Chapter 3
A Thought Experiment

In order to describe certain processes and introduce some useful terms, we will examine in detail the results of a thought experiment. By the term 'experiment' we mean here a series of successive measurements of the same magnitude, under identical experimental conditions. To be specific, let us assume that we are using a certain instrument to measure the distance x between two points. We will assume that the *real* distance between the points is $x_0 = 100\,\text{mm}$, exactly. Obviously, this is the quantity we do not know and which we will try to determine with our measurements, with as small an error as possible. In the example to follow, the results of our measurements have been selected in such a way that they are subjected to some limitations, which will be taken into account in our arithmetic results to be given below, but which the reader is not required to know at this stage understanding these conditions is, after all, one of the main aims of this book! The 'results' of our measurements in this thought experiment were found using random numbers for the purpose of deciding what the numerical result of each one of these assumed measurements should be. The simulation of experiments using random numbers in order to decide what the outcome of a certain process (e.g. a measurement) should be, is one of the many applications of the so-called *Monte Carlo method*.

3.1 The Thought Experiment

Commencing our measurements, let us assume that the first measurement gives a result of $x_1 = 101.82\,\text{mm}$. This is the only measurement we have of x and is, therefore, the best estimate we can have for x_0. We have, however, no estimate of the amount by which this value is *possible* to differ from x_0. We repeat the measurement under exactly identical conditions which, let us assume, gives the result of $x_2 = 100.49\,\text{mm}$. The values of x_1 and x_2 differ from each other by 0.33 mm. The first thing we notice is that the reading error of our instrument, i.e. 0.01 mm, is

C. Christodoulides and G. Christodoulides, *Analysis and Presentation of Experimental Results*, Undergraduate Lecture Notes in Physics,
DOI 10.1007/978-3-319-53345-2_3

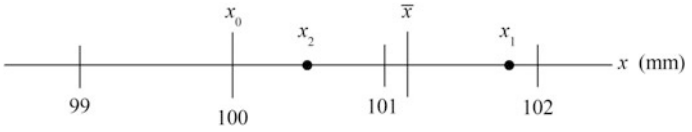


Fig. 3.1 The positions of the first 2 measurements of the magnitude x on the x-axis. Also shown is the real value $x_0 = 100\,\text{mm}$ of the magnitude being measured and the mean value $\overline{x} = 101.16\,\text{mm}$ of the 2 measurements

much smaller than needed, given that our measurements differ from each other by a difference of the order of 1 mm. This makes it clear to us that giving the results with two decimal digits is rather unjustified. Let us ignore this fact for the moment. Something else we can say is that, since the two measurements have a difference of the order of 1 mm, their difference from the real value x_0 is, most probably, of the same order of magnitude. The reader must have noticed that we continuously refer to possible estimates for the various values. This is natural and will be a characteristic of all our conclusions; we may, however, expect that, as we increase the number of our measurements, the confidence we will have in the various numerical values will increase. We mark the numerical results of our two measurements, x_1 and x_2, on the x-axis, as shown in Fig. 3.1.

Having only two measurements, which differ from each other, the only thing we can be almost certain of is that one of them will be nearer to the real value than the other, without knowing which one this value is. It is also certain that, if we take the *mean value* of these two measurements,

$$\overline{x} = \frac{1}{2}(x_1 + x_2), \tag{3.1}$$

this will differ from the real value x_0 by less than the worst one of the two measurements, whichever that might be. We can prove this as follows: The *error* in the measurement x_i is

$$e_i \equiv x_i - x_0. \tag{3.2}$$

For the two measurements it is $x_1 = x_0 + e_1$, $x_2 = x_0 + e_2$ and

$$\overline{x} = \frac{1}{2}(x_1 + x_2) = x_0 + \frac{1}{2}(e_1 + e_2). \tag{3.3}$$

The difference between the mean and the real value is

$$e_\mu = e_{\overline{x}} = \overline{x} - x_0 = \frac{1}{2}(e_1 + e_2) \tag{3.4}$$

which has a lower absolute value than the larger of the absolute values of e_1 and e_2. There is also the possibility for $\overline{x}$ to be nearer to x_0 than either of x_1 and x_2. For the

Table 3.1 Ten measurements of length

i	1	2	3	4	5	6	7	8	9	10
x_i (mm)	101.82	100.49	100.05	99.69	101.31	98.70	98.68	98.75	98.51	100.87

two values of Fig. 3.1 the mean is $\bar{x} = 101.16$ mm and this is also marked in the figure.

If we continue taking measurements, we may have a table of the successive results x_i, such as Table 3.1 in which the results of the first 10 measurements are given, in the order in which they were acquired.

The *mean* of N measurements x_i $(i = 1, 2, \ldots, N)$ is defined as

$$\bar{x} \equiv \frac{1}{N} \sum_{i=1}^{N} x_i. \tag{3.5}$$

For the 10 values of our example we have

$$\bar{x} = \frac{1}{10} \sum_{i=1}^{10} x_i = \frac{1}{10} (101.82 + 100.49 + 100.05 + 99.69 + 101.31 + 98.70 + 98.68 + 98.75 + 98.51 + 100.87) = 99.89 \text{ mm}.$$

As we will explain later, in the particular example we are justified in giving the results with only one decimal digit, so we write:

$$\bar{x} = 99.9 \text{ mm}.$$

In Fig. 3.2 these 10 measurements are marked on the x-axis just as was done in Fig. 3.1 for the first two.

The distribution of the results is better visualized, especially if their number is large, if they are plotted in a *histogram*. To construct a histogram, we follow the procedure described below (with reference to Fig. 3.2): we divide the x-axis into equal intervals of some width Δx, we measure the number ΔN of the measurements found in each interval and raise in each such bin a column with height proportional to ΔN. Figure 3.3 shows the histogram of the measurements of Table 3.1 and Fig. 3.2. A bin width of $\Delta x = 0.1$ mm was chosen. The centers of the intervals were

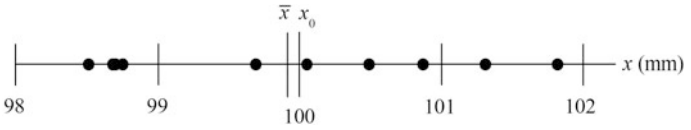


Fig. 3.2 The distribution on the x-axis of the results of the first 10 measurements of quantity x. The real value x_0 of the magnitude being measured is also shown, as well as the mean $\bar{x}$ of the 10 measurements

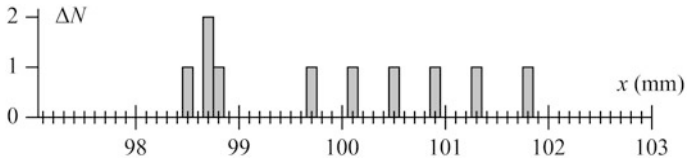


Fig. 3.3 A histogram of the 10 measurements x_i of Table 3.1. The bin width has been taken equal to $\Delta x = 0.1$ mm

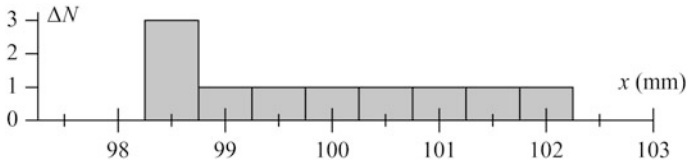


Fig. 3.4 A histogram of the 10 measurements x_i of Table 3.1. The bin width has been taken equal to $\Delta x = 0.5$ mm

taken to be those points of the x-axis which correspond to tenths of the unit. Thus, the first interval in which a measurement appears is the interval $98.45 < x_i \leq 98.55$. The next two measurements appear in the interval $98.65 < x_i \leq 98.75$ and so on.

If the histogram of Fig. 3.3 is redrawn with a bin width of $\Delta x = 0.5$ mm, we obtain the histogram of Fig. 3.4. The choice of the suitable bin width is dictated by the total number of measurements and their distribution. If the width is very small, one only measurement will appear in most bins, as in Fig. 3.3. At the other end, a Δx which is too large would result in almost all the measurements being classified in one bin. In both these extreme cases, the resulting histogram does not convey the maximum possible information about the distribution of the values. It is desirable to choose an intermediate value of Δx, so that the accumulation of events in one region is shown by columns of adequate height, but in such a way that the information about the detailed distribution of the measurements on the x-axis is not lost. These will become apparent in the histograms shown below.

The advantages of a histogram are apparent when the number of measurements is large. To demonstrate this fact, we suppose that we have performed 2500 measurements of the quantity x. The results x_i of the first 1000 measurements are recorded in Fig. 3.5 as a function of the order in which each measurement was performed, i. If we assume that the time interval between successive measurements is constant, the horizontal axis of the figure is that of time.

[For the reader who already has the necessary knowledge, we mention that the choice of the values was done, using random numbers, in such a way that the statistical distribution of the values of x_i is *Gaussian*, with mean $\bar{x} = 100$ mm and parent standard deviation $\sigma = 1$ mm. For reference, the values $\bar{x} - \sigma = 99$ mm, $\bar{x} = 100$ mm and $\bar{x} + \sigma = 101$ mm are marked in Fig. 3.5 by horizontal lines.

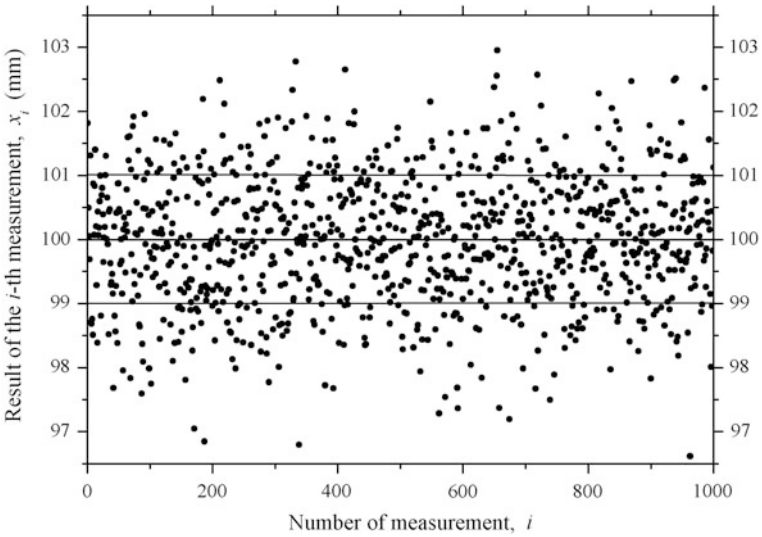


Fig. 3.5 1000 measurements of the magnitude x. The results x_i of the measurements are recorded as a function of the increasing number of the order, i, in which they were acquired. The mean of the measurements is 100 mm

The conditions under which real values behave in this way will be examined in a subsequent chapter.]

In Table 3.2 the first 100 measurements of Fig. 3.5 are given.

Histograms of the first 50, 100, 500 and 1000 measurements are shown in Figs. 3.6, 3.7, 3.8 and 3.9. In these histograms, the bin width was taken equal to $\Delta x = 0.2$ mm.

From Figs. 3.6, 3.7, 3.8 and 3.9, it becomes evident that, as the number of measurements increases, the nature of the statistical distribution of the measurements becomes clearer. Of course, in our example, the shape of the distribution was adopted when the results of the measurements were chosen. However, the same distribution applies for real measurements, for which we can only make assumptions about the statistical distribution of their parent population, which may be one of the required magnitudes to be derived from the series of measurements. In order to make the comparison of the histograms of Figs. 3.6, 3.7, 3.8 and 3.9 with each other easier, the total (shaded) area covered by the columns is approximately the same in all the figures. Consequently, the scales of the histograms may differ from each other, but in all cases the sum of the heights of all the columns is, equal to the total number of measurements N. If we plot all our 2500 'theoretical' measurements in a histogram, as in Fig. 3.10, we have the maximum possible information that can be extracted from the given experiment. Obviously, one has the ability to choose a small bin width Δx, in order to bring out a fine structure in the distribution, if it exists.

Table 3.2 100 measurements x_i of the magnitude x, in mm

101.82	100.49	100.05	99.69	101.31	98.70	98.68	98.75	98.51	100.87
100.07	100.83	101.40	100.20	99.29	98.39	100.43	100.07	100.18	100.60
100.99	98.81	100.69	100.80	100.39	99.64	100.03	100.19	101.00	101.30
100.01	100.65	99.92	98.52	100.44	100.47	99.63	99.29	99.33	99.46
99.15	97.68	100.20	98.80	98.56	99.27	100.41	98.38	100.87	100.87
100.47	100.43	99.62	100.00	99.66	99.15	97.95	100.16	100.34	100.03
99.48	100.24	100.78	100.12	101.60	101.23	100.41	101.61	97.83	101.09
99.76	99.03	101.77	101.92	99.51	99.30	100.10	100.97	100.57	100.07
99.12	101.39	98.67	98.64	99.44	100.15	97.59	98.37	98.09	100.48
100.93	101.96	99.92	100.48	99.56	101.04	101.24	100.38	97.98	99.80

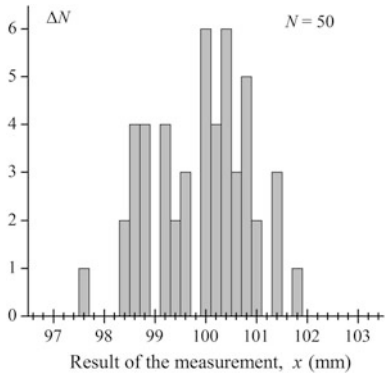


Fig. 3.6 A histogram of the first 50 measurements x_i of Fig. 3.5, with a bin width of $\Delta x = 0.2\,\text{mm}$

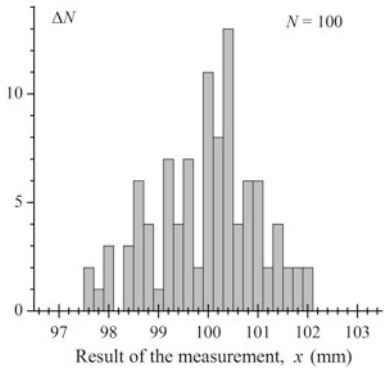


Fig. 3.7 A histogram of the first 100 measurements x_i of Fig. 3.5, with a bin width of $\Delta x = 0.2\,\text{mm}$

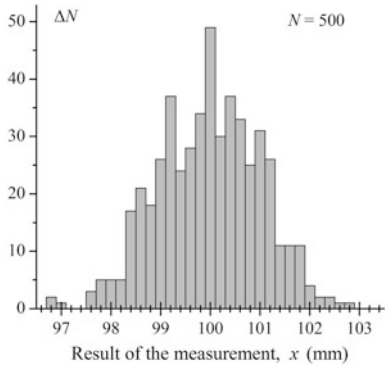


Fig. 3.8 A histogram of the first 500 measurements x_i of Fig. 3.5, with a bin width of $\Delta x = 0.2\,\text{mm}$

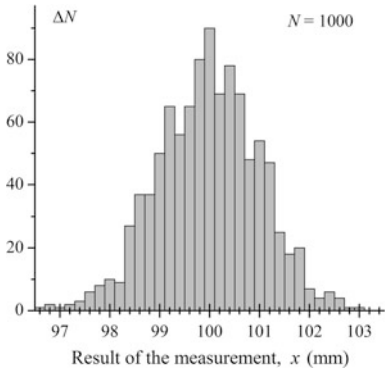


Fig. 3.9 A histogram of all the 1000 measurements x_i of Fig. 3.5, with a bin width of $\Delta x = 0.2\,\text{mm}$

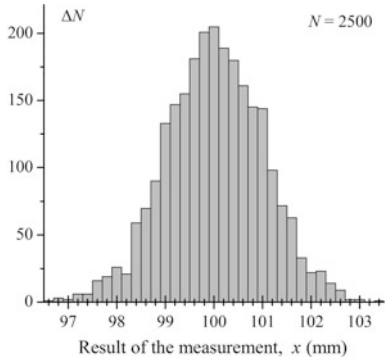


Fig. 3.10 A histogram of 2500 measurements x_i, with a bin width of $\Delta x = 0.2\,\text{mm}$

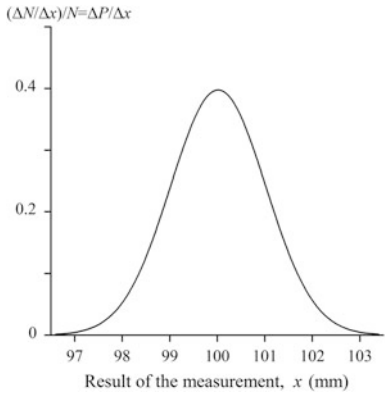


Fig. 3.11 The normalized 'histogram' of a very large number of measurements x_i

If it were possible to have a much larger number of measurements of the same quantity, our histogram would have the form of the curve in Fig. 3.11. In this histogram, the division into columns is not visible, due to their large number. The curve of Fig. 3.11 differs from the others in a few very important characteristics. Having a very large number of measurements at our disposal, we have the freedom to choose a very small interval Δx. This will have as a consequence the histogram to be transformed into a continuous curve. In fact, the general shape of the curve is not changed if we choose an even smaller Δx. Given these facts, we will choose the scale of ΔN in a different manner, so that the graph is easier to use. If for some value of Δx, we have between $x - \Delta x/2$ and $x + \Delta x/2$ a number of measurements equal to ΔN and the total number of measurements is N, then in the interval Δx around the value of x we have a proportion $\Delta N/N$ of the measurements. We may, therefore, say that the *probability* for a measurement to be found between the values $x - \Delta x/2$ and $x + \Delta x/2$ is

$$\Delta P(x) = \frac{\Delta N}{N}. \tag{3.6}$$

If we divide by Δx, we have

$$\frac{\Delta P}{\Delta x} = \frac{1}{N}\frac{\Delta N}{\Delta x} \tag{3.7}$$

as *the probability per unit x-interval* for a measurement to have a value near a certain x. This quantity has been plotted as a function of x in Fig. 3.11. The use of the curve is now much easier, being independent of N. For example, for $x = 100$ we read from the curve the value of $\frac{\Delta P}{\Delta x} = \frac{1}{N}\frac{\Delta N}{\Delta x} = 0.4$, approximately. Because it is $N = 2500$, if we take the interval $\Delta x = 0.2$ mm about $x = 100$, we will have $\Delta N = 0.4 \times 2500 \times 0.2 = 200$ as the number of measurements with values in the region $99.9 < x < 100.1$. This is in satisfactory agreement with the indication of Fig. 3.10 for $x = 100$.

In the limit $\Delta x \to 0$, this *probability density* is equal to

$$f(x) \equiv \frac{\mathrm{d}P}{\mathrm{d}x} = \frac{1}{N}\frac{\mathrm{d}N}{\mathrm{d}x}. \tag{3.8}$$

Plotting the curve is easily done as follows: we count the number $\mathrm{d}N$ of the measurements found in the small interval $\mathrm{d}x$ around the value x and divide by $\mathrm{d}x$ and by the total number N of the measurements. The result gives us $\mathrm{d}P/\mathrm{d}x$ and so we have the point $(x,\ \mathrm{d}P/\mathrm{d}x)$ of the curve. Figure 3.12 gives one such *probability density function* for the variable x. Its shape will not concern us at the moment. We will only mention that it has the general shape of the histograms we examined, with

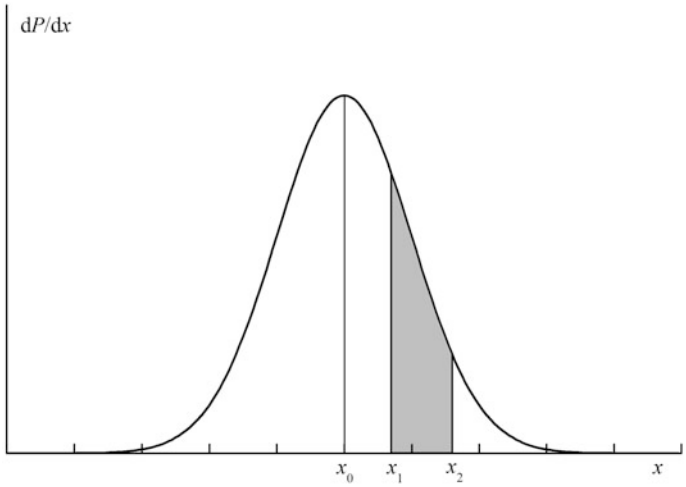


Fig. 3.12 A probability density function for the variable x

maximum at some value x_0 and most values of x being found in a region near the maximum.

The relation (3.8) may be used in the extraction of some conclusions concerning the properties of a curve such as that of Fig. 3.12. If the probability density $f(x) = \mathrm{d}P/\mathrm{d}x$ is multiplied by the width $\mathrm{d}x$ of an interval of x values, the product

$$f(x)\mathrm{d}x = \left(\frac{\mathrm{d}P}{\mathrm{d}x}\right)\mathrm{d}x = \mathrm{d}P \tag{3.9}$$

gives the probability for the variable to take a value in an interval of width $\mathrm{d}x$ about x, say between x and $x + \mathrm{d}x$. The probability for x to have a value between x_1 and x_2 ($x_2 > x_1$) is, therefore,

$$P\{x_1 \leq x \leq x_2\} = \int_{x_1}^{x_2} \mathrm{d}P = \int_{x_1}^{x_2} \left(\frac{\mathrm{d}P}{\mathrm{d}x}\right)\mathrm{d}x = \int_{x_1}^{x_2} f(x)\mathrm{d}x. \tag{3.10}$$

In Fig. 3.12, this probability is given by the area of the shaded surface under the curve between the points x_1 and x_2. The probability for x to have any value is, obviously, equal to unity and is given by the integral

$$P\{-\infty < x < +\infty\} = \int_{-\infty}^{+\infty} f(x)\mathrm{d}x = 1. \tag{3.11}$$

This is seen from

$$P\{-\infty < x < +\infty\} = \int_{-\infty}^{+\infty} \left(\frac{\mathrm{d}P}{\mathrm{d}x}\right)\mathrm{d}x = \int_{-\infty}^{+\infty} \frac{1}{N_{\mathrm{tot}}}\frac{\mathrm{d}N}{\mathrm{d}x}\mathrm{d}x = \frac{1}{N_{\mathrm{tot}}}\int_{x=-\infty}^{x=+\infty} \mathrm{d}N$$
$$= \frac{1}{N_{\mathrm{tot}}} N_{\mathrm{tot}} = 1. \tag{3.12}$$

Equation (3.11) simply states the fact that the probability density function $f(x)$ is, due to the way it was defined, *normalized*. The division by N in Eqs. (3.6) and (3.7) was done for exactly this purpose.

For the measurements we performed to be useful, certain conditions must be satisfied. The measurements must have been all made under identical conditions and the magnitude being measured must have remained unchanged during the duration of the measurements. Systematic errors must also be absent. This last condition is very difficult to verify and we must bear in mind that even in what appear to be the best of measurements, considerable systematic errors may be present.

If there are variations with time, in the magnitude being measured or the systematic errors that may be present, the series of measurements shown in Fig. 3.5 will be modified. Two such examples are shown in Figs. 3.13 and 3.14. In Fig. 3.13, apart from the random errors, there is also present in the measurements a

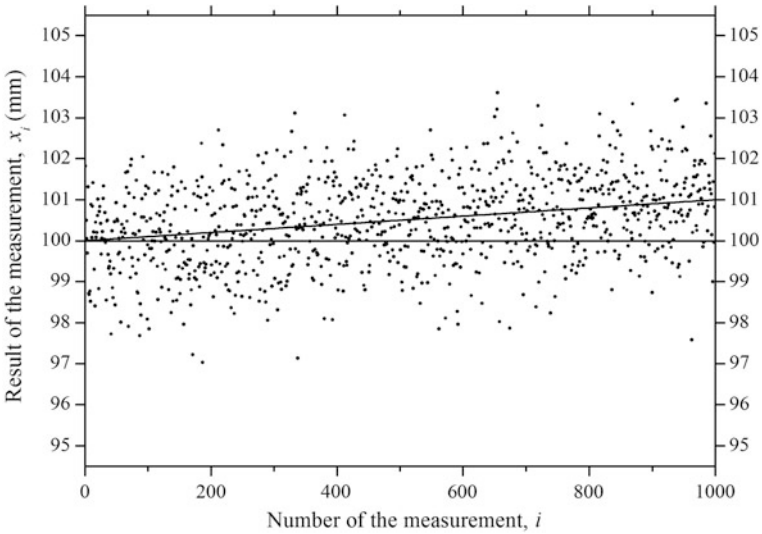


Fig. 3.13 The measurements of Fig. 3.5, to which a systematic error increasing linearly with time has been added

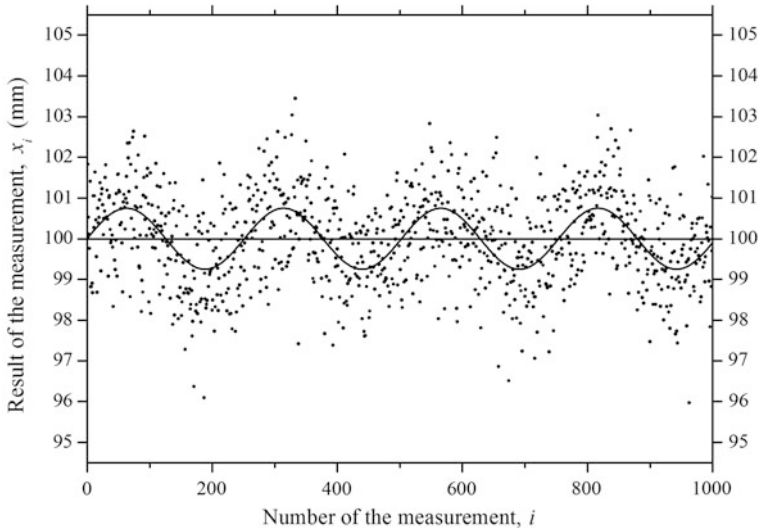


Fig. 3.14 The measurements of Fig. 3.5, to which a systematic error varying periodically with time has been added

systematic error which increases linearly with time. If the same time interval intervenes between successive measurements, the results will have the distribution shown in the figure. Something similar is true in the measurements of Fig. 3.14, with the difference that in this case the systematic error varies periodically with time. Situations like these are very difficult to detect, especially since the number of measurements is never as large as that of our examples.

We will now use the measurements made in order to derive an estimate for the value of the magnitude measured. Let the *real value* of the measured quantity be x_0, which of course is unknown to us. If the result of the i-th measurement is x_i, the *error* in this measurement is defined as

$$e_i \equiv x_i - x_0. \tag{3.13}$$

If a total of N measurements have been made, the *mean* of their results, x_i $(i = 1, 2, \ldots, N)$, is defined as

$$\overline{x} \equiv \frac{1}{N} \sum_{i=1}^{N} x_i. \tag{3.14}$$

Which is the behavior of the mean $\overline{x}$ as the number of measurements increases? To answer this question, we will examine the variation with the number of measurements of the difference between the mean and the real value, which is defined as

$$e_\mu \equiv \bar{x} - x_0. \tag{3.15}$$

Combining Eqs. (3.13) and (3.14), we have

$$e_\mu = \bar{x} - x_0 = \frac{1}{N}\sum_{i=1}^{N} x_i - x_0 = \frac{1}{N}\sum_{i=1}^{N} (x_0 + e_i) - x_0 = \frac{1}{N}\sum_{i=1}^{N} e_i. \tag{3.16}$$

The error in the mean,

$$e_\mu = \frac{1}{N}(e_1 + e_2 + \ldots + e_i + \ldots + e_N), \tag{3.17}$$

is, therefore, equal to the mean value of the errors in the measurements. Given that the errors e_i are assumed to be positive or negative with the same probability, the absolute value of e_μ will be smaller than the largest absolute value of the errors e_i. In general, therefore, the mean $\bar{x}$ will be nearer to x_0 than the worst measurement. Statistically, we expect that the mean value e_μ of the errors will be smaller the bigger the number of measurements N. Thus we accept that the mean $\bar{x}$ is the best estimate we have for x_0.

We will demonstrate what we have said above, using again the measurements of our thought experiment. Figure 3.15 shows the variation with n of the mean $\bar{x}(n)$

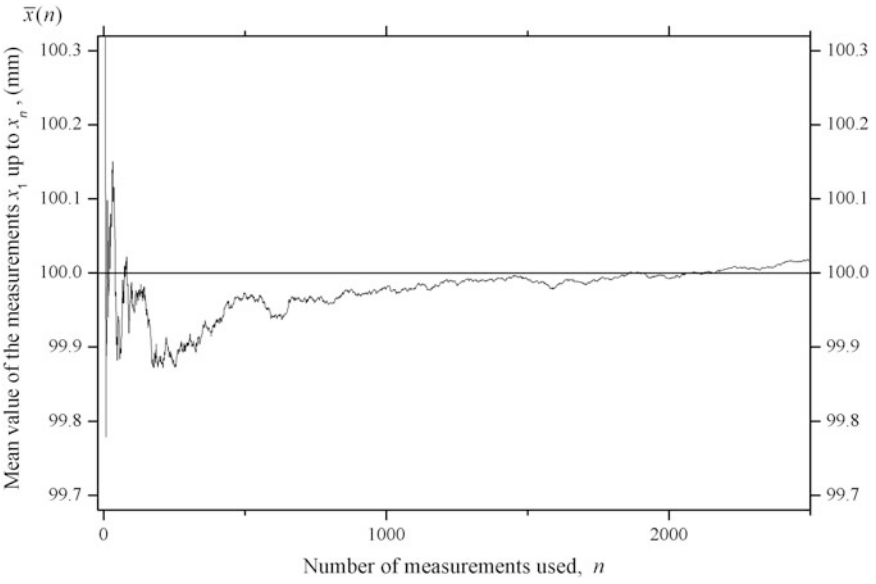


Fig. 3.15 The variation with n of the mean $\bar{x}(n)$ evaluated using the first n results x_i of the measurements of our thought experiment (*running average*). The real value of the magnitude being measured was taken to be $x_0 = 100$ mm

evaluated using the first n results x_i of the measurements of our thought experiment (a *running average*). The first value is, according to Table 3.1, $\overline{x}(1) = 101.82$ mm and this is the first point of the graph (which happens to be off scale). The next point is evaluated using the first two measurements and so on. As expected, it is seen in the figure that the mean has a general tendency to approach the real value (taken to be $x_0 = 100$ mm in the example) as the number of measurements increases. After the expected relatively large initial fluctuations, the variations in $\overline{x}(n)$ become smaller and smaller. This is natural, since for a large number of measurements, $\overline{x}(n)$ has acquired a kind of 'inertia', resisting change. This is due to the fact that the addition of one more measurement (with error e_i) to the n which are already present does not affect significantly neither the mean $\overline{x}(n)$ nor the error $e_\mu(n)$ of the mean.

Although these effects will be examined rigorously in the next chapter, we can give a short mathematical interpretation of what we stated above. By definition, it is

$$\overline{x}(n) = \frac{1}{n}(x_1 + x_2 + \ldots + x_i + \ldots + x_n). \tag{3.18}$$

Since

$$x_i = x_0 + e_i, \tag{3.19}$$

we have

$$\overline{x}(n) = x_0 + \frac{1}{n}(e_1 + e_2 + \ldots + e_i + \ldots + e_n) = x_0 + e_\mu(n) \tag{3.20}$$

where

$$e_\mu(n) = \frac{1}{n}(e_1 + e_2 + \ldots + e_i + \ldots + e_n) \tag{3.21}$$

is the error in $\overline{x}(n)$. Adding one more measurement will change $e_\mu(n)$ and, therefore $\overline{x}(n)$ as well. The error in the mean when $n+1$ values have been used in its evaluation will be

$$\begin{aligned} e_\mu(n+1) &= \frac{1}{n+1}(e_1 + e_2 + \ldots + e_i + \ldots + e_{n+1}) \\ &= \frac{1}{n+1}(e_1 + e_2 + \ldots + e_i + \ldots + e_n) + \frac{e_{n+1}}{n+1}. \end{aligned} \tag{3.22}$$

Therefore, it is

$$e_\mu(n+1) = \frac{n}{n+1} e_\mu(n) + \frac{e_{n+1}}{n+1}. \tag{3.23}$$

The fractional change in the error in the mean caused by the addition of one more measurement to the n already used is

$$\frac{\Delta e_\mu(n)}{e_\mu(n)} = \frac{e_\mu(n+1) - e_\mu(n)}{e_\mu(n)} = \frac{1}{n+1}\left(\frac{e_{n+1}}{e_\mu(n)} - 1\right). \tag{3.24}$$

The value of the expression in brackets is of order unity:

$$\left(\frac{e_{n+1}}{e_\mu(n)} - 1\right) \sim 1. \tag{3.25}$$

Therefore,

$$\frac{\Delta e_\mu(n)}{e_\mu(n)} \sim \frac{1}{n+1} \tag{3.26}$$

and

$$\Delta e_\mu(n) \sim \frac{e_\mu(n)}{n+1}. \tag{3.27}$$

The fractional change in the error in the mean, $e_\mu(n)$, caused by the addition of one more measurement in the evaluation of the mean, is of the order of $1/(n+1)$. As n increases, this change tends to zero. The change in $\overline{x}(n)$ also tends to zero and $\overline{x}(n) \to x_0$, something that is obvious in Fig. 3.15. We must remember that for the values of our numerical example the errors e_i in x_i are of the order of 1 mm (see Fig. 3.5). It will be proved below that, if the dispersion of the measurements is such that they differ from the real value by 1 mm on the average, then the difference of the mean from the real value will be, approximately, equal to 1 mm/$\sqrt{N}$, where N is the number of measurements in the calculation of the mean. In Fig. 3.15, for the maximum value $N = 2500$, the deviation of the mean from the real value is expected to be 1 mm/$\sqrt{2500} = 0.02$ mm, approximately. The deviations seen in the figure for N near 2500 are actually of this order of magnitude.

The variation with n of the standard deviation $s_x(n)$ from the mean $\overline{x}(n)$ evaluated using the first n results x_i of the measurements of our thought experiment is shown in Fig. 3.16. The standard deviation of the measurements is seen to tend to that of their parent population, which was chosen to be 1 mm in our thought experiment.

Qualitatively, we may make the following comments: as the number of the measurements is increased, the dispersion of the results about the mean does not vary drastically. This is seen in the histograms of Figs. 3.4, 3.6, 3.7, 3.8, 3.9 and 3.10, in which the width of the distribution remains virtually constant. This, in any case, depends on the distribution of the random errors, which in its turn depends on

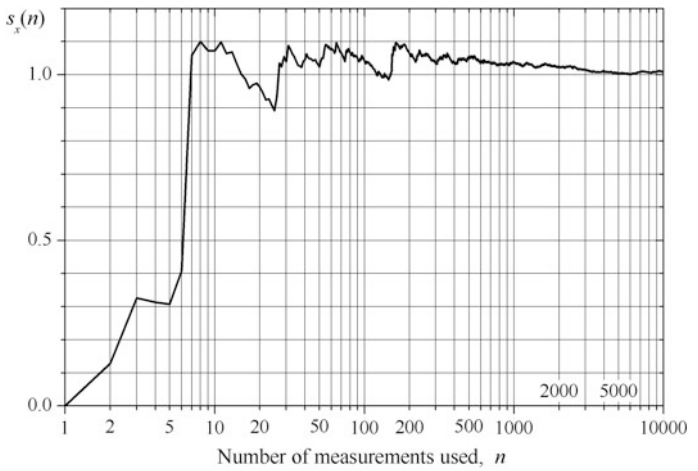
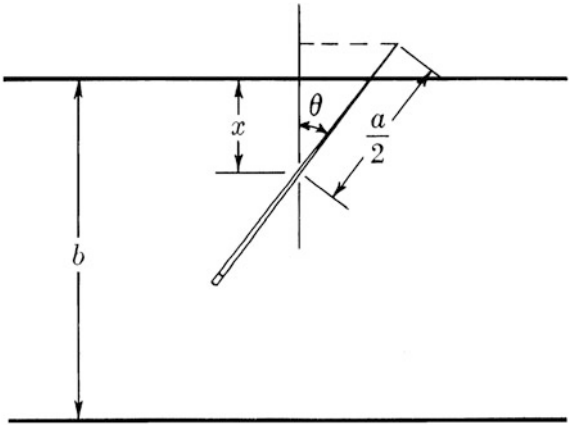


Fig. 3.16 The variation with n of the standard deviation $s_x(n)$ from the mean $\bar{x}(n)$ evaluated using the first n results x_i of the measurements of our thought experiment. The results of the measurements were selected so that $\sigma = 1$ mm. The scale for n is logarithmic

the accuracy of the method used. As the number of measurements increases, the shape of the distribution is defined with greater accuracy. This makes possible the more accurate determination of the mean. For the measurements of our thought experiment, in which negative and positive deviations of the measurements from the real value were taken to be equally probable, the determination of the mean actually depends on the determination of the axis of symmetry of the histogram of the distribution of the results. As the shape of the histogram is established with greater accuracy, the mean is known with smaller deviation from the real value. The quantitative demonstration of this effect will be given in the next chapter.

Example 3.1

Buffon's needle. For an experimental determination of the value of π using the method of Buffon's needle, many parallel straight lines are drawn on a horizontal surface, at distances b from each other. A needle having length equal to $a < b$ is dropped on the surface, many times, successively (see figure). It is proved that the probability that the needle will fall in such a position and orientation that it cuts one of the lines is $P = 2a/\pi b$. In a 'thought experiment', with $b = a$, Ukrainian *idiot savant* (learned idiot) Kerenii 'threw' the needle 355 times and observed that it cut the lines 226 times. The value for π he found by this method is $\pi = 2 \times 355/226 = 3.141\,592\,920$, which is correct in its first 7 significant figures. Discuss the 'experiment'.



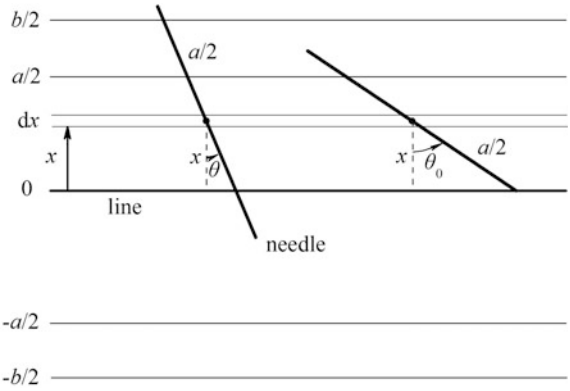
The Probability that the Needle Cuts a Line

Shown in the figure below is a line at $x = 0$. For the needle to be possible to cut the line, its center must fall at points for which it is $-a/2 \leq x \leq a/2$. The probability of this happening is $P_1 = a/b$. For a needle whose center falls in the range $-a/2 \leq x \leq a/2$, the probability of having its center fall in a strip between x and $x + \mathrm{d}x$ is $\mathrm{d}P_2 = \mathrm{d}x/a$. Let the angle the needle forms with the normal on the line be θ. Now, for a needle with its center in the strip between x and $x + \mathrm{d}x$ to cut the nearest line, it must form an angle θ in the range $-\theta_0 \leq \theta \leq \theta_0$ or $\pi/2 - \theta_0 \leq \theta \leq \pi/2 + \theta_0$, where $\cos\,\theta_0 = x/(a/2)$. The probability, therefore, that a needle whose center lies in the strip between x and $x + \mathrm{d}x$ cuts the line is $P_3 = 4\theta_0/2\pi = 2\theta_0/\pi$. The combined probability that the needle's center lies in the region $-a/2 \leq x \leq a/2$ and between x and $x + \mathrm{d}x$ and cuts the line is

$$\mathrm{d}P = P_1\,\mathrm{d}P_2\,P_3 = \frac{a}{b}\frac{\mathrm{d}x}{a}\frac{2\theta_0}{\pi} = \frac{2}{\pi b}\arccos(2x/a)\mathrm{d}x$$

The total probability for the needle to cut a line is

$$P = \frac{2}{\pi b}\int_{-a/2}^{a/2} \arccos\,(2x/a)\,\mathrm{d}x = \frac{a}{\pi b}\int_{-1}^{1} \arccos\,z\,\mathrm{d}z = \frac{2a}{\pi b}.$$



Discussion of the Experiment

To find out what the accuracy of the 'experiment' is, we will assume that the needle (for which it is $a = b$) is thrown one more time. There are two possibilities:

(a) If the needle cuts a line, the new value of π will be $\pi = 2 \times 356/227 = 3.136\,56 \approx 3.14$, which, even after being rounded, is correct to only 3 significant figures.

(b) If the needle does not cut a line, the new value of π will be $\pi = 2 \times 356/226 = 3.150\,44$, which is correct only to 2 significant figures.

Obviously, the great accuracy of the result is due to the fact that Kerenii 'happened to observe' 226 events in which the needle cut the lines in 355 tries and that he stopped the 'experiment' exactly at that point. The proximity of its value to the real one is, therefore, due to a coincidence. Besides, if we had no knowledge of the real value of π and we calculated the expected error $\delta\pi$ in π, this would have been much larger than the one achieved by Kerenii with the numbers he chose (see below).

It should be mentioned that the Italian mathematician Lazarini had estimated [1], before Kerenii, that with a needle of length 5 cm and a line distance of 6 cm, if the needle cuts the lines in 1808 out of 3408 trials, then it follows that

$$\pi = 2 \times \frac{5}{6} \times \frac{3408}{1808} = 3.1415929,$$

a value that is exact to 7 significant figures. Kerenii achieved the same result with a tenth of the trials.

The reliability of the value of π evaluated in this way may be found. From

$$P = \frac{2a}{\pi b} = \frac{N_1}{N}, \text{ it follows that } \pi = \frac{2a}{b}\frac{N}{N_1}.$$

If we assume that a, b and N are known with a satisfactory accuracy, then π may be considered to be a function of only one variable and its error, $\delta\pi$, will be due solely on the error in N_1. Therefore,

$$d\pi = \frac{2a}{b} N \left(-\frac{dN_1}{N_1^2} \right), \quad \frac{\delta\pi}{\pi} = -\frac{\delta N_1}{N_1}$$

if $\delta\pi$ is the error in π due to the error δN_1 in N_1. It will be proved later that if the expected number of events is N_1, the deviations of the measured values from N_1 is of the order of $\sqrt{N_1}$ (Poisson statistics). Then

$$\frac{\delta\pi}{\pi} = \frac{\delta N_1}{N_1} = \frac{\sqrt{N_1}}{N_1} = \frac{1}{\sqrt{N_1}}.$$

For the $N_1 = 226$ of the Kerenii 'experiment' we expect $\delta\pi/\pi = 1/\sqrt{N_1} = 1/\sqrt{226} = 0.067$ and $\delta\pi = 0.067 \times \pi = 0.21$ or 6.7%. For the $N_1 = 1808$ of the Lazarini 'experiment' we expect $\delta\pi/\pi = 1/\sqrt{N_1} = 1/\sqrt{1808} = 0.024$ and $\delta\pi = 0.024 \times \pi = 0.074$ or 2.4%. The accuracy which can be achieved by the method is not enough to justify the results of Kerenii and Lazarini, which, obviously are due to numerical coincidences.

From a purely experimental point of view, if we wish to determine π with an accuracy of 1 part in 3×10^6 using the method of Buffon's needle, we must measure the lengths a and b with at least this accuracy. We must also be able to judge that the needle cuts a line with such a power of discrimination, i.e. on the length scale of the order of one third of a millionth of a or b. If these lengths are about 3 cm, this resolving power must be better than $3 \times 10^{-2}/3 \times 10^6 = 10^{-8}$ m or 10 nm. This length is approximately equal to 100 atomic diameters. In addition, the thickness of the lines must be of the same order of magnitude, or smaller, and the lines must be straight to this accuracy. Anyone planning such an experiment must achieve experimental conditions such as these (in the real world and not the 'thought' world).

Reference

1. T.H. O'Beirne, *Puzzles and Paradoxes* (Oxford University Press, 1965)