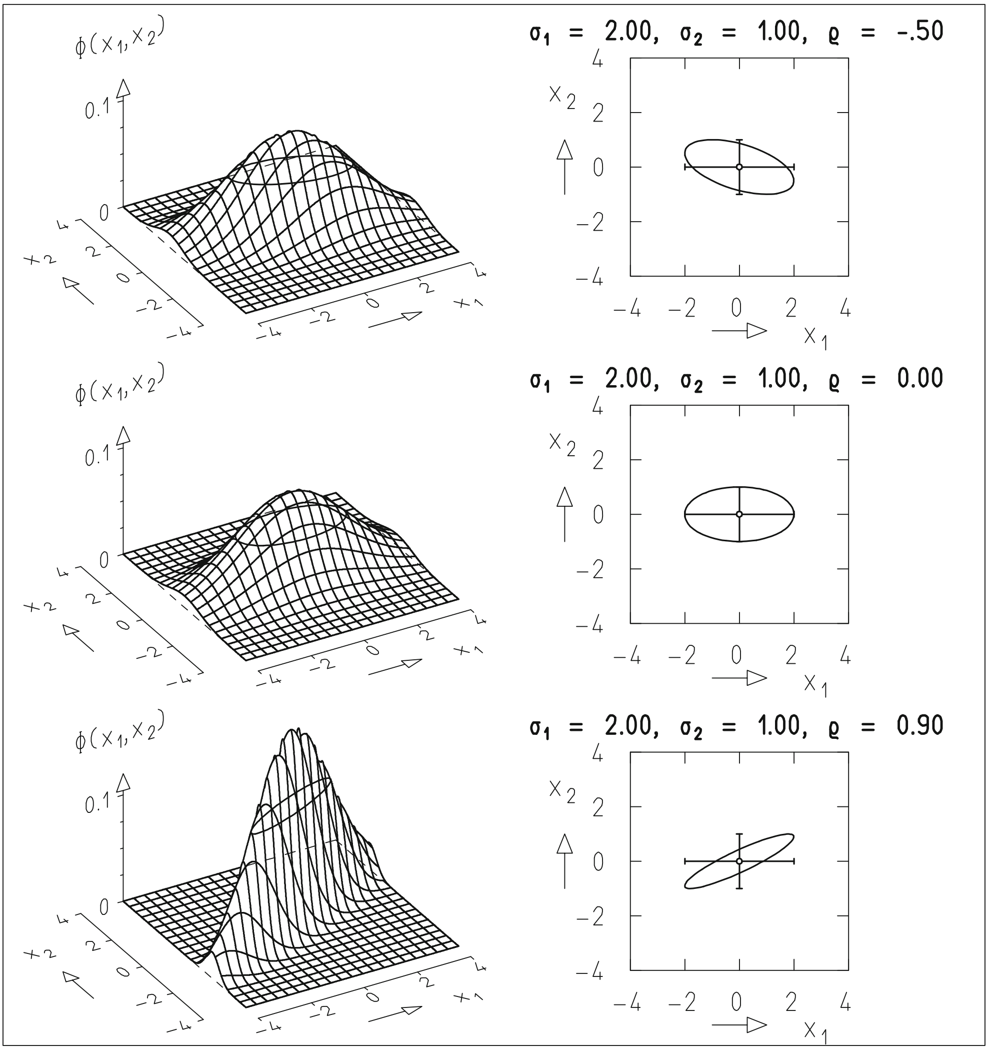

Fig. 5.12: Probability density of a bivariate Gaussian distribution (*left*) and the corresponding covariance ellipse (*right*). The three rows of the figure differ only in the numerical value of the correlation coefficient ρ.

normal distribution therefore corresponds to a surface in the three-dimensional space (x_1, x_2, ϕ) (Fig. 5.12), whose horizontal sections are concentric ellipses. For the largest probability this ellipse collapses to the point (a_1, a_2). The vertical sections through the center have the form of a Gaussian distribution whose width is directly proportional to the diameter of the covariance ellipse along which the section extends. The probability of observing a pair x_1, x_2 of random variables inside the covariance ellipse is equal to the integral

$$\int_A \phi(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1 - \mathrm{e}^{-\frac{1}{2}} = \mathrm{const} \quad , \tag{5.10.18}$$

where the region of integration A is given by the area within the covariance ellipse (5.10.14). The relation (5.10.18) is obtained by application of the transformation of variables $\mathbf{y} = T\,\mathbf{x}$ with $T = B^{-1}$ to the distribution $\phi(\mathbf{x})$. The resulting distribution has the properties $\sigma(y_1) = \sigma(y_2) = 1$ and $\mathrm{cov}(y_1, y_2) = 0$, i.e., it is of the form of (5.10.9). In this way the region of integration is transformed to a unit circle centered about (a_1, a_2).

In our consideration of the normal distribution of measurement errors of a variable we found the interval $a - \sigma \le x \le a + \sigma$ to be the region in which the probability density $f(x)$ exceeded a given fraction, namely $e^{-1/2}$ of its maximum value. The integral over this region was independent of σ. In the case of two variables, the role of this region is taken over by the covariance ellipse determined by σ_1, σ_2, and ρ, and not – as is sometimes incorrectly assumed – by the rectangle that circumscribes the ellipse in Fig. 5.11. The meaning of the covariance ellipse can also be seen from Fig. 5.13. Points 1 and 2, which lie on the covariance ellipse, correspond to equal probabilities ($P(1) = P(2) = P_e$), although the distance of point 1 from the middle is less in both coordinate directions. In addition, point 3 is more probable, and point 4 less probable ($P(4) < P_e$, $P(3) > P_e$), even though point 4 even closer is to (a_1, a_2) than point 3.

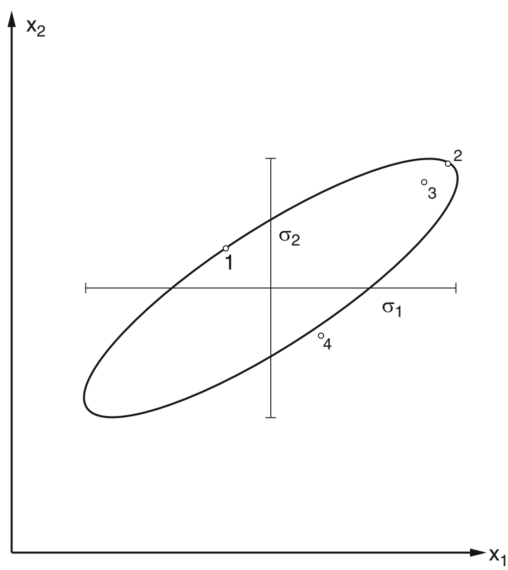

Fig. 5.13: Relative probability for various points from a bivariate Gaussian distribution ($P_1 = P_2 = P_e$, $P_3 > P_e$, $P_4 < P_e$).

For three variables one obtains instead of the covariance ellipse a *covariance ellipsoid*, for n variables a hyperellipsoid in an n-dimensional space (see also Sect. A.11). According to our construction, the covariance ellipsoid is the hypersurface in the n-dimensional space on which the function $g(\mathbf{x})$ in the exponent of the normal distribution (5.10.1) has the constant value $g(\mathbf{x}) = 1$. For other values $g(\mathbf{x}) = \text{const}$ one obtains similar ellipsoids which lie inside ($g < 1$) or outside ($g > 1$) of the covariance ellipsoid. In Sect. 6.6 it will be shown that the function $g(\mathbf{x})$ follows a χ^2-distribution with n degrees of freedom if $\mathbf{x}$ follows the normal distribution (5.10.1). The probability to find $\mathbf{x}$ inside the ellipsoid $g = \text{const}$ is therefore

$$W = \int_0^g f(\chi^2; n)\,\mathrm{d}\chi^2 = P\left(\frac{n}{2}, \frac{g}{2}\right) \quad . \tag{5.10.19}$$

Here P is the incomplete gamma function given in Sect. D.5. For $g = 1$, that is, for the covariance ellipsoid in n dimensions, this probability is

$$W_n = P\left(\frac{n}{2}, \frac{1}{2}\right) \quad . \tag{5.10.20}$$

Numerical values for small n are

$$\begin{aligned} W_1 &= 0.68269 \quad , & W_2 &= 0.39347 \quad , & W_3 &= 0.19875 \quad , \\ W_4 &= 0.09020 \quad , & W_5 &= 0.03734 \quad , & W_6 &= 0.01439 \quad . \end{aligned}$$

The probability decreases rapidly as n increases. In order to be able to give regions for various n which correspond to equal probability content, one specifies a value W on the left-hand side of (5.10.19) and determines the corresponding value of g. Then g is the quantile with probability W of the χ^2-distribution with n degrees of freedom (see also Appendix C.5),

$$g = \chi^2_W(n) \quad . \tag{5.10.21}$$

The ellipsoid that corresponds to the value of g that contains $\mathbf{x}$ with the probability W is called the *confidence ellipsoid* of probability W. This expression can be understood to mean that, e.g., for $W = 0.9$ one should have 90% confidence that $\mathbf{x}$ lies within the confidence ellipsoid.

The variances σ_i^2 or the standard deviations $\Delta_i = \sigma_i$ also have a certain meaning for n variables. The probability to observe the variable x_i in the region $a_i - \sigma_i < x_i < a_i + \sigma_i$ is, as before, 68.3%, independent of the number n of the variables. This only holds, however, when one places no requirements on the positions of any of the other variables x_j, $j \neq i$.

5.11 Convolutions of Distributions

5.11.1 Folding Integrals

On various occasions we have already discussed sums of random variables, and in the derivation of the Central Limit Theorem, for example, we found the characteristic function to be a useful tool in such considerations. We would now like to discuss the distribution of the sum of two quantities, but for greater clarity we will not make use of the characteristic function.

A sum of two distributions is often observed in experiments. One could be interested, for example, in the angular distribution of secondary particles from the decay of an elementary particle. This can often be used to determine the spin of the particle. The observed angle is the distribution of a sum of random quantities, namely the decay angle and its measurement error. One speaks of the *convolution* of two distributions.

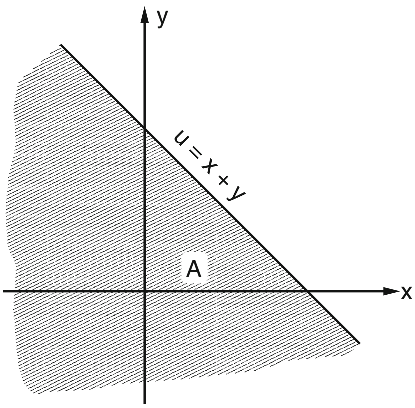

Fig. 5.14: Integration region for (5.11.4).

Let the original quantities be x and y and the sum

$$\mathsf{u} = \mathsf{x} + \mathsf{y} \quad . \tag{5.11.1}$$

A requirement for further treatment is that the original variables must be independent. In this case, the joint probability density is the product of simple densities,

$$f(x,y) = f_x(x) f_y(y) \quad . \tag{5.11.2}$$

If we now ask for the distribution function of u, i.e., for

$$F(u) = P(\mathsf{u} < u) = P(\mathsf{x}+\mathsf{y} < u) \quad , \tag{5.11.3}$$

then this is obtained by integration of (5.11.2) over the hatched region A in Fig. 5.14;

$$\begin{aligned} F(u) &= \iint_A f_x(x) f_y(y)\,\mathrm{d}x\,\mathrm{d}y = \int_{-\infty}^{\infty} f_x(x)\,\mathrm{d}x \int_{-\infty}^{u-x} f_y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} f_y(y)\,\mathrm{d}y \int_{-\infty}^{u-y} f_x(x)\,\mathrm{d}x \quad . \end{aligned} \tag{5.11.4}$$

By differentiation one obtains the probability density for u,

$$f(u) = \frac{\mathrm{d}F(u)}{\mathrm{d}u} = \int_{-\infty}^{\infty} f_x(x) f_y(u-x)\,\mathrm{d}x = \int_{-\infty}^{\infty} f_y(y) f_x(u-y)\,\mathrm{d}y \quad . \tag{5.11.5}$$

If x or y or both are only defined in a restricted region, then (5.11.5) is still true. The limits of integration, however, may be limited. We will consider various cases:

(a) $0 \le \mathsf{x} < \infty\,, \quad -\infty < \mathsf{y} < \infty$:

$$f(u) = \int_{-\infty}^{u} f_x(u-y) f_y(y)\,\mathrm{d}y \quad . \tag{5.11.6}$$

(Because $y = u - x$ and since for $x_{\mathrm{min}} = 0$ one has $y_{\mathrm{max}} = u$.)

(b) $0 \le \mathsf{x} < \infty\,, \quad 0 \le \mathsf{y} < \infty$:

$$f(u) = \int_{0}^{u} f_x(u-y) f_y(y)\,\mathrm{d}y \quad . \tag{5.11.7}$$

(c) $a \le \mathsf{x} < b\,, \quad -\infty < \mathsf{y} < \infty$:

$$f(u) = \int_{a}^{b} f_x(x) f_y(u-x)\,\mathrm{d}x \quad . \tag{5.11.8}$$

We will demonstrate case (d) in the following example, in which both x and y are bounded from below and from above.

Example 5.8: Convolution of uniform distributions
With

$$f_x(x) = \left\{ \begin{array}{l} 1,\ 0 \le x < 1 \\ 0\,\text{otherwise} \end{array} \right. \qquad \text{and} \qquad f_y(y) = \left\{ \begin{array}{l} 1,\ 0 \le y < 1 \\ 0\,\text{otherwise} \end{array} \right.$$

and Eq. (5.11.8) we obtain

$$f(u) = \int_{0}^{1} f_y(u-x)\,\mathrm{d}x \quad .$$

We substitute $v = u - x$, $\mathrm{d}v = -\mathrm{d}x$ and obtain

$$f(u) = -\int_{u}^{u-1} f_y(v)\,\mathrm{d}v = \int_{u-1}^{u} f_y(v)\,\mathrm{d}v \quad . \tag{5.11.9}$$

Clearly one has $0 < u < 2$. We now consider separately the two cases

$$\begin{array}{llll} \text{(a)} & 0 \le \mathsf{u} < 1 & : & f_1(u) = \displaystyle\int_{0}^{u} f_y(v)\,\mathrm{d}v = \int_{0}^{u} \mathrm{d}v = u \quad , \\ \text{(b)} & 1 \le \mathsf{u} < 2 & : & f_2(u) = \displaystyle\int_{u-1}^{1} f_y(v)\,\mathrm{d}v = \int_{u-1}^{1} \mathrm{d}v = 2 - u \quad . \end{array} \tag{5.11.10}$$

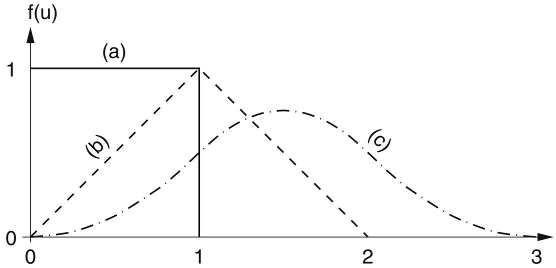

Fig. 5.15: Convolution of uniform distributions. Probability density of the sum u of uniformly distributed random variables x (*a*) $u = x$, (*b*) $u = x + x$, (*c*) $u = x + x + x$.

Note that the lower (upper) limit of integration is not lower (higher) than the value 0 (1). The result is a triangular distribution (Fig. 5.15).

If this result is folded again with a uniform distribution, i.e., if u is the sum of three independent uniformly distributed variables, then one obtains

$$f(u) = \begin{cases} \frac{1}{2}u^2 , & 0 \leq u < 1 \quad , \\ \frac{1}{2}(-2u^2 + 6u - 3) , & 1 \leq u < 2 \quad , \\ \frac{1}{2}(u-3)^2 , & 2 \leq u < 3 \quad . \end{cases} \tag{5.11.11}$$

The proof is left to the reader. The distribution consists of three parabolic sections (Fig. 5.15) and is similar already to the Gaussian distribution predicted by the Central Limit Theorem. ■

5.11.2 Convolutions with the Normal Distribution

Suppose a quantity of experimental interest x can be considered to be a random variable with probability density $f_x(x)$. It is measured with a measurement error y, which follows a normal distribution with a mean of zero and a variance of σ^2. The result of the measurement is then the sum

$$u = x + y \quad . \tag{5.11.12}$$

Its probability density is [see also (5.11.4)]

$$f(u) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} f_x(x) \exp[-(u-x)^2/2\sigma^2]\, dx \quad . \tag{5.11.13}$$

By carrying out many measurements, $f(u)$ can be experimentally determined. The experimenter is interested, however, in the function $f_x(x)$. Unfortunately, Eq. (5.11.13) cannot in general be solved for $f_x(x)$. This is only possible for a restricted class of functions $f(u)$. Therefore one usually approaches the problem in a different way. From earlier measurements or theoretical considerations one possesses knowledge about the form of $f_x(x)$, e.g., one might assume that $f_x(x)$ is described by a uniform distribution, without, however, knowing its boundaries a and b. One then carries out the convolution (5.11.13), compares the resulting function $f(u)$ with the experiment and in this way determines the unknown parameters (in our example a and b).

In many cases it is not even possible to perform the integration (5.11.13) analytically. Numerical procedures, e.g., the Monte Carlo method, then have to be used. Sometimes approximations (cf. Example 5.11) give useful results. Because of the importance in many experiments of convolution with the normal distribution we will study some examples.

Example 5.9: Convolution of uniform and normal distributions

Using Eqs. (3.3.26) and (5.11.8) and substituting $v = (x-u)/\sigma$ we obtain

$$\begin{aligned} f(u) &= \frac{1}{b-a}\frac{1}{\sqrt{2\pi}\sigma}\int_a^b \exp[-(u-x)^2/2\sigma^2]\,\mathrm{d}x \\ &= \frac{1}{b-a}\frac{1}{\sqrt{2\pi}}\int_{(a-u)/\sigma}^{(b-u)/\sigma} \exp(-\frac{1}{2}v^2)\,\mathrm{d}v \quad , \\ f(u) &= \frac{1}{b-a}\left\{\psi_0\left(\frac{b-u}{\sigma}\right)-\psi_0\left(\frac{a-u}{\sigma}\right)\right\} \quad . \end{aligned} \tag{5.11.14}$$

The function ψ has already been defined in (5.6.5). Figure 5.16 shows the result for $a=0$, $b=6$, $\sigma=1$. If one has $|b-a| \gg \sigma$ (as is the case in Fig. 5.16), one of the terms in parentheses in (5.11.14) is either 0 or 1. The rising edge of the uniform distribution at $u=a$ is replaced by the distribution function of the normal distribution with standard deviation σ (see also Fig. 5.7). The falling edge at $u=b$ is its "mirror image". ■

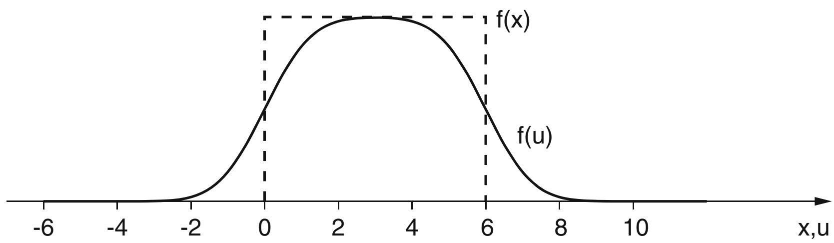

Fig. 5.16: Convolution of a uniform and Gaussian distribution.

Example 5.10: Convolution of two normal distributions. "Quadratic addition of errors"

If one convolutes two normal distributions with mean values 0 and variances σ_x^2 and σ_y^2, one obtains

$$f(u) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-u^2/2\sigma^2) \quad , \qquad \sigma^2 = \sigma_x^2 + \sigma_y^2 \quad . \tag{5.11.15}$$

The proof has already been shown in Sect. 5.7 with the help of the characteristic function. It can also be obtained by computation of the folding integral (5.11.5). If the distributions $f_x(x)$ and $f_y(y)$ describe two independent sources of measurement errors, the result (5.11.15) is known as the *"quadratic addition of errors"*. ■

Example 5.11: Convolution of exponential and normal distributions

With

$$\begin{aligned} f_x(x) &= \frac{1}{\tau}\exp(-x/\tau) \quad , \qquad x > 0 \quad , \\ f_y(y) &= \frac{1}{\sqrt{2\pi}\sigma}\exp(-y^2/2\sigma^2) \quad , \end{aligned}$$

Eq. (5.11.6) takes on the following form:

$$f(u) = \frac{1}{\sqrt{2\pi}\sigma\tau}\int_{-\infty}^{u}\exp[-(u-y)/\tau]\exp(-y^2/2\sigma^2)\,\mathrm{d}y \quad .$$

We can rewrite the exponent

$$\begin{aligned} &-\frac{1}{2\sigma^2\tau}[2\sigma^2(u-y)+\tau y^2] \\ &\qquad = -\frac{1}{2\sigma^2\tau}\left[2\sigma^2 u - 2\sigma^2 y + \tau y^2 + \frac{\sigma^4}{\tau} - \frac{\sigma^4}{\tau}\right] \\ &\qquad = -\frac{u}{\tau} + \frac{\sigma^2}{2\tau^2} - \frac{1}{2\sigma^2}\left(y - \frac{\sigma^2}{\tau}\right)^2 \end{aligned}$$

and obtain

$$f(u) = \frac{1}{\sqrt{2\pi}\sigma\tau}\exp\left\{\frac{\sigma^2}{2\tau^2} - \frac{u}{\tau}\right\}\int_{-\infty}^{u-\sigma^2/\tau}\exp\left(\frac{-v^2}{2\sigma^2}\right)\mathrm{d}v \quad .$$

We now require that $\sigma \ll \tau$, i.e., that the measurement error is much smaller than the typical value (width) of the exponential distribution. In addition, we only consider values of u for which $u - \sigma^2/\tau \gg \sigma$, i.e., $u \gg \sigma$. The integral is then approximately equal to $\sqrt{2\pi}\sigma$ or

$$f(u) \approx \frac{1}{\tau}\exp\left\{-\frac{u}{\tau} + \frac{\sigma^2}{2\tau^2}\right\} \quad .$$

In a semi-logarithmic representation, i.e., in a plot of $\ln f(u)$ versus u, the curve $f(u)$ lies above the curve $f_x(x)$, by an amount $\sigma^2/2\tau^2$, since

$$\ln f(u) = \ln\frac{1}{\tau} + \frac{\sigma^2}{2\tau^2} - \frac{u}{\tau} = \ln f_x(x) + \frac{\sigma^2}{2\tau^2} \quad .$$

This is plotted in Fig. 5.17. The result can be qualitatively understood in the following way. For each small x interval of the exponential distribution, the convolution leads with equal probability to a shift to the left or to the right. Since, however, the exponential distribution for a given u is greater for small values of x, contributions to the convolution $f(u)$ originate with greater probability from the left than from the right. This leads to an overall shift to the right of $f(u)$ with respect to $f_x(x)$. ■

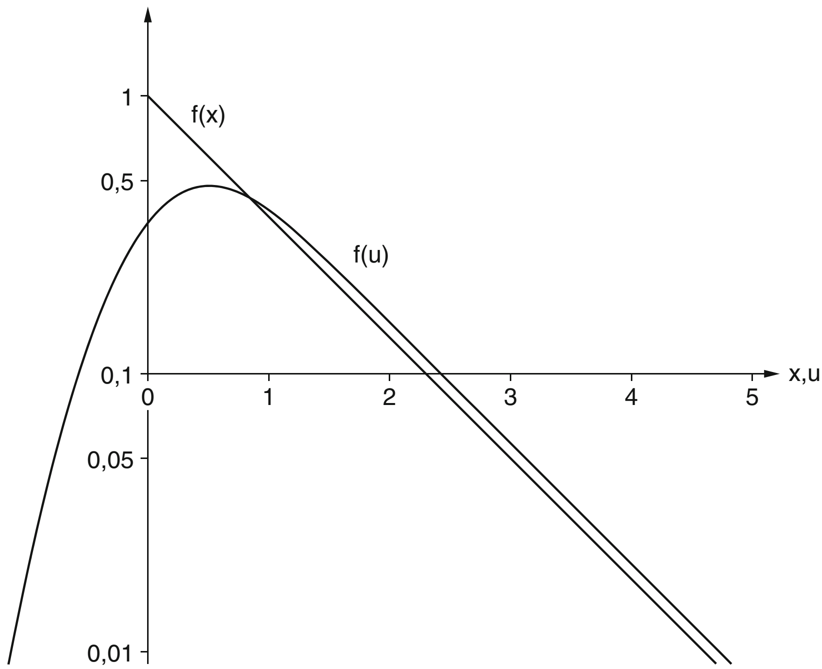

Fig. 5.17: Convolution of exponential and normal distributions.

5.12 Example Programs

Example Program 5.1: Class E1Distrib to simulate empirical frequency and demonstrate statistical fluctuations

The program simulates the problem of Example 5.1. It allows input of values for $n_{\rm exp}$, $n_{\rm fly}$, and $P(A)$, and then consecutively performs $n_{\rm exp}$ simulated experiments. In each experiment $n_{\rm fly}$ objects are analyzed. Each object has a probability $P(A)$ to have the property A. For each experiment one line of output is produced containing the current number $i_{\rm exp}$ of the experiment, the number N_A of objects with the property A and the frequency $h_A = N_A/n_{\rm fly}$ with which the property A was found. The fluctuation of h_A around the known input value $P(A)$ in the individual experiments gives a good impression of the statistical error of an experiment.

Example Program 5.2: Class E2Distrib to simulate the experiment of Rutherford and Geiger

The principle of the experiment of Rutherford and Geiger is described in Example 5.3. It is simulated as follows. Input quantities are the number N of decays observed and the number n_{int} of partial intervals ΔT of the total observation time T. For simplicity the length of each partial interval is set equal to one. A total of N random events are simulated by simply generating N random numbers uniformly distributed between 0 and T. They are entered into a histogram with n_{int} intervals. The histogram is first displayed graphically and then analyzed numerically. For each number $k = 0, 1, \ldots, N_{\mathrm{int}}$ the program determines how many intervals $N(k)$ of the histogram have k entries. The numbers $N(k)$ themselves are presented in the form of another histogram.

Show that for the process simulated in this example program one obtains in the limit $N \to \infty$

$$N(k) = n_{\mathrm{int}} W_k^N (p = 1/n_{\mathrm{int}}) \quad .$$

If N is increased step by step and at the same time $\lambda = N_p = N/n_{\mathrm{int}}$ is kept constant, then for large N one has

$$W_k^N (p = \lambda/N) \to \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda}$$

and, in the limit $N \to \infty$,

$$N(k) = n_{\mathrm{int}} \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} \quad .$$

Check the above statements by running the program with suitable pairs of numbers, e.g., $(N, n_{\mathrm{int}}) = (4, 2)$, $(40, 20)$, $\ldots$, $(2000, 1000)$, by reading the numbers $N(k)$ from the graphics display and by comparing them with the statements above.

Example Program 5.3: Class E3Distrib to simulate Galton's board

Galton's board is a simple implementation of Laplace's model described in Example 5.7. The vertical board contains rows of horizontally oriented nails as shown in Fig. 5.18. The rows of nails are labeled $j = 1, 2, \ldots, n$, and row j has j nails. One by one a total of N_{exp} balls fall onto the nail in row 1. There each ball is deflected with probability p to the right and with probability $(1-p)$ to the left. (In a realistic board one has $p = 1/2$.) The distance between the nails is chosen in such a way that in each case the ball hits one of the two nails in row 2 and there again it is deflected with the probability p to the right. After falling through n rows each ball assumes one of $n+1$ places, which we denote by $k = 0$ (on the left), $k = 1, \ldots, k = n$ (on the right). After a total of N_{exp} experiments (i.e., balls) one finds $N(k)$ balls for each value k.

The program allows input of numerical values for N_{exp}, n, and p. For each experiment the number k is first set to zero and n random numbers r_j are generated from a uniform distribution and analyzed. For each $r_j < p$ (corresponding to a deflection to the right in row j) the number k is increased by 1. For each experiment the value of k is entered into a histogram. After all experiments are simulated the histogram is displayed.

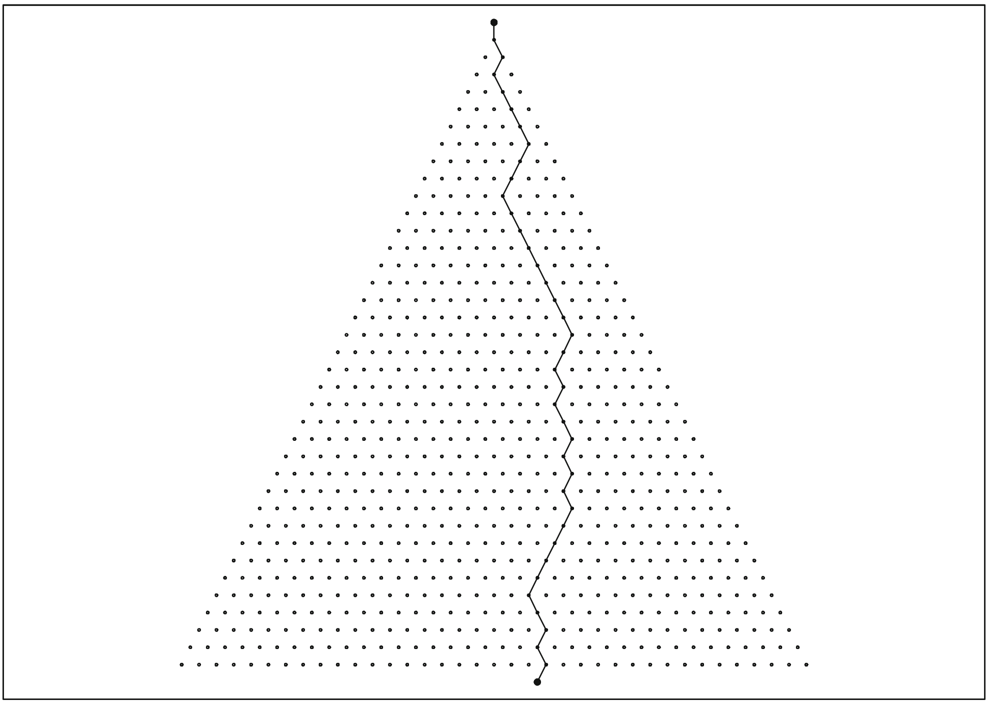

Fig. 5.18: Arrangement of nails in Galton's board and one possible trajectory of a ball.

Show that in the limit $N \to \infty$ one has

$$N(k) = N\, W_k^n(p) \quad .$$

By choosing and entering suitable pairs of numbers (n, p), e.g., $(n, p) = (1, 0.5)$, $(2, 0.25)$, $(10, 0.05)$, $(100, 0.005)$ approximate the Poisson limit

$$N(k) = N \frac{\lambda^k}{k!} \mathrm{e}^{-\lambda} \quad , \qquad \lambda = np \quad ,$$

and compare these predictions with the results of your simulations.