CHAPTER IV

FURTHER DEVELOPMENTS

All that has been said so far about probability and probability distributions has established a foundation on which we can now build several techniques of great practical importance in handling experimental data. This chapter contains several very powerful tools which are developed from the principles we have learned so far.

12 | Standard Deviation of the Mean

We return to an important question which was raised in Sec. 3. This is: What is the relation between the standard deviation of a set of measurements and the precision of the *mean* of the set?

We answer this question by a straightforward extension of the ideas which have already been introduced. First, suppose that we take N measurements having random errors which follow the Gauss distribution. We calculate the mean and σ of this set of measurements. Now suppose we take *another* set of N measurements and calculate the mean and σ of this set. This mean will not

in general be exactly equal to the mean of the first set, although we expect intuitively that the difference of the means on the average will be considerably smaller than the difference between the individual measurements. The values of σ will be somewhat different also. Another way of saying the same thing is to say that the mean and variance of a *sample* of N observations are not in general equal to the mean and variance of the parent distribution. We continue this process until we have taken many sets, say M, each with its own mean and σ. We now ask: What is the *standard deviation of the means?* It is clear that this standard deviation provides an indication of how reliable any one of the means is.

To facilitate our calculation of the standard deviation of the means, for which we shall use the symbol σ_m, we introduce some new notation. We shall take M sets of measurements with N measurements in each set. There will then be MN readings in all. We use a Greek index μ to indicate which set of measurements we are talking about and i, as always, to designate a particular measurement within a set. Let

$x_{\mu i}$ = measurement i in set μ

$\bar{x}_\mu$ = mean of set μ

X = mean of all measurements

$d_{\mu i} = x_{\mu i} - X$ = deviation of $x_{\mu i}$

$D_\mu = \bar{x}_\mu - X$ = deviation of mean $\bar{x}_\mu$

The variance of the individual measurements is given by

$$\sigma^2 = \frac{1}{MN} \sum_{\mu=1}^{M} \sum_{i=1}^{N} d_{\mu i}{}^2 \tag{12.1}$$

The variance of the means is given by

$$\sigma_m{}^2 = \frac{1}{M} \sum_{\mu=1}^{M} D_\mu{}^2 \tag{12.2}$$

Now the deviations D_μ of the means can be expressed in terms of the deviations $d_{\mu i}$ of the individual observations, as follows:

$$D_\mu = \bar{x}_\mu - \bar{X} = \frac{1}{N} \sum_{i=1}^{N} x_{\mu i} - \bar{X}$$

$$= \frac{1}{N} \sum_{i=1}^{N} (x_{\mu i} - \bar{X}) = \frac{1}{N} \sum_{i=1}^{N} d_{\mu i} \tag{12.3}$$

Inserting Eq. (12.3) into Eq. (12.2), we obtain

$$\sigma_m{}^2 = \frac{1}{M} \sum_{\mu=1}^{M} \left(\frac{1}{N} \sum_{i=1}^{N} d_{\mu i} \right)^2 = \frac{1}{MN^2} \sum_{\mu=1}^{M} \left(\sum_{i=1}^{N} d_{\mu i} \right)^2 \tag{12.4}$$

Now let us squint for a moment at Eq. (12.4). The double sum at the right side of this equation, when evaluated, contains two different kinds of terms. There are terms in which one of the $d_{\mu i}$ is squared, and other terms containing products of two different $d_{\mu i}$. Now, because of the symmetry of the Gauss distribution function with respect to positive and negative deviations, the ds are as likely to be positive as negative. So in the limit, when we take a very large set of observations

94

MN, the products of two different ds will tend to cancel each other out. In Eq. (12.4) it is therefore legitimate to replace

$$\sum_{\mu=1}^{M} \left(\sum_{i=1}^{N} d_{\mu i} \right)^2 \qquad \text{by} \qquad \sum_{\mu=1}^{M} \sum_{i=1}^{N} d_{\mu i}{}^2$$

which contains only the d^2 terms. This argument for eliminating the cross terms is intended to be a plausibility argument rather than a rigorous one. It is quite possible but somewhat involved to put it on a more firm mathematical basis.

Equation (12.4) now becomes

$$\sigma_m{}^2 = \frac{1}{MN^2} \sum_{\mu=1}^{M} \sum_{i=1}^{N} d_{\mu i}{}^2 \tag{12.5}$$

This is closely related to Eq. (12.1). Combining Eq. (12.1) and Eq. (12.5), we obtain

$$\sigma_m{}^2 = \frac{\sigma^2}{N} \qquad \text{or} \qquad \sigma_m = \frac{\sigma}{\sqrt{N}} \tag{12.6}$$

The variance of the *mean* of a set of N measurements is simply the variance of the individual measurements divided by the number of measurements!

The standard deviation of the mean is used universally to describe the precision of the *mean* of the set of measurements; we now have available a method of *calculating* the standard deviation of the mean from the measurements themselves. In the surveyor's problem at the end of Sec. 9, for example, we find that the standard deviation of the mean is $0.026 \text{ ft}/\sqrt{10} = 0.008 \text{ ft}$. There

95

is about a 68% chance that the mean is within 0.008 ft of the mean of the parent distribution, which is presumably the true value of the quantity.

A word of caution about Eq. (12.6) is in order. It should not be thought that this equation is valid for measurements taken from *every* parent distribution. We have assumed that the cross terms in the sum used to find σ_m^2 are negligibly small. This is true for the Gauss distribution, as may be proved from a theorem known as the *central limit theorem*. Because experimental measurements so often obey the Gauss distribution, this is a useful formula. But it is quite possible to dream up strange distributions for which Eq. (12.6) is not true. For the Cauchy distribution, which will not be discussed here, σ_m^2 may be infinite!

13 | Propagation of Errors

We now return to the question raised in Sec. 2—the effect which errors in measurements have on the error of the result of a calculation which incorporates these measurements. We consider a quantity Q which is to be calculated from several observed quantities $a, b, c, \ldots$:

$$Q = f(a, b, c, \ldots) \tag{13.1}$$

Suppose that $a, b, c, \ldots$ are all measured N times. We can then calculate N different values of Q. We can also calculate the mean and variance for the set of measurements of a,

$$\sigma_a^2 = \frac{1}{N} \sum_{i=1}^{N} (\Delta a_i)^2 \tag{13.2}$$

where $\Delta a_i = a_i - \bar{a}$, and also the variance of Q,

$$\sigma_Q^2 = \frac{1}{N} \sum_{i=1}^{N} (\Delta Q_i)^2 \tag{13.3}$$

where $\bar{Q} = f(\bar{a}, \bar{b}, \ldots)$, $Q_i = f(a_i, b_i, \ldots)$, and $\Delta Q_i = Q_i - \bar{Q}$. The ΔQ_i can be approximated by the same methods used in Sec. 2, Eq. (2.8):

$$\Delta Q_i \cong \frac{\partial Q}{\partial a} \Delta a_i + \frac{\partial Q}{\partial b} \Delta b_i + \cdots \tag{13.4}$$

Inserting Eq. (13.4) into Eq. (13.3),

$$\sigma_Q^2 = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\partial Q}{\partial a} \Delta a_i + \frac{\partial Q}{\partial b} \Delta b_i + \cdots \right)^2 \tag{13.5}$$

When the quantity in the parentheses in Eq. (13.5) is squared, two kinds of terms appear. The first are squares, a typical one of which is

$$\left(\frac{\partial Q}{\partial a} \Delta a_i \right)^2$$

The other terms are cross terms of the form

$$\frac{\partial Q}{\partial a} \frac{\partial Q}{\partial b} \Delta a_i \Delta b_i$$

Now, we use exactly the same argument as used in Sec. 12 to obtain Eq. (12.5). The cross terms, since they contain quantities which are equally likely to be positive or negative, add up to very nearly zero, or at

least something very much smaller than the sum of the squared terms. We therefore drop them from the sum; the remaining terms are

$$\sigma_Q{}^2 = \frac{1}{N} \sum_{i=1}^{N} \left[\left(\frac{\partial Q}{\partial a}\right)^2 (\Delta a_i)^2 + \left(\frac{\partial Q}{\partial b}\right)^2 (\Delta b_i)^2 + \cdots \right]$$

(13.6)

This can be rewritten:

$$\sigma_Q{}^2 = \left(\frac{\partial Q}{\partial a}\right)^2 \frac{1}{N} \sum_{i=1}^{N} \Delta a_i{}^2 + \left(\frac{\partial Q}{\partial b}\right)^2 \frac{1}{N} \sum_{i=1}^{N} \Delta b_i{}^2 + \cdots$$

$$= \left(\frac{\partial Q}{\partial a}\right)^2 \sigma_a{}^2 + \left(\frac{\partial Q}{\partial b}\right)^2 \sigma_b{}^2 + \cdots \qquad (13.7)$$

This important result gives us a relation between the variances of the individual observations and the variance of the quantity Q calculated from these observations.

Usually, we are interested not in the variance of the individual observations, but in the variance of the *mean*. Assuming that the errors are normally distributed, we can convert Eq. (13.7) into one containing variances of the means by using Eq. (12.6). The result is

$$\sigma_{mQ}{}^2 = \left(\frac{\partial Q}{\partial a}\right)^2 \sigma_{ma}{}^2 + \left(\frac{\partial Q}{\partial b}\right)^2 \sigma_{mb}{}^2 + \cdots \qquad (13.8)$$

where $\sigma_{mQ}{}^2$ is the variance of the mean of Q, $\sigma_{ma}{}^2$ the variance of the mean of a, and so forth.

This is the result referred to at the end of Sec. 2; it is of much greater usefulness than Eq. (2.8) because it is the correct formula to use when the standard deviations

of the means of a, b, . . . are known. The corresponding formula for the fractional standard deviation of the mean, obtained by simply dividing Eq. (13.8) by Q^2, is

$$\left(\frac{\sigma_{mQ}}{Q}\right)^2 = \left(\frac{1}{Q}\frac{\partial Q}{\partial a}\right)^2 \sigma_{ma}{}^2 + \left(\frac{1}{Q}\frac{\partial Q}{\partial b}\right) \sigma_{mb}{}^2 + \cdots$$

(13.9)

A further remark needs to be made concerning Eq. (13.8), the truth of which is not exhibited clearly by the nonrigorous derivation which we have given. Equation (13.8) can be shown to be true even if *different numbers of observations* are made on the quantities a, b, c. So Eq. (13.8) actually has a much wider range of applicability than has been demonstrated. In the case of unequal numbers of observations, however, Eq. (13.7) must be modified, and the derivation of Eq. (13.8) is a little more involved. In what follows we shall make use of this more general validity of Eq. (13.8), although the proof has not been given here.

Here is an example of the foregoing analysis. Suppose the quantity Q is the area of a rectangle, whose dimensions are a and b; then $Q = ab$. Using Eq. (13.8), we find

$$\sigma_Q{}^2 = b^2 \sigma_a{}^2 + a^2 \sigma_b{}^2$$

(13.10)

In Eq. (13.10) and in the remainder of this section the subscript m is dropped from the standard deviations, but it is understood that each standard deviation is that of the *mean*, unless otherwise noted. Thus, σ_a is the standard deviation of the *mean* of a.

Introducing the *fractional* standard deviations of the means, σ_a/a, etc., we obtain

$$\left(\frac{\sigma_Q}{Q}\right)^2 = \left(\frac{\sigma_a}{a}\right)^2 + \left(\frac{\sigma_b}{b}\right)^2 \tag{13.11}$$

More generally, if $Q = a^m b^n$, then it is easy to show that

$$\left(\frac{\sigma_Q}{Q}\right)^2 = m^2\left(\frac{\sigma_a}{a}\right)^2 + n^2\left(\frac{\sigma_b}{b}\right)^2 \tag{13.12}$$

This Pythagorean sort of addition of fractional standard deviations makes them very convenient for practical calculations.

It is important to note the difference between Eq. (13.8) derived in this section and the much more naïve result, Eq. (2.8). If even crude estimates of the standard deviations of the means of the measurements are available, Eq. (13.8) *always* gives a more reliable estimate of the precision of the result than Eq. (2.8); therefore Eq. (13.8) should always be used in such cases. Only if the actual errors are known is Eq. (2.8) used.

Here is an example of the methods developed in this section. Suppose we have a horizontal beam of length l, supported at its ends and loaded in the center with a weight w. It can be shown that the deflection Y at the center of the beam is given by

$$Y = \frac{wl^3}{48EI}$$

where E is an elastic modulus and I is the moment of inertia of the cross section about its center of area.

Now it may happen that the characteristics of the

beam, E and I, are known very well, but that one has only crude measurements of w and l:

$$w = 100 \text{ tons} \pm 1 \text{ ton}$$
$$l = 50 \text{ ft} \pm 0.5 \text{ ft}$$

where the "$\pm$" in each case refers to the standard deviation of the mean. What is the resulting fractional standard deviation of Y? This is just the question answered by Eq. (13.12). We have

$$\left(\frac{\sigma_Y}{Y}\right)^2 = \left(\frac{\sigma_w}{w}\right)^2 + 3^2\left(\frac{\sigma_l}{l}\right)^2$$
$$= (0.01)^2 + 3^2(0.01)^2 = 0.001$$

$$\frac{\sigma_Y}{Y} = 0.032$$

Note that although w and l have equal fractional standard deviations, σ_l has a much more important effect because l appears to the third power. Also, the fractional standard deviation in Y is considerably larger than that in either w or l.

14 | Method of Least Squares

We now come to a very powerful method for obtaining the most reliable possible information from a set of experimental observations. We first state the *principle of least squares* for a set of measurements on one quantity, and then discuss how the principle can be derived from the *principle of maximum likelihood* if the errors follow the

Gauss distribution. In the next section we discuss applications of the principle to observations of more than one unknown quantity.

The principle of least squares can be stated as follows: The most probable value of a quantity is obtained from a set of measurements by choosing the value which minimizes the sum of the squares of the deviations of these measurements. For a set of measurements x_i the most probable value of x is that which minimizes the quantity

$$\sum_{i=1}^{N} (x - x_i)^2 \tag{14.1}$$

in which x is regarded as a variable which can be varied to obtain the minimum value of the function (14.1).

We note in passing that expression (14.1) is just N times the variance of the x_i, computed on the basis of the most probable value x. Thus an equivalent statement of the principle of least squares is: The most probable value of a quantity is that value which minimizes the variance (or alternately the standard deviation) of the measurements.

We know that the condition which must be satisfied for the function (14.1) to be a minimum is

$$\frac{d}{dx} \sum_{i=1}^{N} (x - x_i)^2 = 0 \tag{14.2}$$

This is a derivative of a sum of terms; we evaluate it by differentiating each term in turn:

$$\frac{d}{dx} \sum_{i=1}^{N} (x - x_i)^2 = \sum_{i=1}^{N} \frac{d}{dx}(x - x_i)^2$$

$$= \sum_{i=i}^{N} 2(x - x_i) = 2Nx - 2 \sum_{i=1}^{N} x_i \tag{14.3}$$

The condition which must be satisfied is therefore

$$2Nx - 2 \sum_{i=1}^{N} x_i = 0$$

or

$$x = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{14.4}$$

The proper value of x to use is just the average of the observations! This is the result which we *guessed* to be correct in Sec. 3.

Now, why should it be desirable to minimize the sum of the squares of the deviations? To answer this question, let us consider first the probability of occurrence of the set of measurements x_i which we obtained. Assuming that the measurements are distributed normally (according to the Gauss distribution), the probability of obtaining a measurement within an interval dx of x_i is

$$P_i = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-x_i)^2/2\sigma^2} dx \tag{14.5}$$

where σ characterizes the parent distribution from which x_i is obtained. The probability of obtaining the whole

set of N measurements is the *product* of the separate probabilities:

$$P = P_1 P_2 \cdots P_N$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-x_1)^2/2\sigma^2} dx \right)$$

$$\cdots \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-x_N)^2/2\sigma^2} dx \right)$$

$$= \left(\frac{dx}{\sigma\sqrt{2\pi}} \right)^N \exp\left[-\sum_{i=1}^{N} (x-x_i)^2/2\sigma^2 \right] \qquad (14.6)$$

Now the plot is beginning to thicken! The probability P of observing the whole set of measurements x_i depends upon the value of x, of course. If x is a number vastly different from any of the x_i, then the exponent in the last form of Eq. (14.6) is a very large negative quantity, and P will be very small. That is, it is very unlikely that we obtain a set of measurements *all* of which are very far from the true value of the quantity.

We now make a basic assumption, called the *principle of maximum likelihood;* we assume that the set of measurements which we obtain is actually *the most probable* set of measurements. According to this assumption, the proper value of x to choose is that which gives P the largest possible value. We want to *maximize* the probability of obtaining the particular set of measurements which we actually obtained. We then call the value of x so obtained the *most probable value of x.*

Clearly, the way to maximize P is to minimize the

value of the exponent in Eq. (14.6). We shall refer to the sum in this exponent as the *least-squares sum* and denote it by $M(x)$. Thus

$$M(x) = \sum \frac{(x_i - x)^2}{2\sigma^2} \qquad (14.7)$$

The principle of maximum likelihood thus leads to the conclusion that we should minimize $M(x)$, which is of course equivalent to minimizing $\sum (x_i - x)^2$, in accordance with our original statement.

To summarize what has been said so far: We have assumed that the best value of the observed quantity which we can obtain is the value which maximizes the probability of the set of observations which we have obtained, and we have called this the *most probable value.* If the observations are distributed normally, we maximize the probability by minimizing the sum of the squares of the deviations. For the case of observations on one quantity, this leads to the conclusion that the most probable value of the observed quantity is simply the arithmetic mean of the series of observations. Saying the same thing in slightly different language, we want to find the mean of the infinite parent distribution, which we regard as the true value of the quantity. The best estimate we can make of this mean is the mean of the sample of N measurements.

The standard deviation of the most probable value of x obtained above can be found easily by using the propagation of errors formula, Eq. (13.8). The quan-

tity x is regarded as a function of all the x_i, each of which has a standard deviation equal to that of the parent distribution, that is, σ. Therefore,

$$\sigma_m{}^2 = \left(\frac{\partial x}{\partial x_1}\right)^2 \sigma^2 + \left(\frac{\partial x}{\partial x_2}\right)^2 \sigma^2 + \cdots = \sum_{i=1}^{N} \left(\frac{\partial x}{\partial x_i}\right)^2 \sigma^2$$

(14.8)

From Eq. (14.4) we find

$$\frac{\partial x}{\partial x_i} = \frac{1}{N}$$

(14.9)

so

$$\sigma_m{}^2 = \sum \frac{\sigma^2}{N^2} = \frac{\sigma^2}{N}$$

(14.10)

This result should not be surprising; it is the same conclusion we reached in Sec. 12, Eq. (12.6), from a slightly different point of view. The difference between Eq. (14.10) and Eq. (12.6) is that Eq. (14.10) contains the variance of the infinite parent distribution, while Eq. (12.6) contains the variance of a sample of N measurements, which is used as an *estimate* of the variance of the parent distribution. The error of this estimate is thrown away when we discard the cross terms in Eq. (12.4).

The variance of the parent distribution is of course not known. All that can be done is to estimate it by computing the variance of the sample, and this is ordinarily sufficient. In extremely critical work it is occasionally desirable to inquire into the precision of the sample variance, that is, to ask how well it is likely to approximate the variance of the parent distribution.

This can be investigated in a straightforward way by computing the variance *of the variance*. We shall not discuss this calculation here; it is rarely needed.

In the foregoing discussion we have assumed that all the x_i belong to the same infinite parent distribution and that this is a normal distribution. But one can easily think of cases where this is not true. If one makes a series of measurements with an ordinary meter stick, and then measures with a good-quality steel scale, the random errors will in general be distributed differently in the two cases. There may of course also be systematic errors; we assume here that these have been either eliminated or corrected.

How shall we handle the case when the x_i come from different parent distributions? Specifically, suppose that x_i comes from a normal parent distribution characterized by variance $\sigma_i{}^2$. Referring to Eq. (14.6), we see that the probability of the set of measurements must be written as

$$P = \frac{(dx)^N}{\sigma_1 \sigma_2 \cdots \sigma_N (\sqrt{2\pi})^N} \exp\left[-\sum \frac{(x - x_i)^2}{2\sigma_i{}^2} \right]$$

(14.11)

The "least-squares sum" in this case is $\sum (x - x_i)^2 / 2\sigma_i{}^2$. To maximize P, according to the principle of maximum likelihood, we minimize this sum, leading to the condition

$$\frac{d}{dx} \sum \frac{(x - x_i)^2}{2\sigma_i{}^2} = 0$$

(14.12)

Carrying out the differentiation and rearranging the result, we find

$$x = \frac{\Sigma \, x_i/\sigma_i{}^2}{\Sigma \, 1/\sigma_i{}^2} \qquad (14.13)$$

We have found that the most probable value of x in this case is not the simple mean of the x_i, but a *weighted mean*, in which each weight w_i is the reciprocal of the variance of the corresponding parent distribution.

We have thus obtained an important and very useful result: In computing the average of several quantities whose variances are known, the most probable value is a weighted average in which each weight w_i is given by

$$w_i = \frac{1}{\sigma_i{}^2} \qquad (14.14)$$

The variance of the value of x obtained from Eq. (14.13) can be found by exactly the same procedure used to derive Eq. (14.10). From the propagation-of-errors formula, Eq. (13.8), we have, using Eq. (14.14),

$$\sigma_m{}^2 = \sum_j \left(\frac{\partial x}{\partial x_j}\right)^2 \sigma_j{}^2 = \sum_j \frac{w_j{}^2}{\left(\sum_i w_i\right)^2} \sigma_j{}^2$$

$$= \frac{1}{(\Sigma \, 1/\sigma_i{}^2)^2} \sum \frac{1}{\sigma_i{}^2} \qquad (14.15)$$

Thus we find

$$\frac{1}{\sigma_m{}^2} = \sum_i \frac{1}{\sigma_i{}^2} \qquad (14.16)$$

Clearly, the variance of the weighted mean is smaller

than any of the individual variances. We note also that in the special case where all the variances are equal $1/\sigma_m{}^2 = N/\sigma^2$ and Eq. (14.16) reduces to Eq. (14.10).

As an example of the use of the methods just outlined, suppose that two experimenters have measured the velocity of light in vacuum and have obtained the following results:

1. $c = 299{,}774 \pm 2$ km/sec
2. $c = 299{,}778 \pm 4$ km/sec

where the errors are standard deviations of the means. What is the most probable value of c, based on these two determinations, and what is its standard deviation? According to Eq. (14.13), we should weight each observation according to $1/\sigma^2$. Clearly, it is immaterial whether the weights are *equal* to their respective values of $1/\sigma^2$ or simply *proportional* to them. Thus it is correct to give the first determination a weight of 4, and the second a weight of unity. The most probable value is then

$$c = \frac{4 \times 299{,}774 + 1 \times 299{,}778}{4 + 1} = 299{,}774.8 \text{ km/sec}$$

Its standard deviation is given by Eq. (14.16):

$$\frac{1}{\sigma^2} = \frac{1}{(2 \text{ km/sec})^2} + \frac{1}{(4 \text{ km/sec})^2}$$

or

$$\sigma = 1.7 \text{ km/sec}$$

In using Eqs. (14.13) and (14.16), one should keep in mind that the variance associated with each x_i also

provides a means of testing whether the values are *consistent* in a statistical sense. Suppose, for example, that on two different days one makes measurements on the melting temperature of a certain alloy. One day's result yields the value $736 \pm 1°C$, and the other day's result is $749 \pm 2°C$, where in each case the figures after the $\pm$ sign are standard deviations. The difference is very much larger than the standard deviation in either result; and the probability of this occurring by chance is infinitesimally small. Thus we suspect that in one or both determinations there is a *systematic* error. Perhaps the composition of the alloy has changed. Considerations of this sort are an important weapon in detecting systematic errors. Of course, one can devise more quantitative tests of consistency; we shall not go into any further detail here.

The result given by Eq. (14.16), and some other results to be derived later, can be obtained somewhat more simply if one is willing to accept a statement which can be put on firm theoretical ground but which we cannot discuss in detail. The statement is this: In Eq. (14.11), which gives the probability of the set of observations as a function of x, P is approximately a *Gauss function of x* if the number of observations is large. That is, P can be represented by

$$P = \text{const} \times e^{-(x-x_0)^2/2\sigma_m^2} \tag{14.17}$$

in which x_0 is the value of x which maximizes P, which we have shown to be equal to the weighted mean, Eq. (14.13), and σ_m^2 is the variance of the mean, which

we should like to find. To find σ_m^2, we compare Eq. (14.17) with Eq. (14.11). The sum in Eq. (14.11) is again called the least-squares sum and denoted by $M(x)$. We make a Taylor series expansion of $M(x)$ about the most probable value x_0:

$$M(x) = M(x_0) + (x - x_0)\left(\frac{dM}{dx}\right)_{x_0}$$
$$+ \tfrac{1}{2}(x - x_0)^2\left(\frac{d^2M}{dx^2}\right)_{x_0} + \cdots \tag{14.18}$$

The derivatives are evaluated at the most probable value, and thus $(dM/dx)_{x_0} = 0$. Comparing Eq. (14.18) with the exponent in Eq. (14.17), we see that they are equal only if

$$\frac{1}{2}\frac{d^2M}{dx^2} = \frac{1}{2\sigma_m^2} \tag{14.19}$$

Thus we conclude that the variance of the mean is related to the least-squares sum by the simple equation

$$\frac{1}{\sigma_m^2} = \frac{d^2M}{dx^2} \tag{14.20}$$

where the derivative is evaluated at the point $x = x_0$. This result can also be used to simplify some derivations in which the maximum likelihood principle is used for the determination of several unknowns.

We conclude this section by considering another application of the method of least squares in a situation slightly different from the simple one of making a series of measurements on a single observable. This example deals instead with determining an unknown quantity

indirectly from *pairs* of observations on two other quantities.

Suppose we want to determine the force constant of a spring (or Young's modulus of a wire). Suppose also that there is reason to believe that the spring obeys Hooke's law, so that the relation of force F to elongation y is

$$F = ky \qquad (14.21)$$

where k is the spring constant to be determined. We apply several different forces F_i to the spring by hanging accurately calibrated weights on the end. For each, we measure the elongation y_i. The observations are shown graphically in Fig. 14.1.

The y_i all have random errors; we assume that the errors all have the same distribution and thus the same variance. If there were no errors in y_i, we would have $y_i - F_i/k = 0$. As it is, the quantity $d_i = y_i - F_i/k$ represents the error in y_i. Therefore in the principle of maximum likelihood, the correct least-squares sum to use is

$$M(k) = \sum \frac{(y_i - F_i/k)^2}{2\sigma^2} \qquad (14.22)$$

Taking dM/dk and setting it equal to zero,

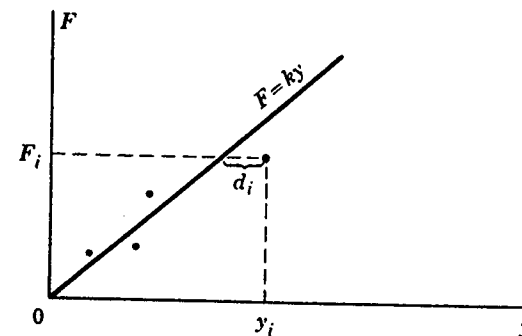$$\frac{dM}{dk} = \frac{1}{\sigma^2 k^2} \sum F_i (y_i - F_i/k) = 0 \qquad (14.23)$$

or

$$k = \frac{\sum F_i^2}{\sum F_i y_i} \qquad (14.24)$$

It is enlightening to compare this result with the procedure one might naïvely be tempted to use, namely, the average of the ratios F_i/y_i, or

$$k = \frac{1}{N} \sum \frac{F_i}{y_i}$$

The correct least-squares result, Eq. (14.24), is quite different.

Fig. 14.1. Each point represents a pair of observations of F and y. The deviation d_i corresponding to (F_i, y_i) is shown. The line represents the result of the least-squares calculation.

The variance σ_k^2 of k may be found by use of Eq. (14.20). The derivative of the left side of Eq. (14.23) is d^2M/dk^2; this is

$$\frac{d^2M}{dk^2} = -\frac{2}{\sigma^2 k^3} \sum F_i \left(y_i - \frac{F_i}{k} \right) + \frac{1}{\sigma^2 k^4} \sum F_i^2 \qquad (14.25)$$

The first sum in this equation is just $(2/k)$ times the first derivative, and this is zero. Thus

$$\frac{1}{\sigma_k{}^2} = \frac{1}{\sigma^2 k^4} \sum_i F_i{}^2 \tag{14.26}$$

The value of k used is, of course, the most probable value just found.

The variance σ^2 of the measurements y_i can be estimated in various ways.

The most straightforward procedure is to compute the deviation $d_i = y_i - F_i/k$ for each pair of observations, using the most probable value of k as given by Eq. (14.24). Then the variance of the y_i is given by

$$\sigma^2 = \frac{1}{N} \sum d_i{}^2$$

$$= \frac{1}{N} \sum \left(y_i{}^2 - \frac{2 F_i y_i}{k} + \frac{F_i{}^2}{k^2} \right) \tag{14.27}$$

Inserting Eq. (14.24) into Eq. (14.27),

$$\sigma^2 = \frac{1}{N} \left[\sum y_i{}^2 - \frac{(\sum F_i y_i)^2}{\sum F_i{}^2} \right]$$

or

$$\sigma^2 = \frac{1}{N} \left(\sum y_i{}^2 - \frac{1}{k} \sum F_i y_i \right) \tag{14.28}$$

Since $\sum F_i y_i$ has already been computed, evaluation of Eq. (14.28) involves relatively little additional work. In fact, when such calculations are done by machine, it is usual to compute $\sum F_i{}^2$, $\sum F_i y_i$, and $\sum y_i{}^2$ simultaneously.

A direct estimate of σ^2 can be obtained, of course, by repeating the observation several times with the same force. In practical cases, if one wants only a rough

estimate of $\sigma_k{}^2$, an estimate or shrewd guess of σ^2, based on inspection of the instruments, may suffice.

15 | Least Squares with Several Unknowns

The method of least squares is also useful when more than one quantity is to be determined from a series of measurements. We start with an example. We know that for an object moving in a straight line with constant acceleration the velocity varies with time according to the relation

$$v = v_0 + at \tag{15.1}$$

where v_0 is the velocity at time $t = 0$. Now suppose that in a particular situation we want to determine the values of v_0 and a by measuring the velocity v_i at each of a succession of times t_i. The measurements might be made, for example, with a speedometer and a stop watch. Furthermore, suppose that the times can be measured very accurately, so that the principal experimental errors are in the velocity measurements.

If we merely measure the velocity at two times, obtaining two pairs of observations (v_1, t_1) and (v_2, t_2), we obtain two simultaneous equations:
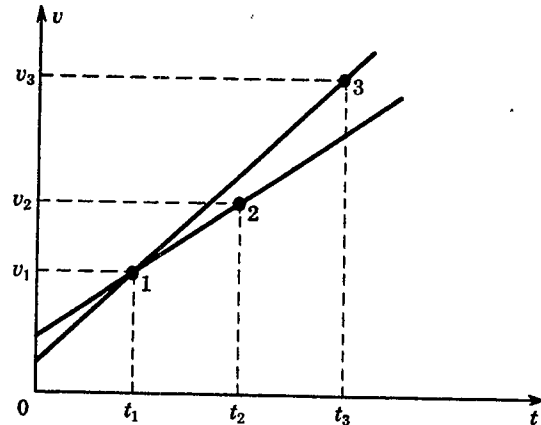
$$\begin{aligned} v_1 &= v_0 + a t_1 \\ v_2 &= v_0 + a t_2 \end{aligned} \tag{15.2}$$

which can be solved to determine v_0 and a. But now suppose that, in order to increase the precision of our results, we take a *series* of pairs of observations, say N

pairs in all, of which a typical pair is (v_i,t_i). The resulting set of equations is

$$v_1 = v_0 + at_1$$
$$v_2 = v_0 + at_2$$
$$\cdots\cdots\cdots$$
$$v_N = v_0 + at_N$$

(15.3)

Fig. 15.1. The line which fits observations (v_1,t_1) and (v_2,t_2) has, in general, a different slope (a) and intercept (v_0) from that which fits (v_1,t_1) and (v_3,t_3). There is no straight line which fits all three points.

Now these equations are not in general consistent; if we take different pairs of equations, and solve them for v_0 and a, we obtain different values of v_0 and a. Graphical solutions for v_0 and a are shown in Fig. 15.1. The reason for the various values of v_0 and a, of course, is that there are experimental errors in the v_i. Equations (15.3) should be regarded therefore not as true

equalities but as *observation equations* whose two sides are not quite equal. We use the symbol $\overset{\circ}{=}$, which is read "observed to be equal to," and write

$$v_i \overset{\circ}{=} v_0 + at_i$$

(15.4)

In actual fact the two sides of the equation are not exactly equal, but differ by an amount d_i:

$$d_i = v_0 + at_i - v_i$$

(15.5)

where d_i is the deviation corresponding to equation i and the *pair* of observations (v_i,t_i).

Since the d_i are the results of the experimental errors, we shall assume that they are distributed according to the Gauss distribution function. The most probable values of v_0 and a can now be obtained from the principle of maximum likelihood.

For a set of observations (v_i,t_i), they are the values which make this set *most probable*. As in the previous example, this probability is *maximized* when we *minimize* the sum of the squares of the deviations.

Here is the *principle of least squares* operating again. That is, we want to minimize the quantity

$$\Sigma d_i^2 = \Sigma (v_0 + at_i - v_i)^2$$

(15.6)

by choosing v_0 and a properly.

To minimize a function of two variables, we take the partial derivative of the function with respect to each of the variables in turn and set each derivative equal to zero. Thus the conditions which determine v_0 and a are

$$\frac{\partial}{\partial v_0} \sum d_i{}^2 = 0 \qquad \text{and} \qquad \frac{\partial}{\partial a} \sum d_i{}^2 = 0 \qquad (15.7)$$

When we evaluate these derivatives, we obtain two equations which then can be solved simultaneously to find v_0 and a. Notice that in general this procedure always gives us as many equations as we have unknown quantities to determine. These equations, of which Eqs. (15.7) are examples, are sometimes called *normal equations*.

Inserting Eq. (15.6) into Eq. (15.7),

$$\frac{\partial}{\partial v_0} \sum d_i{}^2 = \sum 2(v_0 + at_i - v_i) = 0$$

or

$$v_0 N + a \sum t_i = \sum v_i \qquad (15.8a)$$

and

$$\frac{\partial}{\partial a} \sum d_i{}^2 = \sum 2t_i(v_0 + at_i - v_i) = 0$$

or

$$v_0 \sum t_i + a \sum t_i{}^2 = \sum v_i t_i \qquad (15.8b)$$

We now have a pair of simultaneous equations for v_0 and a:

$$
\begin{aligned}
v_0 N \;\;+ a \sum t_i &= \sum v_i \\
v_0 \sum t_i + a \sum t_i{}^2 &= \sum v_i t_i
\end{aligned}
\qquad (15.8)
$$

These equations are the *normal equations* for this problem. The number of normal equations is equal to the number of unknowns. Equations (15.8) can be solved in a straightforward manner for v_0 and a, using determinants:

$$v_0 = \frac{\begin{vmatrix} \sum v_i & \sum t_i \\ \sum v_i t_i & \sum t_i{}^2 \end{vmatrix}}{\begin{vmatrix} N & \sum t_i \\ \sum t_i & \sum t_i{}^2 \end{vmatrix}} = \frac{(\sum v_i)(\sum t_i{}^2) - (\sum v_i t_i)(\sum t_i)}{N \sum t_i{}^2 - (\sum t_i)^2}$$

$$(15.9)$$

$$a = \frac{\begin{vmatrix} N & \sum v_i \\ \sum t_i & \sum v_i t_i \end{vmatrix}}{\begin{vmatrix} N & \sum t_i \\ \sum t_i & \sum t_i{}^2 \end{vmatrix}} = \frac{N \sum v_i t_i - (\sum v_i)(\sum t_i)}{N \sum t_i{}^2 - (\sum t_i)^2}$$

Before proceeding further, it is worthwhile to stop to consider what we have done. If there had not been any experimental errors, all the pairs of observations (v_i, t_i) would have obeyed Eq. (15.1). A graphical rep-
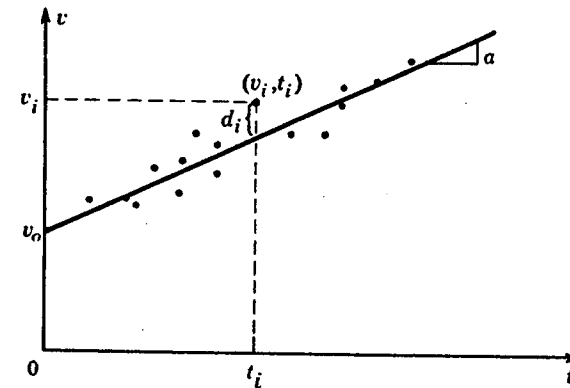
Fig. 15.2. Graph illustrating least-squares calculation of v_0 and a. Each point represents a pair of observations; a typical one is labeled (v_i, t_i), with its deviation. The line is drawn with the values of v_0 and a given by Eqs. (15.9). In general this line need not pass exactly through *any* of the points.

resentation of this statement is that if there were no experimental errors, all the points represented by the coordinates (v_i, t_i) would lie on a straight line whose slope is a and whose intercept on the v axis is v_0. Such a line is shown in Fig. 15.2.

Because of the random errors in the v_i, the actual observations are represented by points which lie somewhat above and below the line, as shown in the figure. Reference to Eq. (15.5), which defines the deviations, shows that the graphical significance of d_i is that its magnitude is the vertical distance between point (v_i, t_i) and the line. The method of least squares selects a line which *minimizes* the sum of squares of these vertical distances. We have used vertical distances rather than horizontal distances or some combination because of the assumption that only the v_i, not the t_i, contain errors.

For reference, we restate the results just obtained, in slightly more general language. If two variables x and y are known to be related by a linear equation of the form $y = mx + b$, where m is the slope of the line and b is its y intercept, if a series of N observations (x_i, y_i) are made, in which random errors occur only in the y_i measurements, and if these errors all belong to the same parent distribution, then the normal equations are

$$m \Sigma x_i + bN = \Sigma y_i$$
$$m \Sigma x_i{}^2 + b \Sigma x_i = \Sigma x_i y_i \qquad (15.10)$$

and the most probable values of m and b are given by

$$m = \frac{N \Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)}{N \Sigma x_i{}^2 - (\Sigma x_i)^2}$$

$$b = \frac{(\Sigma y_i)(\Sigma x_i{}^2) - (\Sigma x_i y_i)(\Sigma x_i)}{N \Sigma x_i{}^2 - (\Sigma x_i)^2} \qquad (15.11)$$

These expressions have been obtained directly from Eqs. (15.9) by appropriate substitution of symbols.

The next logical step is to try to calculate the standard deviations of the values of m and b which have been obtained from the method of least squares. This can be done by exactly the same methods as used for the case of one unknown. Errors in m and b are produced by errors in the y_i, which we assume are taken all from the same parent distribution with variance σ^2. Thus we may use Eq. (13.8) to compute the variances of m and b in terms of the variance of the parent distribution. Then the y_i themselves can be used to estimate the variance of the parent distribution.

We proceed as follows: From Eq. (13.8) we obtain

$$\sigma_m{}^2 = \sum_j \left(\frac{\partial m}{\partial y_j}\right)^2 \sigma^2 \qquad (15.12)$$

The partial derivatives are evaluated by use of Eq. (15.11), in which we abbreviate the denominators by the symbol $\Delta = N \Sigma x_i{}^2 - (\Sigma x_i)^2$. To evaluate

$$\frac{\partial}{\partial y_j} \sum_i x_i y_i$$

we note that there is only one term in the sum in which

y_i is the same as y_j, the variable we are differentiating. Therefore,

$$\frac{\partial}{\partial y_j} \sum_i x_i y_i = x_j$$

Similarly,

$$\frac{\partial}{\partial y_j} \sum_i y_i = 1 \qquad .$$

Thus we find

$$\frac{\partial m}{\partial y_j} = \frac{N x_j - \sum x_i}{\Delta} \qquad (15.13)$$

$$\left(\frac{\partial m}{\partial y_j}\right)^2 = \frac{N^2 x_j^2 - 2N x_j \sum x_i + (\sum x_i)^2}{\Delta^2} \qquad (15.14)$$

Inserting Eq. (15.14) into Eq. (15.12),

$$\sigma_m{}^2 = \frac{\sigma^2}{\Delta^2}\left[N^2 \sum x_j{}^2 - 2N\left(\sum x_j\right)\left(\sum x_i\right) + N\left(\sum_i x_i\right)^2\right]$$

$$= \frac{\sigma^2}{\Delta^2}\left[N^2 \sum x_j{}^2 - N\left(\sum x_j\right)^2\right] \qquad (15.15)$$

where we have used the obvious fact that $\sum x_i = \sum x_j$. Finally, recalling the definition of Δ,

$$\sigma_m{}^2 = \frac{N\sigma^2}{\Delta} \qquad (15.16)$$

Using precisely the same procedure to find the variance of b, we obtain

$$\sigma_b{}^2 = \frac{\sigma^2 \sum x_i{}^2}{\Delta} \qquad (15.17)$$

All that remains now is to estimate the variance σ^2

of the parent distribution, and this is easy to do. We recall that the deviation of each observation equation is given by

$$d_i = mx_i + b - y_i \qquad (15.18)$$

The variance of the sample is then

$$\sigma^2 = \frac{1}{N} \sum d_i{}^2 = \frac{1}{N} \sum (mx_i + b - y_i)^2 \qquad (15.19)$$

in which the values of m and b are those given by Eq. (15.11).

We have now solved, at least *in principle*, the problem of finding the variances of m and b, in that we have shown how they may be computed from the observed data by means of Eqs. (15.16), (15.17), and (15.19). In practice, the calculations are rather long and complicated. For this reason it is important to ask, in any particular problem, whether the variances are needed badly enough to justify the labor of obtaining them. If a large number of data are to be used in an all-out effort to determine constants with the greatest possible precision, then of course one wants to know what the precision is. In this case, the necessary calculations are often done with a high-speed digital computer.

The theory of least squares can be generalized in at least four ways, which we shall discuss only very briefly.

1. It can be used to determine constants in equations when there are more than two unknowns. One has an observation equation for each set of observations. There is a deviation for each observation equation, and

the most probable values of the constants are determined by minimizing the sum of squares of deviations. This procedure involves taking the partial derivative of the sum with respect to each of the unknowns and setting the derivative equal to zero. This gives a number of normal equations equal to the number of unknowns. Simultaneous solution of these equations then gives the most probable values of the unknowns. The computational labor increases very rapidly, of course, as the number of unknowns increases.

2. The theory can be used when the observations are not all samples of the same parent distribution. In this case, as with one unknown, the deviations are weighted inversely as the variances of their parent distributions. For example, if the observation equations are $y_i = mx_i + b$, and the various parent distributions for the y_i are characterized by their variances σ_i^2, then the correct procedure is to minimize the quantity

$$\sum w_i d_i^2 = \sum \frac{d_i^2}{\sigma_i^2} = \sum w_i(mx_i + b - y_i)^2 \qquad (15.20)$$

where we have again used $w_i = 1/\sigma_i^2$. It is then easy to show (and is, in fact, almost obvious) that the normal equations for this example are

$$\begin{aligned} m \sum w_i x_i + b \sum w_i &= \sum w_i y_i \\ m \sum w_i x_i^2 + b \sum w_i x_i &= \sum w_i x_i y_i \end{aligned} \qquad (15.21)$$

As in the unweighted case, we can next calculate the variances of m and b by straightforward extensions of the methods already presented.

3. The method of least squares can be used when the observation equations are nonlinear. As a simple example, consider a capacitor of capacitance C which is initially charged to a potential V_0 and allowed to discharge through a resistance R. It can be shown that the potential difference across the capacitor is given by

$$V = V_0 e^{-t/RC} \qquad (15.22)$$

Suppose we want to determine the quantity RC by making a series of observations of V at various times. Suppose further that we have very accurate time-measuring instruments, so that the only significant random errors are in V, and that we have carefully eliminated any systematic errors in these measurements.

We write an observation equation for each pair of observations:

$$V_i \overset{\circ}{=} V_0 e^{-t_i/RC} \qquad (15.23)$$

and a corresponding deviation

$$d_i = V_i - V_0 e^{-t_i/RC} \qquad (15.24)$$

Assuming that the errors in the V_i are normally distributed, all with the same variance, we determine V_0 and RC using the principle of maximum likelihood by computing Σd_i^2 and minimizing it. The normal equations for V_0 and RC are, however, nonlinear, and can be solved only by numerical methods.

If the voltmeter happens to have a logarithmic scale, as some electronic voltmeters do, the problem becomes much simpler. We take logarithms of both sides of Eq. (15.23):

$$\ln V_i \overset{\circ}{=} \ln V_0 - \frac{t_i}{RC} \qquad (15.25)$$

Introducing a new variable y, defined by $y = \ln V$, we now have a *linear* observation equation in y_i and t_i. Furthermore, because of the logarithmic scale, it is reasonable to assume that the errors in the y_i all have the same variance; so we may proceed with exactly the same methods which led to Eqs. (15.11). If the scale is not logarithmic, the y_i will not have the same variance, however.

It is not always possible to reduce an equation to linear form by a simple substitution. In more complicated cases it may be expedient to calculate approximate values of the unknown quantities and then represent the nonlinear equations by linear approximations, using Taylor series expansions.

4. It sometimes happens that we do not know the form of the observation equations or, indeed, whether the observed quantities are related at all. We then need a systematic method of investigating whether there *is* any relationship between two variables. This leads to the theory of correlations, a simple example of which is given in Sec. 16.

16 | Correlations

In Sec. 15, we discussed the problem of determining the constants in a linear equation relating two variables (in this case x and y) by using pairs of observations (x_i, y_i)

of these variables; it was known in advance that such a linear relationship existed.

Sometimes it happens, however, that we do not know in advance whether two variables, say x and y, are related. Furthermore, if we make pairs of observations (x_i, y_i) as before, the data may be scattered so widely because of experimental errors that it is not clear whether
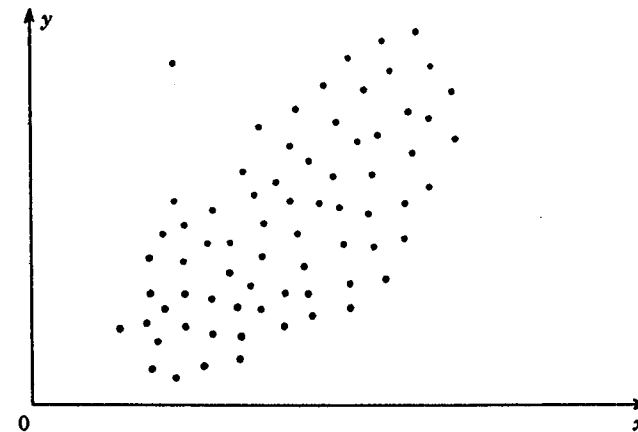
Fig. 16.1. To what extent are x and y related?

or not there is any relation between x and y. Representing the observations (x_i, y_i) graphically, we might obtain a picture similar to Fig. 16.1. Are x and y related, or are they not? Is there a *correlation* between x and y?

Of course, there is no end to the variety of possible functional relationships between x and y. There is no general way of investigating all possible relationships,

but it is fairly easy to check some simple ones. The simplest possible one, of course, is a linear equation. So a reasonable place to start is to ask whether there is a *linear* relationship between x and y or, in other words, a *linear correlation*.

We can answer this question at least partially by a slight extension of the method of least squares for two unknowns, introduced in Sec. 15. We assume a linear relationship between x and y in the form

$$y = mx + b \qquad (16.1)$$

and proceed to determine the constants m and b from the observations (x_i, y_i) in exactly the same manner as in Sec. 15, Eq. (15.11). In particular,

$$m = \frac{N \sum xy - \sum x \sum y}{N \sum x^2 - (\sum x)^2} \qquad (16.2)$$

In this expression, and in those which follow, we have dropped the limits of the summation, which are always 1 to N, and also have omitted the summation indices on x and y. Thus, $\sum xy$ is an abbreviation for

$$\sum_{i=1}^{N} x_i y_i$$

The graphical interpretation of the procedure just described is as follows: We are trying to represent the scattered points in Fig. 16.1 by drawing the *best straight line* through the points. The slope of this line is m, and its intercept on the y axis is b. Since the deviations we have used in the method of least squares are

$$d_i = mx_i + b - y_i \qquad (16.3)$$

d_i represents the vertical distance between the point (x_i, y_i) and the straight line described by constants m and b. In this case, then, the method of least squares minimizes the sum of the squares of the vertical distances between the point and the straight line. The line determined by this procedure is sometimes called the *line of regression of y on x*.

If there is no correlation at all between x and y, this sum of squares will be minimized by a horizontal line; we shall find therefore in the case of *no* correlation that $m = 0$, a line with zero slope.

Now let us back up slightly. There is no particular reason for writing our assumed linear relationship between x and y in the particular form of Eq. (16.1). We might equally well have written instead

$$x = m'y + b' \qquad (16.4)$$

in which the roles of x and y have been reversed. In this case, the deviations we use in the method of least squares are given by

$$d_i' = m'y_i + b' - x_i \qquad (16.5)$$

The method of least squares now minimizes the sum of the squares of the *horizontal* distances between the line described by Eq. (16.4) and the points (x_i, y_i) representing the observations. The result is the *line of regression of x on y*. The expression for m' is obtained simply by reversing the roles of x and y in Eq. (16.2) and is

$$m' = \frac{N \sum xy - \sum x \sum y}{N \sum y^2 - (\sum y)^2} \qquad (16.6)$$

Note that m' is not the slope of the line given by Eq. (16.4) but rather its reciprocal. This is easy to see if we solve Eq. (16.4) for y.

$$y = \frac{x}{m'} - \frac{b'}{m'} \tag{16.7}$$

We see that the slope of this line is $1/m'$, and its intercept with the y axis is $-b'/m'$.

Using Eq. (16.4), if there is no correlation between x and y, the method of least squares will give the value $m' = 0$, a *vertical* line. If, on the other hand, all the points lie exactly on the line, so that the correlation is perfect, then this method must give us the same line as the previous one, Eq. (16.1). That is, in the case of perfect correlation, we must find that $1/m' = m$. Thus if there is *no* correlation between x and y, $mm' = 0$, while if the correlation is *perfect*, $mm' = 1$. Clearly, the value of the product mm' has something to do with the extent to which the variables x and y are correlated.

It is therefore natural to define a *correlation coefficient* r as follows:

$$r = \sqrt{mm'} = \frac{N \sum xy - \sum x \sum y}{[N \sum x^2 - (\sum x)^2]^{1/2} [N \sum y^2 - (\sum y)^2]^{1/2}} \tag{16.8}$$

Thus $r = 1$ means perfect correlation, and $r = 0$ means no correlation. If there is imperfect correlation, we expect a value of r somewhere between 0 and 1. In fact, it can be shown that Eq. (16.8) must *always* have a value between -1 and 1.

Suppose now that we have calculated r for a set of

observations. How do we interpret the result? In other words, how large must r be in order to indicate a significant correlation between the variables x and y? Clearly, because of random fluctuations, we will not in general get *exactly* $r = 0$ even if there is no real connection between x and y. And if a linear relationship exists, we will not get *exactly* $r = 1$, especially if the experimental errors are large. Given a value of r, then, the question to ask is: What is the probability of obtaining a value of r as large as this purely by *chance* from observations on two variables which are not really related? This situation is similar to the one which arose in interpreting the results of a χ^2 calculation in Sec. 11.

Tables have been calculated which give the probability of obtaining a given value of r for various numbers N of pairs of observations. Table V gives a few values of this probability. A reference to more extensive tables is also given.

Here is an example of the use of this table. Suppose we make 10 observations; then $N = 10$. The table says that there is a probability $P = 0.10$ of finding a correlation coefficient of 0.549 or larger by chance, and a probability $P = 0.01$ of finding $r \geq 0.765$, if the variables are not really related. If for our 10 observations we find $r = 0.9$, we can be reasonably sure that this indicates a true correlation and not an accident. But if we find only $r = 0.5$ we cannot be sure, because there is more than 10% chance that this value will occur by chance.

A commonly used rule of thumb in interpreting values of r is to regard the correlation as significant if there is less than 1 chance in 20 ($P = 0.05$) that the value will occur by chance. According to this rule of thumb, we find from Table V that for 10 sets of observations, any value of r greater than 0.632 should be regarded as showing a significant correlation. For five sets, r must be greater than 0.878 to be significant.

The theory of correlations can be extended in several directions. First, there may exist a functional relationship between x and y which is not linear and which is not detected by our linear correlation coefficient. For example, if the graph of x versus y is a circle, the correlation coefficient will be zero even if there are no experimental errors. To take such possibilities as this into account, we can assume a quadratic, cubic, or more complicated functional relationship and use the theory of least squares to determine the constants in the equations. Such an analysis gives us *nonlinear correlations*.

It is also necessary at times to consider correlations among more than two variables, so-called *multiple correlations*. These extensions of the theory of correlations have rather specialized applications, and we shall not consider them here.

PROBLEMS

1. Find the standard deviation of the mean in Prob. 18, Chap. I.

2. Find the standard deviation of the mean in Prob. 27, Chap. III. Compare this value with the *change* in the mean which results from the rejection of unlikely data.

3. A certain quantity was measured N times, and the mean and its standard deviation were computed. If it is desired to increase the precision of the result (decrease σ) by a factor of 2, how many *additional* measurements should be made?

4. In Prob. 3, discuss how the mean of the first N measurements should be combined with the mean of the additional measurements, and how the standard deviation of the result should be computed from the standard deviations of the two sets.

5. Show that the standard deviation of a weighted mean is always smaller than any individual standard deviation. Is this a reasonable result?

6. Two different measurements of the speed of light using the same general method (a rotating mirror) yielded the following results:

$$299,796 \pm 4 \text{ km/sec}$$
$$299,774 \pm 2 \text{ km/sec}$$

Are these results consistent? (Assume that the errors given are standard deviations of the means.)

7. In Prob. 13, Chap. I, suppose that the "errors" referred to are standard deviations of the means. Find the standard deviation in g. Compare with the result of Prob. 13, Chap. I. Which is more significant?

8. For some obscure reason an individual wants to make an accurate determination of the area of a sheet of typewriting paper. The following measurements are made on the length and width:

Length, in.	Width, in.
11.03	8.51
11.00	8.48
10.97	8.49
10.98	8.50
11.02	8.53
	8.48
	8.51

a. Determine the standard deviation of each set of measurements.

b. Determine the most probable value of each quantity and its standard deviation.

c. Determine the most probable value of the area and its standard deviation.

9. In a centripetal-force experiment the force exerted on a body moving in a circle with constant speed is measured directly and is also computed from the equation

$$F = \frac{mv^2}{R} = \frac{4\pi^2 mR}{T^2}$$

The following data were obtained:

F, dynes	T, sec	R, cm
6.92×10^5	0.200	5.13
6.82	0.198	5.06
6.87	0.202	5.05
6.87	0.199	5.09
6.92	0.201	5.10

The mass is known very accurately: $m = 140.00$ g. Do the measured and calculated values of F agree? Discuss.

10. In the Bohr theory of the structure of the hydrogen atom, the energies of the various quantum states are given by

$$E_n = -\frac{1}{2}\frac{me^4}{n^2\hbar^2}$$

where m is the mass of the electron, e is its charge, $\hbar$ is Planck's constant divided by 2π, and $n = 1, 2, 3, \cdots$. If the mass is known with a fractional standard deviation of 0.1%, the charge with 0.2%, and Planck's constant with 0.1%, what is the per cent standard deviation in E_n for the state for which $n = 1$? For the $n = 2$ state? If the accuracy is to be improved, which quantity (m, e, or $\hbar$) should be determined more accurately?

11. The phase angle ϕ between voltage V and current I supplied to an electric motor (or any other device) is related to the electrical power input P by the equation $P = EI \cos\phi$. The quantities P, E, and I are measured, with the following results:

$P = 515 \pm 50$ watts

$E = 110 \pm 2$ volts

$I = 5.20 \pm 0.20$ amp

a. The quantity $\cos\phi$ is called the *power factor*. Calculate the power factor and its standard deviation.

b. Calculate ϕ and its standard deviation.

12. The number of radioactive decays occurring in a given interval has been shown to follow the Poisson distribution. Often the parameter a is not known in advance, but is determined by counting for several intervals. Suppose N intervals are used, and n_i counts are observed in interval i (where $i = 1, 2, \cdots, N$). Apply the principle of maximum likelihood to determine a. That is, find the value of a which maximizes the probability of occurrence of the set of observa-

tions n_i. Specifically, show directly that the most probable value of a is the *average* of the n_i.

13. What is the standard deviation of the most probable value of a obtained in Prob. 12?

14. The value of a resistor R is to be found by passing several different currents I through it, measuring the corresponding voltage drop V, and using the relation $V = IR$. The values of V are measured very precisely with a potentiometer, while I is measured with an ordinary ammeter, resulting in normally distributed random errors. Using the method of least squares, derive an expression for the most probable value of R in terms of the pairs of observations (V_i, I_i).

15. From the set of observation equations given, find the most probable values of x and y, using the method of least squares, assuming all the observational errors to belong to the same normal distribution.

$$3x + y \overset{\circ}{=} 2.9$$
$$x - 2y \overset{\circ}{=} 0.9$$
$$2x - 3y \overset{\circ}{=} 1.9$$

16. The three interior angles of a triangle were observed to be

$$31° \qquad 62° \qquad 88°$$

Using the method of least squares and the fact that the sum of the angles must be 180°, find the most probable values of the angles. Does your method make any assumption about relative accuracies of the measurements of the angles?

17. In an experiment to measure the acceleration of a freely falling object, a tuning fork is set into vibration and allowed to drop, scratching a wavy line on a strip of waxed paper as it falls. From this trace, the positions at a succession

of times separated by equal time intervals can be determined. The theoretical relation between distance and time is

$$s = s_0 + v_0 t + \tfrac{1}{2}gt^2$$

Assuming that the times are known much more precisely than the positions, use the method of least squares to derive expressions for the initial position s_0, the initial velocity v_0, and the acceleration g in terms of the pairs of observations (s_i, t_i).

18. Use the method of least squares to find the best straight line for the four points (4,5), (6,8), (8,10), (9,12). Are any assumptions regarding the errors necessary?

19. An experimenter wanted to determine the ratio of inches to centimeters by using a yardstick and a meter stick, side by side. His procedure is to observe the centimeter corresponding to each of a succession of inch marks. Unfortunately, the centimeters are not subdivided, so he reads only to the nearest centimeter. Use the method of least squares to derive a formula for the conversion factor.

20. In Prob. 19, the result is considerably simpler if an odd number of inch marks are used, and if they are renumbered so that zero is in the middle. That is, if there are $2N + 1$ marks, they are labeled from $-N$ to N. Obtain the simplified result, using this scheme. Useful information: The sum of the first N integers is

$$1 + 2 + \cdots + N = \frac{N(N + 1)}{2}$$

and the sum of their squares is

$$1^2 + 2^2 + \cdots + N^2 = \frac{N(N + 1)(2N + 1)}{6}$$

21. Using the data of Prob. 18, calculate the line of regres-

Further Developments

sion of x on y, the line of regression of y on x, and the correlation coefficient. Is the correlation significant?

22. Following are two sets of pairs of observations on variables x and y:

x	y	x	y
1	2	1	2
2	3	2	3
3	3	3	4
4	2	4	5
5	4	5	5

Determine whether either of these sets exhibits a significant correlation between x and y.

APPENDIX A
SUMMARY OF FORMULAS

Following is a summary of important and useful formulas which have been developed in the text. The numbered equations are given the same numbers as in the text to facilitate reference to appropriate parts of the text.

Approximations

If a quantity Q is determined from quantities $a, b, \ldots$ by a relation $Q = f(a, b, \ldots)$, then the change ΔQ of the quantity produced by changes $\Delta a, \Delta b, \ldots$ is

$$\Delta Q = \frac{\partial Q}{\partial a} \Delta a + \frac{\partial Q}{\partial b} \Delta b + \frac{\partial Q}{\partial c} \Delta c + \cdots \qquad (2.8)$$

The Mean and Dispersion

The *mean* (or arithmetic mean or average) of a set of N numbers, of which a typical one is x_i, is

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \qquad (3.2)$$

The *weighted mean* of a set of N numbers, of which a typical one is x_i with weight w_i, is