

Chapter 9

Elements from the Theory of Errors

We will now examine the general principles of the Theory of Errors. Apart from the a posteriori justification of some assumptions we have already made and used, this chapter will give us new theoretical tools which we will use in the development of new techniques in the analysis of experimental results.

9.1 The Normal or Gaussian Law of Errors

All believe in the exponential law of errors; the experimentalists because they think it can be proved with mathematics and the mathematicians because they believe it has been established experimentally.

E. Whittaker and G. Robinson [1]

The law known as *Gauss' law of errors*, was first formulated by Laplace in 1783. Laplace based his derivation on the assumption that the deviation of a measurement belonging to a group of measurements from the group's mean is due to a large number of small deviations, due to causes which are independent of each other. He assumed that deviations of the same magnitude are equally probable to be positive or negative. Gauss, later, proved the law based on the assumption that the numerical mean is the most probable value of a number of equivalent measurements.

The mathematical form of the law defines the *normal or Gaussian probability density function*, which gives the distribution of the results x of the measurements of a physical magnitude x , as

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-(x-\mu)^2/2\sigma^2}. \quad (9.1)$$

The probability density depends on two parameters, μ and σ . It is proved that μ is the mean of a large number of measurements or the value towards which the mean \bar{x} of a series of measurements tends as the number of measurements tends to infinity and σ is the standard deviation of the measurements from μ . The function (9.1) describes the distribution of a very large number ($N \rightarrow \infty$) of measurement results, i.e. what we call the *parent population (or universe) of all the possible results x of the measurements*. A series of measurements of x constitutes a sample taken at random from this population. From this finite number of experimental results, we may deduce the best estimates for μ and σ , according to what has been said in Chap. 4.

Having in mind what we have shown for the binomial and the Gaussian distributions in Chap. 7, the extraction of the expression (9.1) for the probability density is simple. Let us suppose that the magnitude being measured has real value equal to x_0 and that the final deviation of a measurement x from x_0 is due to N independent sources of error. To simplify the arguments, we assume that the random errors from these sources all have the same magnitude ε . This limitation is not necessary and the law may also be extracted for a general distribution of these elementary errors [2]. The probabilities for the errors to have values $\pm\varepsilon$ are both equal to $p = \frac{1}{2}$. For N errors of magnitude ε , the final results will lie between the values $x = x_0 - N\varepsilon$ and $x = x_0 + N\varepsilon$.

If, in a measurement, n errors are positive and, therefore, $N - n$ errors are negative, the result of the measurement will be

$$x = x_0 + n\varepsilon - (N - n)\varepsilon = x_0 + (2n - N)\varepsilon. \quad (9.2)$$

The probability of this occurring is, according to the binomial distribution,

$$P_N(n) = \frac{N!}{n!(N - n)!} p^n (1 - p)^{N - n} = \frac{N!}{n!(N - n)!} \left(\frac{1}{2}\right)^N, \quad (9.3)$$

and this is the probability of a measurement having as its result the value $x = x_0 + (2n - N)\varepsilon$.

In the limit, in which the binomial distribution approaches the Gaussian distribution, this probability becomes (see Sect. 7.4)

$$P_N(n) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(n-\mu)^2/2\sigma^2} \quad (9.4)$$

with

$$\mu = Np = \frac{N}{2}, \quad \sigma = \sqrt{Npq} = \frac{\sqrt{N}}{2}. \quad (9.5)$$

When we take as variable the continuous variable x [see Eqs. (7.54)–(7.56)], we have the probability distribution

$$f(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-(x-\mu)^2/2\sigma^2}, \tag{9.6}$$

with

$$\mu = x_0, \quad \sigma = \sqrt{Npq}\varepsilon = \frac{\sqrt{N}}{2}\varepsilon. \tag{9.7}$$

The Gaussian curve is shown in Fig. 9.1. This distribution was, as we have already mentioned, originally derived by de Moivre for the results of coin tossing.

The relation between the binomial distribution and the error of a measurement, according to the assumptions we have made, is seen in Fig. 9.2. A large enough number of small balls fall, colliding with regularly arranged cylindrical obstacles, such that they force the balls to deviate to the left or to the right by a given constant step $\varepsilon/2$. The probabilities for a deviation to the right and for a deviation to the left are equal, $p = \frac{1}{2}$. The balls do not interact with each other during their descent. Shown in the figure is the distribution of the final positions of the balls after $N = 24$ such deviations. The distribution is binomial and, in the limit, Gaussian. The symmetry of the distribution is seen, as well as the fact that small deviations are more probable than large ones, since they are attained in a larger number of possible ways (paths).

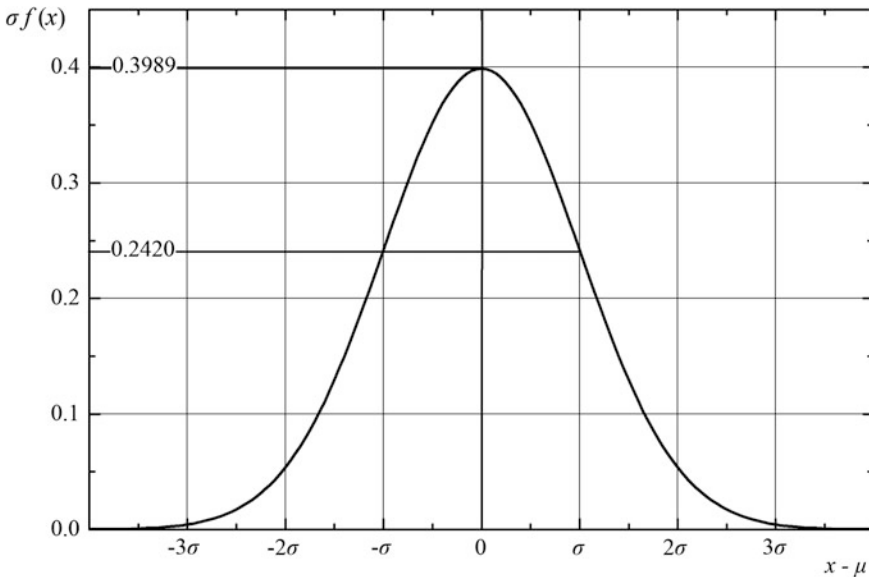


Fig. 9.1 The Gaussian function

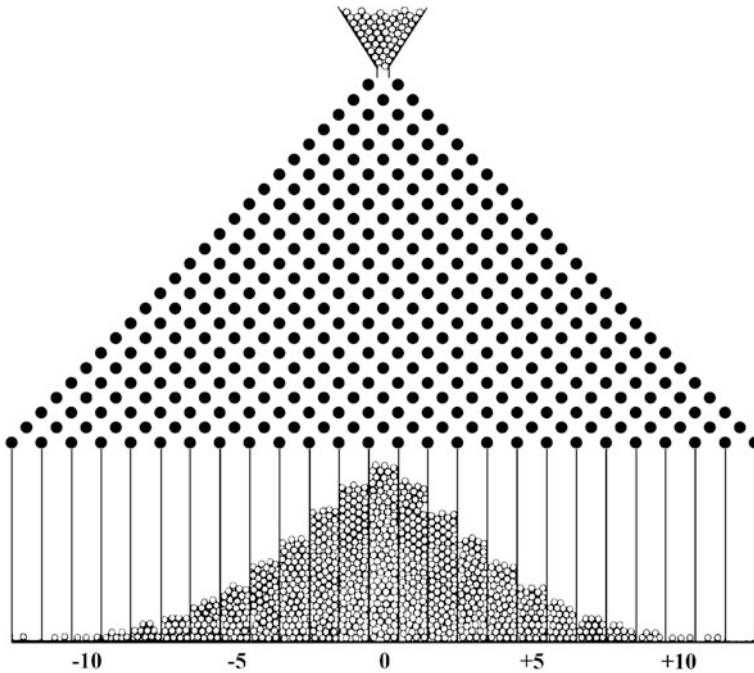


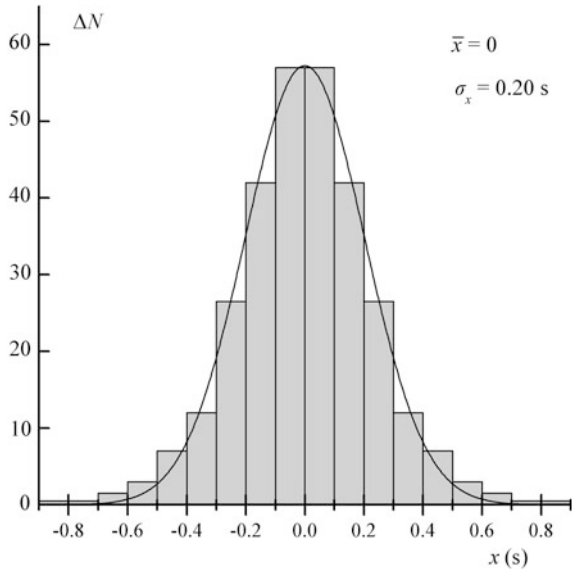
Fig. 9.2 The distribution of the deviations suffered by falling small balls which are forced 24 times to deviate to the *left* or to the *right*, by equal steps and with equal probabilities. Their number in each region of deviation is given by the binomial distribution

The arrangement is due to Francis Galton and is known by its Latin name *quincunx*, which describes the way fruit-bearing trees are planted. Since in the figure it is $N = 24$, the deviations are even multiples of $\varepsilon/2$. Thus, we see that the deviations have values which are integral multiples of ε between -12ε and $+12\varepsilon$. Each ball may be considered to be the result of a measurement, which is subjected to 24 small deviations of magnitude $\varepsilon/2$, with equally probable positive or negative sign.

Laplace and Gauss assumed that the normal distribution has universal validity, based on their own studies of experimental results. Today, this is known not to be true. Apart from the cases in which the deviations from the normal distribution are too striking to neglect, the distribution is used because, as put by Jeffreys [3]: ‘*The normal law of errors cannot be proved theoretically. The justification for its use is that the law represents measurements of many kinds without being too wrong and that it is much easier to use than other laws, which would be more accurate in certain cases*’.

Despite all these, even when the parent population does not have a normal distribution, the distribution of the means of a series of a finite number of measurements is nearer to the normal distribution than the parent population. This follows from a very important theorem, the *central limit theorem*, which we will discuss in detail below.

Fig. 9.3 The histogram of Bessel for the errors of 300 measurements of the right ascension of stars. The Gaussian fitted to the histogram has a mean $\bar{x} = 0$ and a standard deviation $\sigma_x = 0.20$ s



The first example of measurements appearing to be normally distributed was given by Bessel, who grouped the results of 300 measurements of the right ascension of stars. The errors were given in seconds of arc and lied between -1 and $+1$ s. The histogram of the errors (Fig. 9.3) is symmetrical with respect to zero, since Bessel grouped together the positive and the negative errors. As shown in the figure, a Gaussian curve with mean $\bar{x} = 0$ (as expected) and standard deviation from the mean $\sigma_x = 0.20$ s, is fitted to the histogram. The fit of the Gaussian to the data is very good. In fact, it so good that there were suggestions that a selection of values was made, in order to get a better agreement with the normal distribution of errors.

Birge [4] performed, with a spectrometer, a series of 500 adjustments in which he placed the vertical wire of the cross wire in the optical field of the instrument as near as he could to the center of a wide but symmetrical spectral line in the solar spectrum. This is the procedure followed in the measurement of the wavelength of the spectral line. He recorded the readings on the instrument's scale, in μm . The frequencies of the measurements' residuals from their mean value, $v_i = x_i - \bar{x}$, are presented in the histogram of Fig. 9.4. The agreement of the distribution of the errors with a normal distribution is very good. We may check whether the deviations of the values of Birge's histogram are near the expected ones. If ΔN_G are the values of the Gaussian curve fitted to the histogram of Fig. 9.4 (thick curve), then the expected deviations will be, according to the Poisson distribution, of the order of $\sqrt{\Delta N_G}$. The curves ΔN_G and $\Delta N_G \pm \sqrt{\Delta N_G}$ are also drawn in the figure. We note that the deviations of the histogram's columns from the Gaussian curve are within or near the expected limits.

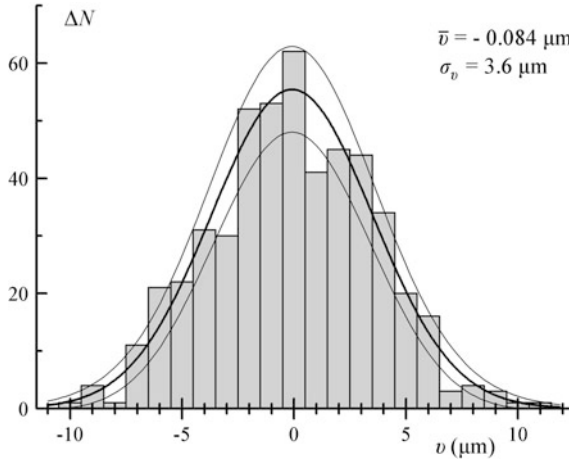


Fig. 9.4 Histogram of the residuals $v_i = x_i - \bar{x}$ of the 500 measurements performed by Birge with a spectrometer in order to test the normal law of errors. The Gaussian fitted to the histogram has a standard deviation of $\sigma_v = 3.6 \mu\text{m}$. Apart from the Gaussian curve ΔN_G (thick line) the curves $\Delta N_G \pm \sqrt{\Delta N_G}$ are also drawn. The expected standard deviation of ΔN_G is, according to the Poisson distribution, equal to $\sqrt{\Delta N_G}$

In other cases, the errors in the measurements of other experiments are not described so satisfactorily by a Gaussian curve. In some cases, the data required the use of a sum of two normal curves with different standard deviations. Naturally, it is obvious that the use of two curves, with four parameters to be determined instead of the two of a single curve, will always give better agreement than a single curve. On the other hand, the need for the use of two curves for a better fit, might be an indication of either that the errors are due to two widely different sources or that the measurements were not performed under exactly the same experimental conditions (i.e. they are derived from two different parent populations).

9.2 The Lyapunov Central Limit Theorem

A theorem of very great importance in the Theory of Probability and Statistics is *the Lyapunov central limit theorem*, which we will now discuss, without proving it [5]. An elementary formulation of the theorem, which is adequate for the purposes of this book, is the following:

If x_1, x_2, \dots, x_N are the N values of a random sample taken from a parent population of the random variable \mathbf{x} , which has mean μ and standard deviation σ , then, as the number N tends to infinity, the distribution of the means \bar{x} of the x_i approaches a normal distribution with mean and standard deviation

$$\overline{(\bar{x})} = \mu_{\bar{x}} = \mu \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}, \tag{9.8}$$

respectively. In other words, the probability density of the means $\bar{x} = \frac{1}{N}(x_1 + x_2 + \dots + x_N)$ tends to

$$f(\bar{x}) = \frac{1}{\sqrt{2\pi} \sigma_{\bar{x}}} e^{-\frac{(\bar{x} - \mu_{\bar{x}})^2}{2\sigma_{\bar{x}}^2}} = \frac{\sqrt{N}}{\sqrt{2\pi} \sigma} e^{-N\frac{(\bar{x} - \mu)^2}{2\sigma^2}}. \tag{9.9}$$

It must be noted that the distribution of the parent population does not need to be normal.

We will explain the meaning of the theorem with the aid of Fig. 9.5. Figure 9.5a shows the probability density of the parent distribution of all the possible results of the measurements of x . To make the description of the sampling method easier, the area under the curve has been divided into a finite number of identical rectangular cells. In the limit, this number will be considered to tend to infinity. To each cell there corresponds a small region of x values, given by the projection of the cell on the x -axis. The process of *sampling* is simply the random picking of a number N of these cells. The cells have the same probability of being picked during the sampling. At points of large $f(x)$, the vertical column consists of a larger number of cells and the values of x corresponding to this column are more likely to be the result of a measurement. The 10 black cells in the figure could be the ones selected in a sampling with $N = 10$ values (measurements). Their projection on the x -axis results in the histogram of the 10 measurements (Fig. 9.5b). The mean \bar{x} of these N values is evaluated. The central limit theorem states that, independently of the shape of the distribution of the parent population, these means, \bar{x} , which result from different series of N measurements each, have a distribution which, for large N , tends to a normal distribution with mean equal to the mean of the parent population, $\overline{(\bar{x})} = \mu_{\bar{x}} = \mu$, and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$, where σ is the standard deviation of the parent population.

We will demonstrate the central limit theorem with a few examples, in which the sampling is done from parent populations of known distributions.

Example 9.1

Use the Monte Carlo method to check the validity of the central limit theorem for measurement numbers $N = 1, 2, 4, 8$ and 16 , when the parent distribution of the measurements has probability density: $f(x) = 0$ everywhere, except in the interval $0 \leq x \leq 1$, in which it is $f(x) = 1$.

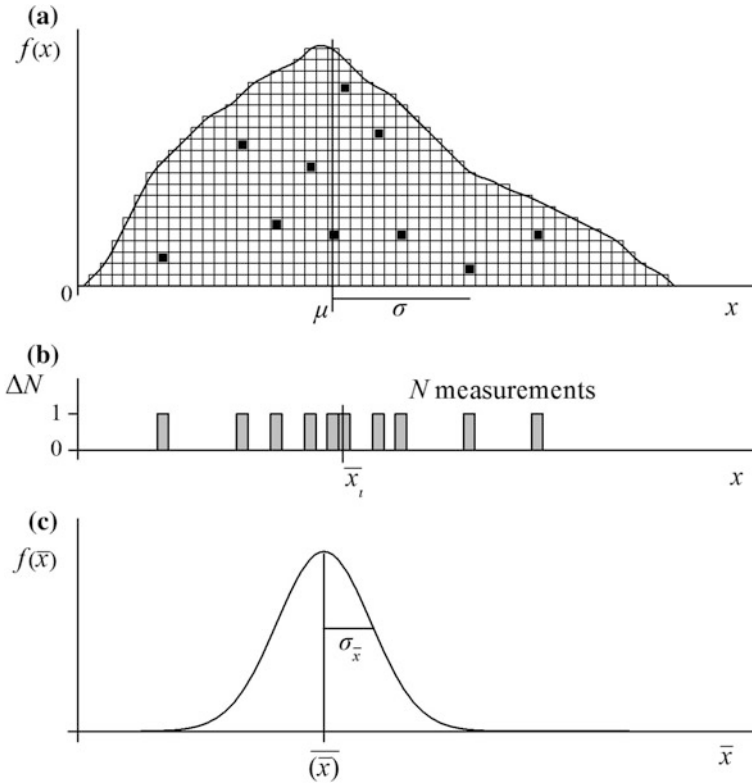
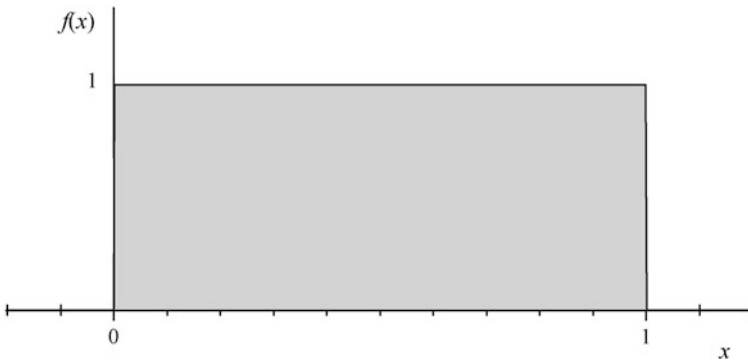


Fig. 9.5 The central limit theorem. **a** The parent population, with mean μ and standard deviation σ . **b** One series of N measurements with mean \bar{x}_i . **c** The distribution of the means, \bar{x}_i . It has a mean $(\bar{x}) = \mu_{\bar{x}}$, which tends to μ for large N and a standard deviation $\sigma_{\bar{x}}$, which tends to σ/\sqrt{N}

The probability density $f(x)$ has been drawn in the figure that follows.



It has a mean value $\mu = 0.5$ and a standard deviation given by the relation

$$\sigma = \sqrt{\int_0^1 (x - 0.5)^2 f(x) dx} = \sqrt{\left[\frac{1}{3} (x - 0.5)^3 \right]_0^1} = \sqrt{\frac{1}{12}} = 0.289.$$

The Monte Carlo method will be used in a simulation of the experimental process in order to ‘find’ the results x_i of the measurements. The first 50 000 of the decimal digits of π were used as the random numbers required for the application of the method. They were divided into groups of 5 digits and divided by 10^5 , thus giving 10 000 numbers between 0 and 1, with 5 significant figures each (from 0.00000 to 0.99999).

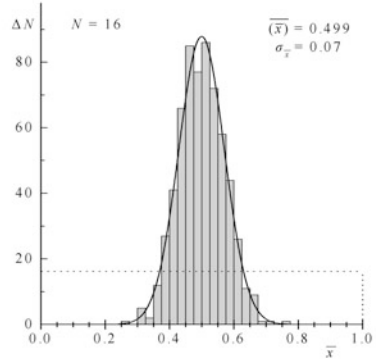
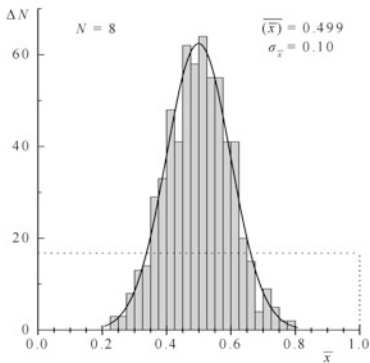
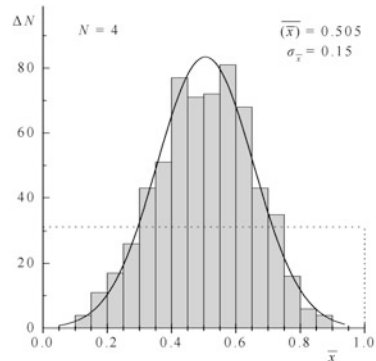
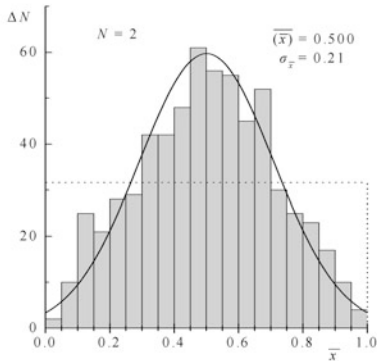
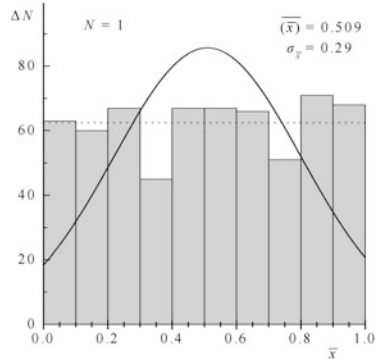
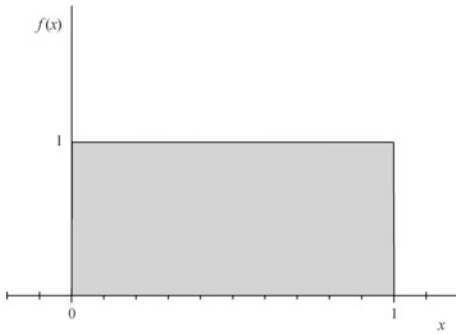
Since the results x_i are uniformly distributed between 0 and 1 (constant probability density) and the same must be true for random numbers, if they are indeed random, the random numbers found as described above are taken directly to be the values of x_i .

Thus, the first 625 numbers gave 625 results x_i , which are recorded in the first histogram of the figure given below ($N = 1$). The variable is denoted by \bar{x} for uniformity with the rest of the histograms, but these ‘mean’ values consist of one measurement each ($N = 1$). As a consequence, this histogram must reproduce, approximately, the probability density $f(x)$. The Gaussian which is fitted to the histogram, with some dose of exaggeration, has a mean of 0.509 and $\sigma_{\bar{x}} = 0.29$.

Next, the first $2 \times 625 = 1250$ random numbers gave 625 pairs of values, the means of which, \bar{x} , were found and are given in the second histogram ($N = 2$). Even with N being just two, the fitting of a Gaussian curve to the distribution of the means is very satisfactory.

The procedure is continued for $N = 4, 8$ and 16. For $N = 16$, the $16 \times 625 = 10000$ random numbers give 625 series of 16 results of measurement x_i each. The distribution of the means of these groups of 16 values is given by the last histogram. The Gaussian approximation to this histogram is seen to be very good.

Given with all the histograms, in dashed line, is the mean number of values corresponding to each class, taking into account the width of the classes and the total number of means \bar{x} , which is 625 in all cases.



The table below gives, for comparison, the theoretically expected values of the means $\overline{(\bar{x})}$ of the \bar{x} and the standard deviations $\sigma_{\bar{x}} = \sigma/\sqrt{N}$, as well as the values determined with the simulation of the experiment which was performed.

| N | $\overline{(\bar{x})}$ | | $\sigma_{\bar{x}}$ | |
|----|------------------------|-----------------------|------------------------|-----------------------|
| | Theoretically expected | 'Experimental' result | Theoretically expected | 'Experimental' result |
| 1 | 0.5 | 0.509 | 0.289 | 0.29 |
| 2 | 0.5 | 0.500 | 0.204 | 0.21 |
| 4 | 0.5 | 0.505 | 0.144 | 0.15 |
| 8 | 0.5 | 0.499 | 0.102 | 0.10 |
| 16 | 0.5 | 0.499 | 0.072 | 0.07 |

As a conclusion, we note that the mean value of the means, $\overline{(\bar{x})}$, tends very quickly towards the expected value according to the central limit theorem (as should be expected for the relatively large number of 625 measurements!). The distribution of \bar{x} also tends towards a normal distribution very quickly, with its standard deviation of the Gaussian being $\sigma_{\bar{x}} \propto 1/\sqrt{N}$.

In the next example, a distribution of x will be used which is very far from being normal, a kind of 'anti-Gaussian' distribution. The example will also give us the opportunity of a better understanding of the Monte Carlo method.

Example 9.2

Use the Monte Carlo method in order to test the validity of the central limit theorem for numbers of measurements $N = 1, 2, 4, 8$ and 16 , when the parent distribution of the measurements is given by the probability density:

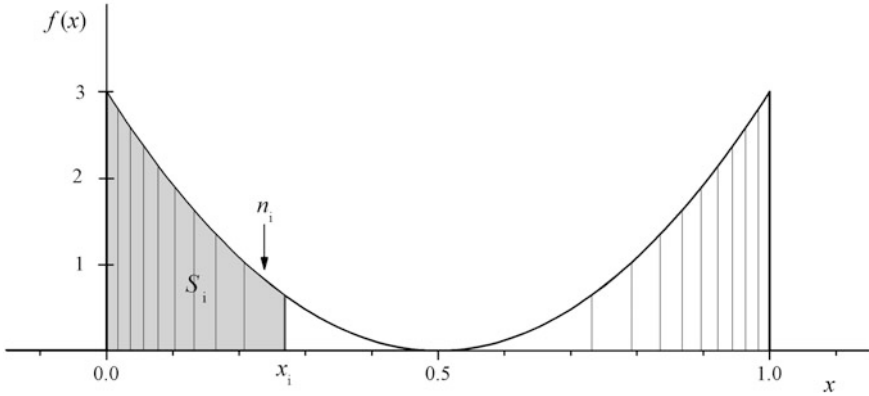
$$f(x) = 0 \text{ everywhere except in the region } 0 \leq x \leq 1, \text{ where it is } f(x) = 3(1 - 2x)^2.$$

The function of the probability density is normalized. It has a parabolic shape (see figure below) with a minimum equal to 0 at $x = 0.5$. Due to the symmetry of the distribution, the mean value of x is $\bar{x} = 0.5$. This distribution was chosen as an example of an *anti-Gaussian* distribution since results near the mean have very small probability of being observed, while the opposite happens for values near the edges of the distribution.

The standard deviation of this distribution is

$$\begin{aligned} \sigma &= \sqrt{\int_0^1 (x - 0.5)^2 f(x) dx} = \sqrt{12 \int_0^1 (x - 0.5)^4 dx} = \sqrt{\left[\frac{12}{5} (x - 0.5)^5 \right]_0^1} \\ &= \sqrt{\frac{3}{20}} = 0.387. \end{aligned}$$

The Monte Carlo method will be used again in the simulation of the measurement procedure. 10 000 random numbers, n_i , between 0 and 1, with 5 significant figures each, were found exactly as in the last example, using the first 50 000 decimal digits of π . The correspondence of these random numbers n_i to values of x_i is a little more difficult than in the last example, because now the values of x_i are not uniformly distributed between 0 and 1. As the way in which this will be achieved is of general importance for the Monte Carlo method, we will describe it in some detail.



In the present example we need to attribute values x_i of x to 10^4 random numbers. It should be noted that the 10^4 random numbers may take any one of 10^4 possible values, between 0.0000 and 0.9999. We divide the interval $[0, 1]$ into 10^4 strips (see figure), each with different width Δx , choosing these Δx in such a way that the area under the curve between x and $x + \Delta x$ has the value of $f(x) \Delta x = 10^{-4}$. The area corresponds to the probability that a value lies between x and $x + \Delta x$. We now have 10^4 surface elements of equal areas and 10^4 possible values of the random numbers. The 10^4 random numbers each have the same probability to appear and the result of a measurement is equally probable to lie in one of the 10^4 surface elements.

We will correspond the 10^4 surface elements to the 10^4 random numbers in increasing order, so that the random number 0.0000 corresponds to the first Δx interval, the number 0.0001 to the second and so on, up to the number 0.9999 which corresponds to the 10 000th interval. We may then say that the appearance of a random number is equivalent to the result of the corresponding measurement lying in the interval between x and $x + \Delta x$ that is covered by the corresponding surface element. Let the increasing order numbers of the strips be N_i , between 0 and 9999. The appearance of a random number n_i , which according to the convention we adopted belongs to the strip with order number $N_i \equiv n_i \times 10^4$, is interpreted to mean that one 'measurement' gave a result lying in this strip. The area of the surface under the curve and to the left of the N_i -th strip is equal to $S_i = N_i \times 10^{-4} = n_i$. We conclude that the corresponding value x_i of x is such that the area under the curve from $x = 0$ to x_i is equal to $S_i = n_i$. Thus, the random number n_i uniquely defines a probability between 0 and 1, which corresponds to an area S_i which, in its turn, corresponds to a value x_i such that

$$S_i = \int_0^{x_i} f(x) dx = n_i.$$

This equation may be solved for $x_i(n_i)$.

For the probability density $f(x) = 3(1 - 2x)^2$ ($0 \leq x \leq 1$), we find that it is $S_i = \frac{1}{2} \left[1 + (2x_i - 1)^3 \right]$ and, therefore, $x_i = \frac{1}{2} + \frac{1}{2} (2S_i - 1)^{1/3}$, from which relation we have the correspondence of x_i to the random number n_i :

$$x_i = \frac{1}{2} + \frac{1}{2} (2n_i - 1)^{1/3}.$$

In order to apply the Monte Carlo method to the present example, the random numbers of the previous example were used. Thus, the first 625 numbers gave, via the last relation, 625 results x_i , which are recorded in the first histogram of the figure that follows (for $N = 1$). Again the variable is denoted by \bar{x} for the purposes of uniformity with the rest of the histograms, although these ‘mean values’ are actually single values. This histogram must reproduce the probability density $f(x)$, something which is seen to happen to a satisfactory degree. The process is simply sampling and the result shows the degree to which a sample of 625 measurements may determine the parent population.

The next histogram (for $N = 2$) is based on the first $2 \times 625 = 1250$ random numbers, which gave 625 pairs of values, the means of which, \bar{x} , were found. As expected, a maximum is observed in the region of $x = 0.5$, which has its origin in the means of pairs of values, one of which originates from the region around the center of the distribution (≈ 0) and the other from the regions of its edges (≈ 1). Striking maxima and minima due to similar combinations are still visible in the histogram for $N = 4$, but the Gaussian shape of the distribution is already clearly visible. The mean for $N = 4$ is 0.511 (instead of the expected 0.5) and the standard deviation is 0.22 (instead of the expected 0.19).

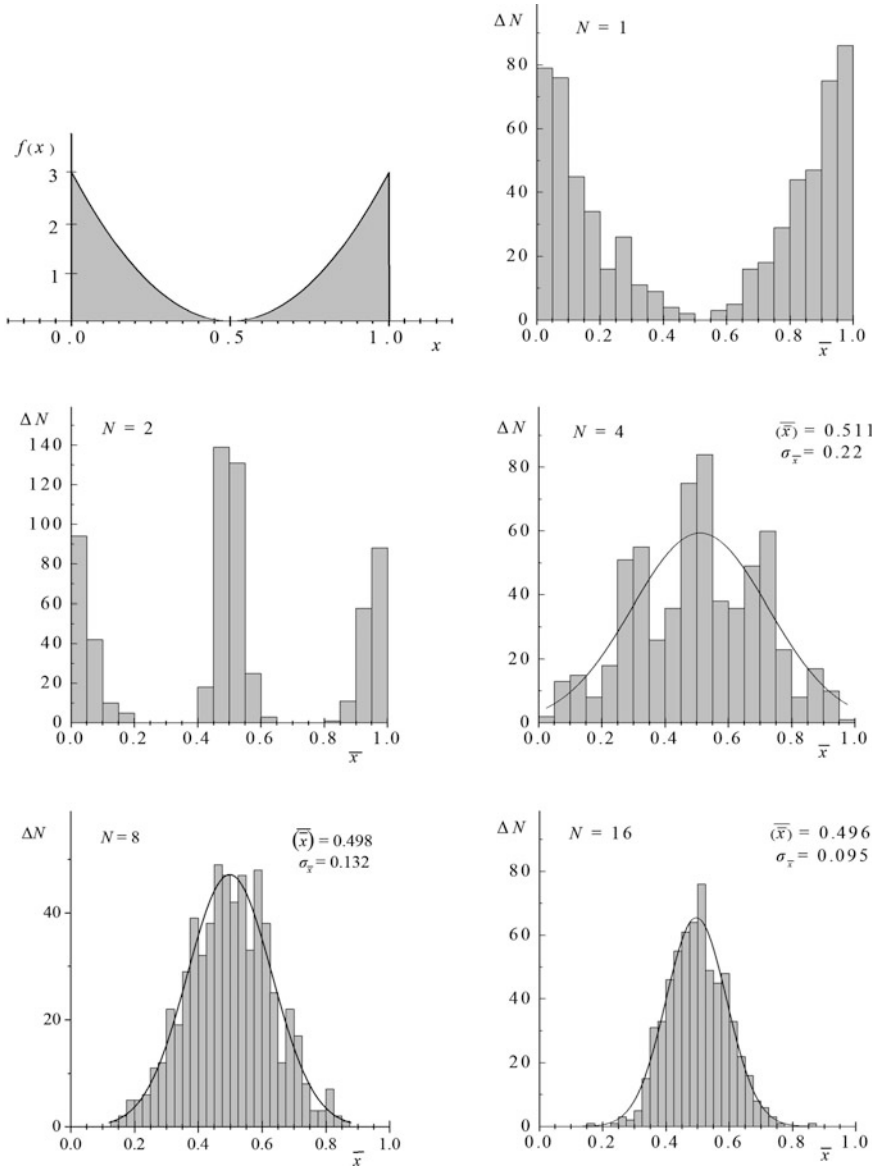
The normal shape of the histograms is more evident for means evaluated from $N = 8$ and $N = 16$ values of x_i (last two histograms).

The table that follows gives the theoretically expected values of the mean values ($\overline{\bar{x}}$) of the means \bar{x} and of the standard deviations $\sigma_{\bar{x}} = \sigma/\sqrt{N}$, as well as the values which were determined through the simulation of the experiment we have performed.

| N | $\overline{\bar{x}}$ | | $\sigma_{\bar{x}}$ | |
|----|------------------------|-----------------------|------------------------|-----------------------|
| | Theoretically expected | ‘Experimental’ result | Theoretically expected | ‘Experimental’ result |
| 1 | 0.5 | – | 0.387 | – |
| 2 | 0.5 | – | 0.274 | – |
| 4 | 0.5 | 0.511 | 0.194 | 0.220 |
| 8 | 0.5 | 0.498 | 0.137 | 0.132 |
| 16 | 0.5 | 0.496 | 0.097 | 0.095 |

We note that, although the distribution of the measurements of the parent population is far from being normal, the distribution of the means \bar{x} tends, relatively

quickly, to the normal form, in agreement with the central limit theorem. The standard deviations are again observed to be in satisfactory agreement with the theoretical relation $\sigma_{\bar{x}} = \sigma/\sqrt{N}$.



One more (purely theoretical) example demonstrating the validity of the central limit theorem will be given in Sect. 9.6 (Example 9.5), in the study of convolution and the calculation of the means and the standard deviations of sums of numbers picked from a certain distribution.

A note regarding random numbers. The subject of random numbers is of enormous importance in applications of the Monte Carlo method and in simulation in general. There are large tables of random numbers [6], as well as algorithms for their production [7] (pseudorandom numbers). In this book, we usually choose to use the decimal digits of π as a source of random numbers. The absolute randomness of a series of digits is not easy to prove and is actually impossible to prove beyond any doubt. The only thing one can say is that the series of these particular digits have passed successfully certain basic tests, such as, for example, that the variations of the frequencies of appearance of the 10 digits (0, 1, ..., 9) are within the statistically expected limits, the same holds for the 100 two-digit combinations (00, 01, 02, ..., 99) etc. The digits of π have successfully passed these tests [8]. Of course, a large number of people, ranging from professional mathematicians to amateurs interested in the theory of numbers, continuously search and find many coincidences, the occurrence of which is expected to be much rarer than observed in practice. Yasumasa Kanada, for example, having calculated the first 206.1 billion decimal digits of π , found out that the sequence 01234567891 appears 5 times, instead of the expected two times. Not paying any attention to such ‘strange phenomena’, the decimal digits of π are considered to be adequately random for the purposes of this book.

9.3 The Best Estimate that May Be Made for the Real Value of a Magnitude, Based on the Results of N Measurements of It

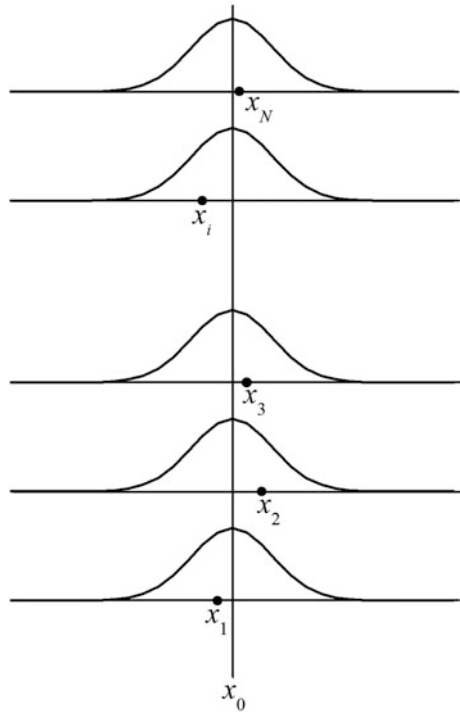
Assume that we have N values (measurements), x_i ($i = 1, 2, \dots, N$), of a random variable, which we will denote by \mathbf{x} . We take these values to be normally distributed about the real value x_0 of \mathbf{x} , with standard deviation σ . Then, referring to Fig. 9.6, we may say that

the probability for the first value of \mathbf{x} to lie between x_1 and $x_1 + dx_1$ is $\frac{1}{\sqrt{2\pi}\sigma} e^{-(x_1-x_0)^2/2\sigma^2} dx_1$,

the probability for the second value of \mathbf{x} to lie between x_2 and $x_2 + dx_2$ is $\frac{1}{\sqrt{2\pi}\sigma} e^{-(x_2-x_0)^2/2\sigma^2} dx_2$ etc., and

the probability for the N -th value of \mathbf{x} to lie between x_N and $x_N + dx_N$ is $\frac{1}{\sqrt{2\pi}\sigma} e^{-(x_N-x_0)^2/2\sigma^2} dx_N$.

Fig. 9.6 The probability density function for a result x , and the results x_i of N measurements



The compound probability for all the measurements to lie within the limits mentioned is:

$$d^N P = \frac{1}{(\sqrt{2\pi}\sigma)^N} \exp\left\{-\frac{1}{2\sigma^2} [(x_1 - x_0)^2 + (x_2 - x_0)^2 + \dots + (x_N - x_0)^2]\right\} \\ \times dx_1 dx_2 \cdots dx_N, \quad (9.10)$$

or

$$d^N P = \frac{1}{(\sqrt{2\pi}\sigma)^N} e^{-\chi^2/2} d^N v, \quad (9.11)$$

where

$$\chi^2 \equiv \frac{1}{\sigma^2} [(x_1 - x_0)^2 + (x_2 - x_0)^2 + \dots + (x_N - x_0)^2] \quad (9.12)$$

and

$$d^N v = dx_1 dx_2 \dots dx_N \tag{9.13}$$

may be considered to be the element of the N -dimensional volume about the point (x_1, x_2, \dots, x_N) .

The real value x_0 is not known. We consider that the best estimate we can make for it is the value \hat{x}_0 of x_0 which maximizes the probability for the results of the measurements of \mathbf{x} we have to occur. For given limits dx_1, dx_2, \dots, dx_N , this happens when the quantity χ^2 is minimum, i.e. for

$$\frac{\partial \chi^2}{\partial x_0} = \frac{2}{\sigma^2} [(x_1 - x_0) + (x_2 - x_0) + \dots + (x_N - x_0)]_{x_0=\hat{x}_0} = 0 \tag{9.14}$$

or

$$N\hat{x}_0 - (x_1 + x_2 + \dots + x_N) = 0. \tag{9.15}$$

The best estimate for x_0 is, therefore,

$$\hat{x}_0 = \frac{1}{N} (x_1 + x_2 + \dots + x_N), \tag{9.16}$$

i.e. the mean of the N values of x .

The result may be considered to be a proof of *the principle of least squares*, first formulated by Legendre. The principle states that:

The most probable value of a magnitude being measured is that which minimizes the sum of the squares of the deviations of the results of the measurements from this value.

Both Gauss and Laplace studied this principle. Gauss, assuming the mean of the measurements to be the most probable value of the magnitude being measured, derived the normal law of errors. Inversely, the normal law of errors may be used in order to prove the principle of the most probable value, as we have done.

9.4 The Weighting of Values

If the values we have at our disposal come from different parent populations with different standard deviations (i.e. distributions relative to the real value) due to the different accuracies in the determination of each value, then we will have

$$d^N P = \frac{1}{(\sqrt{2\pi})^N \sigma_1 \sigma_2 \dots \sigma_N} \exp \left\{ - \left[\frac{(x_1 - x_0)^2}{2\sigma_1^2} + \frac{(x_2 - x_0)^2}{2\sigma_2^2} + \dots + \frac{(x_N - x_0)^2}{2\sigma_N^2} \right] \right\} dx_1 dx_2 \dots dx_N, \tag{9.17}$$

and the minimization of

$$\chi^2 = \frac{(x_1 - x_0)^2}{\sigma_1^2} + \frac{(x_2 - x_0)^2}{\sigma_2^2} + \dots + \frac{(x_N - x_0)^2}{\sigma_N^2} \quad (9.18)$$

gives the relation

$$\hat{x}_0 = \frac{(x_1/\sigma_1^2) + (x_2/\sigma_2^2) + \dots + (x_N/\sigma_N^2)}{1/\sigma_1^2 + 1/\sigma_2^2 + \dots + 1/\sigma_N^2} \quad (9.19)$$

as the best estimate for the real value x_0 .

Equation (9.19) may also be written as

$$\hat{x}_0 = \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i} = \bar{x} \quad (9.20)$$

which is the *weighted mean* of x , with *statistical weight* for value x_i equal to

$$w_i = \frac{1}{\sigma_i^2}. \quad (9.21)$$

The importance of a value in the determination of the mean is, therefore, inversely proportional to the square of its standard deviation. The bigger the standard deviation of a measurement, the smaller the weight it is given in the determination of the mean, something that appears qualitatively reasonable. When all the measurements have the same weight, it is $w_i / \sum_i w_i = 1/N$ and the equations reduce to the known ones.

More generally, if the weights w_1, w_2, \dots, w_N are attributed, for whatever reason, to the values of measurements x_1, x_2, \dots, x_N , respectively, the weighted mean of these values is given by

$$\bar{x} = \frac{\sum_{i=1}^N w_i x_i}{\sum_{i=1}^N w_i}. \quad (9.22)$$

Equations (9.18) and (9.19) show that the magnitude $\sum w_i (x_i - x_0)^2$ has a minimum when $\bar{x} = x_0$. In order to evaluate the standard deviation of the weighted values x_1, x_2, \dots, x_N , we normalize the statistical weights w_i so that their sum is equal to unity, by dividing each one with $\sum w_i$. Defining the *normalized statistical weights*

$$\beta_i \equiv \frac{w_i}{\sum_{i=1}^N w_i}, \tag{9.23}$$

for which it is true that

$$\sum_i \beta_i = 1, \tag{9.24}$$

we have the weighted mean of x ,

$$\bar{x} = \sum_{i=1}^N \beta_i x_i. \tag{9.25}$$

When all the measurements have the same weight, it is $\beta_i = \frac{1}{N}$.

The *weighted standard deviation* of the values x_i is defined as

$$s_x \equiv \sqrt{\sum_{i=1}^N \beta_i (x_i - \bar{x})^2} \tag{9.26}$$

and the *weighted standard deviation of the mean* \bar{x} is given by

$$\sigma_{\bar{x}} \equiv \sqrt{\frac{1}{(N-1)} \sum_{i=1}^N \beta_i (x_i - \bar{x})^2}, \tag{9.27}$$

where N here is the number of x values with non-zero weight. It should be noted that in the evaluation of the standard deviation, β_i is used as statistical weight and not β_i^2 .

Example 9.3

The results of 5 measurements, x_i , with their statistical weights w_i are given in columns 2 and 3 of the table below. Find the weighted mean of the results and its standard deviation.

| i | x_i | w_i | β_i | $\beta_i x_i$ | $x_i - \bar{x}$ | $(x_i - \bar{x})^2$ | $\beta_i (x_i - \bar{x})^2$ |
|------|-------|-------|-----------|---------------|-----------------|---------------------|-----------------------------|
| 1 | 5.05 | 2 | 0.182 | 0.919 | -0.074 | 0.00548 | 0.000997 |
| 2 | 5.25 | 1 | 0.091 | 0.478 | 0.126 | 0.01588 | 0.001445 |
| 3 | 5.16 | 3 | 0.273 | 1.409 | 0.036 | 0.00130 | 0.000355 |
| 4 | 5.09 | 4 | 0.363 | 1.848 | -0.034 | 0.00116 | 0.000420 |
| 5 | 5.17 | 1 | 0.091 | 0.470 | 0.046 | 0.00212 | 0.000193 |
| Sums | | 11 | 1 | 5.124 | | | 0.003410 |

The weighted mean is $\bar{x} = \sum_{i=1}^N \beta_i x_i = 5.124$.

The weighted standard deviation of the measurements is

$$s_x = \sqrt{\sum_{i=1}^N \beta_i (x_i - \bar{x})^2} = \sqrt{0.003410} = 0.0584$$

and the weighted standard deviation of the mean is

$$\sigma_{\bar{x}} = \sqrt{\frac{1}{(N-1)} \sum_{i=1}^N \beta_i (x_i - \bar{x})^2} = \frac{s_x}{\sqrt{N-1}} = \frac{0.0584}{\sqrt{4}} = 0.0292.$$

The final result is:

$$x = 5.124 \pm 0.029.$$

Without weighting, these quantities are $\bar{x} = 5.144$, $s_x = 0.069$ and $\sigma_{\bar{x}} = 0.035$.

Example 9.4 [E]

Solve Example 9.3 using Excel®.

We will first evaluate the weighted mean. We enter the values of x_i and w_i in columns A and B, respectively. Highlight an empty cell, say E1. Left click on cell E1 and type:

$$=\text{SUMPRODUCT}(A1:A5; B1:B5)/\text{SUM}(B1:B5)$$

Pressing **ENTER** will return the number 5.123636 in cell E1. This is the required mean, $\bar{x} = 5.1236$ mm.

We will give this number the name **M**. To do this, we right click on cell E1. In the dialog box that opens, we select **Define Name...** and in the cell for **Name** we write **M**. Press **ENTER**.

We will now evaluate the weighted standard deviation. We first evaluate the terms $(x_i - \bar{x})^2$. We highlight cell C1 and type: $=(A1-M)^2$. Pressing **ENTER** returns the number 0.005422 in cell C1. To fill cells C1 to C5 with the values of $(x_i - \bar{x})^2$, we highlight cells C1-C5 and press **Fill > Down**.

To evaluate the standard deviation, we highlight an empty cell, say E2 and type

$$=\text{SQRT}(\text{SUMPRODUCT}(B1:B5; C1:C5)/\text{SUM}(B1:B5))$$

Pressing **ENTER** returns the number 0.058352.

The weighted standard deviation of the measurements is $s_x = \sqrt{\frac{1}{\sum_{i=1}^N w_i} \sum_{i=1}^N w_i (x_i - \bar{x})^2} =$
0.058352 and the weighted standard deviation of the mean is

$$\sigma_{\bar{x}} = \frac{s_x}{\sqrt{N-1}} = \frac{0.058352}{\sqrt{4}} = 0.029176.$$

The final result is:

$$x = 5.124 \pm 0.029.$$

Example 9.5 [O]

Solve Example 9.3 using Origin[®].

We enter x_i and w_i in columns A(X) and B(Y). Highlight column A by left-clicking on its label. Then

Statistics > Descriptive Statistics > Statistics on Columns > Open Dialog...

In the window that opens, in **Input Data, Range 1, Data Range**, column A is already selected. In **Weighting Range**, we select column B(Y).

In **Quantities**, we click **Mean** and **Standard Deviation**.

We open the window **Computation Control**. We select **Weight Method, Direct Weight** and **Variation Divisor of Moment, WS**. We press **OK**. The results are:

Mean = \bar{x} = 5.12364, Standard Deviation = s_x = 0.05835.

We calculate $\sigma_{\bar{x}} = \frac{s_x}{\sqrt{N-1}} = \frac{0.05835}{\sqrt{5-1}} = 0.02918$

Summarizing, \bar{x} = 5.124, s_x = 0.058 and $\sigma_{\bar{x}}$ = 0.029, in agreement with the results of Example 9.3.

Example 9.6 [P]

Solve Example 9.3 using Python.

```
from __future__ import division
import numpy as np
import math

# Enter the values given as the components of the vector x:
x = np.array([5.05, 5.25, 5.16, 5.09, 5.17])

# Enter the corresponding weights w of the x values:
w = np.array([2, 1, 3, 4, 1])

# Evaluation
N = len(x)
wmean = np.average(x, weights = w)
variance = np.average((x-wmean)**2, weights = w)
stdev = math.sqrt(variance)
```

```
# Presentation of the results:
print ("Number of values, N =", N)
print ("Weighted mean = ", wmean)
print ("Weighted standard deviation of the sample =", stdev)
print ("Weighted standard deviation of the mean =", stdev/math.sqrt(N-1))

# Results:
Number of values, N = 5
Weighted mean = 5.12363636364
Weighted standard deviation of the sample = 0.058352023766840865
Weighted standard deviation of the mean = 0.029176011883420432
```

Example 9.7 [R]

Solve Example 9.3 using R.

The vectors x and w have as their components the values of x and w , respectively:

```
> x <- c(5.05, 5.25, 5.16, 5.09, 5.17)
> w <- c(2, 1, 3, 4, 1)
```

The weighted mean is found as

```
> wmean = weighted.mean(x, w)
> wmean
[1] 5.123636
```

The variance of the sample, s_x^2 , is the weighted mean of the quantity $(x_i - \bar{x})^2$. This is found to be:

```
> variance = weighted.mean((x-wmean)^2, w)
> variance
[1] 0.003404959
```

The standard deviation of the sample is $s_x = \sqrt{s_x^2}$

```
> sqrt(variance)
[1] 0.05835202
```

and the standard deviation of the mean is $\sigma_{\bar{x}} = \frac{s_x}{\sqrt{N-1}} = \frac{0.0584}{\sqrt{4}} = 0.0292$.

Summarizing, $\bar{x} = 5.124$, $s_x = 0.058$ and $\sigma_{\bar{x}} = 0.029$, in agreement with the results of Example 9.3.

Let us suppose that we have N results of measurements, x_i , each with its weight w_i and that they can be grouped in a number of K classes each of which consists of measurements with the same values of x and w :

$$\underbrace{w_1x_1 + w_1x_1 + \dots + w_1x_1}_{k=1} + \underbrace{w_2x_2}_{k=2} + \underbrace{w_3x_3 + \dots + w_3x_3}_{k=3} + \dots + \underbrace{w_Kx_K + w_Kx_K + \dots + w_Kx_K}_{k=K} \tag{9.28}$$

It should be noted that any two x 's or any two w 's may be the same in two different groups but not *both* x and w may be the same in any two different groups. In such a case the two terms would be placed in the same group.

In this case, the numerator of Eq. (9.22) may be written as

$$\sum_{i=1}^N w_i x_i = n_1 w_1 x_1 + n_2 w_2 x_2 + \dots + n_K w_K x_K = \sum_{k=1}^K n_k w_k x_k. \tag{9.29}$$

Similarly,

$$\sum_{i=1}^N w_i = n_1 w_1 + n_2 w_2 + \dots + n_K w_K = \sum_{k=1}^K n_k w_k. \tag{9.30}$$

Therefore, the *weighted mean* is

$$\bar{x} = \frac{\sum_{k=1}^K n_k w_k x_k}{\sum_{k=1}^K n_k w_k} \tag{9.31}$$

It is seen that the product $n_k w_k$ replaces the weight w_i in estimating the weighted mean. In this sense it may be considered to be an *active weight*.

In a similar way, the *weighted sample standard deviation* is

$$s_x = \sqrt{\frac{\sum_{k=1}^K n_k w_k (x_k - \bar{x})^2}{\sum_{k=1}^K n_k w_k}} \tag{9.32}$$

The *weighted population standard deviation* is

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum_{k=1}^K n_k w_k (x_k - \bar{x})^2}{(N - 1) \sum_{k=1}^K n_k w_k}}, \tag{9.33}$$

where N is the number of x values with non-zero weight.

In a way similar to that of Eq. (9.23) we may define the *normalized statistical weights*

$$\beta_k \equiv \frac{n_k w_k}{\sum_{k=1}^K n_k w_k}, \quad (9.34)$$

for which it is true that

$$\sum_k \beta_k = 1, \quad (9.35)$$

Equations (9.31)–(9.33) reduce to Eqs. (9.25)–(9.27), respectively.

Example 9.8

The results of $N = 30$ measurements x_i , with their statistical weights w_i , are grouped in $K = 9$ classes as shown in the table below.

| | | | | | | | | | |
|-------|------|------|------|------|------|------|------|------|------|
| k | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| n_k | 2 | 1 | 4 | 7 | 6 | 3 | 4 | 1 | 2 |
| x_k | 4.60 | 4.70 | 4.80 | 4.90 | 5.00 | 5.10 | 5.20 | 5.30 | 5.40 |
| w_k | 3 | 2 | 4 | 4 | 3 | 4 | 1 | 2 | 2 |

Find the weighted mean of the results and its standard deviation.

We construct a table with the quantities k , n_k , x_k , w_k , $n_k w_k$, $n_k w_k x_k$, $(x_k - \bar{x})^2$ and $n_k w_k (x_k - \bar{x})^2$.

| k | n_k | x_k | w_k | $n_k w_k$ | $n_k w_k x_k$ | $(x_k - \bar{x})^2$ | $n_k w_k (x_k - \bar{x})^2$ |
|-------|-------|-------|-------|-----------|---------------|---------------------|-----------------------------|
| 1 | 2 | 4.6 | 3 | 6 | 27.6 | 0.1267 | 0.7602 |
| 2 | 1 | 4.7 | 2 | 2 | 9.40 | 0.0655 | 0.1310 |
| 3 | 4 | 4.8 | 4 | 16 | 76.8 | 0.0243 | 0.3888 |
| 4 | 7 | 4.9 | 4 | 28 | 137.2 | 0.0031 | 0.0868 |
| 5 | 6 | 5.0 | 3 | 18 | 90.0 | 0.0019 | 0.0342 |
| 6 | 3 | 5.1 | 4 | 12 | 61.2 | 0.0208 | 0.2496 |
| 7 | 4 | 5.2 | 1 | 4 | 20.8 | 0.0596 | 0.2384 |
| 8 | 1 | 5.3 | 2 | 2 | 10.6 | 0.1184 | 0.2368 |
| 9 | 2 | 5.4 | 2 | 4 | 21.6 | 0.1972 | 0.7888 |
| Sums: | 30 | | | 92 | 455.2 | | 2.9146 |

The weighted mean is $\bar{x} = \frac{\sum_{k=1}^K n_k w_k x_k}{\sum_{k=1}^K n_k w_k} = \frac{455.2}{92} = 4.948$.

The weighted standard deviation of the sample is

$$s_x = \sqrt{\frac{\sum_{k=1}^K n_k w_k (x_k - \bar{x})^2}{\sum_{k=1}^K n_k w_k}} = \sqrt{\frac{2.9146}{92}} = 0.1778.$$

The weighted standard deviation of the mean is $\sigma_{\bar{x}} = \frac{s_x}{\sqrt{N-1}}$, where $N = \sum_{k=1}^9 n_k = 30$. Therefore, $\sigma_{\bar{x}} = \frac{0.1778}{\sqrt{29}} = 0.0330$. The final result is: $x = 4.948 \pm 0.033$.

Example 9.9 [E]

Solve Example 9.8 using Excel®.

Comparing Eqs. (9.31) and (9.32) with Eqs. (9.22) and (9.26), it is obvious that this example is the same as Example 9.4 [E] if we replace w_i with $n_i w_i$. We enter the values of n_i , x_i and w_i in columns A, B and C, respectively. We calculate the values of $n_i w_i$; In cell D1 we type = A1*C1 and press **ENTER**. We fill down to cell D9. Column D now contains the values of $n_i w_i$.

We will first evaluate the weighted mean. Highlight an empty cell, say E1. Left click on cell E1 and write:

$$=\text{SUMPRODUCT}(B1:B9; D1:D9)/\text{SUM}(D1:D9)$$

Pressing **ENTER** will return the number 4.9478 in cell E1. This is the required mean, $\bar{x} = 4.9478$ mm.

We will give this number the name **M**. To do this, we right click on cell E1. In the dialog box that opens, we select **Define Name...** and in the cell for **Name** we write **M**.

We will now evaluate the weighted standard deviation. We first evaluate the terms $(x_r - \bar{x})^2$. We highlight cell F1 and type: =(B1-M)^2. Pressing **ENTER** returns the number 0.120983 in cell F1. To fill cells F1 to F9 with the values of $(x_r - \bar{x})^2$, we highlight cells F1-F9 and press **Fill > Down**.

To evaluate the standard deviation, we highlight an empty cell, say G1 and type

$$=\text{SQRT}(\text{SUMPRODUCT}(D1:D9; F1:F9)/\text{SUM}(D1:D9))$$

Pressing **ENTER** returns the number 0.177836.

The weighted standard deviation of the measurements is $s_x =$

$$\sqrt{\frac{1}{\sum_{i=1}^N w_i} \sum_{i=1}^N w_i (x_i - \bar{x})^2} = 0.177836$$

and the weighted standard deviation of the

mean is $\sigma_{\bar{x}} = \frac{s_x}{\sqrt{N-1}}$, where $N = \sum_{k=1}^9 n_k = 30$. Therefore, $\sigma_{\bar{x}} = \frac{0.1778}{\sqrt{29}} = 0.0330$. The final result is: $x = 4.948 \pm 0.033$.

Example 9.10 [O]

Solve Example 9.8 using Origin[®].

We enter n_i , x_i and w_i in columns A(X), B(Y) and C(Y). Highlight column B by left-clicking on its label. Then

Statistics > Descriptive Statistics > Statistics on Columns > Open Dialog...

In the window that opens, in **Input Data, Range 1, Data Range**, column B is already selected. In **Weighting Range**, we select column C(Y).

In **Quantities**, we click **Mean and Standard Deviation**.

We open the window **Computation Control**. We select **Weight Method, Direct Weight and Variance Divisor of Moment, WS**. We press **OK**. The results are:

Weighted Mean $\bar{x} = 4.94783$, Weighted Standard Deviation of the Sample $s_x = 0.17784$.

We calculate the weighted standard deviation of the mean using the equation

$\sigma_{\bar{x}} = \frac{s_x}{\sqrt{N-1}}$, where $N = \sum_{k=1}^9 n_k = 30$. Therefore, $\sigma_{\bar{x}} = \frac{0.1778}{\sqrt{29}} = 0.0330$. The final result is: $x = 4.948 \pm 0.033$.

Example 9.11 [P]

Solve Example 9.8 using Python.

```
from __future__ import division
import numpy as np
import math

# Enter values of members of the groups:
n = np.array([2, 1, 4, 7, 6, 3, 4, 1, 2])

# Enter the values given as the components of the vector x:
x = np.array([4.6, 4.7, 4.8, 4.9, 5, 5.1, 5.2, 5.3, 5.4])

# Enter the corresponding weights w of the x values:
wt = np.array([3, 2, 4, 4, 3, 4, 1, 2, 2])

# "Active" weights:
w = n*wt
```

```
# Evaluation
G = len(x)
N = sum(n)
wmean = np.average(x, weights = w)
variance = np.average((x-wmean)**2, weights = w)
stdev = math.sqrt(variance)

# Presentation of the results
print ("Number of groups, G =", G)
print ("Number of measurements, N =", N)
print ("Weighted mean =", wmean)
print ("Weighted standard deviation of the sample =", stdev)
print ("Weighted standard deviation of the mean =", stdev/math.sqrt(N-1))

# Results:
Number of groups, G = 9
Number of measurements, N = 30
Weighted mean = 4.94782608696
Weighted standard deviation of the sample = 0.17783618553232663
Weighted standard deviation of the mean = 0.033023350612542336
```

Example 9.12 [R]

Solve Example 9.8 using R.

Comparing Eqs. (9.31) and (9.32) with Eqs. (9.22) and (9.26), it is obvious that this example is the same as Example 9.4 [E] if we use as weights the values $W_k = n_k w_k$.

| | | | | | | | | | |
|-----------------|------|------|------|------|------|------|------|------|------|
| k | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| n_k | 2 | 1 | 4 | 7 | 6 | 3 | 4 | 1 | 2 |
| x_k | 4.60 | 4.70 | 4.80 | 4.90 | 5.00 | 5.10 | 5.20 | 5.30 | 5.40 |
| w_k | 3 | 2 | 4 | 4 | 3 | 4 | 1 | 2 | 2 |
| $W_k = n_k w_k$ | 6 | 2 | 16 | 28 | 18 | 12 | 4 | 2 | 4 |

We define the vectors

```
> x <- c(4.60, 4.70, 4.80, 4.90, 5, 5.10, 5.20, 5.30, 5.40)
> W <- c(6, 2, 16, 28, 18, 12, 4, 2, 4)
>
```

and find the weighted mean

```
> Wmean = weighted.mean(x, W)
> Wmean
[1] 4.947826
>
```

The variance of the sample, s_x^2 , is the weighted mean of the quantity $(x_i - \bar{x})^2$. This is found to be:

```
> variance = weighted.mean((x-Wmean)^2, W)
> variance
[1] 0.03162571
>
```

The standard deviation of the sample is $s_x = \sqrt{s_x^2}$

```
> sqrt(variance)
[1] 0.1778362
```

We calculate the weighted standard deviation of the mean using the equation

$\sigma_{\bar{x}} = \frac{s_x}{\sqrt{N-1}}$, where $N = \sum_{k=1}^9 n_k = 30$. Therefore, $\sigma_{\bar{x}} = \frac{0.1778}{\sqrt{29}} = 0.0330$.

Summarizing the results: $\bar{x} = 4.948$, $s_x = 0.1778$ and $\sigma_{\bar{x}} = 0.033$.

In the case when the statistical weight of measurement x_i is, according to Eq. (9.21), equal to $w_i = \frac{1}{\sigma_i^2}$, then, from Eq. (9.22) we have for the weighted mean

$$\bar{x} = \frac{\sum_{i=1}^N x_i / \sigma_i^2}{\sum_{i=1}^N 1 / \sigma_i^2}, \quad (9.28)$$

for the weighted standard deviation of the measurements

$$s_x = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2 / \sigma_i^2}{\sum_{i=1}^N 1 / \sigma_i^2}} \quad (9.29)$$

and for the weighted standard deviation of the mean

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2 / \sigma_i^2}{(N - 1) \sum_{i=1}^N 1 / \sigma_i^2}} \tag{9.30}$$

The same is true when we have the mean values $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_r, \dots, \bar{x}_M$, of M different series of measurements, which have standard deviations $\sigma_{\bar{x}_1}, \sigma_{\bar{x}_2}, \dots, \sigma_{\bar{x}_r}, \dots, \sigma_{\bar{x}_M}$, respectively. In this case, the statistical weight of each mean is the inverse of the square of its standard deviation. The means have a (general) mean

$$\overline{(\bar{x})} = \frac{\sum_{r=1}^M \bar{x}_r / \sigma_{\bar{x}_r}^2}{\sum_{r=1}^M 1 / \sigma_{\bar{x}_r}^2} \tag{9.31}$$

while the standard deviation of this general mean is

$$\sigma_{\overline{(\bar{x})}} = \sigma\left(\overline{(\bar{x})}\right) = \sqrt{\frac{\sum_{r=1}^M [\bar{x}_r - \overline{(\bar{x})}]^2 / \sigma_{\bar{x}_r}^2}{\sum_{r=1}^M 1 / \sigma_{\bar{x}_r}^2}}. \tag{9.32}$$

Example 9.13

Three experiments for the determination of the speed of light in vacuum gave the following results, in m/s:

$$c_1 = 299\,792\,459.3 \pm 1.6, \quad c_2 = 299\,792\,457.82 \pm 0.86, \\ c_3 = 299\,792\,458.4 \pm 1.1.$$

Find the weighted mean of these results and its standard deviation, taking as weights the inverses of the square of the standard deviation in each case.

Since the numbers are given with many digits, to avoid loss of accuracy, we subtract from all of them the number $c_0 = 299\,792\,40$ m/s and work with the smaller numbers that remain, x .

| i | c_i (m/s) | x_i (m/s) | σ_i (m/s) | $1/\sigma_i^2$ (m/s) ² | β_i | $\beta_i x_i$ (m/s) | $c_i - \bar{c}$ (m/s) | $(c_i - \bar{c})^2$ | $\beta_i (c_i - \bar{c})^2$ (m/s) ² |
|------|----------------|----------------|---------------------|--------------------------------------|-----------|------------------------|--------------------------|---------------------|---|
| 1 | 299 792 459.3 | 9.3 | 1.60 | 0.3906 | 0.1520 | 1.414 | 1.068 | 1.141 | 0.1734 |
| 2 | 299 792 457.82 | 7.82 | 0.86 | 1.3521 | 0.5263 | 4.116 | -0.412 | 0.170 | 0.0895 |
| 3 | 299 792 458.4 | 8.4 | 1.10 | 0.8264 | 0.3217 | 2.702 | -0.168 | 0.0282 | 0.0091 |
| Sums | | | | 2.5691 | 1 | 8.232 | | | 0.272 |

The weighted mean of the results is $\bar{c} = c_0 + \bar{x} = c_0 + \sum_i \beta_i x_i = 299\,792\,450 + 8.232 = 299\,792\,458.232$ m/s.

From the sum $\sum_i \beta_i (c_i - \bar{c})^2 = 0.272$ (m/s)², we find that $s_c = \sqrt{\sum_i \beta_i (c_i - \bar{c})^2} = \sqrt{0.272} = 0.52$ m/s.

By Eq. 9.27, the standard deviation of the mean is $\sigma_{\bar{c}} = \frac{s_c}{\sqrt{N-1}} = \sqrt{\frac{0.272}{2}} = 0.37$ m/s.

Therefore, $c = 299\,792\,458.23 \pm 0.37$ m/s.

Example 9.14 [E]

Solve Example 9.13 using Excel[®].

| i | c_i (m/s) | x_i (m/s) | σ_i (m/s) |
|-----|----------------|-------------|------------------|
| 1 | 299 792 459.3 | 9.3 | 1.60 |
| 2 | 299 792 457.82 | 7.82 | 0.86 |
| 3 | 299 792 458.4 | 8.4 | 1.10 |

Acting as above, we subtract from all the values the quantity $c_0 = 299\,792\,450$ m/s and work with the smaller numbers that remain, x . We enter x_i and σ_i in cells A2-A4 and B2-B4, respectively. We will evaluate the weights to be used, $w_i = 1/\sigma_i^2$. We highlight cell C2 and type in it $=1/(B2)^2$. We **Fill Down** to cell C4. Column C now contains the values of w_i .

We will first evaluate the weighted mean. Highlight an empty cell, say E2. Left click on cell E2 and write:

$$=\text{SUMPRODUCT}(A2:A4; C2:C4)/\text{SUM}(C2:C4)$$

Pressing **ENTER** will return the number 8.2316 in cell E2. We will give this number the name **M**. To do this, we right click on cell E2. In the dialog box that opens, we select **Define Name...** and in the cell for **Name** we write **M**.

The weighted mean of the results is $\bar{c} = c_0 + \bar{x} = 299\,792\,450 + 8.2316 = 299\,792\,458.232$ m/s.

We will now evaluate the weighted standard deviation. We first evaluate the terms $(x_i - \bar{x})^2$. We highlight cell F2 and type: **=(A2-M)^2**. Pressing **ENTER** returns the number 0.141478 in cell F2. To fill cells F2 to F4 with the values of $(x_i - \bar{x})^2$, we highlight cells F2-F4 and press **Fill > Down**.

To evaluate the standard deviation, we highlight an empty cell, say G2 and type

$$=\text{SQRT}(\text{SUMPRODUCT}(\text{C2:C4;F2:F4})/\text{SUM}(\text{C2:C4}))$$

Pressing **ENTER** returns the number 0.5214. The weighted standard deviation of

the measurements is $s_c = \sqrt{\frac{1}{N} \sum_{i=1}^N w_i (x_i - \bar{x})^2} = 0.5214$. The standard deviation

of the mean is $\sigma_{\bar{c}} = \frac{s_c}{\sqrt{N-1}} = \frac{0.5214}{\sqrt{2}} = 0.3687$ m/s. Therefore, $c = 299\,792\,458.23 \pm 0.37$ m/s.

Example 9.15 [O]

Solve Example 9.13 using Origin®.

| <i>i</i> | <i>c_i</i> (m/s) | <i>x_i</i> (m/s) | <i>σ_i</i> (m/s) |
|----------|----------------------------|----------------------------|----------------------------|
| 1 | 299 792 459.3 | 9.3 | 1.60 |
| 2 | 299 792 457.82 | 7.82 | 0.86 |
| 3 | 299 792 458.4 | 8.4 | 1.10 |

Acting as above, we subtract from all the measurements the quantity $c_0 = 299\,792\,450$ m/s and work with the smaller numbers that remain, x . We enter x_i and σ_i in columns A and B, respectively.

Highlight column A and then: **Column > Set As > Y**

Highlight column B and then: **Column > Set As > Y Error**

Highlight columns A and B and then,

Statistics > Descriptive Statistics > Statistics on Columns > Open Dialog . .

In the window that opens,

Input > Input Data > Range 1 > Weighting Range > B(E)

Open the **Quantities** window and tick: **Mean, Standard Deviation**

Open the **Computation Control** window and **Weight Method > Instrumental**

The last choice sets the weight of each measurement x_i equal to $w_i = 1/\sigma_i^2$, where σ_i is the error in x_i . Then,

Variance Divisor of Moment > WS

The last choice sets the denominator of Eq. (9.29) equal to $w = \sum_i 1/\sigma_i^2$.

Pressing **OK** we obtain the results (for column A):

[Mean] = $\bar{x} = 8.2316$ m/s, [Standard Deviation] = $s_c = s_x = 0.52138$ m/s

By Eq. 9.27, the standard deviation of the mean is $\sigma_{\bar{c}} = \frac{s_c}{\sqrt{N-1}} = \frac{0.52138}{\sqrt{2}} = 0.36867$ m/s.

The final result is $c = 299\,792\,458.23 \pm 0.37$ m/s, in agreement with the results of Example 9.4.

Example 9.16 [P]

Three experiments for the determination of the speed of light in vacuum gave the following results, in m/s:

$$c_1 = 299\,792\,459.3 \pm 1.6, \quad c_2 = 299\,792\,457.82 \pm 0.86, \\ c_3 = 299\,792\,458.4 \pm 1.1.$$

Find the weighted mean of these results and its standard deviation, taking as weights the inverses of the square of the standard deviation in each case.

```
from __future__ import division
import numpy as np
import math

# Enter the values given as the components of the vector x:
x = np.array([299792459.3, 299792457.82, 299792458.4])

# Enter the values of the errors corresponding to the values of x:
s = np.array([1.6, 0.86, 1.1])

# Evaluation:
# Evaluate the corresponding weights w of the x values:
w = 1/(s*s)
N = len(x)
wmean = np.average(x, weights = w)
variance = np.average((x-wmean)**2, weights = w)
stdev = math.sqrt(variance)
```



```
# Presentation of the results:
print ("Number of values, N =", N)
print ("Weighted mean =", wmean)
print ("Weighted standard deviation of the mean =", stdev/math.sqrt(N-1))

# Results:
Number of values, N = 3
Weighted mean = 299792458.232
Weighted standard deviation of the mean = 0.36867082350704317
```

The final result is $c = 299\,792\,458.23 \pm 0.37$ m/s.

Example 9.17 [R]

Three experiments for the determination of the speed of light in vacuum gave the following results, in m/s:

$$c_1 = 299\,792\,459.3 \pm 1.6, \quad c_2 = 299\,792\,457.82 \pm 0.86, \\ c_3 = 299\,792\,458.4 \pm 1.1.$$

Find the weighted mean of these results and its standard deviation, taking as weights the inverses of the square of the standard deviation in each case.

| i | c_i (m/s) | x_i (m/s) | σ_i (m/s) |
|-----|----------------|-------------|------------------|
| 1 | 299 792 459.3 | 9.3 | 1.60 |
| 2 | 299 792 457.82 | 7.82 | 0.86 |
| 3 | 299 792 458.4 | 8.4 | 1.10 |

We form the vector x with the values of $x_i = c_i - c_0$ as components and s with the probable errors in c_i (or x_i).

```
> x <- c(9.3, 7.82, 8.4)
> s <- c(1.60, 0.86, 1.10)
> w <- c(1/s^2)
> w
[1] 0.3906250 1.3520822 0.8264463
```

The weighted mean of x is:

```
> weighted.mean(x, w)
[1] 8.2316
> wmean = weighted.mean(x, w)
```

The variance of the sample, s_x^2 , is the weighted mean of the quantity $(x_i - \bar{x})^2$, and the standard deviation of the sample is $s_x = \sqrt{s_x^2}$:

```

> variance <- weighted.mean((x-wmean)^2, w)
> variance
[1] 0.2718363
> stdev <- sqrt(variance)
> stdev
[1] 0.5213793

```

We calculate the weighted standard deviation of the mean using the equation $\sigma_{\bar{x}} = \frac{s_x}{\sqrt{N-1}}$. Therefore, $\sigma_{\bar{x}} = \frac{0.5213793}{\sqrt{2}} = 0.368671$. The final result is $c = 299\,792\,458.23 \pm 0.37$ m/s.

9.5 The Joint Probability Density for Two Random Variables

We will now examine the probability density of a random variable which is a function of two other random variables. To avoid confusion, we will adopt the following notation [9]:

A random variable is denoted by a bold letter, \mathbf{x} , and the values it takes by italic x . The probability density of the random variable \mathbf{x} is denoted by $f_{\mathbf{x}}(x)$.

The probability for the random variable \mathbf{x} to take a value which is smaller than or equal to x is denoted by $P\{\mathbf{x} \leq x\}$.

The distribution function of the random variable \mathbf{x} is denoted by $F_{\mathbf{x}}(x)$ and is equal to the probability for the random variable \mathbf{x} to take a value which is equal to or smaller than x . Obviously,

$$F_{\mathbf{x}}(x) = P\{\mathbf{x} \leq x\} \quad (9.33)$$

It is

$$f_{\mathbf{x}}(x) = \frac{dF_{\mathbf{x}}(x)}{dx}. \quad (9.34)$$

The probability for the random variable \mathbf{x} to have a value larger than x_1 and smaller than or equal to x_2 , where $x_1 < x_2$, is denoted by $P\{x_1 < \mathbf{x} \leq x_2\}$. Obviously, it is

$$P\{x_1 < \mathbf{x} \leq x_2\} = P\{\mathbf{x} \leq x_2\} - P\{\mathbf{x} \leq x_1\}. \quad (9.35)$$

The *joint or common probability density function* of the variables \mathbf{x} and \mathbf{y} , denoted by $f_{\mathbf{x},\mathbf{y}}(x, y)$, is such that the probability for the random variable \mathbf{x} to have a

value between x and $x + dx$ and the random variable \mathbf{y} to have a value between y and $y + dy$ is equal to $f_{\mathbf{x},\mathbf{y}}(x, y) dx dy$.

The function $f_{\mathbf{x},\mathbf{y}}(x, y)$ is said to be *normalized* if it is

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\mathbf{x},\mathbf{y}}(x, y) dx dy = 1. \tag{9.36}$$

The *joint or common distribution function* of the variables \mathbf{x} and \mathbf{y} is denoted by $F_{\mathbf{x},\mathbf{y}}(x, y)$ and is equal to the probability for the random variable \mathbf{x} to have a value smaller than or equal to x and the random variable \mathbf{y} to have a value smaller than or equal to y . It is

$$F_{\mathbf{x},\mathbf{y}}(x, y) \equiv P\{\mathbf{x} \leq x, \mathbf{y} \leq y\}. \tag{9.37}$$

The following relations are considered to be obvious [10]:

$$f_{\mathbf{x},\mathbf{y}}(x, y) = \frac{\partial^2 F_{\mathbf{x},\mathbf{y}}(x, y)}{\partial x \partial y} \tag{9.38}$$

$$f_{\mathbf{x}}(x) = \int_{-\infty}^{\infty} f_{\mathbf{x},\mathbf{y}}(x, y) dy \quad f_{\mathbf{y}}(y) = \int_{-\infty}^{\infty} f_{\mathbf{x},\mathbf{y}}(x, y) dx \tag{9.39}$$

$$F_{\mathbf{x}}(x) = F_{\mathbf{x},\mathbf{y}}(x, \infty) = \int_{-\infty}^{\infty} dy \int_{-\infty}^x f_{\mathbf{x},\mathbf{y}}(\chi, y) d\chi \tag{9.40}$$

$$F_{\mathbf{y}}(y) = F_{\mathbf{x},\mathbf{y}}(\infty, y) = \int_{-\infty}^{\infty} dx \int_{-\infty}^y f_{\mathbf{x},\mathbf{y}}(x, \psi) d\psi. \tag{9.41}$$

Let the random variable \mathbf{x} have probability density $f_{\mathbf{x}}(x)$ and the variable \mathbf{y} have probability density $f_{\mathbf{y}}(y)$. These are known as *marginal probability densities*. If the variables \mathbf{x} and \mathbf{y} are independent of each other, then the probability for the random variable \mathbf{x} to have a value between x and $x + dx$ and the random variable \mathbf{y} to have value between y and $y + dy$ is equal to the product of the two separate probabilities, i.e.

$$f_{\mathbf{x},\mathbf{y}}(x, y) dx dy = f_{\mathbf{x}}(x)f_{\mathbf{y}}(y) dx dy. \tag{9.42}$$

If \mathbf{x} and \mathbf{y} are normally distributed, with means and standard deviations μ_x, σ_x and μ_y, σ_y , respectively, then

$$f_{\mathbf{x},\mathbf{y}}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2}}. \tag{9.43}$$

This function of the two variables, x and y , has been drawn, in contour form, in Fig. 9.7.

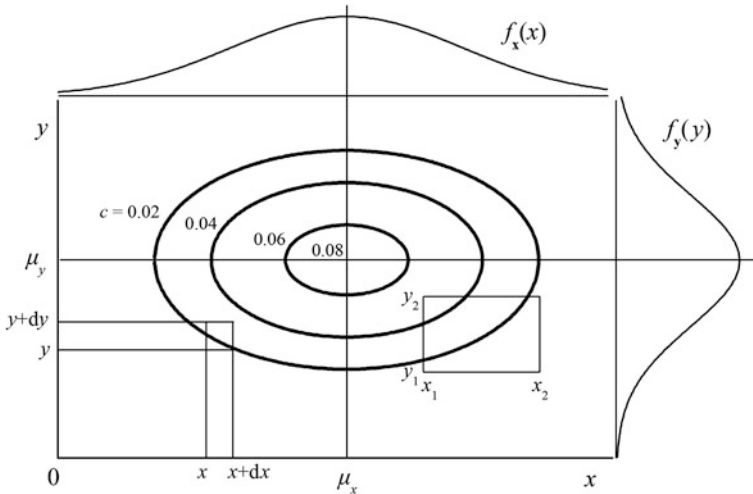


Fig. 9.7 The joint probability density function $f_{x,y}(x,y)$ for the normally distributed random variables \mathbf{x} and \mathbf{y} , which have means and standard deviations μ_x, σ_x and μ_y, σ_y , respectively. Three ellipses of constant $f_{x,y}(x,y) = c$ are shown in the figure. Also shown are the marginal probability densities $f_x(x)$ and $f_y(y)$

Drawn in the figure are:

1. Curves of constant values $f_{x,y}(x,y) = c$, which are ellipses with center at the point (μ_x, μ_y) . Putting $f_{x,y}(x,y) = c$ in Eq. (9.43), we find the equations of these ellipses to be

$$\frac{(x - \mu_x)^2}{\sigma_x^2} + \frac{(y - \mu_y)^2}{\sigma_y^2} = -2 \ln(2\pi\sigma_x\sigma_y c). \tag{9.44}$$

2. The marginal probability densities

$$f_x(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \quad \text{and} \quad f_y(y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}}, \tag{9.45}$$

situated at the upper part of the figure and the right-hand side, respectively.

3. The surface element $dx dy$ used in the evaluation of the probability $f_{x,y}(x,y) dx dy$ for the random variable \mathbf{x} to have a value between x and $x + dx$ and the random variable \mathbf{y} to have a value between y and $y + dy$.
4. The parallelogram lying between the values of x_1 and x_2 , and y_1 and y_2 , over which the function $f_{x,y}(x,y)$ must be integrated for the evaluation of the probability for the random variable \mathbf{x} to have a value greater than x_1 and smaller than

or equal to x_2 and the random variable \mathbf{y} to have a value greater than y_1 and smaller than or equal to y_2 . This probability is equal to

$$P\{x_1 < \mathbf{x} \leq x_2, y_1 < \mathbf{y} \leq y_2\} = \frac{1}{2\pi\sigma_x\sigma_y} \int_{x_1}^{x_2} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} dx \int_{y_1}^{y_2} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}} dy. \quad (9.46)$$

Example 9.18

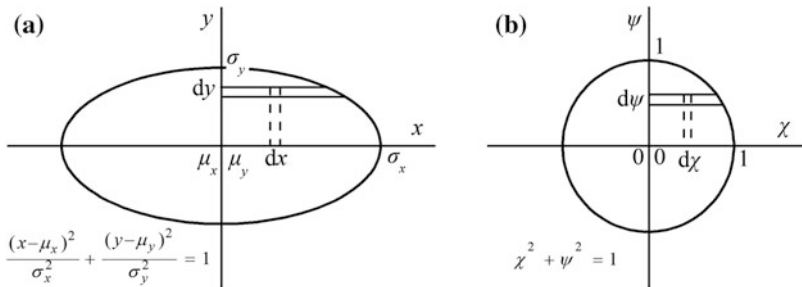
For the distribution of Eq. (9.43) and Fig. 9.7, find the probability that a point (x, y) lies within the ellipse with center the point (μ_x, μ_y) and semi-axes equal to σ_x and σ_y , along the respective axes.

The joint probability density function is $f_{\mathbf{x},\mathbf{y}}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2}}$. The probability that a point lies in the surface element $dx dy$ about the point (x, y) is

$$d^2P = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2}} dx dy.$$

The probability that a point (x, y) lies within the ellipse with center the point (μ_x, μ_y) and semi-axes equal to σ_x and σ_y , along the respective axes is found by integrating this over the surface of the ellipse, as shown in figure (a):

$$P_{\sigma_x, \sigma_y} = \frac{1}{2\pi\sigma_x\sigma_y} \iint_{\text{ellipse}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(y-\mu_y)^2}{2\sigma_y^2}} dx dy.$$



We change the variables to $\chi = \frac{x-\mu_x}{\sigma_x}$ and $\psi = \frac{y-\mu_y}{\sigma_y}$. Then

$$d^2P = \frac{1}{2\pi} e^{-\chi^2/2 - \psi^2/2} d\chi d\psi$$

and the ellipse transforms into the circle $\chi^2 + \psi^2 = 1$. The surface integral is now, as shown in figure (b),

$$P_{\sigma_x, \sigma_y} = \frac{1}{2\pi} \iint_{\text{circle}} e^{-\chi^2/2 - \psi^2/2} d\chi d\psi = 4 \times \frac{1}{2\pi} \int_0^1 e^{-\psi^2/2} d\psi \int_0^{\sqrt{1-\psi^2}} e^{-\chi^2/2} d\chi.$$

Since

$$\int_0^{\sqrt{1-\psi^2}} e^{-\chi^2/2} d\chi = \sqrt{\frac{2}{\pi}} \operatorname{erf} \left(\frac{\sqrt{1-\psi^2}}{\sqrt{2}} \right),$$

we have

$$P_{\sigma_x, \sigma_y} = \sqrt{\frac{2}{\pi}} \int_0^1 e^{-\psi^2/2} \operatorname{erf} \left(\frac{\sqrt{1-\psi^2}}{\sqrt{2}} \right) d\psi.$$

We could not find this integral in the tables, so we resorted to numerical integration. This gave:

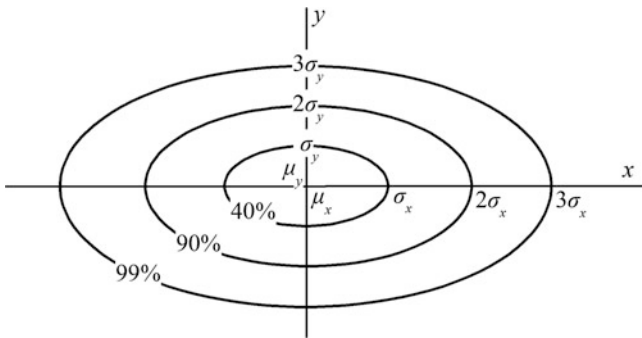
$$P_{\sigma_x, \sigma_y} = 0.394.$$

The probability that a point (x, y) lies within the ellipse with center the point (μ_x, μ_y) and semi-axes equal to σ_x and σ_y , along the respective axes is, therefore 39.4%.

For the $2\sigma_x, 2\sigma_y$ ellipse it is $P_{2\sigma_x, 2\sigma_y} = 0.865$.

For the $3\sigma_x, 3\sigma_y$ ellipse it is $P_{3\sigma_x, 3\sigma_y} = 0.989$.

This last result states that 99% of the points lie within the ellipse with center the point (μ_x, μ_y) and semi-axes equal to $3\sigma_x$ and $3\sigma_y$, along the respective axes. The percentages may be remembered as 40-90-99. These results are illustrated in the figure below.



9.6 The Probability Density of the Sum of Two Random Variables

Let the random variable $\mathbf{z} = \mathbf{x} + \mathbf{y}$ take values $z = x + y$, where the random variables \mathbf{x} and \mathbf{y} are normally distributed. The probability density $f_{\mathbf{z}}(z) = f_{\mathbf{x}+\mathbf{y}}(x + y)$ is required.

Figure 9.8 shows the probability densities $f_{\mathbf{x}}(x)$ and $f_{\mathbf{y}}(y)$ of \mathbf{x} and \mathbf{y} [(a) and (b) respectively]. In Fig. 9.8a, the area of the strip under the curve $f_{\mathbf{x}}(x)$ between x and $x + dx$ gives the probability for \mathbf{x} to have a value between x and $x + dx$. In Fig. 9.8b, the area of the region under the curve $f_{\mathbf{y}}(y)$ and in the region of values $-\infty < \mathbf{y} \leq y$, gives the probability for \mathbf{y} to have a value in the region $-\infty < \mathbf{y} \leq y$.

Due to the independence of the random variables \mathbf{x} and \mathbf{y} from each other, the probability for \mathbf{x} to have a value between x and $x + dx$ and \mathbf{y} to have a value smaller than or equal to y is

$$\begin{aligned}
 P\{x < \mathbf{x} \leq x + dx, \mathbf{y} \leq y\} &= P\{x < \mathbf{x} \leq x + dx\} P\{\mathbf{y} \leq y\} \\
 &= f_{\mathbf{x}}(x) dx F_{\mathbf{y}}(y) = f_{\mathbf{x}}(x) dx \int_{-\infty}^y f_{\mathbf{y}}(y) dy. \quad (9.47)
 \end{aligned}$$

If it is $z = x + y$, this is the probability for the random variable \mathbf{x} to have a value between x and $x + dx$ and the sum $\mathbf{z} = \mathbf{x} + \mathbf{y}$ to have a value smaller than or equal to z :

$$P\{x < \mathbf{x} \leq x + dx, \mathbf{z} \leq z\} = f_{\mathbf{x}}(x) dx \int_{-\infty}^{z-x} f_{\mathbf{y}}(y) dy. \quad (9.48)$$

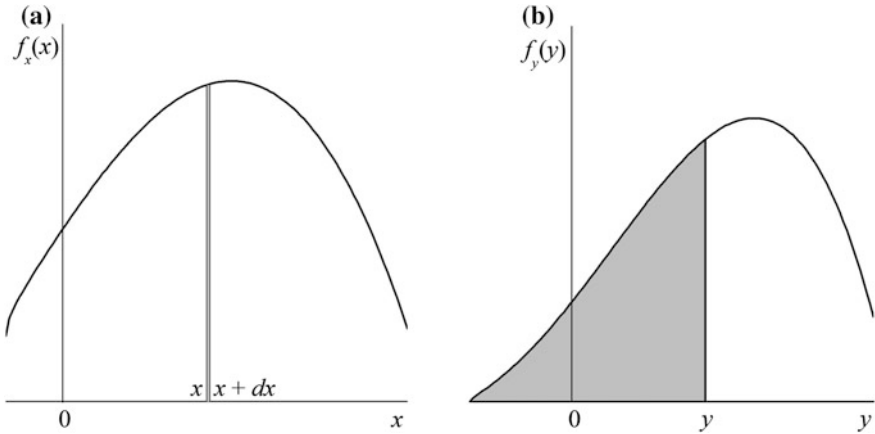


Fig. 9.8 The probability densities $f_x(x)$ and $f_y(y)$ of the random variables x and y

The probability for the random variable x to have *any* value and the sum $z = x + y$ to have a value which is smaller than or equal to z , i.e. the probability for z to have a value which is smaller than or equal to z , is

$$F_z(z) = \int_{-\infty}^{\infty} f_x(x) dx \int_{-\infty}^{z-x} f_y(y) dy. \tag{9.49}$$

The geometrical interpretation of this relation is seen in Fig. 9.9. The double integral of Eq. (9.49) gives the probability which corresponds to the integration of the function $f_x(x)f_y(y)$ over the shaded region and under the straight line $x + y = z$. In the shaded region, it is $x + y \leq z$. The magnitude $dx \int_{-\infty}^{z-x} f_y(y) dy$ is evaluated in the strip of Fig. 9.9 between x and $x + dx$. Integrating then for all the values of x , we cover all the shaded region of the figure ($x + y \leq z$) and we find $F_z(z)$.

The probability density for z is found by differentiating of $F_z(z)$ with respect to z :

$$f_z(z) = \frac{dF_z(z)}{dz} = \int_{-\infty}^{\infty} f_x(x) dx \left\{ \frac{\partial}{\partial z} \int_{-\infty}^{z-x} f_y(y) dy \right\} = \int_{-\infty}^{\infty} f_x(x) dx \{f_y(z - x)\}. \tag{9.50}$$

Therefore,

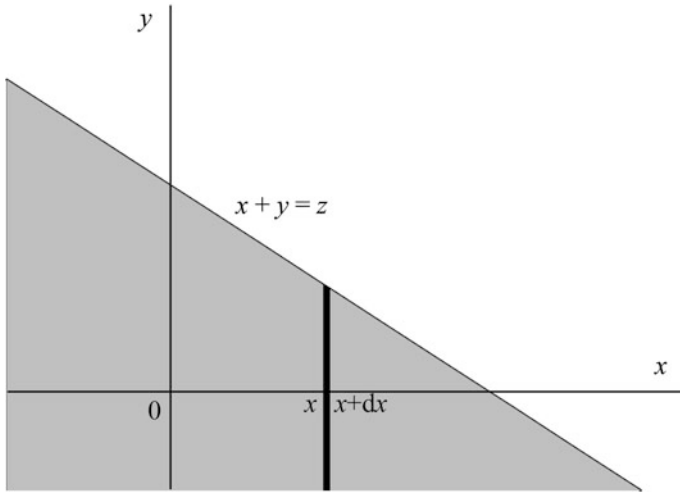


Fig. 9.9 The region of integration of the function $f_x(x)f_y(y)$ (shaded region) for the evaluation of the distribution function $F_z(z)$ of the sum $z = x + y$

$$f_z(z) = \int_{-\infty}^{\infty} f_x(x)f_y(z - x) dx \tag{9.51}$$

and, due to symmetry,

$$f_z(z) = \int_{-\infty}^{\infty} f_x(z - y)f_y(y) dy. \tag{9.52}$$

These integrals express the *convolution* of the functions $f_x(x)$ and $f_y(y)$, which is denoted by $f_x(x) * f_y(y)$ or, equivalently, by $f_y(y) * f_x(x)$.

If x and y take only positive values, then Eqs. (9.51) and (9.52) simplify to

$$f_z(z) = \int_0^z f_x(x)f_y(z - x) dx, \quad (z > 0) \tag{9.53}$$

and

$$f_{\mathbf{z}}(z) = \int_0^z f_{\mathbf{x}}(z-y)f_{\mathbf{y}}(y) dy, \quad (z > 0). \quad (9.54)$$

The proof is simple: writing Eq. (9.51) as

$$f_{\mathbf{z}}(z) = \int_{-\infty}^0 f_{\mathbf{x}}(x)f_{\mathbf{y}}(z-x) dx + \int_0^z f_{\mathbf{x}}(x)f_{\mathbf{y}}(z-x) dx + \int_z^{\infty} f_{\mathbf{x}}(x)f_{\mathbf{y}}(z-x) dx, \quad (9.55)$$

we see that the first integral is equal to zero because in the region of integration $(-\infty, 0)$ the function $f_{\mathbf{x}}(x)$ is equal to zero, while the third integral is also equal to zero because for $x > z$ the function $f_{\mathbf{y}}(z-x)$ is equal to zero. Only the second integral remains, which gives Eq. (9.53). Equation (9.54) is proved in the same way.

9.6.1 The Probability Density of the Sum of Two Normally Distributed Random Variables

Let \mathbf{x} and \mathbf{y} be two mutually independent random variables with means and standard deviations μ_x , σ_x and μ_y , σ_y , respectively. Then, it will be

$$f_{\mathbf{x}}(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \quad \text{and} \quad f_{\mathbf{y}}(y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}} \quad (9.56)$$

and the probability density for their sum, $\mathbf{z} = \mathbf{x} + \mathbf{y}$, will be, according to Eq. (9.53),

$$f_{\mathbf{z}}(z) = \frac{1}{2\pi\sigma_x\sigma_y} \int_{-\infty}^{\infty} \exp\left\{-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(z-x-\mu_y)^2}{2\sigma_y^2}\right\} dx \quad (9.57)$$

After some algebraic manipulation, the exponent may be written in the form:

$$\{\} = -\frac{(z-\mu_x-\mu_y)^2}{2(\sigma_x^2+\sigma_y^2)} - \frac{\left(x - \frac{\sigma_y^2\mu_x + \sigma_x^2(\mu_y-z)}{\sigma_x^2 + \sigma_y^2}\right)^2}{2\frac{\sigma_x^2\sigma_y^2}{\sigma_x^2 + \sigma_y^2}}. \quad (9.58)$$

Therefore,

$$f_z(z) = \frac{1}{2\pi \sigma_x \sigma_y} \exp \left[-\frac{(z - \mu_x - \mu_y)^2}{2(\sigma_x^2 + \sigma_y^2)} \right] \int_{-\infty}^{\infty} \exp \left[-\frac{\left(x - \frac{\sigma_y^2 \mu_x + \sigma_x^2 (\mu_y - z)}{\sigma_x^2 + \sigma_y^2} \right)^2}{2 \frac{\sigma_x^2 \sigma_y^2}{\sigma_x^2 + \sigma_y^2}} \right] dx. \tag{9.59}$$

The value of the integral is simply $\frac{\sqrt{2\pi\sigma_x\sigma_y}}{\sqrt{\sigma_x^2 + \sigma_y^2}}$. Thus, (9.59) becomes

$$f_z(z) = \frac{1}{\sqrt{2\pi} \sqrt{\sigma_x^2 + \sigma_y^2}} \exp \left[-\frac{(z - \mu_x - \mu_y)^2}{2(\sigma_x^2 + \sigma_y^2)} \right], \tag{9.60}$$

or

$$f_z(z) = \frac{1}{\sqrt{2\pi} \sigma_z} e^{-\frac{(z - \mu_z)^2}{2\sigma_z^2}}, \tag{9.61}$$

which is a normal (Gaussian) distribution with mean and standard deviation

$$\mu_z = \mu_x + \mu_y \quad \text{and} \quad \sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2}, \tag{9.62}$$

respectively.

This is the same result as that found in Sect. 6.2.2. The result may be generalized to more added terms and, obviously, \mathbf{x} and \mathbf{y} may also take negative values. The result is valid, therefore, for the algebraic sum of any number of normally distributed variables.

Example 9.19

The magnitude \mathbf{x} has a real value x_0 and a series of measurements of it have a mean value \bar{x} and standard deviation of the mean $\sigma_{\bar{x}}$. The magnitude \mathbf{y} , of the same nature as \mathbf{x} , has a real value y_0 and a series of measurements of it have a mean value \bar{y} and standard deviation of the mean $\sigma_{\bar{y}}$. What is the probability for y_0 to be greater than x_0 ?

The probability density of the *mean* \bar{z} of the difference $\mathbf{z} = \mathbf{y} - \mathbf{x}$ was found to be equal to

$$f_{\bar{z}}(\bar{z}) = \frac{1}{\sqrt{2\pi} \sigma_{\bar{z}}} e^{-\frac{(\bar{z} - \mu_{\bar{z}})^2}{2\sigma_{\bar{z}}^2}}, \quad \text{where} \quad \mu_{\bar{z}} = \bar{y} - \bar{x} \quad \text{and} \quad \sigma_{\bar{z}} = \sqrt{\sigma_{\bar{x}}^2 + \sigma_{\bar{y}}^2}.$$

The probability for the random variable \bar{z} to have a value greater than \bar{z} is (Sect. 4.4.2)

$$\Pr\{\bar{z} > \bar{z}\} = \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left(\frac{\bar{z} - \mu_{\bar{z}}}{\sqrt{2} \sigma_{\bar{z}}} \right) = \frac{1}{2} - \Phi \left(\frac{\bar{z} - \mu_{\bar{z}}}{\sigma_{\bar{z}}} \right).$$

The values x_0 and y_0 are the real values of the magnitudes \mathbf{x} and \mathbf{y} . We wish to find the probability for y_0 to be greater than x_0 . The best estimates we have for x_0 and y_0 are \bar{x} and \bar{y} , respectively. Therefore, the best estimate we can have for the probability for y_0 to be greater than x_0 , is equal to the probability for the value of the magnitude \bar{z} to be greater than 0,

$$\Pr\{y_0 > x_0\} = \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left(\frac{-\mu_{\bar{z}}}{\sqrt{2} \sigma_{\bar{z}}} \right) = \frac{1}{2} - \Phi \left(\frac{-\mu_{\bar{z}}}{\sigma_{\bar{z}}} \right)$$

and finally, since it is $\operatorname{erf}(-z) = -\operatorname{erf}(z)$ and $\Phi(-z) = 1 - \Phi(z)$,

$$\Pr\{y_0 > x_0\} = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left(\frac{\bar{y} - \bar{x}}{\sqrt{2} \sqrt{\sigma_x^2 + \sigma_y^2}} \right) = \frac{1}{2} + \Phi \left(\frac{\bar{y} - \bar{x}}{\sqrt{\sigma_x^2 + \sigma_y^2}} \right).$$

If it is $\bar{y} - \bar{x} = 3\sigma_{\bar{z}}$, then $\Pr\{y_0 > x_0\} = \frac{1}{2} + \Phi(3) = 0.999$.

If it is $\bar{y} - \bar{x} = 2\sigma_{\bar{z}}$, then $\Pr\{y_0 > x_0\} = \frac{1}{2} + \Phi(2) = 0.977$.

If it is $\bar{y} - \bar{x} = \sigma_{\bar{z}}$, then $\Pr\{y_0 > x_0\} = \frac{1}{2} + \Phi(1) = 0.84$.

If it is $\bar{y} - \bar{x} = 0$, then $\Pr\{y_0 > x_0\} = \frac{1}{2} + \Phi(0) = 0.5$.

Also,

If it is $\bar{y} - \bar{x} = -\sigma_{\bar{z}}$, then $\Pr\{y_0 > x_0\} = \frac{1}{2} - \Phi(1) = 0.16$.

If it is $\bar{y} - \bar{x} = -2\sigma_{\bar{z}}$, then $\Pr\{y_0 > x_0\} = \frac{1}{2} - \Phi(2) = 0.023$.

If it is $\bar{y} - \bar{x} = -3\sigma_{\bar{z}}$, then $\Pr\{y_0 > x_0\} = \frac{1}{2} - \Phi(3) = 0.001$.

The example that follows is rather extensive and may be omitted without any consequences in the understanding of what will follow. It is, however, useful, since it deals with many topics we have already discussed and uses the last theoretical results.

Example 9.20

Find the probability densities for the sums of n random values of x , in the case of the distribution of the measurements having a probability density $f_{\mathbf{x}}(x) = 0$ for $x < 0$ and $f_{\mathbf{x}}(x) = \alpha e^{-\alpha x}$ ($0 < \alpha$, $0 \leq x$) and check the validity of the central limit theorem.

The characteristics of the distribution

The probability density is normalized because $\alpha \int_0^\infty e^{-\alpha x} dx = 1$.

We will first find the mean and the standard deviation of x .

The mean is

$$\bar{x} = \alpha \int_0^\infty x e^{-\alpha x} dx = \frac{1}{\alpha}.$$

The standard deviation is

$$\sigma_x = \sqrt{\int_0^\infty \left(x - \frac{1}{\alpha}\right)^2 \alpha e^{-\alpha x} dx}$$

or

$$\begin{aligned} \sigma_x &= \sqrt{\alpha \int_0^\infty x^2 e^{-\alpha x} dx - \frac{2}{\alpha} \int_0^\infty x e^{-\alpha x} dx + \frac{1}{\alpha^2} \int_0^\infty e^{-\alpha x} dx} = \sqrt{\frac{1}{\alpha^2} - \frac{1}{\alpha^2} + \frac{1}{\alpha^2}} = \frac{1}{\alpha} \\ \sigma_x &= \frac{1}{\alpha}. \end{aligned}$$

The probability densities of the sum of n values of \mathbf{x}

We now denote as $\mathbf{z}_n = \mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_i + \dots + \mathbf{x}_n$ the sum of n values of the variable \mathbf{x} (e.g. measurements of the magnitude \mathbf{x}).

The probability density of single values $\mathbf{z}_1 = \mathbf{x}$ is the given function

$$f_{\mathbf{x}}(x) = \alpha e^{-\alpha x} \equiv f_{z_1}(z_1) \quad (0 \leq x).$$

According to Eq. (9.53) the probability density for the sum of two values of \mathbf{x} , is given by the convolution of $f_{z_1}(z_1)$ with itself:

$$f_{z_2}(z_2) = \int_0^{z_2} f_{z_1}(z_1) f_{z_1}(z_2 - z_1) dz_1.$$

Substituting, we find

$$f_{z_2}(z_2) = \int_0^{z_2} (\alpha e^{-\alpha z_1})(\alpha e^{-\alpha(z_2 - z_1)}) dz_1 = \alpha^2 e^{-\alpha z_2} \int_0^{z_2} dz_1 = \alpha^2 z_2 e^{-\alpha z_2}.$$

Knowing the probability density for the sum of two values of \mathbf{x} , we may find the probability density for the sum of three values of \mathbf{x} . This will be equal to the convolution of the probability density for the sum of two values with the probability density for the result of a measurement:

$$\begin{aligned}
 f_{z_3}(z_3) &= \int_0^{z_3} f_{z_2}(z_2) f_{z_1}(z_3 - z_2) dz_2 \\
 f_{z_3}(z_3) &= \int_0^{z_3} (\alpha^2 z_2 e^{-\alpha z_2}) (\alpha e^{-\alpha z_3 + \alpha z_2}) dz_2 \\
 &= \alpha^3 e^{-\alpha z_3} \int_0^{z_3} z_2 dz_2 = \alpha^3 \frac{z_3^2}{2!} e^{-\alpha z_3}.
 \end{aligned}$$

The general relation for the probability density $f_{z_{n+1}}(z_{n+1})$ is given by the convolution of the probability density for n measurements, $f_{z_n}(z_n)$, with the probability density for the result of a measurement, $f_{z_1}(z_1)$,

$$f_{z_{n+1}}(z_{n+1}) = \int_0^{z_{n+1}} f_{z_n}(z_n) f_{z_1}(z_{n+1} - z_n) dz_n.$$

We test the *assumption* that it is $f_{z_n}(z_n) = \alpha^n \frac{z_n^{n-1}}{(n-1)!} e^{-\alpha z_n}$. Substituting in the last relation, we find

$$\begin{aligned}
 f_{z_{n+1}}(z_{n+1}) &= \int_0^{z_{n+1}} \left(\alpha^n \frac{z_n^{n-1}}{(n-1)!} e^{-\alpha z_n} \right) (\alpha e^{-\alpha z_{n+1} + \alpha z_n}) dz_n \\
 f_{z_{n+1}}(z_{n+1}) &= \alpha^{n+1} \frac{e^{-\alpha z_{n+1}}}{(n-1)!} \int_0^{z_{n+1}} z_n^{n-1} dz_n = \alpha^{n+1} \frac{z_{n+1}^n}{n!} e^{-\alpha z_{n+1}},
 \end{aligned}$$

which is in agreement with the assumption we made for the form of $f_{z_n}(z_n)$. Since the formula $f_{z_n}(z_n) = \alpha^n \frac{z_n^{n-1}}{(n-1)!} e^{-\alpha z_n}$ gives the correct results for $n = 1, 2$ and 3 (which we already know) and, since, as we have just proved, if it is valid for $f_{z_n}(z_n)$ then it will be valid for $f_{z_{n+1}}(z_{n+1})$ also, we reach the conclusion that it is valid for all values of n .

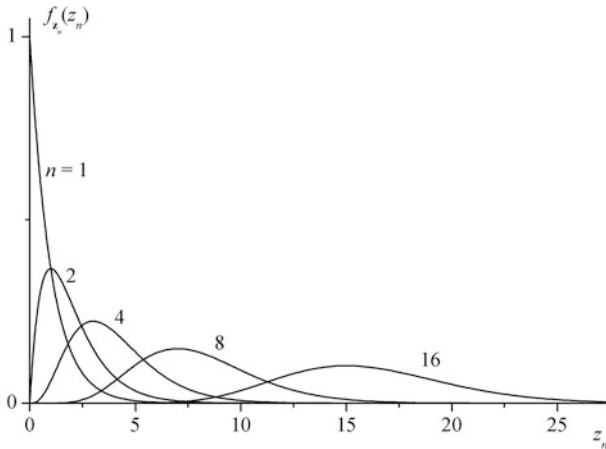
The probability densities for the sums $\mathbf{z}_n = \mathbf{x}_1 + \mathbf{x}_2 + \dots + \mathbf{x}_i + \dots + \mathbf{x}_n$ are, therefore,

$$f_{z_1}(z_1) \equiv f_x(x) = \alpha e^{-\alpha x}, \quad f_{z_2}(z_2) = \alpha^2 z_2 e^{-\alpha z_2}, \quad f_{z_3}(z_3) = \alpha^3 \frac{z_3^2}{2!} e^{-\alpha z_3}, \quad \dots$$

and are given by the general formula

$$f_{z_n}(z_n) = \alpha^n \frac{z_n^{n-1}}{(n-1)!} e^{-\alpha z_n}.$$

In the figure that follows, the curves $f_{z_n}(z_n)$ were drawn for $\alpha = 1$ and the values of $n = 1, 2, 4, 8$ and 16 .



The probability densities for the sum of a number of $n = 2, 4, 8$ or 16 values, taken at random from a parent population with probability density $f_z(z) = e^{-z} (z \geq 0)$

The probability densities of the means \bar{x} of n measurements of \mathbf{x} .

Knowing the probability densities $f_{z_n}(z_n)$ of the sums of n measurements of \mathbf{x} and if $\bar{x} = z_n/n$ is the mean of n measurements of \mathbf{x} , we wish to find the probability density $f_n(\bar{x})$ of the values \bar{x} .

The relation between the functions $f_{z_n}(z_n)$ and $f_n(\bar{x})$ is found as follows: Because it is

$$\begin{aligned} & \text{(Probability for the mean } \bar{x} \text{ of the } n \text{ measurements of } \mathbf{x} \text{ to lie between } \bar{x} \text{ and } \bar{x} + d\bar{x}) \\ & = \text{(Probability for the sum } z_n \text{ of the } n \text{ measurements of } \mathbf{x} \text{ to lie between } z_n \text{ and } z_n + dz_n) \end{aligned}$$

we have

$$f_n(\bar{x}) d\bar{x} = f_{z_n}(z_n) dz_n$$

and

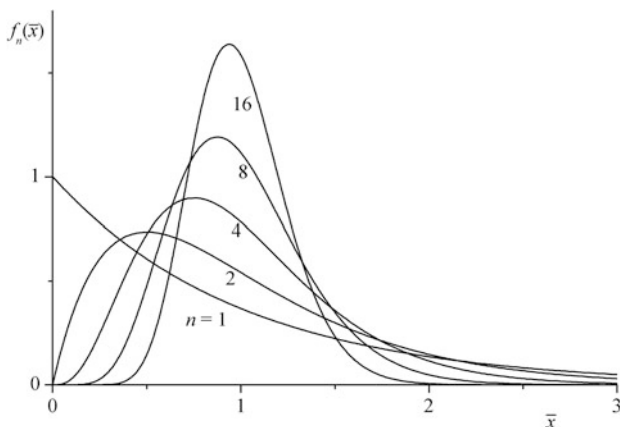
$$f_n(\bar{x}) = f_{z_n}(z_n) \left| \frac{dz_n}{d\bar{x}} \right|,$$

where the absolute value is taken as, by definition, the probability densities are positive.

Taking into account the fact that $\bar{x} = z_n/n$, we have, $\frac{dz_n}{d\bar{x}} = n$, $f_n(\bar{x}) = n f_{z_n}(z_n)$ and so the probability densities of the means \bar{x} of n measurements of \mathbf{x} is:

$$f_n(\bar{x}) = n\alpha^n \frac{(n\bar{x})^{n-1}}{(n-1)!} e^{-n\alpha\bar{x}} \quad \text{or} \quad f_n(\bar{x}) = \frac{(n\alpha)^n}{(n-1)!} \bar{x}^{n-1} e^{-n\alpha\bar{x}}.$$

These functions have been drawn in the figure that follows, for $\alpha = 1$ and $n = 1, 2, 4, 8$ and 16 .



The probability densities of the means of a number $n = 2, 4, 8$ or 16 values, which are taken at random from a parent population with probability density $f_z(z) = e^{-z}$ ($z \geq 0$)

Since

$$\begin{aligned} \int_0^\infty \frac{(n\alpha)^n}{(n-1)!} \bar{x}^{n-1} e^{-n\alpha\bar{x}} d\bar{x} &= \frac{1}{(n-1)!} \int_0^\infty (n\alpha\bar{x})^{n-1} e^{-n\alpha\bar{x}} d(n\alpha\bar{x}) \\ &= \frac{1}{(n-1)!} \int_0^\infty t^{n-1} e^{-t} dt = 1, \end{aligned}$$

the probability densities $f_n(\bar{x})$ are normalized, as expected.

The asymptotic approach to the Gaussian curve

The maximum of $f_n(\bar{x})$ appears at the value of \bar{x} for which it is $df_n(\bar{x})/d\bar{x} = 0$, i.e. for $\bar{x} = \frac{n-1}{n\alpha}$. Substituting in $f_n(\bar{x})$, we find its maximum value:

$$\hat{f}_n = \frac{(n-1)^{n-1}}{(n-1)!} \alpha n e^{-(n-1)}.$$

In terms of \hat{f}_n , we have $f_n(\bar{x}) = \hat{f}_n \left(e \frac{n\alpha\bar{x}}{n-1} \right)^{n-1} e^{-n\alpha\bar{x}}$.

Let

$$\alpha\bar{x} \equiv \frac{n-1}{n} + \delta,$$

where δ expresses, in units of $1/\alpha$, the distance along the \bar{x} -axis from the point $\bar{x} = \frac{n-1}{\alpha n}$ which corresponds to the curve's maximum. Then,

$$f_n(\bar{x}) = \hat{f}_n \left(1 + \frac{n\delta}{n-1}\right)^{n-1} e^{-n\delta}$$

Taking logarithms,

$$\ln(f_n(\bar{x})/\hat{f}_n) = (n-1) \ln\left(1 + \frac{n\delta}{n-1}\right) - n\delta.$$

For small values of δ ,

$$\begin{aligned} \ln(f_n(\bar{x})/\hat{f}_n) &= (n-1) \left[\frac{n}{n-1} \delta - \frac{1}{2} \left(\frac{n}{n-1}\right)^2 \delta^2 + \dots \right] - n\delta \\ \ln(f_n(\bar{x})/\hat{f}_n) &= -\frac{1}{2} \frac{n^2}{n-1} \delta^2 \end{aligned}$$

and, therefore,

$$f_n(\bar{x}) = \hat{f}_n e^{-\frac{n^2\delta^2}{2(n-1)}}.$$

Returning to \bar{x} via the relation $\alpha\bar{x} = \frac{n-1}{n} + \delta$, we have $f_n(\bar{x}) = \hat{f}_n \exp\left\{-\frac{(\bar{x} - \frac{n-1}{\alpha n})^2}{2\left(\frac{\sqrt{n-1}}{\alpha n}\right)^2}\right\}$.

For large n , Stirling's formula gives $\hat{f}_n \approx \frac{1}{\sqrt{2\pi}} \frac{\alpha n}{\sqrt{n-1}}$ and, therefore, it is

$$f_n(\bar{x}) = \frac{1}{\sqrt{2\pi}} \frac{\alpha n}{\sqrt{n-1}} \exp\left\{-\frac{(\bar{x} - \frac{n-1}{\alpha n})^2}{2\left(\frac{\sqrt{n-1}}{\alpha n}\right)^2}\right\},$$

which is a Gaussian with mean $\overline{(\bar{x})} = \frac{n-1}{\alpha n}$ and standard deviation $\sigma_{\bar{x}} = \frac{\sqrt{n-1}}{\alpha n}$.

We notice that $\sigma_{\bar{x}} = \frac{\sqrt{n-1}}{\alpha n} = \frac{1/\alpha}{\sqrt{n}} \sqrt{\frac{n-1}{n}} = \frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{n-1}{n}}$ and $\sigma_{\bar{x}} \rightarrow \frac{\sigma_x}{\sqrt{n}}$ as $n \rightarrow \infty$.

We see that the central limit theorem applies. In the figure it is seen that the curve for $n = 16$ is already very similar to a Gaussian.

Programs

Excel

Ch. 09. Excel—Weighted Mean and Standard Deviations

Origin

Ch. 09. Origin—Weighted Mean and Standard Deviations

Python

Ch. 09. Python—Weighted Mean and Standard Deviations

R

Ch. 09. R—Weighted Mean and Standard Deviations

Problems

9.1 **[E.O.P.R.]** The results of 8 measurements of a magnitude are

$$x_i : 5.24 \quad 5.42 \quad 5.20 \quad 5.00 \quad 5.15 \quad 5.32 \quad 5.24 \quad 5.37.$$

Find the mean \bar{x} of the measurements and its error, $\sigma_{\bar{x}}$,

- if equal weights are attributed to the results,
- if the weights given to the results are, respectively,

$$w_i \quad 2 \quad 1 \quad 1 \quad 3 \quad 3 \quad 2 \quad 1 \quad 2.$$

9.2 **[E.O.P.R.]** A series of 10 measurements of the quantity x gave the result $x_1 = 8.65 \pm 0.12$, while another series of 20 measurements of the same quantity gave $x_2 = 8.45 \pm 0.08$. Find the value of x and its error for the total of the 30 measurements, if to the two values are attributed the weights:

- equal to the number of measurements in each result and
- inversely proportional to the square of the error of each result.

9.3 The probability of observation of the discrete values x_1, x_2, \dots, x_N is proportional to

$$P = \alpha^N e^{-\alpha^2 \sum_i (x_i - \bar{x})^2}$$

where \bar{x} is the mean of the x 's. Find the value of α which maximizes P .

References

1. E. Whittaker, G. Robinson, *Calculus of observations* (Blackie and Son Ltd, London, 1929), p. 179
2. See, for example, Chandrasekhar, Stochastic problems in physics and astronomy, *Reviews of Modern Physics*, **15**, 1–89, 1943. Reprinted in N. Wax (ed.), *Selected Papers on Noise and Stochastic Processes* (Dover, 1954). In this article, which is of quite an advanced level, many problems are examined, among which the problem of particle diffusion in 1, 2 and 3 dimensions, with steps that are not constant but are characterized by any distribution function
3. H. Jeffreys, *Theory of Probability* (Clarendon Press, Oxford, 1948)
4. R.T. Birge, *Phys. Rev.* **40**, 207–227 (1932)
5. For an outline of the proof, see e.g. P.G. Hoel, *Introduction to Mathematical Statistics* (J. Wiley and Son Inc., New York, 1971). 4th ed., p. 359. For a complete proof, see W. Feller, *An Introduction to Probability Theory and its Applications* (J. Wiley and Son Inc., New York, 1968). 3rd ed. vol. 1, p. 244, vol. 2, p. 252
6. See, for example, Rand Corporation, *A Million Random Digits with 100 000 Normal Deviates* (The Free Press, 1955)
7. See, for example, C.A. Whitney, *Random Processes in Physical Systems* (J. Wiley and Sons, 1990) p. 20
8. See, for example, J. Arnt and C. Haenel, *π Unleashed* (Springer, 2000). This book is accompanied by a CD which contains, among other things, a table with the first 400 million digits of π !
9. Athanasios Papoulis and S. Unnikrishna Pillai, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill Book Company, 4th ed., 2002). Sects. 4.2 and 6.1
10. For proofs, see, for example A. Papoulis and S. Unnikrishna Pillai, *op. cit.*, Chap. 6