CHAPTER 10

Elements from the Theory of Statistics

10.1. DEFINITIONS

Statistics is the science that tries to draw inferences from a finite number of observations constituting only a sample, so as to postulate rules that apply to the entire population from which the sample was drawn.

In the field of physics, statistics is needed (a) to fit data—that is, to estimate the parameters of assumed frequency functions; (b) to treat random errors; and (c) to interpret phenomena that are inherently of a statistical nature.

10.1.1. Definition of Probability

The probability of occurrence of an event can be axiomatically defined as equal to one (= 1) if the event occurred, or equal to zero (= 0) if the event did not take place. An alternative definition of probability is based on the *frequency* of occurrence of an event. Suppose that several trials of the same experiment have been made; then the probability of occurrence

of an event A, that is $P(A)$, is given by the number of times event A was obtained divided by the total number of trials (in the limit that the total number of trials approaches infinity). This definition of probability retains its full value even in the case of nonrepetitive experiments, since the one trial can be considered as the first of a series of trials.

10.1.2. Sample Space

Any *set of points* that represents all possible outcomes of an experiment is a sample space. For example, if a coin is tossed twice, the sample space consists of the 4 points indicated in Fig. 10.1. (Sample spaces can be finite or infinite and discrete or continuous.)

Once the sample space for a particular experiment is constructed, we may assign (in the sense of Definition 10.1.1) a probability p_i to each point i of the space. From the definition of probability, we have

$$p_i > 0 \qquad \sum_{\text{all sample space points}} p_i = 1;$$

thus

$$p_i \leq 1$$

and the probability of occurrence of an event A is

$$P(A) = \frac{\sum_A p_i(A)}{\sum p_i} = \sum_A p_i(A),$$

where $\sum_A$ indicates summation over all points that include event A.

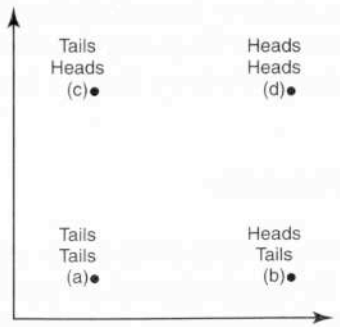

FIGURE 10.1 Simple example of a discrete and finite sample space. Here the sample-space points correspond to all possible outcomes of "tossing a coin" twice.

In most situations treated by statistics, equal probability is assigned to each sample-space point, a condition we will maintain throughout this discussion. Then

$$p_i = \frac{1}{n},$$

n being the total number of sample-space points, and

$$P(A) = \frac{n(A)}{n},$$

where $n(A)$ is the number of sample-space points containing event A.

For example, in the case of the sample space of Fig. 10.1, the probability of obtaining heads at least once is

$$P(\text{heads}) = \frac{n(\text{heads at least once})}{n} = \frac{3}{4}$$

while the probability of obtaining heads once and tails once (irrespective of order) can again be found by counting the appropriate points in the sample space of Fig. 10.1. We obtain

$$P(\text{heads, tails}) = \frac{n(\text{heads, tails})}{n} = \frac{2}{4}.$$

10.1.3. Probability for the Occurrence of a Complex Event

The probability that both events A and B will occur is called the joint probability

$$P[AB] = \frac{n(A \text{ and } B)}{n},$$

where n = total number of sample-space points. The probability that either A or B will occur is called the *either probability*

$$P[A + B] = \frac{n(A \text{ or } B)}{n},$$

and the probability that A will occur when it is certain that B occurred is called the *conditional probability*

$$P[A|B] = \frac{n(A \text{ and } B)}{n(B)}.$$

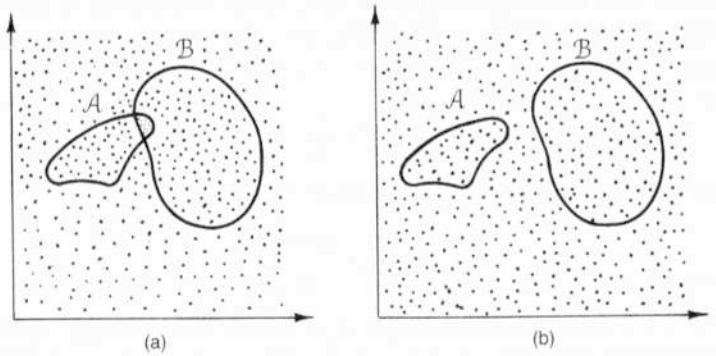

FIGURE 10.2 In the sample spaces shown it is assumed that all sample-space points in domain $\mathcal{A}$ contain event A, whereas all points in domain $\mathcal{B}$ contain event B. (a) There exists a region where both event A and event B can occur simultaneously. (b) No such region exists; events A and B are mutually exclusive.

All these probabilities are defined in the sense of Definition 10.1.2 as the number of sample-space points that contain the stated condition divided by the total number of sample-space points *allowed for* by the statement.

Figures 10.2a and 10.2b illustrate two sample spaces. All points within domain $\mathcal{A}$ include event A while all points within domain $\mathcal{B}$ include events B. The points contained in any intersection of the two domains $\mathcal{A}$ and $\mathcal{B}$ include both events A and B.

If such a common intersection does not exist in sample space, the two events are *mutually exclusive*, and

$$P[AB] = 0.$$

It follows from consideration of Fig. 10.2 that

$$P[A + B] = P[A] + P[B] - P[AB].$$

For the conditional probability

$$P[A|B] = \frac{n(\mathcal{A} \text{ intersection } \mathcal{B})}{n(\mathcal{B})},$$

since the condition that event B occurred restricts our sample within domain $\mathcal{B}$. However,

$$P[B] = \frac{n(\mathcal{B})}{n}$$

and

$$P[AB] = \frac{n(\mathcal{A} \text{ intersection } \mathcal{B})}{n} = P[A|B] \cdot P[B] = P[B|A] \cdot P[A]. \tag{10.1}$$

If $P[A|B] = P[A]$, it means that the occurrence of B does not affect the probability of occurrence of A. We say that the two events A and B are *independent*. It then follows from Eq. (10.1) that

$$P[AB] = P[A] \cdot P[B]. \tag{10.2}$$

Equation (10.2) in turn implies (when combined with Eq. (10.1)) that for independent events

$$P[B|A] = P[B].$$

To illustrate some of the ideas we have just expressed, consider the following. For the sample space of Fig. 10.1 we may define: event A = heads in first throw, and event B = heads in second throw. The domains are shown in Fig. 10.3, and it follows (assigning $p = 1/4$ to each point) that

$$P[A] = \frac{1}{2}; \qquad P[B] = \frac{1}{2}$$

$$P[AB] = \frac{1}{4}$$

$$P[A + B] = P[A] + P[B] - P[AB] = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$$

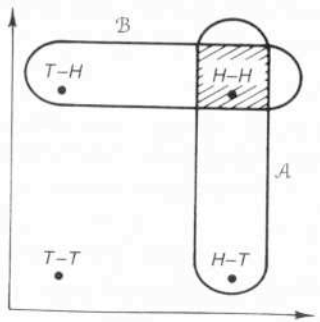

FIGURE 10.3 The sample space of Fig. 10.1 including the domain $\mathcal{A}$ (heads in the first throw) and the domain $\mathcal{B}$ (heads in the second throw).

$$P[A|B] = \frac{1}{2}; \qquad P[B|A] = \frac{1}{2}$$

$$P[AB] = P[A|B] \cdot P[B] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} = P[A] \cdot P[B].$$

Thus events A and B are *not* mutually exclusive but are independent.

10.1.4. Random Variable

To study a sample space analytically (instead of geometrically), it is convenient to use a numerical variable that takes a definite value for each and *every* point of the sample space; however, the same value may be assigned to several points. Thus, a random variable used for the representation of a finite and discrete sample space will have a definite range and will take only discrete values. As an example, for the sample space of Fig. 10.1, we can assign to the random variable x the value 0 for points (b) and (c) (one each of heads and tails), the value -1 for point (a) (both tails), and the value $+1$ for point (d) (both heads).

10.1.5. Frequency Function

A frequency function (of a random variable) is a function $f(x)$ such that $f(x_0)$ is the probability that the random variable x may take the specific value x_0. By Definition 10.1.1, $f(x)$ gives the number of points in the sample space that have been assigned the value of x of the random variable, divided by the total number of sample-space points. The function $f(x)$ is defined only within the range of x and need not have a definite analytic form. For the example considered above (the sample space of Fig. 10.1), $f(x)$ is just a table, as shown in Table 10.1 (see also Fig. 10.4).

TABLE 10.1 Example of a Frequency Function $f(x)$ of the Random Variable x

Sample-space point	x	$f(x)$
(a)	-1	$\frac{1}{4}$
(b,c)	0, 0	$\frac{1}{2}$
(d)	$+1$	$\frac{1}{4}$

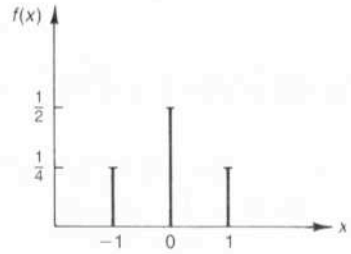

FIGURE 10.4 The distribution function of the discrete random variable x defined in Table 10.1.

The summation of $f(x)$ over the entire range of x must give 1:

$$\sum_{\text{all } x} f(x) = 1.$$

The probability that the random variable may take any value smaller or equal to x is given by

$$F(x) = \sum_{t<x} f(t)$$

and is called the *distribution function* of x (or integral distribution function).

It is sometimes convenient to describe a sample space in terms of two or more random variables, a frequency function existing for each of them. If these random variables are independently distributed in the sense of Eq. (10.2), the joint frequency function is

$$f(x_1, x_2, \ldots) = f(x_1)\, f(x_2) \cdots f(x_n).$$

If the random variable is continuously varying (for example, it describes the height of individuals), the probability of occurrence of the specific value x when a measurement is performed defines the frequency function $f(x)\,dx$ of the random variable x. The random variable may now take any value within the range of its definition. Note, however, that the probability of occurrence of the exact value x is zero, while it is the probability of occurrence of some value in the infinitesimal interval dx about x that exists. For a continuously varying random variable, we have

$$f(x) \geq 0 \qquad \text{and} \qquad \int_{-\infty}^{+\infty} f(x)\,dx = 1.$$

Similarly

$$\int_a^b f(x)\,dx = P[a < x < b] \qquad a < b$$

and

$$F(x) = \int_{-\infty}^{x} f(t)\,dt.$$

10.1.6. Some Definitions from Combinatorial Analysis

(a) *Permutations.* A permutation of n objects in groups of r objects is defined as follows. Consider n objects; any group of r of these objects, when ordered, forms a permutation; the same group of r objects, when ordered in a different fashion, forms a *new* permutation. As an example consider the three objects:

$$\square, \triangle, \bigcirc$$

There are only six possible permutations of three objects in groups of two:

$$\square\triangle, \quad \triangle\square; \quad \square\bigcirc, \bigcirc\square; \quad \triangle\bigcirc, \quad \bigcirc\triangle$$

We state without proof that the number of possible permutations of n objects in groups of r, ${}_nP_r$, is

$${}_nP_r = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}.$$

Then

$${}_nP_n = n!$$

as it must be.

(b) *Combinations.* A combination of n objects in groups of r objects is defined as any grouping of r objects out of the original n. The ordering within the grouping is not relevant. Thus for the previous example there are only three possible combinations

$$\square\triangle; \quad \bigcirc\square; \quad \bigcirc\triangle.$$

The number of possible combinations of n objects in groups of r, $\begin{bmatrix} n \\ r \end{bmatrix}$, is

$$\begin{bmatrix} n \\ r \end{bmatrix} = \frac{{}_nP_r}{{}_rP_r} = \frac{n!}{r!(n-r)!}.$$

(c) *Note.* Note that

$$n! = n \cdot (n-1)! \qquad 1! = 0! = 1.$$

10.2. FREQUENCY FUNCTIONS OF ONE VARIABLE

10.2.1. Definitions

Let us assume that a population (for example, all the possible outcomes of an experiment) can be described by a frequency function; we may attempt to find this function in two ways:

(a) By the use of a mathematical model based on the definitions of the previous section, thus obtaining a "theoretical frequency function."

(b) By observing a sample of the population and determining its "empirical frequency function."

The advantage of obtaining a frequency function for a population is that the few parameters involved in the frequency function suffice to describe completely the *entire* population and thus provide as much information as the most extensive data.

We will now deal only with populations that can be described by a frequency function depending on a single variable. To obtain the empirical frequency function it is best to divide the members of the sample into classes (defined by the random variable) and then make a graphical plot or *histogram* of the sample. If we try to describe the histogram, the first obvious features are its location and its spread.

A very useful set of measures are the *moments* of a histogram, defined in the usual way (moments of forces, electric moments, etc.). Thus, if x_i is the value of the random variable for the class i and if f_i is the number of events in this class, the k^{th} moment of the *empirical* frequency function about the origin is

$$m'_k = \frac{1}{n}\sum_{\text{all } i} x_i^k f_i,$$

where n is the size of the sample. Similarly, the kth moment about any other point x_0 is

$$m_k(x_0) = \frac{1}{n}\sum_{\text{all } i}(x_i - x_0)^k f_i.$$

10.2.2. Mean and Standard Deviation

The first moment about the origin, m_1', is called the mean and will be denoted by m:

$$m = m_1' = \frac{1}{n}\sum_{\text{all } i} x_i\, f_i \tag{10.3}$$

(commonly called the "average" of x). The second moment about the mean, m_2, is called the *variance*; its square root is called the *standard deviation* and is denoted by s: s has the same dimensions as the random variable x:

$$s = \sqrt{m_2} = \sqrt{\frac{1}{n}\sum_{\text{all } i}(x_i - m)^2 f_i}. \tag{10.4}$$

An often used relation pertaining to s is

$$s^2 = \frac{1}{n}\sum_{\text{all } i}(x_i - m)^2 f_i = \frac{1}{n}\sum(x_i^2 - 2mx_i + m^2) f_i$$

$$s^2 = \frac{1}{n}\sum_{\text{all } i} x_i^2 f_i - \frac{2m}{n}\sum(x_i f_i) + m^2$$

$$s^2 = \frac{1}{n}\sum_{\text{all } i}(x_i^2 f_i) - m^2$$

usually written as

$$\overline{\Delta x^2} = \overline{x^2} - (\bar{x})^2. \tag{10.5}$$

In most cases the mean and the standard deviation are the best measures (contain most information) of an empirical frequency function; there are, nevertheless, cases where they are very poor measures, and instead it is much better to give other location measures, such as the median or the geometric mean, and so on; and other variation measures such as the range or the mean variation, $(1/n)\sum |x_i - m|\, f_i$, and so on.

10.2.3. Theoretical Frequency Functions

As mentioned before, a theoretical frequency function $f(x)$ might be of the *discrete* type—that is, the random variable x takes only integer values, or of the "continuous" type. Most of the discrete random variables usually represent the number of successes, or of counts obtained, etc. In going from discrete frequency functions to continuous ones, obviously all summations are replaced by integrals.

Moments are defined as in Eq. (10.3), but instead of the empirical frequencies f_i, the theoretical frequency function $f(x)$ is used; the theoretical moments are designated by Greek letters, Latin letters being reserved for the empirical moments.

Thus, the k^{th} moment about the origin is

$$\mu_k' = \sum_{x=-\infty}^{x=+\infty} x^k f(x).$$

The first moment about the origin gives the mean, and is denoted by $\mu = \mu_1'$. The k^{th} moment of a theoretical frequency function about its mean is

$$\mu_k = \sum_{x=-\infty}^{x=+\infty} (x - \mu)^k f(x).$$

The square root of the second moment about the mean gives the standard deviation and is denoted by $\sigma = \sqrt{\mu_2}$:

$$\mu_2 = \sum_{x=-\infty}^{x=+\infty} (x - \mu)^2 f(x).$$

10.2.4. The Bernoulli or Binomial Frequency Function

This basic frequency function is applicable when there are only *two* possible outcomes of an experiment, as, for example, the occurrence of an event A or its nonoccurrence (we designate this by B). If the experiment is repeated n times, the random variable x describes the number of times event A occurred. The frequency function—that is, the probability of obtaining a certain x—is given by

$$f(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}, \tag{10.6}$$

where p is the probability that event A will occur in this experiment (defined in the sense of Section 10.1.1); and $q = 1 - p$ is the probability that B will happen, namely, that event A will not occur.

To prove Eq. (10.6), consider the probability of obtaining event A, x times in a definite sequence

$$\underbrace{AA\cdots A}_{x} \qquad \underbrace{BB\cdots B}_{n-x};$$

this joint probability of order n is according to Definition 10.1.3

$$\underbrace{pp\cdots p}_{x}\underbrace{qq\cdots q}_{n-x} = p^x q^{n-x}$$

since the outcome of consecutive experiments is independent. However, any other sequence, containing the same number x of occurrences, is also a satisfactory answer, since we are not interested in the order of occurrence of event A. Thus we must sum over all sample-space points that give x occurrences; the number of all such sample-space points is given by the permutations of n objects in groups of n when x of them are alike (have probability p), which is

$$\frac{n!}{x!(n-x)!},$$

completing the proof of Eq. (10.6).

The frequency function fulfills the normalization requirement as it should, since

$$\sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \frac{n!}{x!(n-x)!} p^x q^{n-x} = (p+q)^n = [p+(1-p)]^n = 1. \tag{10.7}$$

10.2.5. Moments of the Binomial Frequency Function

From the definitions of Section 10.2.3, and since the range of x is from 0 to n, we have

$$\begin{aligned}\mu = \mu_1' &= \sum_{x=0}^{n} x f(x) = \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x} \\ &= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x} \\ &= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x}.\end{aligned}$$

If we let $y = x - 1$, it follows that

$$\mu = np \sum_{y=0}^{n-1} \frac{(n-1)!}{y![(n-1)-y]!} p^y q^{[(n-1)-y]},$$

where now the sum is equal to $(p+q)^{n-1} \equiv 1$. Thus

$$\mu = np. \tag{10.8}$$

Next we wish to obtain the second moment about the mean, $\mu_2 = \sigma^2$. We first calculate μ_2', given by

$$\mu_2' = \sum_{x=0}^{n} x^2 \frac{n!}{x!(n-x)!} p^x q^{n-x}.$$

We use

$$x^2 = x(x-1) + x$$

so that

$$\begin{aligned}\mu_2' &= \sum_{x=0}^{n} x(x-1) \frac{n!}{x!(n-x)!} p^x q^{n-x} + \mu \\ &= \sum_{x=2}^{n} x(x-1) \frac{n!}{x!(n-x)!} p^x q^{n-x} + \mu \\ &= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} q^{n-x} + \mu\end{aligned}$$

and letting $y = x - 2$, as before, the sum is equal to $(p+q)^{n-2} \equiv 1$ and we obtain

$$\mu_2' = n(n-1)p^2 + \mu = n^2 p^2 - np^2 + np.$$

Next we use Eq. (10.5) to obtain

$$\mu_2 = \sigma^2 = \mu_2' - \mu^2 = -np^2 + np = np(1-p) = npq.$$

Thus

$$\sigma = \sqrt{npq}. \tag{10.9}$$

The binomial frequency function is applicable to many physical situations, but it is cumbersome to calculate with. When n becomes large,

however, the binomial frequency function approaches either the Poisson or the Gaussian frequency function, which will be discussed in Sections 10.2.6 and 10.2.7. In order for the binomial frequency function[1] to approach the

Poisson distribution	n must be large, for example, $n > 100$, but $\mu = np$ must be finite and small, for example, $p < 0.05$.
Gaussian distribution	n must be large, for example, $n > 30$, and also p must be large, for example, $p > 0.05$.

10.2.6. The Poisson Frequency Function

This is still a frequency function for the discrete random variable x, which describes, as in Section 10.2.4, the number of times event A will be obtained if the experiment is repeated n times when $n \to \infty$ for (large n). Contrary to Eq. (10.6), however, neither n nor p appears explicitly in the analytic expression of the frequency function, but instead only their product

$$y = np, \tag{10.10}$$

which remains finite despite $n \to \infty$, since $p \to 0$. The Poisson frequency function is given by

$$f(x) = \frac{y^x e^{-y}}{x!}, \tag{10.11}$$

and it is shown in the next section that y is the mean of the distribution governed by Eq. (10.11).

To prove Eq. (10.11), let us first note that since n is large, it (but not x) may be treated as a continuous variable; second, we will assume that for a small (differential) number of trials dn, the probability of obtaining event A once is proportional to this number of trials: that is,

$$P\{1, dn\} = \lambda\, dn, \tag{10.12}$$

where λ is a constant. Note that Eq. (10.6) fulfills this requirement for $x = 1$ in the limit that $p \to 0$ or $q \to 1$. In terms of sample space our assumption means that the density of sample-space points containing event A is uniform in the limit of a differential element of sample-space area.

[1] See, however, the detailed discussion in Section 10.2.9.

The Poisson frequency function then follows for all populations for which assumption (10.12) is valid.

Let $P\{x, n\}$ be the probability of obtaining event A, x times in n trials, so that $P\{0, n\}$ is the probability of obtaining no events A in n trials. Then the probability of obtaining no events in $n + dn$ trials is

$$P\{0, n + dn\} = P\{0, n\} \cdot [1 - P\{1, dn\}]$$

since the events are independent.[2] Using Eq. (10.12) we obtain

$$\frac{P\{0, n + dn\} - P\{0, n\}}{dn} = -P\{0, n\} \cdot \lambda$$

or

$$-\frac{dP\{0, n\}}{dn} = P\{0, n\} \cdot \lambda,$$

which has the solution

$$\ln P\{0, n\} = -n\lambda$$

$$P\{0, n\} = e^{-n\lambda} \tag{10.13}$$

and use has been made of the initial condition that for $n = 0$

$$P\{0, 0\} = 1.$$

In a similar manner we obtain

$$P\{1, n + dn\} = P\{1, n\}P\{0, dn\} + P\{0, n\}P\{1, dn\},$$

where the two possible *either* probabilities are summed. Making use again of Eq. (10.12), we may write the above result as

$$P\{1, n + dn\} = P\{1, n\} \cdot [1 - \lambda dn] + P\{0, n\} \cdot \lambda dn$$

by further transforming and using Eq. (10.13) as well,

$$\frac{dP\{1, n\}}{dn} + \lambda P\{1, n\} - \lambda e^{-n\lambda} = 0.$$

The solution of this linear first-order equation is straightforward, leading to

$$P\{1, n\} = e^{-n\lambda}\left[\int e^{n\lambda} \lambda e^{-n\lambda} dn + C\right] = (n\lambda)e^{-n\lambda}, \tag{10.14}$$

making use of the initial condition $P\{1, 0\} = 0$.

[2] Since the increase in the number of trials dn is differential, the possibility of obtaining more than one event in dn is excluded.

In general the following recursion formula holds

$$\frac{dP\{x,n\}}{dn} + \lambda P\{x,n\} - \lambda P\{(x-1),n\} = 0,$$

which is satisfied by

$$f(x) = P\{x,n\} = \frac{(\lambda n)^x e^{-n\lambda}}{x!}, \tag{10.15}$$

as can be verified by substitution.

Thus Eq. (10.11) has been proven, and we can identify the proportionality constant λ as the probability that event A will occur in one trial.[3] As pointed out before, however, it is only the product $y = \lambda n = pn$ that may be properly defined: it is the theoretical mean of the discrete random variable x when the same (large) number of n trials is repeated many times.

Equation (10.11) correctly fulfills the normalization requirement

$$\sum_{x=0}^{n=\infty} f(x) = e^{-y} \sum_{x=0}^{\infty} \frac{y^x}{x!} = e^{-y} e^{y} = 1.$$

It is shown in Section 10.2.9 that Eq. (10.11) is the limiting form of Eq. (10.6) when $p \to 0$ and $n \to \infty$.

10.2.7. Moments of the Poisson Frequency Function

Following the approach used in Section 10.2.5, the moments of the Poisson frequency function will be obtained by direct evaluation of the defining equations; note that as $n \to \infty$ the upper limit of x is also ∞:

$$\mu = \mu_1' = \sum_{x=0}^{x=n\to\infty} x \frac{y^x e^{-y}}{x!} = \sum_{x=1}^{\infty} \frac{y^x e^{-y}}{(x-1)!}$$

$$= e^{-y} y \sum_{x=1}^{\infty} \frac{y^{(x-1)}}{(x-1)!} = e^{-y} y e^{y} = y.$$

Thus

$$\mu = y \tag{10.16}$$

[3] $P\{1,1\} = \lambda e^{-\lambda} \to \lambda$ when $\lambda \ll 1$.

as expected from our previous discussion. We see that through Eq. (10.16) we obtain the physical significance for the parameter y. Further,

$$\mu_2' = \sum_{x=0}^{\infty} x^2 \frac{y^x e^{-y}}{x!} = \sum_{x=0}^{\infty} \left(x(x-1) \frac{y^x e^{-y}}{x!} \right) + y$$

$$= e^{-y} \sum_{x=2}^{\infty} \left(\frac{y^x}{(x-2)!} \right) + y = e^{-y} y^2 \sum_{x=2}^{\infty} \frac{y^{(x-2)}}{(x-2)!} + y = y^2 + y,$$

and using Eq. (10.5) we obtain

$$\mu_2 = \sigma^2 = \mu_2' - \mu^2 = y^2 + y - y^2 = y.$$

Thus

$$\sigma = \sqrt{y}. \tag{10.17}$$

The close analogy of Eq. (10.16) to Eq. (10.8) and of Eq. (10.17) to Eq. (10.9) should be clear; also the derivation of these equations is completely analogous.

10.2.8. The Gaussian or Normal Frequency Function and Its Moments

This is indeed a most important frequency function because (a) it is a limiting case that many frequency functions approach; (b) the distribution of most physical observables is satisfactorily described by it; and (c) measurements containing *random* errors are distributed normally about the true value of the measured quantity.

The Gaussian distribution gives the frequency of the continuous random variable x in terms of two parameters a and b, which are the first and second moments of the frequency function. In its normalized form, the Gaussian distribution is given by

$$f(x)\,dx = \frac{1}{b\sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{x-a}{b} \right)^2 \right] dx \tag{10.18}$$

and is shown in Fig. 10.5. The range of the variable x is from $-\infty$ to $+\infty$. In order to show the normalization of Eq. (10.18), as well as to find the moments, it is useful to know the values of the integral of $x^n e^{-ax^2}$,

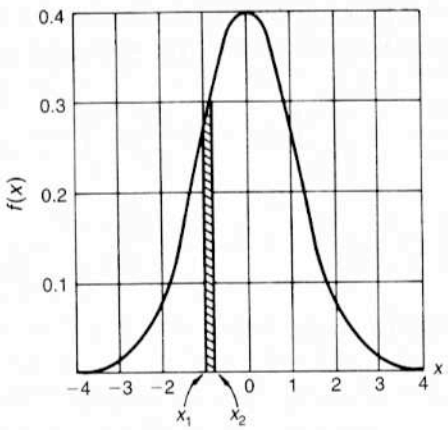

FIGURE 10.5 The Gaussian frequency function normalized to zero mean and unit variance $f(x)\,dx = (1/\sqrt{2\pi})e^{-x^2/2}\,dx$. Note that the probability of finding a value of x between x_1 and x_2 is proportional to the corresponding area under the Gaussian.

TABLE 10.2 Value of the Integral $f(n) = \int_0^\infty x^n \exp(-ax^2)\,dx$

n	$f(n)$	n	$f(n)$
0	$\frac{1}{2}\sqrt{\pi/a}$	1	$1/2a$
2	$\frac{1}{4}\sqrt{\pi/a^3}$	3	$1/2a^2$
4	$\frac{3}{8}\sqrt{\pi/a^5}$	5	$1/a^3$

$$f(n) = \int_{-\infty}^{\infty} x^n \exp(-ax^2)\,dx = \begin{cases} 2f(n) & \text{when } n \text{ is even} \\ 0 & \text{when } n \text{ is odd} \end{cases}$$

which are summarized in Table 10.2. To obtain the moments we proceed as before

$$\mu = \mu_1' = \frac{1}{b\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \exp\left[-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right] dx.$$

We let $x = tb + a$, $dx = b\,dt$; thus

$$\mu = \frac{1}{\sqrt{2\pi}}\left[\int_{-\infty}^{+\infty} bte^{-(t^2/2)}\,dt + \int_{-\infty}^{+\infty} ae^{-(t^2/2)}\,dt\right].$$

According to Table 10.2, integrals with odd powers of t vanish, thus

$$\mu = a. \tag{10.19}$$

Similarly

$$\mu_2' = \frac{1}{b\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 \exp\left[-\frac{1}{2}\left(\frac{x-a}{b}\right)^2\right] dx$$

with the same substitution

$$\mu_2' = \frac{1}{\sqrt{2\pi}}\left[\int_{-\infty}^{+\infty} b^2t^2e^{-(t^2/2)}\,dt + \int_{-\infty}^{+\infty} 2abte^{-(t^2/2)}\,dt + \int_{-\infty}^{+\infty} a^2e^{-(t^2/2)}\,dt\right],$$

so that by using Table 10.2 we obtain

$$\mu_2' = \frac{1}{\sqrt{2\pi}}\left[b^2\frac{1}{2}\sqrt{8\pi} + a^2\sqrt{2\pi}\right] = a^2 + b^2$$

and, using Eq. (10.5),

$$\mu_2 = \sigma^2 = \mu_2' - \mu^2 = b^2.$$

Thus

$$\sigma = b. \tag{10.20}$$

We see that through Eqs. (10.19) and (10.20), we obtain the physical significance of the parameters a and b of Eq. (10.18). Thus, Eq. (10.18) takes the form

$$f(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\mu-x}{\sigma}\right)^2\right] dx. \tag{10.21}$$

It is sometimes useful to transform the random variable linearly so as to obtain a frequency function with zero mean and unit standard deviation; the transformation is

$$y = \frac{x-\mu}{\sigma}; \qquad dy = \frac{dx}{\sigma},$$

and Eq. (10.18) becomes (as shown in Fig. 10.5)

$$f(y)dy = \frac{1}{\sqrt{2\pi}}e^{-(y^2/2)}\,dy. \tag{10.22}$$

10.2.9. The Gaussian Frequency Function as a Limiting Case

In the previous section we gave Eq. (10.18) without proof. We will now show that it can be obtained from the binomial frequency function, Eq. (10.6), in the limit of $n \to$ large and $|np - x| \ll np$.

Consider Eq. (10.6):

$$f(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}.$$

If $n \to \infty$ but $np \to \mu$ remains finite, we may write

$$f(x) = \frac{n(n-1)\cdots(n-x+1)}{n^x} \cdot \frac{(np)^x}{x!} \cdot (1-p)^{n-x}$$

$$f(x) = \frac{1[1-(1/n)]\cdots[1-(x-1)/n]}{(1-p)^x} \cdot \frac{(np)^x}{x!} \cdot (1-p)^n. \quad (10.23)$$

However,

$$(1-p)^n = [(1-p)^{-(1/p)}]^{-np} \to e^{-\mu}$$

since from the definition of e,

$$\lim_{z\to 0}(1+z)^{1/z} = e$$

and in the present case we have $p \to 0$. Further

$$\lim_{n\to\infty} \frac{1[1-(1/n)]\cdots[1-(x-1)/n]}{(1-p)^x} = 1$$

because $p \to 0$ and x is finite; by substituting the last two expressions into Eq. (10.23) we obtain the Poisson frequency function, Eq. (10.11):

$$f(x) = \frac{\mu^x e^{-\mu}}{x!}.$$

We now use the further condition that x be a continuous variable and $|np - x| \ll np$, namely, its deviations from the mean μ be small; then the following approximate expression is valid:

$$\ln \frac{\mu}{x} = \ln\left(1 + \frac{\mu - x}{x}\right) = \left(\frac{\mu - x}{x}\right) - \frac{1}{2}\left(\frac{\mu - x}{x}\right)^2 + \cdots.$$

Hence

$$\frac{\mu}{x} \approx \exp\left(\frac{\mu - x}{x}\right) \exp\left[-\frac{1}{2}\left(\frac{\mu - x}{x}\right)^2\right]$$

and

$$\mu^x \approx x^x \exp(\mu - x) \exp\left[-\frac{1}{2}\frac{(\mu - x)^2}{x}\right].$$

From Stirling's formula we have

$$x! \simeq \sqrt{2\pi x} x^x e^{-x}$$

and by substituting $(\mu)^x$ and $x!$ into Eq. (10.11) we obtain

$$f(x) = \frac{\mu^x e^{-\mu}}{x!} = \frac{e^{-\mu} x^x e^{(\mu - x)} \exp\left\{-\frac{1}{2}[(\mu - x)^2/x]\right\}}{\sqrt{2\pi x} x^x e^{-x}}$$

$$= \frac{1}{\sqrt{2\pi x}} \exp\left[-\frac{1}{2}\left(\frac{\mu - x}{\sqrt{x}}\right)^2\right]. \quad (10.24)$$

Thus the binomial frequency function in its limit approaches a Gaussian frequency function with

$$\begin{array}{ll} \text{mean} & \mu = np \\ \text{standard deviation} & \sigma = \sqrt{x} \approx \sqrt{npq}, \end{array} \quad (10.25)$$

where $x \approx npq$ follows from $|\mu - x| \ll \mu$ and $p \to 0$. From Eq. (10.25) we see that the moments of the limiting Gaussian frequency function are the limits of the moments of the original binomial frequency function.

10.2.10. Properties of the Gaussian Frequency Function

Let us now interpret the frequency function given by Eq. (10.18). We could refer to our original example of obtaining event A, x times when a choice between A or B is made n times; x then can vary from 0 to n in integer values. It is easier, however, to consider the measurement with a ruler of the length of a rod; we let the continuous random variable x represent the result of *one* measurement. If the true length of the rod is x_0, Eq. (10.18) specifies that a result between x and $x + dx$ will be obtained

with a frequency

$$f(x)\,dx = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_0 - x}{\sigma}\right)^2\right] dx. \tag{10.26}$$

One may also say that the probability that the measurement will "yield a result x" between x and $x + dx$ is given by Eq. (10.26). In simpler words, if N measurements are performed, a result between x_1 and x_2 is likely to be obtained in $n(x_1, x_2)$ of these measurements, where

$$n(x_1, x_2) = N \cdot F(x_1, x_2) = \frac{N}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} \exp\left[-\frac{1}{2}\left(\frac{x_0 - x}{\sigma}\right)^2\right] dx \tag{10.27}$$

as shown in Fig. 10.5.

Note that in Eqs. (10.26) and (10.27) the standard deviation σ is determined by the conditions of the measurement. The applicability of the Gaussian distribution to the results obtained from such measurements lies in the fact that: (a) n, the number of (least) divisions of the ruler, is large and (b) the errors in measurement $|x_0 - x|$ are small as compared to x.

In Table 10.3 are given the values of $f(x)$ and its integral, $F(c)$, for the normalized Gaussian function (Eq. (10.22)).

From Table 10.3, for example, we see that half of the measurements do yield a result x between

$$x_0 - 0.69\sigma < x < x_0 + 0.69\sigma$$

or that only 2.23% of the results may yield x, such that

$$x > x_0 + 2\sigma.$$

TABLE 10.3 Some Numerical Values of the Normalized Gaussian Function

$f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$	$F(-c, c) = \int_{-c}^{+c} f(x)\,dx$
$f(0) = 0.3989$	$F(-1, 1) = 0.6826$
$f(1) = f(-1) = 0.2420$	$F(-2, 2) = 0.9554$
$f(2) = f(-2) = 0.0540$	$F(-3, 3) = 0.9974$
	$F(-0.69, 0.69) = 0.5000$

As another example we see that a result x in the small interval Δx about x_0, will be obtained $(0.3989)/(0.0540) = 7.4$ times more frequently than a result in the same small interval Δx about $x_0 + 2\sigma$.

10.3. ESTIMATION OF PARAMETERS AND FITTING OF DATA

In Section 10.1 the basic definitions were given; in Section 10.2, analytic expressions for some frequency functions were obtained. We will now see how statistics can be applied to the interpretation of a measurement or an experiment.

We can consider one or more measurements to form a sample of a population that obeys a certain frequency function; we are then faced with one of two estimation problems:

(a) Given the frequency function and its parameters, what is the probability of obtaining from a measurement the result x?

(b) Given the result x of a measurement, what are the parameters of the frequency function (or the frequency function itself)?

In physics we are usually faced with estimation of type (b), since a set of experimental data are obtained, and it is then desired to reduce them to a few parameters that should describe the whole population and therefore, also any new measurement that may be performed.

There are several methods for obtaining "estimators" to an unknown parameter. Some of these methods are almost subconsciously applied, but most of them can be derived from the principle of "maximum likelihood" introduced by R. A. Fisher in 1920.

10.3.1. Maximum Likelihood

To apply this principle we must have knowledge of the normalized frequency functions of the variables x_i that form the data,

$$f(x_i, \theta),$$

where θ is the parameter to be estimated and upon which the frequency function depends. We may then form the product of the frequency functions for all observed variables,

$$\mathcal{L}(x_1, x_2, \ldots x_n, \theta) = f(x_1, \theta) f(x_2, \theta) \cdots f(x_n, \theta), \tag{10.28}$$

which is called the likelihood function for the parameter θ (note that $\mathcal{L}$ is *not* a frequency function for the parameter θ). The theorem of maximum likelihood then states that the value of θ, θ^*, that maximizes $\mathcal{L}$ (for the set of observed data) is the best estimator of θ:

$$\left.\frac{\partial \mathcal{L}(x_1, x_2, \ldots x_n, \theta)}{\partial \theta}\right|_{\theta=\theta^*} = 0.$$

In practice, it is almost always convenient to work with the logarithm of $\mathcal{L}$, since when $W = \log \mathcal{L}$ is maximum, so will also be $\mathcal{L}$.

As an example, we consider a set of n data x_i that obey a normal frequency function about a, with a standard deviation σ; let us seek the best value for the parameter a:

$$f(x_i, a) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{a - x_i}{\sigma}\right)^2\right]. \tag{10.29}$$

Then

$$\mathcal{L} = \prod_{i=1}^{n} f(x_i, a)$$

and

$$W = \log \mathcal{L} = -n \log\left(\sigma\sqrt{2\pi}\right) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{a - x_i}{\sigma}\right)^2. \tag{10.30}$$

$$\frac{\partial W}{\partial a} = -\sum_{i=1}^{n}\frac{a - x_i}{\sigma^2}.$$

Setting $(\partial W)/(\partial a) = 0$ leads to the estimator a^*;

$$\sum_{i}^{n}\frac{a^* - x_i}{\sigma^2} = 0 \qquad \frac{na^*}{\sigma^2} - \sum_{i}^{n}\frac{x_i}{\sigma^2} = 0$$

or

$$a^* = \frac{1}{n}\sum_{1}^{n} x_i. \tag{10.31}$$

Thus if a set of measurements is distributed *normally*, the best estimator for the true value of the parameter is the *mean* of the measurements (first moment).

Similarly we may obtain the estimator, σ^*, for σ, by differentiating Eq. (10.30) with respect to σ

$$\frac{\partial W}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{i}^{n}\left[\left(\frac{a - x_i}{\sigma}\right)\left(\frac{a - x_i}{\sigma^2}\right)\right]$$

and setting $\partial W/\partial\sigma = 0$ gives

$$(\sigma^*)^2 = \frac{1}{n}\sum_{i}^{n}(a - x_i)^2, \tag{10.32}$$

where, in Eq. (10.32), a should be replaced by its estimator a^* given by Eq. (10.31). Again we obtain the familiar result that the best estimator for the standard deviation of the theoretical frequency function is given by the second moment (about the mean) of the observed measurements.

The principle of maximum likelihood can be further extended to give the variance S^2 of the estimator θ^*; that is, if the determination of estimators θ^* is repeated, the values so obtained will have a standard deviation S, where

$$\frac{1}{S^2} = -\frac{\partial^2 W}{\partial \theta^2}. \tag{10.33}$$

We may apply Eq. (10.33) to our sample of measurements that obeys a normal frequency function, where W was given by Eq. (10.30). We obtain

$$\frac{1}{S^2} = -\frac{\partial^2 W}{\partial a^2} = \sum_{i}^{n}\frac{1}{\sigma^2} = \frac{n}{\sigma^2}.$$

Thus the standard deviation of the estimator will be

$$S = \frac{\sigma}{\sqrt{n}}, \tag{10.34}$$

where n is the number of measurements used for obtaining each estimator. Equation (10.34) is a well-known result that we will obtain again when we discuss the combination of errors in Section 10.4.

10.3.2. The Least-Squares Method

Until now we have discussed the case where all n measurements are made on the same physical quantity whose true value is a, for example, the data of

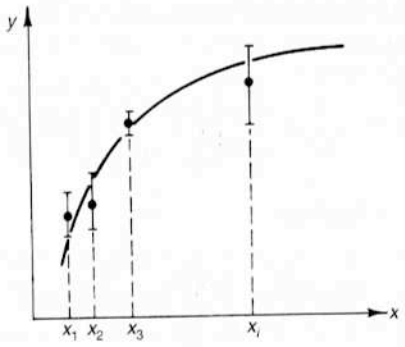

FIGURE 10.6 Least-squares fit of a two-dimensional curve to a set of data points obtained for different values of x. Note that each data point has associated with it a different error as indicated by the flags; this is taken into account when forming the least-squares sum.

Eq. (10.29). However, consider now a set of measurements yielding values $y_1, y_2, \ldots, y_n$ depending on another variable x; the corresponding true values of y, which we designate by $\bar{y}$, are assumed to be a function of x and of one or more parameters α_ν common to the whole sample. Thus we write

$$\bar{y}_i = y(x_i; a_\alpha, \ldots, a_\nu). \tag{10.35}$$

Further, each measurement y_i has associated with it a standard deviation σ_i, which is not the same for each point. This situation is shown in Fig. 10.6.

It is possible that the form of Eq. (10.35) is known or may be correctly inferred from the physics of the process under investigation, in which case the estimation is reduced to finding the best estimators for the parameters a_ν. If, however, the form of Eq. (10.35) is not known, various functional relationships must be assumed, for example, a polynomial of order k. We then speak of fitting a curve to the data. Even though special techniques are developed in Section 10.3.4 to ascertain which curve fits best, the following discussion is generally applicable.

The method of least squares follows directly from the assumption that each individual measurement y_i is a member of a *Gaussian* population with a mean given by the true value of y_i, $\bar{y}(x_i; a_\lambda)$; for the standard deviation of this Gaussian we use the experimental error σ_i of each measurement. Then in analogy to Eq. (10.29) we write for the frequency function of y_i

$$f(y_i; x_i; a_\lambda) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left[\frac{y_i - \bar{y}(x_i; a_\lambda)}{\sigma_i}\right]^2\right\}, \tag{10.36}$$

and in analogy with Eq. (10.28) we form the likelihood function

$$\mathcal{L}(y_1 \cdots y_n; x_1 \cdots x_n; a_\lambda) = \prod_{i=1}^{n} f(y_i; x_i; a_\lambda).$$

We seek the estimators a_λ^* that maximize this function, or its logarithm W

$$W = \log \mathcal{L}$$

$$= -\sum_{i=1}^{n} \log\left(\sigma_i \sqrt{2\pi}\right) - \frac{1}{2}\sum_{i=1}^{n}\left[\frac{y_i - \bar{y}(x_i; a_\lambda)}{\sigma_i}\right]^2. \tag{10.37}$$

Since the values of σ_i are fixed by the measurement, the estimators a_λ^* are those values of a_λ that *minimize* the sum

$$\mathcal{M} = \sum_{i=1}^{n} \frac{[y_i - \bar{y}(x_i; a_\lambda)]^2}{\sigma_i^2}, \tag{10.38}$$

that is, those that give the "least-squares sum." They are obtained by solving the simultaneous equations

$$\frac{\partial \mathcal{M}}{\partial a_\lambda} = 0 \qquad \lambda = 1 \text{ to } \nu.$$

10.3.3. Application of the Least-Squares Method to a Linear Functional Dependence

The simplest case of functional dependence $y(x)$ is the linear one:

$$y = ax + b.$$

If we assume that every measurement y_i has the *same standard deviation* (statistical weight), we may obtain the estimators a^* and b^* that minimize Eq. (10.38) in closed form.

Since $\sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma$, instead of Eq. (10.38) we need only minimize

$$\mathcal{R} = \sum_{i=1}^{n} [y_i - (a + bx_i)]^2. \tag{10.39}$$

Hence

$$\frac{\partial \mathcal{R}}{\partial a} = -2\sum_{i=1}^{n}[y_i - (a + bx_i)] = 0$$
$$\frac{\partial \mathcal{R}}{\partial b} = -2\sum_{i=1}^{n}\{[y_i - (a + bx_i)]x_i\} = 0, \tag{10.40}$$

which after some manipulation[4] leads to

$$a^* = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum (x_i y_i)}{n\sum x_i^2 - \sum x_i \sum x_i}$$
$$b^* = \frac{n\sum (x_i y_i) - \sum y_i \sum x_i}{n\sum x_i^2 - \sum x_i \sum x_i}. \tag{10.41}$$

The standard deviations for the above estimators may be obtained by an extension of Eq. (10.33), which now yields a symmetric square matrix

$$\mathbf{H}_{\lambda\nu} = -\frac{\partial^2 W}{\partial a_\lambda \partial a_\mu} = \frac{1}{2\sigma^2}\frac{\partial^2 \mathcal{M}}{\partial a_\lambda \partial a_\mu}. \tag{10.42}$$

The elements of the *inverse* matrix give the variance of the estimators a^*. A complete discussion of this error matrix is given in Section 10.4; suffice it to say here that the usually given expressions (Eqs. (10.43)) for the standard deviation of the estimators (Eqs. (10.41)) are the square roots of the diagonal elements of $\mathbf{H}^{-1}$ (see Eq. (10.63)). We then obtain

$$\sigma_{a^*} = \sqrt{(\mathbf{H}^{-1})_{aa}} = \sigma\sqrt{\frac{\sum x_i^2}{n\sum x_i^2 - \sum x_i \sum x_i}}$$
$$\sigma_{b^*} = \sqrt{(\mathbf{H}^{-1})_{bb}} = \sigma\sqrt{\frac{n}{n\sum x_i^2 - \sum x_i \sum x_i}}. \tag{10.43}$$

In case $\sigma_1 \neq \sigma_2 \neq \cdots \neq \sigma_n$, it is $\mathcal{M}$ and not $\mathcal{R}$ that must be minimized.

Clearly, such calculations are best done using computer programs. In fact, many packages and self-contained programs that are designed to handle these kinds of problems are available (both commercially and

[4]Note that the second of the above equations is by no means equal to the first one multiplied by x_i.

through "shareware"). In this book, we default to MATLAB (see Appendix B), which is in fact well suited for dealing with problems formulated in terms of matrices. For the problem of linear (or, more generally, polynomial) function fitting with equally weighted data points, MATLAB provides the polyfit utility for exactly this purpose.

For more general problems, the reader is referred to other textbooks on the subject of data analysis. For example, the problem of linear fitting with unequally weighted data points is discussed in Chapter 5 of *Numerical Methods for Physics*, 2nd ed., by Alejandro Garcia (Prentice-Hall, Englewood Cliffs, NJ, 2000). A program linreg for this task, is described and the code is available online from the publisher as a MATLAB m-file, as well as in the languages C++ and FORTRAN.

10.3.4. Goodness of Fit; the χ^2 Distribution

We have seen how the least-squares method, as a consequence of the principle of maximum likelihood, may be used to fit a curve to a set of data. Once the curve has been found, however, the necessity to ascertain quantitatively how good the fit is arises. This is important especially if the functional dependence is not known, a poor fit might indicate the necessity for fitting with a curve of higher order, or a poor fit might indicate inconsistencies in the data.

Similarly, we may wish to test whether a certain hypothesis is supported by the data, in which case the goodness of the fit may establish the level of confidence with which the hypothesis should be accepted.

Let us first suppose that we know the true functional relationship of y to x, that is, $\bar{y}(x) = f(x)$; we may then form the least-squares sum

$$\mathcal{M} = \sum_{i=1}^{n}\frac{[y_i - \bar{y}(x_i)]^2}{\sigma_i^2}. \tag{10.38}$$

The range of $\mathcal{M}$ is $0 < \mathcal{M} < +\infty$ but we would be surprised if $\mathcal{M} = 0$ and would be equally surprised if $\mathcal{M}$ was extremely large. Thus we have already a quantitative indication as to how well the data fit the known (or assumed) curve $y = f(x)$.

If a new set of data pertaining to the same experimental situation is obtained, and Eq. (10.38) is again formed, a new value $\mathcal{M}$ will result. Clearly, if enough such measurements are repeated, each time yielding a value for $\mathcal{M}$, we will obtain the frequency function for $\mathcal{M}$. Once the

frequency function is known, it is then easy to tell what the probability of obtaining a specific $\mathcal{M}$ is. We may, for example, calculate that in 95% of the cases $\mathcal{M} < \mathcal{M}_0$; if then a specific set of data yields $\mathcal{M}_s \geq \mathcal{M}_0$, we know that such data should be obtained only in 5% of the experiments and can therefore be rejected.

Obtaining the frequency function for the least-squares sum in this way is obviously impractical. Nevertheless, it is true that the *distribution* of $\mathcal{M}$ is *independent* of the curve $y = f(x)$ and of σ_i, and can therefore be calculated theoretically; it depends only on the number n of points that are compared, and is called the χ^2 distribution (pronounced "chi-squared")

$$f(\mathcal{M})\,d\mathcal{M} = \frac{\mathcal{M}^{(\nu/2)-1}\exp(-\mathcal{M}/2)}{2^{\nu/2}\Gamma(\nu/2)}\,d\mathcal{M} \equiv f(\chi^2)\,d\chi^2, \qquad (10.44)$$

where ν is the number of "degrees of freedom" of $\mathcal{M}$. In the present case we set

$$\nu = n$$

because this is the number of truly independent points being compared. In Eq. (10.44) $\Gamma(x)$ is the "gamma function," which for positive integer arguments[5] is simply

$$\Gamma(n) = (n-1)!.$$

Consider next that $y = f(x)$ is not known, but that a two-parameter curve is fitted to n data points, yielding estimators a^* and b^*. Then one forms again the least-squares sum $\mathcal{M}$ using $\bar{y} = f(x; a^*, b^*)$ but now the frequency function for the $\mathcal{M}$ values is given by Eq. (10.44) with the n degrees of freedom reduced by the number of estimators obtained from the data, that is,

$$\nu = n - 2.$$

The χ^2 distribution may also be used for comparing the frequency of occurrence of a class of events with the theoretical frequency (function). Let us consider, for example, 100 measurements of a radioactive sample, and divide the sample into seven classes, with mean value $\overline{N} = 85$ counts/min

[5]The general definition of the gamma function is

$$\Gamma(z) = \int_0^\infty t^{z-1}\exp(-t)\,dt;$$

for more details see any text on advanced calculus.

TABLE 10.4 Observed and Expected Frequencies of the Results of 100 Measurements of a Radioactive Sample

Class	0–75	75–79	79–83	83–87	87–91	91–95	95–∞	Counts/min
o_i	15	11	15	15	18	12	14	Observed freq
e_i	13	12	15	16	16	13	15	Expected freq
$(e_i - o_i)^2/e_i^2$	0.307	0.083	0	0.062	0.25	0.077	0.067	χ^2

and approximately equal expected frequencies; the resulting frequency of the experimental observations o_i in each class is given in Table 10.4. Next we obtain from the data the estimators for the parameters of a Gaussian (1) $\mu^* = \overline{N}$, (2) $\sigma^* = \sqrt{\overline{N}}$, and (3) the overall normalization, namely, $\sum o_i = \sum e_i$; thus the degrees of freedom of χ^2 are four, corresponding to seven classes less three estimators. From the Gaussian distribution we calculate the expected frequencies e_i for each class; they are also given in Table 10.4.

In complete analogy with the least-squares sum, Eq. (10.38), we form the χ^2 sum

$$\chi^2 = \sum_{i=1}^{n} \frac{(e_i - o_i)^2}{e_i^2}.$$

Note that χ^2 is now a discrete variable, since frequencies of classes are compared; however, Eq. (10.44), which holds for a continuously variable χ^2, is valid provided the number of classes $n \gtrsim 5$ and the expected frequencies $e_i \geq 5$.

For this experiment we obtain

$$\chi^2 = 0.846,$$

and we explained before that $\nu = 4$. From a table of the χ^2 distribution we find that in 93% of the cases the χ^2 distribution would be larger than the result obtained here. Thus one may suspect that the data are "too good" a fit to the estimated Gaussian.

The χ^2 distribution of Eq. (10.44) for different degrees of freedom is shown in Fig. 10.7. Tables of this distribution may be found in reference manuals, or easily calculated in any number of computer programs. It should not be surprising that when the number of degrees of freedom

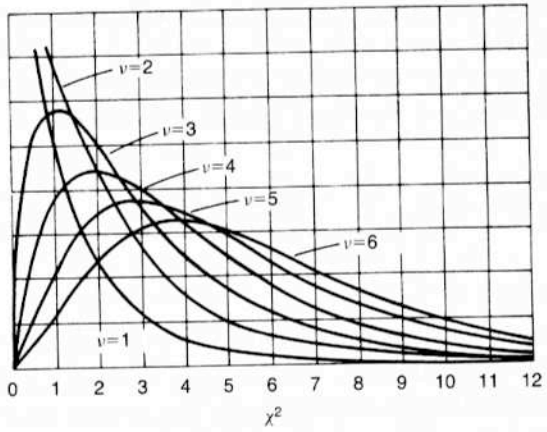

FIGURE 10.7 The frequency function for the distribution of χ^2, for different degrees of freedom. All curves are normalized to the same unit area. Note that for large ν the χ^2 distribution approaches a Gaussian.

increases $\nu > 30$, the χ^2 distribution approaches a Gaussian[6] with mean $\mu = \nu - 1/2$.

10.4. ERRORS AND THEIR PROPAGATION

10.4.1. Introduction

When we perform a measurement of a physical quantity x, it can be expected that the result obtained, x_1, will differ from x; this difference is the *error* of the measurement and consists of a *systematic* and a *random* contribution. Suppose, now, that the measurement is repeated under the same conditions n times; then the results x_n will be distributed (in most cases) normally about a mean $\bar{x}$ with a standard deviation σ. The difference between $\bar{x}$ and the true value x is then the systematic error, and the standard deviation σ of the Gaussian is a measure of the *dispersion* of the results due to the random error.

The object of the measurement, however, is the determination of the unknown true value x; since this is not possible, we seek to find whether x lies between certain limits, or whether the true value x is distributed

[6] It is really the distribution of $\sqrt{2\chi^2}$ that approaches the Gaussian with mean $\mu = \sqrt{(2\nu - 1)}$ and unit standard deviation (R. A. Fisher's approximation).

about some mean x^* with a standard deviation σ^*. Note that in a rigorous sense, this statement is incorrect, since the unknown true value x is not distributed, but is fixed; what we mean is that the probability, $x = x^*$, $x > x^*$, etc., is given by the normal frequency function with mean $\bar{x}$ and $\sigma = \mu_2$, the second moment of the measured data about their mean $\bar{x}$.

Thus, by repeating the measurement several times, it is possible in principle to circumvent the random errors because (a) a knowledge of $\bar{x}$ and σ contains all possible information about the unknown true value x, and (b) as n increases, the second moment should decrease as $1/\sqrt{n}$ and may be made arbitrarily small. On the other hand, the systematic errors cannot be extracted from a set of identical measurements. They can either be estimated by the observer or be judged from a performance of the same measurement with a different technique. Therefore, it is unadvisable to reduce the random errors much below the expected limits of the systematic errors. In what follows we will discuss only the treatment of random errors and work under the assumption that the results of the measurements follow a normal distribution.

Until now we have considered the simple case where the unknown value x is directly measured and an error σ_x can be associated with the measurement; that is, the frequency function of x depends only on one variable:

$$f(x) = \frac{a}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{1}{2}\left(\frac{\bar{x} - x}{\sigma}\right)^2\right].$$

Most frequently, however, the unknown value x is not directly measured, and we distinguish two cases:

(a) x is an explicit function of the quantities $y_1, y_2, \ldots, y_n$ that are measured and have with them associated errors $\sigma_1, \sigma_2, \ldots, \sigma_n$. Namely,

$$x = \phi(y_1, y_2, \ldots, y_n), \tag{10.45}$$

and it is desired to find the estimator x^* and its standard deviation σ_x.

(b) x is an implicit function of other unknown variables $u_1, u_2, \ldots, u_m$, and of the quantities $y_1, y_2, \ldots, y_n$ that are measured and have with them associated errors $\sigma_1, \sigma_2, \ldots, \sigma_n$. Namely,

$$\phi(x; u_1, u_2, \ldots, u_m; y_1, y_2, \ldots, y_n) = 0, \tag{10.46}$$

and it is desired to find the estimators $x^*, u_1^*, u_2^*, \ldots, u_m^*$ and the symmetric error matrix $\sigma_{ij}(i, j = 1, \ldots, m+1)$. Such an example was treated in Section 10.3.3, and we know that at least $m+1$ sets of measurements are required to obtain the $m+1$ estimators.

The techniques for obtaining the best estimators were discussed in Section 10.3. In this section we will discuss how the random error of x may be determined from knowledge of the errors of the independent variables y_n; this procedure is frequently referred to as the combination or the propagation of the errors of the measured values y_n.

10.4.2. Propagation of Errors

Let us first assume x to be an explicit function of the measured y_n as discussed previously (Section (10.4.1)):

$$x = \phi(y_1, y_2, \ldots, y_n). \tag{10.45}$$

By applying the maximum likelihood method, it can be shown that the estimator x^* is obtained by using the mean values, μ_n, of the measured y_n (provided[7] the y_n are distributed normally). Here the mean values μ_n are obtained from r different measurements

$$\mu_n = \frac{1}{r}\sum_{i=1}^{r}(y_n)^i.$$

Thus

$$x^* = \phi(\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_n) = \phi(\mu_1, \mu_2, \ldots, \mu_n). \tag{10.47}$$

Next we make a Taylor expansion of Eq. (10.45) about x^*, through first order

$$x = \phi(\mu_1, \mu_2, \ldots, \mu_n) + \left[\frac{\partial\phi}{\partial y_1}\right]_\mu (\mu_1 - y_1) + \left[\frac{\partial\phi}{\partial y_2}\right]_\mu (\mu_2 - y_2) + \cdots + \left[\frac{\partial\phi}{\partial y_n}\right]_\mu (\mu_n - y_n).$$

[7] Clearly if x is variable, all measurements y_n^i are made so as to correspond to the same point x.

where $[\partial\phi/\partial y_n]_\mu$ means evaluation of the derivative at the point about which we expand—that is, $(\mu_1, \mu_2, \ldots, \mu_n)$. We can now form the second moment of the distribution of the x^i values as they result from the observed ${y_n}^i$ values. The superscript i here refers to the r different sets of measurements:

$$\begin{aligned}
\sigma_x^2 &= \frac{1}{r}\sum_{i=1}^{r}(\bar{x} - x^i)^2 \\
&= \frac{1}{r}\sum_{i=1}^{r}\left[\left(\frac{\partial\phi}{\partial y_1}\right)_\mu(\mu_1 - y_1^i) + \cdots + \left(\frac{\partial\phi}{\partial y_n}\right)_\mu(\mu_n - y_n^i)\right]^2 \\
&= \left(\frac{\partial\phi}{\partial y_1}\right)_\mu^2 \frac{1}{r}\sum_{i=1}^{r}(\mu_1 - y_1^i)^2 + \left(\frac{\partial\phi}{\partial y_2}\right)_\mu^2 \frac{1}{r}\sum_{i=1}^{r}(\mu_2 - y_2^i)^2 + \cdots \\
&\quad + 2\left(\frac{\partial\phi}{\partial y_1}\right)_\mu\left(\frac{\partial\phi}{\partial y_2}\right)_\mu \frac{1}{r}\sum_{i=1}^{r}(\mu_1 - y_1^i)(\mu_2 - y_2^i) + \cdots \\
\sigma_x^2 &= \left(\frac{\partial\phi}{\partial y_1}\right)_\mu^2\sigma_1^2 + \left(\frac{\partial\phi}{\partial y_2}\right)_\mu^2\sigma_2^2 + \cdots + 2\left(\frac{\partial\phi}{\partial y_1}\right)_\mu\left(\frac{\partial\phi}{\partial y_2}\right)_\mu\sigma_{12}^2 + \cdots .
\end{aligned} \tag{10.48}$$

Equation (10.48) is the most general expression for the propagation of errors. If we assume that the errors are uncorrelated, namely, $\sigma_{ij} = 0$ when $i \neq j$, we can obtain the results for the simplest functional relationships:

(a) Addition

$$\begin{aligned}
x &= y_1 + y_2 + \cdots + y_n \\
\sigma_x &= \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2}.
\end{aligned} \tag{10.49}$$

(b) Subtraction

$$\begin{aligned}
x &= y_1 - y_2 \\
\sigma_x &= \sqrt{\sigma_1^2 + \sigma_2^2}.
\end{aligned} \tag{10.50}$$

(c) Multiplication

$$x = y_1 \times y_2 \times \cdots \times y_n$$

$$\left(\frac{\partial\phi}{\partial y_1}\right)_\mu = \mu_2 \times \cdots \mu_n$$

$$\sigma_x = \sqrt{\sigma_1^2 \times (\mu_2 \cdots \mu_n)^2 + \cdots + \sigma_n^2 \times (\mu_1\mu_2 \cdots)^2} \qquad (10.51)$$

$$= x^* \sqrt{\left(\frac{\sigma_1}{\mu_1}\right)^2 + \left(\frac{\sigma_2}{\mu_2}\right)^2 + \cdots + \left(\frac{\sigma_n}{\mu_n}\right)^2}.$$

(d) Division

$$x = \frac{y_1}{y_2}$$

$$\left(\frac{\partial\phi}{\partial y_1}\right)_\mu = \frac{1}{\mu_2}, \qquad \left(\frac{\partial\phi}{\partial y_2}\right)_\mu = \frac{-\mu_1}{(\mu_2)^2} \qquad (10.52)$$

$$\sigma_x = \sqrt{\frac{\sigma_1^2}{(\mu_2)^2} + \frac{\sigma_2^2(\mu_1)^2}{(\mu_2)^4}} = x^* \sqrt{\left(\frac{\sigma_1}{\mu_1}\right)^2 + \left(\frac{\sigma_2}{\mu_2}\right)^2}. \qquad (10.53)$$

From the above examples we see that in general the errors are combined in quadrature—that is, it is their squares that are added. Consequently, if the error in one of the variables σ_i is large, it will dominate all other terms and the error of x, σ_x, will be almost equal to σ_i, despite good measurements made on the other independent variables.

Our simple rule for the case of addition, Eq. (10.49), may be used to obtain in a different way the result derived in Eq. (10.34). Let a variable x be measured and let the mean of a set of measurements be $\bar{x}_i$, with a standard deviation σ_i; if this set of measurements is repeated under identical conditions, a new mean result $\bar{x}_j \neq \bar{x}_i$ will be obtained, but let the standard deviations be equal, that is, $\sigma_j = \sigma_i$. If n such sets of measurements are performed, the new estimator for x will be

$$x^* = \frac{1}{n}(\bar{x}_1 + \bar{x}_2 + \cdots \bar{x}_n),$$

and thus

$$\left(\frac{\partial\phi}{\partial\bar{x}_i}\right) = \frac{1}{n}.$$

Hence, from Eq. (10.48) or (10.49),

$$\sigma_x^* = \sqrt{\left(\frac{\sigma_1}{n}\right)^2 + \left(\frac{\sigma_2}{n}\right)^2 + \cdots + \left(\frac{\sigma_n}{n}\right)^2} = \sqrt{n\frac{\sigma^2}{n^2}} = \frac{\sigma}{\sqrt{n}}. \qquad (10.54)$$

Namely, the standard deviation of the mean of n measurements of a Gaussian distribution is $\sigma/\sqrt{n}$, where σ is the standard deviation of the individual measurements.

10.4.3. Example of Calculation of Error Propagation

As an example, let us consider an experiment to determine Stefan's constant b, from the relation

$$E = bT^4,$$

where the following values of E and T were obtained with the indicated standard deviations:

T (K)	E (W/m^2)
$800(1 \pm 0.02)$	$(3.0 \pm 0.3) \times 10^4$
$1000(1 \pm 0.02)$	$(8.0 \pm 0.8) \times 10^4$
$1200(1 \pm 0.02)$	$(15.6 \pm 0.6) \times 10^4$

We wish to calculate the estimator b^* and its standard deviation σ_b.

There are two ways to proceed in this case. We either may calculate b_j^* from each of the three sets of measurements and then combine these values to obtain $b^* = \bar{b}_j^*$, but weighing each b_j^* according to its standard deviation, or we may use least squares in the observed variables E and T^4. Note that a mean of T or E of the three listed measurements makes no sense whatsoever since each measurement is made for a *different* T.

We will follow the first procedure, and we first obtain the error on T^4 from the known error on T. For this we should use the general expression, Eq. (10.48), but since $\phi = T^4$ is a function of only one variable,[8] simple differentiation gives the desired result directly

$$\frac{d\phi}{dT} = 4T^3 \qquad \frac{\Delta\phi}{\phi} = 4\frac{\Delta T}{T}. \qquad (10.55)$$

[8] If we choose to write $\phi = T \times T \times T \times T$, we may *not* apply Eq. (10.51), since these variables are correlated; use of Eq. (10.48) and $\sigma_{TT} = \sigma_T$ gives back the result of Eq. (10.55).

TABLE 10.5 An Example of a Calculation of Propagation of Errors

Set of data	T^4	$E/T^4 = b_j^*$	$\sigma(T^4)/T^4$	$\sigma(b_j)/b_j^*$
1	0.41×10^{12}	7.3×10^{-8}	0.08	0.13
2	1.0×10^{12}	8.0×10^{-8}	0.08	0.13
3	2.0×10^{12}	7.8×10^{-8}	0.04	0.06

We note from Eq. (10.54) that it is easier to work with relative errors, and we thus form Table 10.5, where

$$\frac{\sigma(b)}{b} = \sqrt{\left[\frac{\sigma(T^4)}{T^4}\right]^2 + \left[\frac{\sigma(E)}{E}\right]^2}$$

since the errors in T and E are uncorrelated.

For the best estimator of b, we will use the mean of the three measurements but weighed in inverse proportion to the square of their standard deviation (see Section 10.3.3). Thus

$$\bar{b} = \frac{1}{6}(7.3 + 8.0 + 4 \times 7.8) \times 10^{-8} = 7.75 \times 10^{-8};$$

for $\sigma(\bar{b})$ we used Eq. (10.49),

$$\sigma(\bar{b}) = \frac{1}{6}\sqrt{\sigma^2(b_1) + \sigma^2(b_2) + 4\sigma^2(b_3)}$$

or the convenient approximation

$$\frac{\sigma(\bar{b})}{\bar{b}} = \frac{1}{6}\sqrt{\left[\frac{\sigma(b_1)}{b_1}\right]^2 + \left[\frac{\sigma(b_2)}{b_2}\right]^2 + 4\left[\frac{\sigma(b_3)}{b_3}\right]^2} = 0.043,$$

so that the final result is

$$b^* = 7.75(1 \pm 0.043) \times 10^{-8}\ \mathrm{W/^\circ K^4\text{-}m^2}.$$

10.4.4. Evaluation of the Error Matrix

In the two previous sections we have discussed the case where only one unknown variable x was sought. We will now consider the random errors when several unknown variables are simultaneously estimated or measured.

When only one variable is measured, we know how to obtain from the data the second moment about the mean

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(\bar{x} - x_i)^2.$$

If now p variables are simultaneously measured in an experiment, we must form the $p(p+1)/2$ second moments about the mean; for example, if we measure x, y, and z, we must calculate the six expressions

$$\sigma_{xx} = \frac{1}{n}\sum_{i=1}^{n}(\bar{x} - x_i)(\bar{x} - x_i); \qquad \sigma_{yy} = \cdots; \qquad \sigma_{zz} = \cdots;$$

$$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{n}(\bar{x} - x_i)(\bar{y} - y_i) = \sigma_{yx}; \tag{10.56}$$

$$\sigma_{xz} = \cdots = \sigma_{zx}; \qquad \sigma_{yz} = \cdots = \sigma_{zy}.$$

(In this notation, the dimensionality of a quantity σ_{pq} is that of the product pq. Hence, σ_x^2 has the same dimensions as σ_{xx}. We avoid the notation σ_{xx}^2, etc., because it misleads one to think that σ_{xy}, for example, is positive definite.) If the distribution of the variables x, y, and z is normal, then these six moments form the symmetric error matrix; if the variables are uncorrelated, the matrix is diagonal.

Clearly, the error matrix must be known if it is desired to apply Eq. (10.48). Consider, for example, that from the measured variables x, y, and z we wish to obtain a new unknown u and its standard deviations $\sigma(u)$, where

$$u = \phi(x, y, z). \tag{10.57}$$

Then the values of σ_{ij}^2 that were obtained from the data with the help of Eq. (10.56) are substituted in Eq. (10.48) along with the partial derivatives of u, which are obtained from Eq. (10.57).

Conversely, if the frequency function of the three variables x, y, and z, and thus of u, is known,

$$f(u) = f[\phi(x, y, x)]$$

it is possible to calculate theoretically the elements of the error matrix through the usual expression

$$\mu_2'(x, y) = \iiint f(x, y, z)xy\,dx\,dy\,dz \tag{10.58}$$

or

$$\mu_2(x, y) = \iiint f(x, y, z)(\mu_x - x)(\mu_y - y)\,dx\,dy\,dz,$$

where

$$\sigma_{xy} = \mu_2(x, y), \qquad \text{etc.}$$

In most practical applications, however, it is difficult to use Eq. (10.56) or (10.58). Equation (10.56) may not be usable because the unknown variables may not be measured directly (although they are measured implicitly); also, extensive data are required to yield meaningful results, and the calculation is cumbersome. Equation (10.58) may not be usable because the multidimensional integrals are frequently too difficult to calculate. Instead, the method of maximum likelihood provides an easy way for obtaining the error matrix.

As already discussed in Section 10.3, if the set of data $x_i, y_i, \ldots, z_i$ has been measured, and the estimators for the m unknown variables $\theta_a, \theta_b, \ldots, \theta_m$ are sought, we may form the likelihood function

$$\begin{aligned}\mathcal{L}(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n, \ldots z_1, z_2, \ldots, z_n; \theta_a, \theta_b, \ldots, \theta_m) \\ = f(x_1, y_1, \ldots, z_1; \theta_a, \theta_b, \ldots, \theta_m)\, f(x_2, y_2, \ldots, z_2; \theta_a, \theta_b, \ldots, \theta_m) \cdots \\ \times f(x_n, y_n, \ldots, z_n; \theta_a, \theta_b, \ldots, \theta_m),\end{aligned}$$

where f is the frequency function of the measured variables and is usually assumed to be a product of Gaussians. Then the estimators $\theta_a^*, \theta_b^*, \ldots, \theta_m^*$ are given by the values that *simultaneously* maximize $\mathcal{L}$, namely,

$$\left.\frac{\partial \mathcal{L}}{\partial \theta_a}\right]_{\theta_a^*, \theta_b^*, \ldots, \theta_m^*} = \cdots = \left.\frac{\partial \mathcal{L}}{\partial \theta_m}\right]_{\theta_a^*, \theta_b^*, \ldots, \theta_m^*} = 0, \tag{10.59}$$

requiring the solution of m coupled equations. Equation (10.41) is a simple example of such a solution of Eq. (10.59). We note that the number of independent data points taken, n, must be larger than or equal to m.

The elements of the error matrix can be obtained from the inverse of the matrix

$$\mathbf{H}_{kl} = \left.\frac{\partial^2 W}{\partial \theta_k \partial \theta_l}\right]_{\theta_a^*, \theta_b^*, \ldots, \theta_m^*}, \tag{10.60}$$

where the second-order partial derivatives must be calculated at the values of the estimators, and $W = \log \mathcal{L}$. We have

$$\sigma_{kl} = (\mathbf{H})_{kl}^{-1},$$

where the rule for matrix inversion is

$$(\mathbf{H}^{-1})_{ij} = (-1)^{i+j} \frac{\text{Det } (ji \text{ minor of } \mathbf{H})}{\text{Det } \mathbf{H}} \tag{10.61}$$

and the minor is the matrix resulting from $\mathbf{H}$ when the jth row and ith column are removed; obviously, the inverse matrix does not exist unless Det $\mathbf{H} \neq 0$.

We will now apply this method of obtaining the error matrix to the simple example treated in Section 10.3.3. The measured variables are x and y, and estimators are sought for the variables a and b; we assume that x is known exactly and that y is distributed normally for each measurement, and related to x through

$$y = a + bx.$$

Using Eq. (10.37), we have

$$\mathcal{L} = \prod_{i=1}^{n} \left[\frac{1}{\sigma_i \sqrt{2\pi}} \exp\left\{ -\frac{1}{2\sigma_i^2} [y_i - \bar{y}(x_i; a, b)]^2 \right\} \right]$$

and

$$W = \log \mathcal{L} = -\frac{n}{2} \log(2\pi) - \sum_{i=1}^{n} \log \sigma_i - \frac{1}{2} \sum_{i=1}^{n} \left[\frac{y_i - (a + bx_i)}{\sigma_i} \right]^2.$$

To simplify the calculations we assume $\sigma_1 = \sigma_2 = \cdots = \sigma_n$, so that

$$-\frac{\partial^2 W}{\partial a^2} = \frac{n}{\sigma^2}; \qquad -\frac{\partial^2 W}{\partial a \partial b} = \frac{\sum x_i}{\sigma^2}; \qquad -\frac{\partial^2 W}{\partial b^2} = \frac{\sum x_i^2}{\sigma^2}.$$

Hence

$$\mathbf{H} = \frac{1}{\sigma^2} \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum (x_i^2) \end{bmatrix} \tag{10.62}$$

and

$$\text{Det } \mathbf{H} = \frac{1}{\sigma^2}\left[n\sum(x_i^2) - \left(\sum x_i\right)^2\right].$$

Thus

$$\sigma_{\nu\mu} = \mathbf{H}^{-1} = \frac{\sigma^2}{n\sum(x_i^2) - \left(\sum x_i\right)^2}\begin{bmatrix}\sum(x_i^2) & -\sum x_i \\ -\sum x_i & n\end{bmatrix}, \qquad (10.63)$$

which gives the results stated in Eq. (10.43); the indices ν, μ stand for a or b.

10.4.5. The Monte Carlo Method

It is clear that the calculation of the propagation of errors may become extremely involved, especially when the frequency functions of the variables cannot be expressed analytically and when intermediate processes of statistical nature take place. It is then preferable to use computer programs based on the so-called "Monte Carlo" method.

By this technique, we follow a particular event through the sequence of processes it may undergo. For each process, all possible outcomes are weighed according to the frequency function and divided into x classes of equal probability. Then, from a table of these classes, one class is selected at random: for example, by looking up a table of x random numbers. The outcome of this process is incorporated in the progress of the event until a new decision point is reached, when again random selection is made. Thus, at the end of the sequence of all processes, certain final conditions will be reached from the initial conditions with which we started and through the intermediary of the random choices made at each decision point.

We follow in this fashion several events, always starting with the same initial conditions, but because of the random choices, the final conditions will be spread over some range. If enough events have been followed through, we are able to find the frequency function of the combined process and of its parameters, namely, the mean and the standard deviation for the final conditions that result from a given set of initial conditions.

For more discussion, including examples with accompanying computer codes, the reader is referred to the material listed at the end of this chapter.

10.5. THE STATISTICS OF NUCLEAR COUNTING

In many experiments related to nuclear physics, we count the particles or photons emitted in the decay of a nucleus. Usually only a very small fraction of the total sample undergoes such decay. The decay of *one* nucleus is a completely random phenomenon, yet from the number of counts in a given time interval, we may determine the decay probability of this species of nuclei or unstable particles. We have already made use of these concepts in Chapters 8 and 9.

10.5.1. The Frequency Function for the Number of Decays

We start with the assumption that the decay of *one* nucleus is purely random and the probability (unnormalized) for decay in a time interval Δt is proportional to Δt and some constant λ with dimensions of inverse time[9]:

$$p_d = \lambda \Delta t. \qquad (10.64)$$

If we have a sample of N nuclei, since the presence of one nucleus does not affect the decay of another, the probability that *one* nucleus out of the *sample* of N nuclei will decay, in time Δt, is

$$P(1, \Delta t) = \lambda N \Delta t. \qquad (10.65)$$

Equation (10.65) is completely analogous to Eq. (10.12) of Section 10.2.6, which leads to the Poisson distribution; the only difference is that the *product* Nt of Eq. (10.65) is the equivalent of the number of trials n of Eq. (10.12). Consequently the probability (frequency function) for obtaining n decays in a time interval t is

$$P(n, t) = \frac{e^{-\lambda N t}(N\lambda t)^n}{n!}. \qquad (10.66)$$

The first moment of Eq. (10.66) (in the discrete unknown variable n), as we know from Eq. (10.16), is

$$\overline{n} = \lambda N t. \qquad (10.67)$$

[9] E. Schweidler, 1905; this assumption has been proven absolutely correct from the agreement of experiment with the deductions following from Eq. (10.64) as developed in the following paragraphs.

Since $\bar{n}/t$ is the average number of decays per unit time (the average decay rate), we find the physical significance of the constant parameter λ. That is, $N\lambda$ gives the average decay rate of the sample; N is the total number of nuclei in the sample.

Similarly, the second moment about the mean of Eq. (10.66), as we know from Eq. (10.17), is

$$\sigma^2 = \lambda N t = \bar{n}.$$

Hence the very frequently used expression,

$$\sigma = \sqrt{n}. \tag{10.68}$$

Note, however, that $\bar{n}/t = N\lambda$ is the theoretical average rate, which is usually unknown (unless λ and N are precisely known for the sample under consideration). The average rate that we measure, $R = n/t$ (counts per unit time), will, in general, differ from the true rate $N\lambda = \bar{n}/t$, but if n is large, R will be distributed normally about $N\lambda$. (See Eq. (10.66a) below.)

From the considerations of Section 10.2.9, it is clear that when the total number of observed counts n is large, Eq. (10.66) is well approximated by a Gaussian with mean $\mu = N\lambda t$ and standard deviation $\sigma = \sqrt{N\lambda t}$:

$$P(n,t) = \frac{1}{\sqrt{2\pi N\lambda t}} \exp\left[-\frac{(N\lambda t - n)^2}{2N\lambda t}\right] \tag{10.66a}$$

$$= \frac{1}{\sqrt{2\pi n}} \exp\left[-\frac{(\bar{n} - n)^2}{2n}\right]. \tag{10.66b}$$

Thus, unless we are dealing with very few counts, Gaussian statistics may be safely applied.

Finally, we summarize here some simple consequences of Eq. (10.64) for a *single* nucleus:

(a) If the probability for decay in dt is

$$p_d(dt) = \lambda\, dt,$$

(b) then the probability for not decaying (survival) in the time interval from $t = 0$ to $t = t$ is

$$p_s(t) = e^{-\lambda t}$$

(for proof see Eq. (10.13)).

(c) The probability for decay in dt at time t is

$$p_d(t, dt) = e^{-\lambda t}\lambda\, dt.$$

(d) The probability for decay in the time interval from $t = 0$ to $t = t$ is

$$p_d(t) = 1 - p_s(t) = 1 - e^{-\lambda t}.$$

Note that only (c) is properly normalized, so that

$$\int_0^\infty p_d(t)\, dt = \int_0^\infty e^{-\lambda t}\lambda\, dt = 1.$$

Expressions (b) and (d) are, correctly, always <1 and reduce to 0 and 1, respectively, as t approaches infinity. As to expression (a), we must keep in mind that it holds only for Δt such that $\lambda\Delta t \ll 1$.

10.5.2. Behavior of Large Samples

Having obtained the frequency functions, we may now examine the behavior of the total sample.[10] From Eq. (10.67) we see that given a sample of N nuclei, on the average, in a time interval Δt there will be

$$n = \lambda N \Delta t$$

decays; that is, the total sample will be decreased by an amount

$$-\Delta N = N\lambda\Delta t. \tag{10.69}$$

Equation (10.69) then leads to the differential equation for the number of nuclei in the sample

$$\frac{dN}{N} = -\lambda\, dt$$

with solution

$$N(t) = N_0 e^{-\lambda t}, \tag{10.70}$$

where N_0 is the number of nuclei at time $t = 0$. Frequently $\tau = 1/\lambda$ is used for the exponent in Eq. (10.70); τ is called the *lifetime* of that particular species of nuclei and is the time in which the population of the sample is

[10] The principles and formulas in this section have already been used in Section 8.6.

reduced to 37% $(1/e)$ of its original value. The *half-life*

$$\tau_{1/2} = \tau \left[\ln_e \frac{1}{2} \right] = 0.693\tau$$

gives the time in which the population of the sample is reduced to half its original value. Using Eq. (10.70) we find, for the *decay rate* as a function of time, that

$$\frac{dN}{dt} = R(t) = -\lambda N(t) = -\lambda N_0 e^{-\lambda t}, \tag{10.71}$$

which has the same time dependence as Eq. (10.70). Experimentally we usually measure $R(t)$ and obtain a curve as shown in Fig. (10.8); from such a plot λ may be obtained. If the sample contains two or more different species of nuclei with different decay constants $\lambda_1, \lambda_2, \ldots$, the time dependence of the decay rate is no longer the simple exponential of Eq. (10.71); instead

$$\frac{dN}{dt} = R(t) = -\lambda_1 N_0^1 e^{-\lambda_1 t} - \lambda_2 N_0^2 e^{-\lambda_2 t} - \cdots .$$

If, however, $\lambda_1 \gg \lambda_2$, then for small t (that is, $t \sim 1/\lambda_1$) $R(t)$ is dominated by the first term; for large t (for example, $t \sim 1/\lambda_2$), $R(t)$ is dominated by

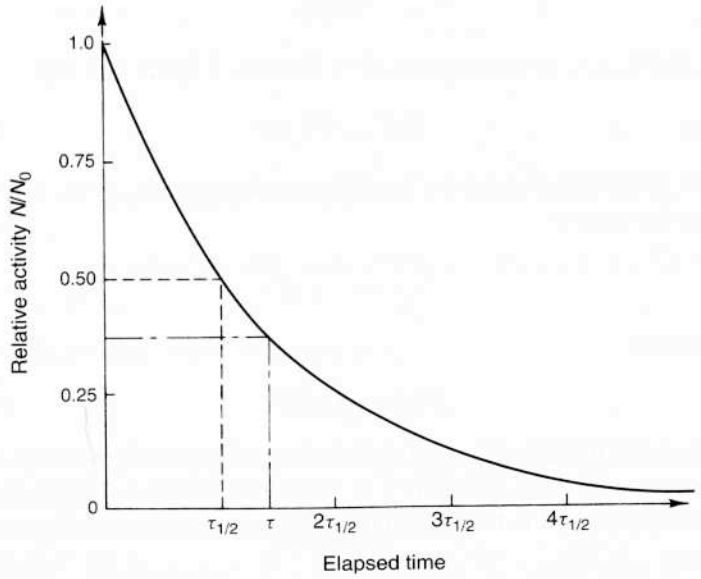

FIGURE 10.8 Exponential decay of a sample of radioactive nuclei. The abscissa is calibrated in units of the half-life of the sample; the lifetime is also indicated.

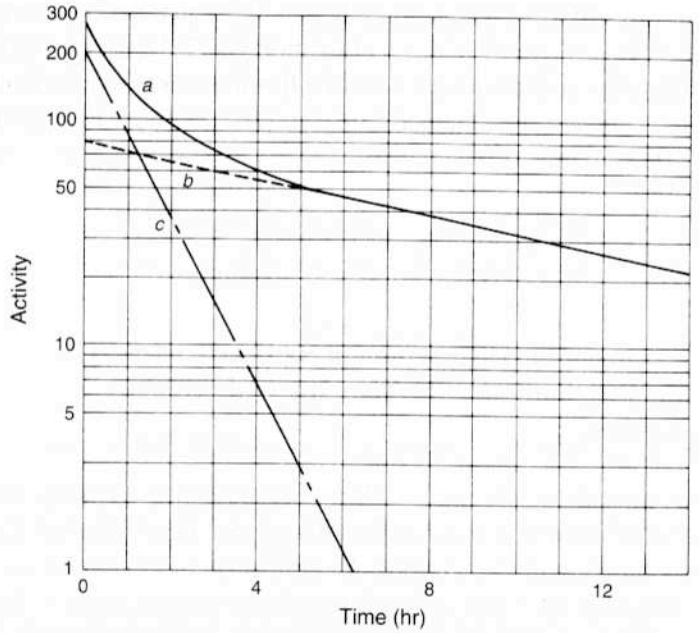

FIGURE 10.9 The decay curve for a sample containing two species of radioactive nuclei, each decaying with a different lifetime. Note that the composite decay curve a is the sum of curves b and c.

the second term. This is shown in Fig. 10.9, which gives the decay curves on a semilogarithmic plot. See also Section 8.6.3, in particular Fig. 8.37.

Another situation of interest arises when nuclei of species A decay into species B with a constant λ_A; nuclei B, however, decay in turn into species C with a constant λ_B. Let, at time $t = 0$, the number of nuclei of species A be N_0 and that of species B be 0.

Then the number of nuclei of species A as a function of time is still given by Eq. (10.70), $N_A = N_0 e^{-\lambda_A t}$. However, for the number of nuclei of species B, the following differential equation holds:

$$\frac{dN_B}{dt} = +\lambda_A N_A - \lambda_B N_B.$$

The solution of this first-order linear differential equation is straightforward, and with the initial condition $N_B(t = 0) = 0$, we have

$$N_B = N_0 \frac{\lambda_A}{\lambda_B - \lambda_A} \left[e^{-\lambda_A t} - e^{-\lambda_B t} \right]. \tag{10.72}$$

Note that Eq. (10.72) always gives $N_B > 0$, as it must be, irrespective of whether $\lambda_A > \lambda_B$ or $\lambda_B > \lambda_A$. Equation (10.72) correctly reduces to $N_B = 0$ for $t = 0$ and $t = \infty$. The two limiting cases for the decay rate from B to C can also be obtained from Eq. (10.72) if we take into account that $R_{BC}(t) = N_B\lambda_B$. Thus

$$\text{for } \lambda_B \gg \lambda_A \qquad R_{BC}(t) \approx N_0\lambda_A e^{-\lambda_A t}$$
$$\text{for } \lambda_A \gg \lambda_B \qquad R_{BC}(t) \approx N_0\lambda_B e^{-\lambda_B t}.$$

10.5.3. Testing of the Distribution of Radioactive Decay; the Distribution of the Time Intervals between Counts

It is frequently desirable to test whether a sample of counting data does indeed come from the decay of radioactive nuclei, that is, that it follows the frequency function of Eq. (10.66). A very sensitive test can be devised if we plot the distribution of the time intervals between successive decays, or every second, third, etc., decay. This method was applied to the distribution of the arrival times of cosmic rays in Section 9.4.2.

First we obtain the distribution of the time intervals between two successive decays. Let $t = 0$ when a decay occurs; we then seek the probability that no decay occurs until $t = t$, but a decay occurs within dt at $t = t$. This probability is given by Eq. (10.66) with $n = 0$, multiplied by Eq. (10.65); namely,[11]

$$P(t, dt) \equiv q_1(t)\,dt = e^{-N\lambda t} N\lambda\,dt. \tag{10.73}$$

Equation (10.73) indicates that the shortest time intervals between two counts are much more frequent than the longer ones; this is true for any *random* events, since they obey Eq. (10.64) and is shown in Fig. 9.22.

Next we consider the distribution of the time intervals between every second, third, etc., mth count. In practice this arises when the counts from the output of a "scaling circuit" are recorded. Consider, therefore, a circuit giving one output count for every m input count. If the true rate is r, then the output rate R is related to r by

$$N\lambda = r = Rm.$$

[11] Compare this equation with the probability for the decay of a single nucleus, as given in Section 10.5.1(c).

Let $t = 0$ when an output pulse arrives, and let $Q_m(t)$ be the probability that another output pulse arrives *in the time interval* t; $q_m(t)\,dt$ will then be the probability that this other output pulse arrives *at* t (between t and $t + dt$).

Another output pulse will arrive if the input counts $n \geq m$, so that

$$Q_m(t) = \sum_{n\geq m}^{\infty} P(n, t) = \sum_{n\geq m}^{\infty} \frac{(rt)^n e^{-rt}}{n!}$$
$$= 1 - \sum_{n=0}^{n=m-1} \frac{(rt)^n e^{-rt}}{n!}, \tag{10.74}$$

where the last equality follows from the normalization of Eq. (10.66)

$$\sum_{n=0}^{\infty} P(n, t) = 1.$$

Now by considering the sample space of Fig. 10.10 we see that the set of points $Q_m(t)$ is a subset of $Q_m(t + dt)$, so that any sample-space point belonging to $Q_m(t + dt)$ but not to $Q_m(t)$ represents an output count between t and $t + dt$. Thus

$$q_m(t)\,dt = Q_m(t + dt) - Q_m(t)$$

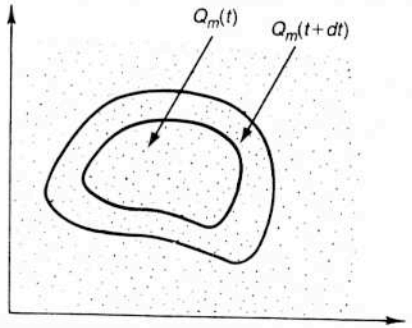

FIGURE 10.10 Sample space indicating the domain $Q_m(t)$, which contains all points corresponding to the arrival of an output count in the time interval from 0 to t after the previous count. This domain forms a subset of $Q_m(t + dt)$, which contains all points corresponding to the arrival of the output count in the time interval from 0 to $t + dt$. The arrival of the output count at t is $q_m(t) = Q_m(t + dt) - Q_m(t)$.

or

$$q_m(t) = \frac{dQ_m(t)}{dt}.$$

Taking the derivative of Eq. (10.74)

$$q_m(t) = -\sum_{n=0}^{n=m-1}\left[\frac{rn(rt)^{n-1}e^{-rt}}{n!} - \frac{r(rt)^n e^{-rt}}{n!}\right]$$

$$= r\sum_{n=0}^{n=m-1}\frac{(rt)^n e^{-rt}}{n!} - r\sum_{n=1}^{n=m-1}\frac{(rt)^{n-1}e^{-rt}}{(n-1)!}.$$

By replacing in the second sum n by $l = n - 1$, we see that only the last term of the first sum survives, so that

$$q_m(t) = r\frac{(rt)^{m-1}e^{-rt}}{(m-1)!}. \tag{10.75}$$

Equation (10.75) correctly reduces to Eq. (10.73) for $m = 1$ (since $r = N\lambda$). For $m \geq 2$, Eq. (10.75) has a maximum at $dq_m(t)/dt = 0$, or

$$[r^2(m-1)(rt)^{m-2}e^{-rt}] - [r^2(rt)^{m-1}e^{-rt}] = 0.$$

Hence $t = (m-1)/r$ and for large m, $t \to m/r = 1/R$. Thus we see that the most probable time interval is not the shortest one, but instead approaches the mean time interval between *output* counts $1/R$; that is, the scaling circuit *regularizes* the counts. Equation (10.75) is shown in Fig. 10.11 for

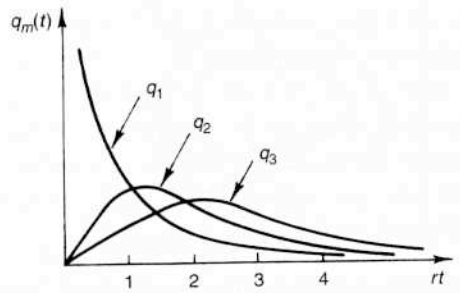

FIGURE 10.11 The probability $q_m(t)$ that the mth count will follow any original count at the time interval t. Note that the abscissa is calibrated in units of rt where r is the unscaled rate of events; for m large the curves approach a Gaussian with mean $\langle rt\rangle = m$ or $\langle t\rangle = m/r$.

different values of m. Comparison of these curves with experimental data has been presented in Section 9.4.2.

10.6. REFERENCES

There are many texts, both elementary and advanced, on the subject of statistics, data fitting, treatment of errors, and computational modeling. The references given below were consulted for the preparation of this chapter.

L. Lyons, *A Practical Guide to Data Analysis for Physical Science Students*, Cambridge Univ. Press, Cambridge, UK, 1994. A succint guide with plenty of examples.

J. R. Taylor, *An Introduction to Error Analysis*, second ed., University Science Books, Sausalito, CA, 1997. A thorough treatment with applications to the physical sciences.

B. P. Roe, *Probability and Statistics in Experimental Physics*, Springer-Verlag, Berlin, 1992. A slightly more advanced and mathematical text.

P. G. Hoel, *Introduction to Mathematical Statistics*, Wiley, New York, 1958. The presentation of Sections 10.1 and 10.2 follows Hoel closely.

A. L. Garcia, *Numerical Methods for Physics*, second ed., Prentice-Hall, Englewood Cliffs, NJ, 2000. A general text including chapters on data analysis and Monte Carlo techniques, with plenty of coding examples in MATLAB, FORTRAN, and C++.

H. Gould and J. Tobochnik, *An Introduction to Computer Simulation Methods: Applications to Physical Systems*, second ed., Addison-Wesley, Reading, MA, 1996. A text devoted to simulations, with extensive use of Monte Carlo methods, with programming examples in BASIC, FORTRAN, C, and PASCAL.