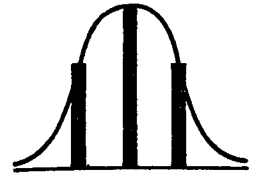
22. A machine cuts out paper rectangles at random. Each dimension is between 1 and 2 in., but all values between these limits are equally likely. What is the probability that the area of a rectangle is greater than 2 in.^{2}?

CHAPTER III



PROBABILITY DISTRIBUTIONS

We have seen in Sec. 4 how some simple probabilities can be computed from elementary considerations. For more detailed analysis of probability we need to consider more efficient ways of dealing with probabilities of whole classes of events. For this purpose we introduce the concept of a *probability distribution*.

6 | The Meaning of a Probability Distribution

To introduce the idea of a probability distribution, suppose that we flip 10 pennies at the same time. We can compute in an elementary way the probability that four will come down heads and the other six tails. But suppose we ask: What is the probability for the appearance of five heads and five tails, or seven heads and three tails, or more generally, for n heads and $(10 - n)$ tails, where n may be any integer between 0 and 10? The answer to this question is a set of numbers, one for each value of n. These numbers can be thought of as forming a *function* of n, $f(n)$. That is, for each n there is a value of $f(n)$ which gives the probability of the event characterized by the number n. Such a function is called a *probability distribution*.

A probability distribution is always defined for a definite range of values of the index n. In the above example, n is an integer between 0 and 10. If, as will usually be the case in our problems, this range of the index includes all the possible events, then the sum of all the probabilities must be unity (certainty). In this case,

$$\sum_n f(n) = 1 \tag{6.1}$$

where the sum extends over the entire range of values of n appropriate to the particular problem under consideration.

An example of a probability distribution which can be obtained using the methods of Sec. 4 is the probability of various results from rolling two dice. The total may be any integer from 2 to 12, but these numbers are not all equally likely. We saw in Sec. 4, in fact, that the probability for 7 was $\frac{1}{6}$, while the probability for 11 was $\frac{1}{18}$. Expressing these facts in the language just introduced, we let n be the total on the two dice, and $f(n)$ be the probability for this number. We have found that $f(7) = \frac{1}{6}$ and $f(11) = \frac{1}{18}$. The other values for this distribution can be obtained similarly; the whole distribution is as follows:

n	$f(n)$
2	$\frac{1}{36}$
3	$\frac{1}{18}$
4	$\frac{1}{12}$
5	$\frac{1}{9}$
6	$\frac{5}{36}$
7	$\frac{1}{6}$
8	$\frac{5}{36}$
9	$\frac{1}{9}$
10	$\frac{1}{12}$
11	$\frac{1}{18}$
12	$\frac{1}{36}$

According to Eq. (6.1), the sum of all the values of $f(n)$ should be unity. The reader is invited to verify that this is in fact the case. The distribution $f(n)$ can be represented graphically by means of a histogram, as shown in Fig. 6.1.

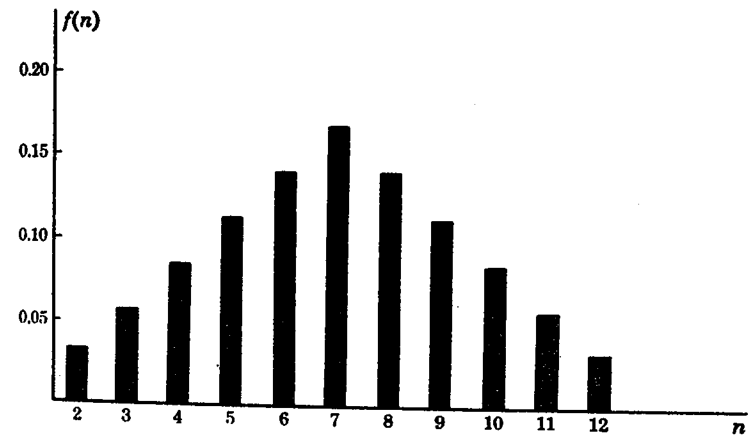


Fig. 6.1. Probability distribution for two dice.

Of course, the probability distribution may have more than one index. If we flip 10 pennies and 4 dimes, we can compute the probability that among the pennies there will be n heads, and that among the dimes there will be r, where n is an integer between 0 and 10, and r is an integer between 0 and 4. We can call the result $f(n,r)$ to indicate that the probability depends on both n and r. We shall not discuss such probability distributions in this text; they are treated by straightforward extensions of the methods to be discussed here.

Returning to the 10-penny problem, suppose that we want to find the *average* or *mean* number of heads in a large number of trials. Suppose we flip the pennies Z times, where Z is a very large number. By definition of the probability distribution, the number of times we obtain n heads is $Zf(n)$. To compute the mean value of n, we should multiply each value of n by the number of times it occurs, add all these products, and divide by Z. That is,

$$\bar{n} = \frac{1}{Z} \sum_n nZf(n) = \sum_n nf(n) \qquad (6.2)$$

The fact that Z cancels out of this expression means, of course, that for a large number of trials, the value $\bar{n}$ is independent of Z.

The expression for $\bar{n}$, given by Eq. (6.2), can be thought of as a *weighted mean* of the values of n, with weights equal to the corresponding probabilities. The sum of the weights in this case is unity.

As an illustration of the meaning of Eq. (6.2), we use the distribution for two dice to compute the *mean* value of the total, in a large number of rolls. We multiply each value of n by its probability and add the results:

n	$f(n)$	$nf(n)$
2	$1/36$	$1/18$
3	$1/18$	$3/18$
4	$1/12$	$1/3$ or $6/18$
5	$1/9$	$5/9$ or $10/18$
6	$5/36$	$5/6$ or $15/18$
7	$1/6$	$7/6$ or $21/18$
8	$5/36$	$10/9$ or $20/18$
9	$1/9$	1 or $18/18$
10	$1/12$	$5/6$ or $15/18$
11	$1/18$	$11/18$
12	$1/36$	$1/3$ or $6/18$
$\bar{n} = \Sigma\, nf(n) = 126/18 = 7$		

The average value of n is $\bar{n} = 7$. This should not be surprising; the probabilities are distributed symmetrically about $n = 7$ so that, roughly speaking, a value of n greater than 7 is as probable as a value smaller than 7 by the same amount.

In the same manner one could calculate the mean value of n^2, which is

$$\overline{n^2} = \sum_n n^2 f(n) \qquad (6.3)$$

More commonly, one is interested in the average value of $(n - \bar{n})^2$, which is of course the *variance* of the values of n which occur. This is given in general by

$$\sigma^2 = \sum_n (n - \bar{n})^2 f(n) \tag{6.4}$$

As an example of the use of Eq. (6.4), we compute the variance of the two-dice distribution. The calculation is conveniently arranged in tabular form as follows:

n	$(n - \bar{n})$	$(n - \bar{n})^2$	$f(n)$	$(n - \bar{n})^2 f(n)$
2	−5	25	1/36	25/36
3	−4	16	1/18	32/36
4	−3	9	1/12	27/36
5	−2	4	1/9	16/36
6	−1	1	5/36	5/36
7	0	0	1/6	0
8	1	1	5/36	5/36
9	2	4	1/9	16/36
10	3	9	1/12	27/36
11	4	16	1/18	32/36
12	5	25	1/36	25/36

$$\sigma^2 = \Sigma\,(n - \bar{n})^2 f(n) = 105/36 = 2\tfrac{11}{12}$$

Thus the root-mean-square spread of the values of n about the mean is $\sigma = (35/12)^{1/2} = 1.71$, which is about what we would guess from looking at the distribution.

So far we have discussed probability distributions based on the definition of probability given in Sec. 4, which in turn is based on the idea of making an indefinitely large number of trials, counting the number of favorable events, and taking the ratio of the two. Our discussion of the mean and standard deviation is based on the assumption that a very large number of trials has been made.

It may not be immediately clear, therefore, how these quantities are related to the results which would be obtained if we made an experiment consisting of a relatively small number of trials. If there is only one trial, for example, the mean is clearly not very likely to equal the mean for an infinite number of trials. The mean $\bar{n}$ for a small number of trials cannot be expected to correspond exactly with the value obtained with an infinite number of trials. The same is true for the standard deviation of a small number of trials.

To describe the distinction between the infinitely large number of trials used to define $f(n)$ and any small number of trials in an actual experiment, we call $f(n)$ the *infinite parent distribution* and the results of any group of trials a *sample* of this distribution. It is clear that the mean of a small sample is only an *estimate* of the mean of the infinite parent distribution. For *some* types of distributions it can be shown that the precision of this estimate increases with the size of the sample, but it is important to remember that it is never more than an estimate. Similarly, the standard deviation of a sample is an estimate of the standard deviation of the infinite parent distribution.

Moreover, there are good theoretical reasons, which we shall not discuss in detail, for stating that Eq. (3.9) does not even give the *best* estimate of the parent distribution standard deviation which can be obtained from a given sample. It turns out that a somewhat better estimate is given by

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{6.5}$$

which differs from Eq. (3.9) in that the sum of the d_i^2 is divided by $(N - 1)$ instead of N. Roughly speaking, the reason for this is that the deviations are not all independent; the same data have been used previously to compute the sample mean which is used to compute the d_i^2, and so the number of *independent* deviations is only $(N - 1)$. Although this modification is of some theoretical significance, it is not usually of any *practical* importance. Ordinarily N is sufficiently large so that the sample standard deviation is affected very little by the choice between N and $(N - 1)$.

Because we shall sometimes want to learn as precisely as possible the characteristics of the infinite parent distribution, it is important to know how well the mean and standard deviation of the sample approximate the mean and the standard deviation of the infinite parent distribution, and how the precision of these approximations depends on the size of the sample. We return to these questions in Chap. IV.

A related question arises if we have a number of trials of some kind and want to ascertain whether the

results of these trials can or cannot be regarded as a sample of some particular infinite parent distribution. The distribution of results of a small sample, as we have just pointed out, will not be identical to that of the infinite parent distribution in any case; but how close should we expect the sample distribution to be to the

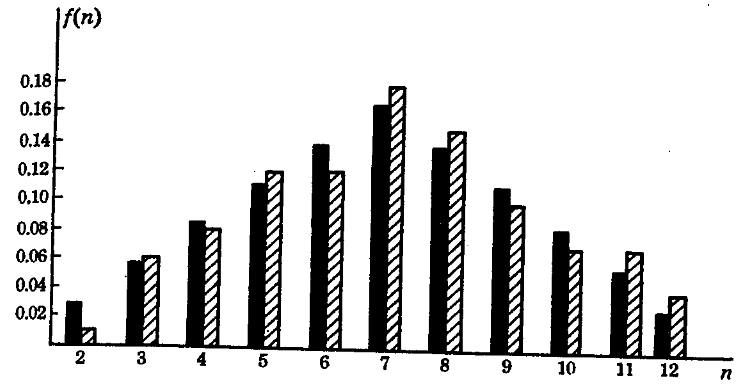


Fig. 6.2. Comparison of parent and sample distributions for two dice. The solid bars represent the parent distribution, the shaded bars the distribution which resulted from a sample of 100 rolls. The mean of the parent distribution is exactly 7, while the mean of the sample is 7.15.

infinite parent distribution in order to conclude that the sample is in fact a sample of this parent distribution? A partial answer to this question is given in Sec. 11.

A very practical example of this kind of question arises in connection with the probability distribution for two dice, shown in Fig. 6.1. Suppose we want to determine whether the dice of a particular pair are loaded. If they are loaded, their parent distribution will not be that of Fig. 6.1, but something different. We roll the dice several times, recording the results. We then need a means of comparing this sample distribution with the parent distribution characteristic of unloaded dice. How much difference between the sample and parent distributions should be expected if the dice are *not* loaded? How much difference should we require as evidence that they *are* loaded? A partial answer to this sort of question is given in Sec. 11.

If some of the above discussion seems somewhat vague and abstract, take heart! It will become clearer as more examples are discussed in the following sections.

7 | Binomial Distribution

We now consider a problem in which we will use all the things we have learned so far about probability and statistics. Suppose that we have N independent events of some kind, each of which has probability p of succeeding and probability $(1 - p)$ of not succeeding. We want to know the probability that exactly n of the events will succeed.

An example may help clarify the situation. Suppose we light five firecrackers. They are supposedly identical, but because of some uncertainty in their manufacture only ¾ of them explode when lighted. In other words, the probability for any one to explode is $p = \frac{3}{4}$, and the probability that it will fizzle is $1 - p = \frac{1}{4}$. In this case the number of independent events, N, is 5.

We now ask for the probability that, of these 5, n will explode when lighted, where n is an integer between 0 and 5.

A few particular cases are easy. If the probability of success in one of these events is p, the probability that all N of them will succeed is p^N. The probability that all N will fail is $(1 - p)^N$. In our example, the probability that all five firecrackers will explode is $(\frac{3}{4})^5 = 0.237$. The probability that none will explode is $(\frac{1}{4})^5 = 0.00098$. In other words, neither of these is very likely; probability favors the other possibilities in which *some* of the firecrackers explode. The various probabilities are shown graphically in Fig. 7.1.

The in-between possibilities are not so simple. If we select a particular group of n events from N, the probability that these n will succeed and all the rest $(N - n)$ will fail is $p^n(1 - p)^{N-n}$. We can shorten the notations slightly by abbreviating $1 - p = q$.

This is not yet the probability that exactly n events will succeed, because we have considered only one particular group or combination of n events. How many combinations of n events can be chosen from N? Just the number of combinations of N things taken n at a time. So the probability that exactly n events will succeed from the group of N, which we denote by $f_{N,p}(n)$, is

$$f_{N,p}(n) = \binom{N}{n} p^n q^{N-n} \tag{7.1}$$

which we can call the probability of n successes in N

trials if the probability of success in one trial is p. This expression $f_{N,p}(n)$, defined by Eq. (7.1), is called the *binomial distribution* because of its close relation to the

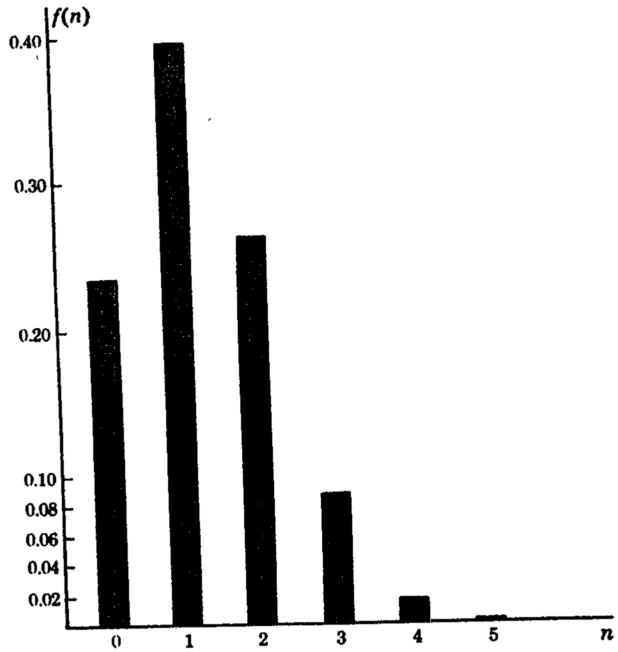


Fig. 7.1. Probability that n firecrackers will explode in a group of five, if the probability for any one to explode is ¾. This is a binomial distribution with $N = 5$, $p = ¾$.

binomial theorem. A few examples of binomial distributions, computed from Eq. (7.1), are shown in Fig. 7.2.

What is the binomial distribution good for? Here is another example. Suppose we roll three dice. We

know that the probability for 5 to come up on a single die is ⅙. What is the probability of 5 coming up on n dice, where n can be 0, 1, 2, or 3? We see that this is

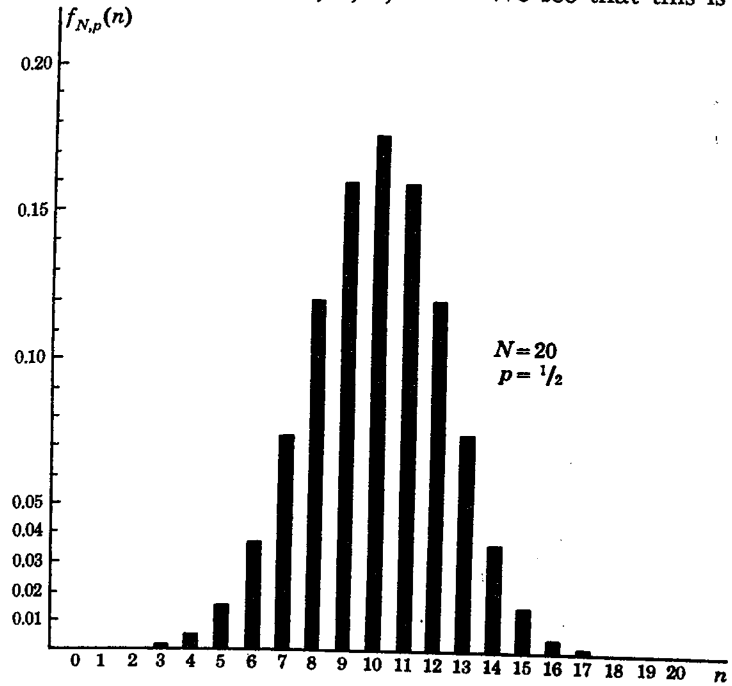


Fig. 7.2a. Example of binomial distribution, with $N = 20$. Distribution for $p = ½$ is symmetric about the mean, $\bar{n} = 10$.

exactly the problem solved by the binomial distribution. The probability of success in a single trial is ⅙ in this case, so that $p = ⅙$. We are asking for the probability of n successes in three trials. This is, according to the binomial distribution,

$$f_{3,1/6}(n) = \binom{3}{n}\left(\frac{1}{6}\right)^n\left(1-\frac{1}{6}\right)^{3-n}$$

The probabilities of 0, 1, 2, and successes in three trials are then

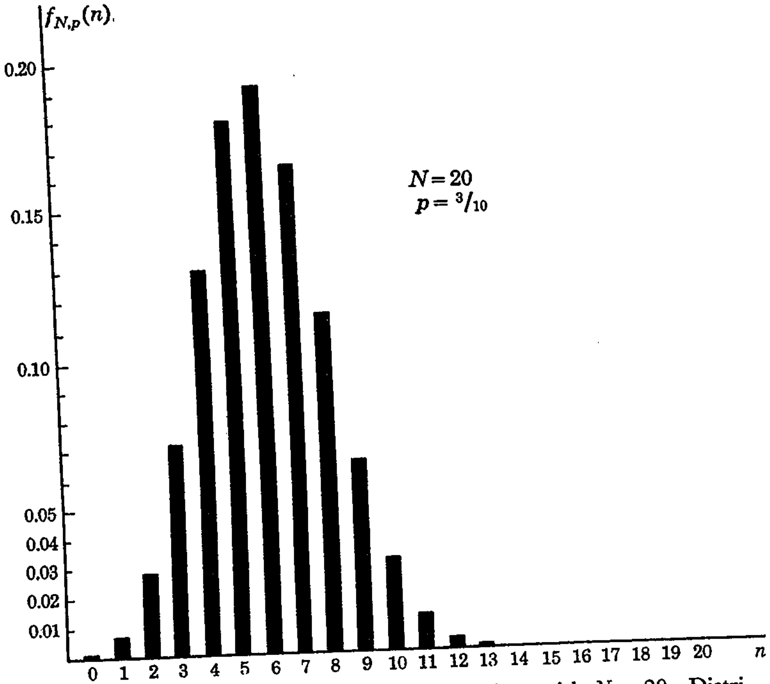


Fig. 7.2b. Example of binomial distribution, with $N = 20$. Distribution for $p = 3/10$ favors smaller values of n, close to $\bar{n} = 6$.

Zero successes: $$f_{3,1/6}(0) = \frac{3!}{(3-0)!\,0!}\left(\frac{5}{6}\right)^3 = \frac{125}{216}$$

One success: $$f_{3,1/6}(1) = \frac{3!}{(3-1)!\,1!}\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^2 = \frac{75}{216}$$

Two successes: $$f_{3,1/6}(2) = \frac{3!}{(3-2)!\,2!}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right) = \frac{15}{216}$$

Three successes: $$f_{3,1/6}(3) = \frac{3!}{(3-3)!\,3!}\left(\frac{1}{6}\right)^3 = \frac{1}{216}$$

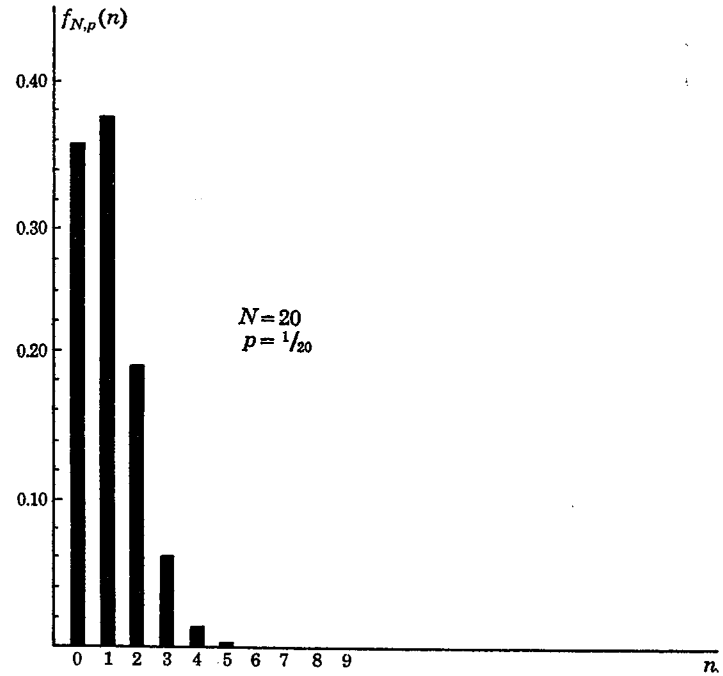


Fig. 7.2c. Example of binomial distribution, with $N = 20$. Distribution is strongly asymmetric. Here $\bar{n} = 1$, and probabilities for $n > 6$ are negligibly small.

As a check on these calculations we note that the total probability for 0, 1, 2, or 3 successes must be one since there are no other possibilities. Thus, the four probabilities which we have calculated above must total unity; this is in fact the case.

More generally, it must be true, for the same reason, that

$$\sum_{n=0}^{N} f_{N,p}(n) = \sum_{n=0}^{N} \binom{N}{n} p^n q^{N-n} = 1 \tag{7.2}$$

To show that this is in fact the case, we note that

$$\sum_{n=0}^{N} \binom{N}{n} p^n q^{N-n} \tag{7.3}$$

is exactly equal to the binomial expansion of $(q + p)^N$, as seen from Eq. (5.6). But $p + q = 1$, so $(q + p)^N = 1$, and Eq. (7.2) is established. Note, incidentally, that it is true for *any* value of p, which is a number between 0 and 1.

Now that we have calculated the probability for any number of successes in N trials, we can calculate the *mean* (or average) number of successes in N trials. The meaning of this mean is the same as in Sec. 6. We make the N trials, observing a certain number n of successes. We make N trials *again*, finding in general a different number of successes. We do this a large number of times, say Z (where Z may stand for a zillion), and then compute the *mean* of all the numbers of successes which we observe.

To do this, we multiply each number of successes n by the number of times $Zf_{N,p}(n)$ that it occurs, and then divide by the total number of sets of trials, Z. The

average number of successes, which we denote by $\bar{n}$, is

$$\bar{n} = \sum_{n=0}^{N} n \binom{N}{n} p^n (1 - p)^{N-n} \tag{7.4}$$

The sum ranges from $n = 0$ to N because in every one of the sets of trials *some* number of successes between 0 and N must occur. To summarize, we have obtained Eq. (7.4) directly from Eq. (6.2) by inserting the expression for the binomial distribution function, Eq. (7.1).

We can calculate the value of $\bar{n}$ if we know the number of trials N and the probability p for success in any one trial. In the example of three dice, we have used the values of N and p given ($N = 3, p = 1/6$) to compute the values of the probability distribution $f_{3,1/6}(n)$. Using these values, we proceed as follows:

n	$f_{3,1/6}(n)$	$nf_{3,1/6}(n)$
0	125/216	0
1	75/216	75/216
2	15/216	30/216
3	1/216	3/216

$$\bar{n} = \sum_{n=0}^{3} f_{3,1/6}(n) = 108/216 = 1/2$$

If we average the numbers of 5s in all the trials, the result is ½. This is not equal to the result of any single trial, of course, and there is no reason to expect it to be. The most probable number is zero, and the probabilities for the others are just such as to make the average ½.

This calculation can be done more simply, but to do it more simply we have to derive an equation which expresses the *result* of performing the sum in Eq. (7.4) in a simple way. Deriving the equation requires some acrobatics, the details of which are given in Appendix B. The result is

$$\bar{n} = \sum_{n=0}^{N} n \binom{N}{n} p^n (1-p)^{N-n} = Np \tag{7.5}$$

This remarkably simple and very reasonable result says that the average number of successes in N trials is just the probability of success in any one trial, multiplied by the number of trials. If we had had to guess at a result, this is probably what we would have guessed!

Applying this result to the three-dice problem, we see that with $N = 3$ and $p = 1/6$, the average number of 5s can be obtained immediately: $\bar{n} = Np = 3 \times 1/6 = 1/2$, in agreement with our previous result.

Just as the mean $\bar{n}$ is defined in general for any distribution $f(n)$, by Eq. (6.2) the variance is obtained by calculating the mean of the squares of the deviations, Eq. (6.4). For the binomial distribution, the variance is given by

$$\sigma^2 = \sum_{n=0}^{N} (n - \bar{n})^2 f_{N,p}(n) = \sum_{n=0}^{N} (n - Np)^2 f_{N,p}(n), \tag{7.6}$$

in which we have used $\bar{n} = Np$, Eq. (7.5).

Evaluation of this sum, as with the evaluation of $\bar{n}$, requires a bit of trickery. The details are again given

in Appendix B so as not to interrupt the continuity of the present discussion. The result of the calculation is

$$\sigma^2 = Np(1-p) = Npq \tag{7.7}$$

or

$$\sigma = \sqrt{Npq} \tag{7.8}$$

another remarkably simple result.

As an illustration of the properties of the binomial distribution just obtained, we return to the example of three dice, for which we have computed the probabilities for the occurrence of any number of 5s between 0 and 3. In this case, $N = 3$, $p = 1/6$. The mean number of 5s was found to be $1/2$. Similarly, we may compute the standard deviation:

$$\sigma = \sqrt{Npq} = \sqrt{3 \times 1/6 \times 5/6} = 0.646$$

which means that the root-mean-square deviation of the values of n about the mean ($\bar{n} = 1/2$) is somewhat less than unity. The deviations of the few events for which $n = 2$ or $n = 3$ are, of course, larger than this.

8 | Poisson Distribution

We consider next a particular application of the binomial distribution which is important in nuclear physics. Suppose that we have N radioactive nuclei. Suppose also that the probability for any one of these to undergo a radioactive decay in a given interval of time (T, for instance) is p. We want to know the probability that

n nuclei will decay in the interval T. The answer is of course the old familiar binomial distribution function $f_{N,p}(n)$. This is, however, somewhat unwieldy for practical calculations; N may be a very large number, such as 10^{23}, and p may be the order of 10^{-20} or so. With numbers of these magnitudes, there is no practical way to evaluate the binomial distribution, Eq. (7.1).

Fortunately, we can make considerable simplifications by using approximations which are valid when N is extremely large and p is extremely small. We therefore consider the *limit* of the binomial distribution function as N grows very large and p grows very small in such a way that the *mean* of the distribution, which is Np, remains finite. We denote this product by

$$Np = a \tag{8.1}$$

We shall introduce the approximations in a manner which will make them seem plausible, but no attempt will be made to attain mathematical rigor.

First of all, we note that if p is a very small quantity, the average number of events will be very much smaller than N so that the values of n which are of interest will be extremely small compared to N. Guided by this observation, we make two approximations in the expression

$$f_{N,p}(n) = \frac{N!}{(N-n)!\,n!}p^n(1-p)^{N-n}$$

Consider first the factor

$$\frac{N!}{(N-n)!} = N(N-1)(N-2)\cdots(N-n+1) \tag{8.2}$$

This is a product of n factors, none of which is significantly different from N. We therefore replace Eq. (8.2) by N^n. We then have approximately

$$f_{N,p}(n) \cong \frac{(Np)^n}{n!}(1-p)^{N-n} = \frac{(Np)^n}{n!}\frac{(1-p)^N}{(1-p)^n} \tag{8.3}$$

Second, we notice that the factor $(1-p)^n$ is very nearly equal to unity because it is a number very close to unity raised to a not terribly large power. We therefore drop this factor. We also eliminate N from the expression, using $a = Np$, and rearrange it to obtain

$$f(n) = \frac{a^n}{n!}(1-p)^{a/p} = \frac{a^n}{n!}[(1-p)^{1/p}]^a \tag{8.4}$$

All that remains now is to evaluate the limit

$$\lim_{p\to 0}(1-p)^{1/p}$$

This limit is discussed in many books on elementary calculus and is shown to have the value $1/e$. Using this fact in Eq. (8.4), we obtain

$$f_a(n) = \frac{a^n e^{-a}}{n!} \tag{8.5}$$

This form is known as the Poisson distribution function. Note that while the binomial distribution contained two independent parameters (N and p), the Poisson distribution has only one (a). The other one disappeared when we took the limit of the binomial distribution as $N \to \infty$.

Using the definition of a, Eq. (8.1), and the general expression for the mean of the binomial distribution, Eq. (7.5), we find that the mean value of n is

$$\bar{n} = a \tag{8.6}$$

That is, if we observe the radioactive material for a series of time intervals T, recording the number of disintegrations taking place in each interval, we find that the *average* number of disintegrations is a.

As with the general form of the binomial distribution, if we add the probabilities for all possible values of n, we must obtain unity (certainty). That is,

$$\sum_{n=0}^{\infty} f_a(n) = 1 \tag{8.7}$$

We extend the summation from zero to infinity because we have let the number of independent events N become indefinitely large. To establish that Eq. (8.7) is in fact true, we insert Eq. (8.5) in Eq. (8.7):

$$\sum_{n=0}^{\infty} f_a(n) = e^{-a} \sum_{n=0}^{\infty} \frac{a^n}{n!} \tag{8.8}$$

But the sum in Eq. (8.8) is nothing but the Maclaurin series expansion of the quantity e^a. Thus the sum in Eq. (8.7) does equal unity, as required.

Any probability distribution which is constructed so that the *sum* of the probabilities of all possible events is unity is said to be *normalized*. It is quite possible to

define a probability distribution differently so that the sum of all the probabilities is a number different from unity. In this case, certainty is represented not by unity, but by some other number. It is usually convenient, however, to construct the probability distribution in such a way that the sum of all the probabilities is unity. This practice is followed everywhere in this book.

As has been stated in Eq. (8.6), the mean value of n for the Poisson distribution is simply $\bar{n} = a$. The standard deviation for the Poisson distribution can also be obtained easily from the expression for the standard deviation of the binomial distribution, Eq. (7.8), by using $Np = a$ and the fact that q is very nearly unity; the result is simply

$$\sigma = \sqrt{a} \quad \text{or} \quad \sigma^2 = a \tag{8.9}$$

Here is an example of the use of the Poisson distribution in radioactive decay. Suppose we have 10^{20} atoms of Shakespeareum, a fictitious radioactive element whose nuclei emit α particles. Shakespeareum might be, for example, a rare unstable isotope of one of the rare-earth elements, with an atomic weight in the vicinity of 150; in this case 10^{20} atoms correspond to about 25 mg of the element. Suppose that the decay constant is 2×10^{-20} per second, which means that the probability for any one nucleus to decay in 1 sec is 2×10^{-20}. This corresponds to a half-life of about 10^{12} years, rather long but not impossibly so.

Now suppose we observe this sample of material for many 1-sec intervals. What is the probability to observe *no* α emissions in an interval? One? Two? The answers are given simply by the Poisson distribution. We are given $N = 10^{20}$ and $p = 2 \times 10^{-20}$; so we have $a = 2$. Substituting this value in Eq. (8.5), we obtain the following values:

n	$f_2(n)$
0	0.135
1	0.271
2	0.271
3	0.180
4	0.090
5	0.036
6	0.012
7	0.003
8	0.001

These results are shown graphically in Fig. 8.1. The mean number of counts in this case is exactly 2, and the standard deviation is $\sqrt{2}$. For comparison, Fig. 8.2 shows a Poisson distribution with $a = 10$.

In many practical applications of the Poisson distribution the problem may be somewhat different, in that the constant a may not be known at the beginning. The problem may be, for example, to determine the value of a from a distribution of experimental data. If it is known that the parent distribution of which the data are a sample is a Poisson distribution, then the best estimate of a is just the mean of the sample distribu-

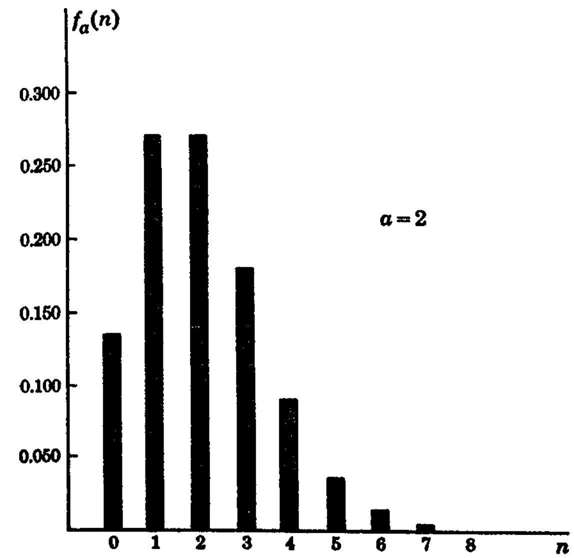


Fig. 8.1. Poisson distribution with $a = 2$.

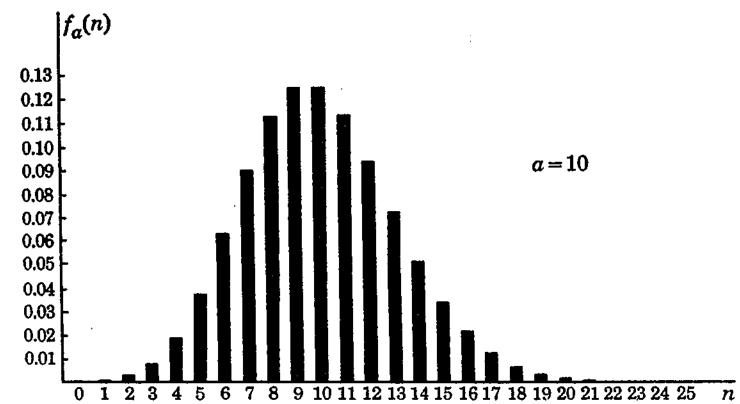


Fig. 8.2. Poisson distribution with $a = 10$.

tion. A little thought will show that the standard deviation of this value is $\sqrt{a}$.

Other cases may arise where it is not certain whether the parent distribution corresponding to a given sample *is* a Poisson distribution. For example, if one observes the number of eggs laid by a flock of chickens on each of several days, one may want to ascertain whether the probability for a given number of eggs on a particular day follows the Poisson distribution. In such a case some test of goodness of fit, such as the test discussed in Sec. 11, may be used.

9 | Gauss Distribution, or Normal Error Function

We now consider another probability distribution which is of great practical importance, the Gauss distribution. It is important for several reasons. (1) It describes the distribution of random errors in many kinds of measurements. (2) It is possible to show that even if individual errors do not follow this distribution, the *averages* of groups of such errors are distributed in a manner which approaches the Gauss distribution for very large groups. We may have, for example, a set of observations which are distributed according to the *xyz* distribution, which may be any distribution at all. If we take groups of N observations and average them, then in the limit of very large N the *averages* will be distributed according to the Gauss distribution. The only condition is that the variance of the *xyz* distribution be finite. This statement is

known as the central-limit theorem; it is very important in more advanced developments in mathematical statistics.

The Gauss distribution can be regarded in two ways: as a result which can be derived mathematically from elementary considerations or as a formula found empirically to agree with random errors which actually occur in a given measurement. Someone has remarked, in fact, that everyone believes that the Gauss distribution describes the distribution of random errors, mathematicians because they think physicists have verified it experimentally, and physicists because they think mathematicians have proved it theoretically!

From a theoretical point of view, we can make the plausible assumption that any random error can be thought of as the result of a large number of elementary errors, all of equal magnitude, and each equally likely to be positive or negative. The Gauss distribution can therefore be associated with a limiting form of the binomial distribution in which the number of independent events N (corresponding to the elementary errors) becomes very large, while the probability p of success in each (the chance of any elementary error being positive) is $\frac{1}{2}$. The derivation of the Gauss distribution from these considerations is given in Appendix C.

Many people feel, however, that the real justification for using the Gauss distribution to describe distribution of random errors is that many sets of experimental observations turn out to obey it. This is a more convincing reason than any mathematical derivation. Hence it is a valid point of view to treat this distribution as an experimental fact, state its formula dogmatically, and then examine what it means and what it is good for.

The Gauss distribution function is often referred to as the *normal error function*, and errors distributed according to the Gauss distribution are said to be *normally distributed*.

The Gauss distribution is

$$f(x) = Ae^{-h^2(x-m)^2} \tag{9.1}$$

where A, h, and m are constants and x is the value obtained from one measurement. This distribution differs from those we have considered previously in that we shall regard x as a continuous variable, rather than an integer as with the binomial and Poisson distributions. This will necessitate some further discussion of the significance of $f(x)$; but first we plot the function $f(x)$ to get a general idea of its behavior.

Figure 9.1 is a graph of the Gauss distribution function, Eq. (9.1). We note that A is the maximum height of the function, m represents the value of x for which the function attains this maximum height, and h has something to do with the broadness or narrowness of the bell-shaped curve. A large value of h corresponds to a narrow, peaked curve, while a small value of h gives a broad, flat curve.

Now, what is the significance of the function $f(x)$? We are tempted to say that $f(x)$ represents the probabil-

ity of observing the value x of the measured quantity. But this is not really correct. Remembering that x is a continuous variable, we realize that the probability for x to have *exactly* any particular value is zero. What we must discuss instead is the probability that x will have a value in a certain region, say between x and $x + \Delta x$.

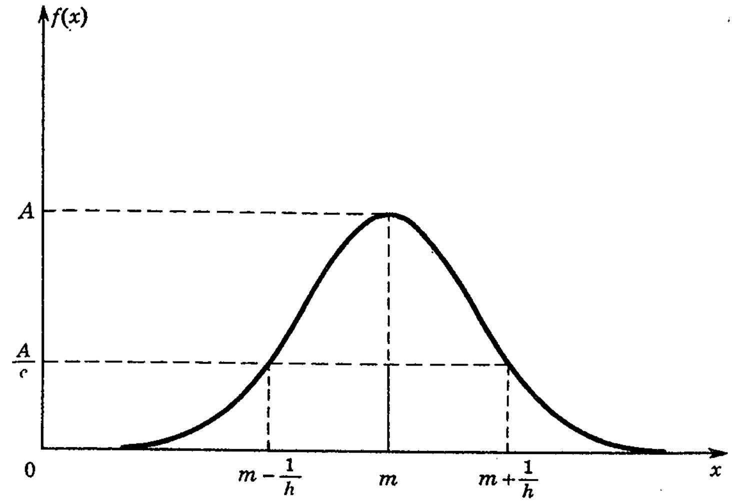


Fig. 9.1. Gauss distribution function. The points $x = m \pm 1/h$, at which the curve has $1/e$ of its maximum height, are shown.

So the proper interpretation of the function $f(x)$ is that for a small interval dx, $f(x)\,dx$ represents the probability of observing a measurement which lies in the interval between x and $x + dx$.

This statement has a simple graphical interpretation. In Fig. 9.2, the area of the shaded strip on the graph represents $f(x)\,dx$. Therefore we can say that the *area* under the curve within the interval dx represents the probability that a measurement will fall in this interval. Similarly,

$$P(a,b) = \int_a^b f(x)\,dx \tag{9.2}$$

is the probability that a measurement will fall somewhere in the interval $a \leq x \leq b$.

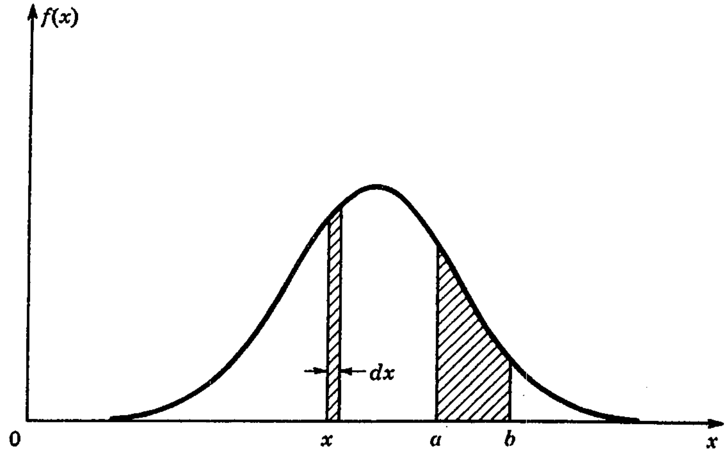


Fig. 9.2. Graphical representation of probabilities. Shaded areas represent probabilities for an observation to fall in the corresponding intervals.

The total probability for a measurement to fall *somewhere* is of course unity; so the total area under curve $f(x)$ must be unity. Analytically, we can say that it must be true that

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \tag{9.3}$$

This is analogous to Eq. (6.1). If this condition is satis-

fied, the function $f(x)$ is said to be normalized in the same sense as the functions in Secs. 7 and 8 were normalized, having a total probability of 1. We have extended the range of integration from $-\infty$ to $+\infty$ because it is necessary to include all possible values of x in the integration.

The requirement that the function $f(x)$ be normalized imposes a restriction on the constants which appear in the function. If we know h and m, then

$$\int_{-\infty}^{\infty} Ae^{-h^2(x-m)^2}\,dx = 1 \tag{9.4}$$

will be satisfied only with one particular value of the constant A. To obtain this value of A, we must actually perform the integration. To simplify the integral we make a change of variable, letting

$$h(x - m) = z \tag{9.5}$$

Equation (9.4) then becomes

$$A\int_{-\infty}^{\infty} e^{-z^2}\,dz = h \tag{9.6}$$

The value of the integral in Eq. (9.6) can be shown to be

$$\int_{-\infty}^{\infty} e^{-z^2}\,dz = \sqrt{\pi} \tag{9.7}$$

Obtaining this value requires a small chicanery, the details of which are given in Appendix D. Inserting Eq. (9.7) in Eq. (9.6),

$$A = \frac{h}{\sqrt{\pi}} \tag{9.8}$$

Thus, we find that in order to satisfy the normalization condition, Eq. (9.4), the constant A must have the value $A = h/\sqrt{\pi}$. From now on, therefore, we write the Gauss distribution function as

$$f(x) = \frac{h}{\sqrt{\pi}} e^{-h^2(x-m)^2} \tag{9.9}$$

which is normalized for every value of h.

Next, we find the mean value of x for this distribution. The meaning of *mean* is the same as always—the average of a very large number of measurements of the quantity x. We could go through the same line of reasoning as in Sec. 6 by introducing the total number of measurements Z and then showing that it divides out of the final result. Instead, we observe simply that $f(x)\,dx$ represents the probability of occurrence of the measurement in the interval dx and that the mean value of x is found simply by integrating the product of this probability and the value of x corresponding to this interval. That is,

$$\bar{x} = \int_{-\infty}^{\infty} xf(x)\,dx \tag{9.10}$$

This expression is completely analogous to Eq. (6.2); we use an integral here rather than a sum because x is a continuous variable rather than a discrete one.

To compute $\bar{x}$ we insert Eq. (9.9) into Eq. (9.10) and make the change of variable given by Eq. (9.5):

$$\begin{aligned} \bar{x} &= \frac{h}{\sqrt{\pi}} \int_{-\infty}^{\infty} xe^{-h^2(x-m)^2}\,dx \\ &= \frac{1}{\sqrt{\pi}} \int \left(\frac{z}{h} + m\right) e^{-z^2}\,dz \end{aligned} \tag{9.11}$$

The first term of this expression integrates to zero because the contributions from negative values of z exactly cancel those from positive values. The part that survives is

$$\begin{aligned} \bar{x} &= \frac{m}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-z^2}\,dz = \frac{m}{\sqrt{\pi}} \sqrt{\pi} \\ &= m \end{aligned} \tag{9.12}$$

a result which we could have guessed in the first place simply by looking at the graph of the function.

The calculation of the variance proceeds in a similar manner. The variance is given by

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} (x - m)^2 f(x)\,dx \\ &= \int_{-\infty}^{\infty} \frac{h}{\sqrt{\pi}} (x - m)^2 e^{-h^2(x-m)^2}\,dx \end{aligned} \tag{9.13}$$

To evaluate this integral we make the change of variable, Eq. (9.5), to obtain

$$\sigma^2 = \frac{1}{h^2\sqrt{\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2}\,dz \tag{9.14}$$

This is a convergent integral; at large values of z, z^2 becomes very large, but e^{-z^2} grows small so rapidly that the product $z^2e^{-z^2}$ also approaches zero very rapidly. The integral in Eq. (9.14) can be integrated by parts to convert it into the form of Eq. (9.7), whose value is known. The final result is

$$\sigma^2 = \frac{1}{2h^2} \qquad \text{or} \qquad \sigma = \frac{1}{\sqrt{2}\,h} \tag{9.15}$$

The standard deviation is inversely proportional to h. This should not be surprising, because larger values of h mean a more sharply peaked curve as well as smaller values of σ. Since h is large for sharply peaked curves,

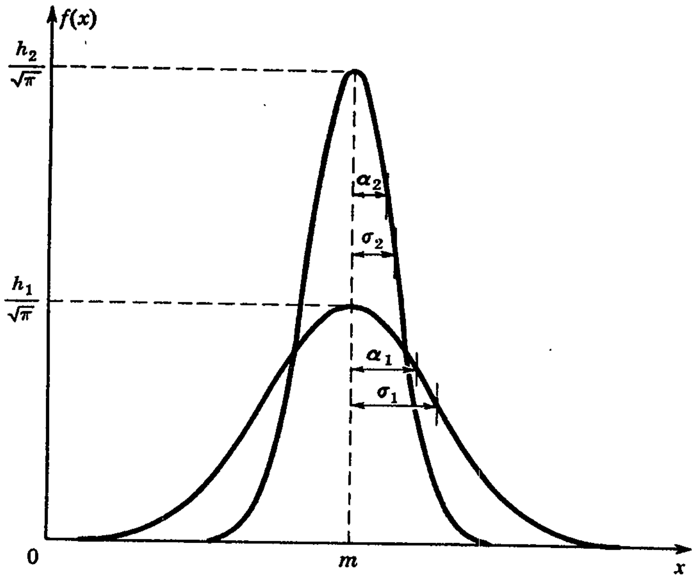


Fig. 9.3. Gauss distributions for two values of h with the same m; $h_2 = 2h_1$. Positions of σ and α for the two curves are shown.

corresponding to small spread of errors, h is sometimes called the *measure of precision* of the distribution. The Gauss distribution is plotted in Fig. 9.3 for two values of h.

It is often useful to write the Gauss distribution in terms of σ rather than h. Using σ, the function becomes

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-m)^2/2\sigma^2} \tag{9.16}$$

The *mean deviation* for the Gauss distribution is even easier to obtain than the variance. It is given by

$$\begin{aligned}\alpha &= \int_{-\infty}^{\infty} |x - m| \frac{h}{\sqrt{\pi}} e^{-h^2(x-m)^2}\,dx \\ &= \frac{2h}{\sqrt{\pi}} \int_0^{\infty} y e^{-h^2y^2}\,dy \end{aligned} \tag{9.17}$$

This integral can easily be evaluated by making the substitution $z = h^2y^2$. It should not be necessary to give the details of this substitution; the result is simply

$$\alpha = \frac{1}{\sqrt{\pi}\,h} \tag{9.18}$$

Comparing this with Eq. (9.15), we see that for the Gauss distribution the standard deviation and mean deviation are proportional, since both are inversely proportional to h. The standard deviation is the larger of the two; the relation is

$$\sigma = \sqrt{\frac{\pi}{2}}\,\alpha = 1.25\,\alpha \tag{9.19}$$

This equation is quite useful when one wants a rough estimate of the standard deviation of a set of observations whose errors are thought to be normally distributed. Instead of calculating σ from the data, one calculates α (which is generally easier since it is not necessary to square all the deviations) and then uses Eq. (9.19). It must be emphasized, though, that this relationship holds only for the Gauss distribution; it is not valid for other distributions.

The Gauss distribution may be used to find the probability that a measurement will fall within any specified limits. In particular, it is of interest to calculate the probability that a measurement will fall within σ of the mean value. This will give a more clear understanding of the significance of the standard deviation. The probability P that a measurement will fall between $m - \sigma$ and $m + \sigma$ is given by

$$P = \int_{m-\sigma}^{m+\sigma} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-m)^2/2\sigma^2}\,dx \tag{9.20}$$

Making a change of variable $t = (x - m)/\sigma$, we find

$$P = \frac{1}{\sqrt{2\pi}} \int_{-1}^{+1} e^{-t^2/2}\,dt \tag{9.21}$$

This integral cannot be evaluated except by making numerical approximations. Fortunately, such integrals are used sufficiently often so that extensive tables of their values have been calculated. A short table of values of the integral

$$\frac{1}{\sqrt{2\pi}} \int_{0}^{T} e^{-t^2/2}\,dt$$

for various values of T is given at the end of the book, Table II. Also included for convenience is a table of

values of the function $(1/\sqrt{2\pi})e^{-t^2/2}$, Table I. References to more extensive tables are also given.

In general, the probability for a measurement to occur in an interval within $T\sigma$ of the mean is

$$P(T) = \frac{1}{\sqrt{2\pi}} \int_{-T}^{T} e^{-t^2/2}\,dt \tag{9.22}$$

The values of this probability for a few values of T are as follows:

$$\begin{aligned} P(1) &= 0.683 \qquad & 1 - P(1) &= 0.317 \\ P(2) &= 0.954 \qquad & 1 - P(2) &= 0.046 \\ P(3) &= 0.997 \qquad & 1 - P(3) &= 0.003 \end{aligned}$$

These figures show that the probability for a measurement to fall within one standard deviation of the mean is about 68%; the probability that it will fall within two standard deviations is about 95%, and the probability that it will be *farther* away from the mean than three standard deviations is only 0.3%.

Here is an example of the use of some of the ideas just discussed. A surveyor runs a line over level ground between two points about 1000 ft apart. He carefully stretches his 100-ft tape to the proper tension for each segment of the measurement and applies the proper temperature correction, to eliminate systematic errors from these sources. He repeats the measurement 10 times. Supposing that the remaining errors are associated with random errors in the individual measurements, and that the resulting errors are randomly distributed, we can make the following calculations:

Observation (x), ft	$\lvert d_i \rvert$, ft
1023.56	0.055
1023.47	0.035
1023.51	0.005
1023.49	0.015
1023.51	0.005
1023.48	0.025
1023.50	0.005
1023.53	0.025
1023.48	0.025
1023.52	0.015
$\bar{x}$ = 1023.505 ft	0.210 ft
	a = 0.021 ft
	$\sigma = 1.25a = 0.026$ ft

The probability for an individual measurement to fall within 0.026 ft of the mean is 0.683, so we expect about 68% of the measurements to lie between 1023.48 and 1023.53 ft. The probability for falling within *two* standard deviations of the mean (1023.45 and 1023.56 ft) is 0.95, and so on. A more important question is: What is the reliability of the *mean?* This question can be answered with the methods introduced in Sec. 12.

10 | Rejection of Data

The question we consider next is a controversial one. It concerns the problem of what to do if, among a set of observations, one or more have deviations so large as to seem unreasonable. If, for example, a set of measurements made with a micrometer caliper has a standard deviation of 0.001 in., but one measurement differs from the mean by 0.010 in., then we are tempted to

regard this large deviation as a blunder or mistake rather than a random error. What shall we do with this observation?

Such an observation creates an awkward situation for the experimenter. If he retains the questionable observation, it can have quite a large effect on the mean. It will also, of course, have an even greater effect on the standard deviation. If on the other hand it is discarded, one runs the risk of throwing away information which might lead to discovery of some unexpected phenomenon in the experiment. Important discoveries have resulted from apparently anomalous data. In any event, it cannot be denied that throwing away an observation constitutes tampering with the data, better known as "fudging."

As has been mentioned, this is a controversial question, and one which has been hotly debated. There is no agreement among authorities as to a definite answer. We therefore present several different points of view, and let the reader take his choice.

At one extreme, there is the point of view that unless there is a definite reason for suspecting that a particular observation is not valid, there is *never* any justification for throwing away data on purely statistical grounds, and that to do so is dishonest. If one takes this point of view, there is nothing more to say, except to advocate taking enough additional data so that the results are not affected much by the questionable observations.

At the other extreme is the point of view that an observation should be rejected if its occurrence is so improbable that it would not reasonably be expected to occur in the given set of data. We reason as follows: Suppose we make N measurements of a quantity; suppose that one of these seems to have an unusually large deviation. We use the Gauss distribution function to calculate the probability that a deviation this large or larger will occur. If this probability is larger than $1/N$, we conclude that it is reasonable to obtain such a deviation. If, on the other hand, the probability of obtaining such a deviation is much smaller than $1/N$, this means that it is very unlikely that such a large deviation should occur even once in a set of N measurements. In this case, we might consider rejecting this measurement as being due probably to a mistake or some anomalous fluctuation of the experimental conditions. We should expect occasionally to obtain deviations whose probabilities of occurrence may be *somewhat* smaller than $1/N$, but not a great deal smaller. One rule of thumb for rejection of data which is sometimes used is to reject an observation if the probability of obtaining it is less than $1/2N$. This criterion is known as *Chauvenet's criterion.*

Here is an example. Suppose we make 10 observations. According to Chauvenet's criterion, an observation should be disregarded if its deviation from the mean is so large that the probability of occurrence of a deviation that large or larger is less than $\frac{1}{20}$. Referring to Eq. (9.22), we want to find the value of T such that $P(T) = 1 - \frac{1}{20}$ or 0.95. Referring to Table II, we find

that the proper value of T is $T = 1.96$. Therefore, after calculating σ for the set of observations, we should discard any observation whose deviation from the mean is larger than 1.96σ.

Table III is a short tabulation of maximum values of $T = d_i/\sigma$ which should be tolerated, according to Chauvenet's criterion. For example, with 20 observations, the maximum value of T is 2.24. If σ for a set of 20 voltage measurements is 0.01 volt, then any observation deviating from the mean by more than 2.24×0.01 volt $= 0.0224$ volt should be discarded.

If we eliminate an observation by Chauvenet's criterion, we should eliminate it *completely*. This means that after the anomalous observations are eliminated, we must recompute the mean and the standard deviation using the remaining observations. If one decides to use Chauvenet's criterion, it should be kept in mind that it may be possible to eliminate most or all of the data by repeated applications. Thus the criterion, of dubious validity at best even on the first round, should certainly not be used more than once.

Between the two extreme views just presented, there are other more moderate views on rejection of data. Some of these unquestionably have better theoretical justification than Chauvenet's criterion. They are also more complicated to use. We shall outline qualitatively one method which is sometimes used.[1]

If there are more observations in the "tails" of

[1] For a full discussion of this method, see H. Jeffreys, "Theory of Probability," sec. 4.41, Oxford University Press, New York, 1948.

the distribution than seems reasonable, one might suspect that the distribution is not quite normal; it may be approximately a Gauss distribution near the mean but somewhat larger than Gaussian in the tails. We can make some simple assumption regarding the small modification to be made in the distribution to represent probabilities for large deviations in agreement with the observations. Then we use the principle of maximum likelihood, which will be used for the theory of least squares in Sec. 14, to determine the most probable value of the observed quantity. This turns out to be a weighted mean, in which the observations far from the mean are given considerably less weight than those in the center. This procedure involves a fair amount of computation, but it is undoubtedly more honest than Chauvenet's criterion.

11 | Goodness of Fit

We now return briefly to a question raised at the end of Sec. 6; that is, if we suspect that a given set of observations comes from some particular parent distribution, how can we test them for agreement with this distribution?

Consider the example of Sec. 6, the probability distribution for the results of rolling two dice. The probability distribution $f(n)$ tabulated on page 41 is computed on the assumption that each die is symmetric, so that the six numbers on each are all equally likely.

Now we roll the dice a large number of times, recording the totals. It is not very likely that 12 will occur on *exactly* $\frac{1}{36}$ of the trials, but we expect the result to be close to $\frac{1}{36}$. If it turns out to be $\frac{1}{2}$ instead, there is probably something very strange about these dice. Now, the problem is: How much disagreement between the parent distribution (in this case the table on page 41) and our sample distribution can we reasonably expect, if the sample *is* taken from this parent distribution? Or, to put the question another way, how great must the disagreement be in order to justify the conclusion that the dice do *not* obey the parent distribution (i.e., that they are loaded)?

What we need is a quantitative index of the difference in the two distributions, and a means of interpreting this index. The sample distribution is expressed most naturally in terms of the *frequencies* of the various events, where the frequency of an event is defined as the total number of times this event occurs among all the trials. Thus it is convenient to express our distributions in terms of frequencies rather than probabilities. Specifically, let $F(n)$ be the frequency of event n (in this case, simply the occurrence of the total n) for the sample, which we shall assume to consist of N trials. If the parent distribution which we are comparing with this sample is $f(n)$, then the frequency predicted by the parent distribution is just $Nf(n)$. The difference $Nf(n) - F(n)$ for each n characterizes the difference in the two frequencies.

The most widely used test for comparing the sample and parent frequencies (or for examining the "goodness of fit" of the sample) consists of computing a weighted mean of the square of the fractional difference of the two frequencies. The resulting quantity is called χ^2; this quantity, together with a suitable interpretation, constitutes the "χ^2 test of goodness of fit."

The quantity $[Nf(n) - F(n)]/Nf(n)$ represents the fractional difference in the frequencies for a given n. Our first impulse is to square this quantity and sum over n. A little thought shows, however, that this would weight the "tails" of the distribution, whose statistical fluctuations are always relatively large, as much as the center. Thus a better criterion for goodness of fit is obtained by multiplying by a weighting factor $Nf(n)$, which then weights the fractional difference according to the importance of the event n in the distribution. Thus it is customary to define a measure of goodness of fit called χ^2 by the equation

$$\chi^2 = \sum_n \frac{[Nf(n) - F(n)]^2}{Nf(n)} \tag{11.1}$$

This discussion is not intended to be a thorough exposition of the reasons for this particular definition of χ^2. To give such an exposition we should relate χ^2 to the idea of the least-squares sum which is introduced in Sec. 14. Such a discussion is beyond the scope of this book; instead, we simply recognize that Eq. (11.1)

seems intuitively to be a reasonable index of goodness of fit.

There remains the question of how to interpret the result of Eq. (11.1). Clearly, if the sample distribution and the assumed parent distribution agree exactly, then $\chi^2 = 0$. This is of course extremely unlikely; even if the sample *is* taken from the assumed parent distribution, one would not expect *exact* agreement in every interval. But, the larger χ^2 is, the more disagreement there is between the two distributions. The proper question to ask is: How large a value of χ^2 is reasonable if the sample *is* taken from the assumed parent? If we obtain a value of χ^2 larger than this reasonable value, then we should assume that the sample *does not* agree with the parent.

Calculating values of χ^2 which can occur simply by chance is quite involved, and we cannot discuss the problem here. Instead, we give a short table which will help interpret χ^2 in specific situations. Table IV lists values of χ^2 for which the probability of occurrence of a χ^2 *larger* than this is a given value P, assuming that the sample *is* taken from the parent distribution used in computing χ^2. This value depends on the number of points at which the theoretical and sample frequencies are compared, which is called ν in the table.

In the dice-rolling example, we are comparing the two frequencies for 11 different events; so in this case $\nu = 11$. For $\nu = 11$, the table lists the value $\chi^2 = 6.989$

under $P = 0.80$, and the value $\chi^2 = 24.725$ under $P = 0.01$. This means that if the sample "fits" the assumed parent distribution, there is an 80% chance that χ^2 will be 6.989 or larger, because of random fluctuations, but only a 1% chance that it will be greater than 24.725. Thus if we calculate χ^2 for a sample and obtain a value around 7, we can say that this is probably due to chance fluctuations, and the sample *does* fit the assumed parent. If on the other hand we obtain $\chi^2 = 40$, then it is very unlikely that this value occurred by chance, and the sample probably *does not* fit the parent. Note that the χ^2 test never gives a cut-and-dried answer "it fits" or "it does not fit." Some judgment is required in all cases.

It is not necessary that the frequencies refer to individual events. They may just as well refer to *groups* of events. Suppose, for example, that we are observing radioactive decays and want to compare the distribution of the number of events in a given time interval with the Poisson distribution with a given value of a. In a particular case it might be expedient to consider the following four groups: $n = 0$; $n = 1$; $n = 2$, 3, or 4; $n > 4$. For these four groups we have four comparisons between the sample and parent frequencies. In this case, then, $\nu = 4$.

An additional complication arises if we must determine the quantity a from the sample distribution. It can be shown that the most probable value of a from the sample is simply the mean of the sample distribu-

tion, as we might have guessed from Eq. (8.6). But in using the sample to determine a, we have *forced* a certain degree of agreement between the two frequencies. Thus the number of real comparisons is reduced from four to three, and we should use $\nu = 3$.

In general, when comparing a sample with a parent distribution using K groups of events, we take $\nu = K$ if the parameters of the parent distribution (such as a for the Poisson distribution, or N and p for the binomial) are specified in advance. If one parameter of the parent distribution (such as a) is determined from the sample, we take $\nu = K - 1$; if two parameters (such as N and p) are determined from the sample, we take $\nu = K - 2$, and so on.

It is easy to extend this method to the case where the observed quantity, say x, is a continuous variable, so that the parent distribution $f(x)$ is a function of a continuous variable. We divide the range of values of x into a series of nonoverlapping intervals which together cover the whole range. Call a typical interval Δx_k, and the value of x at its center x_k. Assume that there are K intervals in all, so that k ranges from 1 to K. The probability P_k given by the parent distribution for a measurement to fall in this interval is

$$P_k = \int_{x_k - \Delta x_k/2}^{x_k + \Delta x_k/2} f(x)\, dx \tag{11.2}$$

In usual practice, the intervals are sufficiently small, except perhaps in the "tails" of the distribution, so that this integral can be approximated by taking the value of $f(x)$ at the center and multiplying by the width of the interval; this is equivalent to assuming that $f(x)$ is approximately constant over any single interval. In this case, we have approximately

$$P_k = f(x_k)\Delta x_k \tag{11.3}$$

The frequency predicted by the parent distribution is then NP_k; the sample frequency F_k is of course the number of times the variable x falls in the interval Δx_k in the given sample. In this case the appropriate definition of χ^2 is

$$\chi^2 = \sum_{k=1}^{K} \frac{(NP_k - F_k)^2}{NP_k} \tag{11.4}$$

PROBLEMS

1. Six pennies are tossed simultaneously. What are the probabilities for *no* heads? One head? Two? Three? Four? Five? Six? ??? ? Would the probabilities be the same if, instead, one penny was tossed six times? Explain.

2. One die (singular of dice) is rolled. What is the probability that 6 will come up? If four dice are rolled, what is the probability for *no* 6s? One? Two? Three? Four? Five?

3. Among a large number of eggs, 1% were found to be rotten. In a dozen eggs, what is the probability that none is rotten? One? More than one?

4. A man starts out for a Sunday afternoon walk, playing the following game. At the starting point he tosses a coin. If the result is heads, he walks north one block; if tails, south one block. Find all the possible positions after four tosses, and the probability for each.

5. In Prob. 4, derive the probability distribution for the possible positions after N tosses of the coin.

6. In Prob. 5, if the man walks on several different Sundays, what is the *average* distance he reaches from the starting point after N tosses? What is the *standard deviation* of his positions after N tosses?

7. The man plays the same game as in Prob. 4, but instead of tossing one coin he tosses two. If *both* are heads, he walks a block north; for any other result he walks a block south. Find the possible positions after four tosses, and the probability for each.

8. In Prob. 7, derive the probability distribution for the possible positions after N tosses of the coin.

9. Answer the questions asked in Prob. 6, for the distribution obtained in Prob. 8.

10. The scores in a final examination were found to be distributed according to the following table:

Score	Distribution, %	Score	Distribution, %
95–100	4	65–69	14
90– 94	6	60–64	10
85– 89	8	55–59	6
80– 84	12	50–54	2
75– 79	16	40–49	2
70– 74	18	9–39	2

a. Draw a histogram illustrating this distribution.

b. Calculate approximate values for the mean and variance of the distribution.

c. If 15% of the students failed the examination, what was the lowest passing grade?

11. Observations on 200 litters of cocker spaniel puppies revealed the following statistics:

Litters	*Puppies in each litter*
5	4
17	5
34	6
47	7
31	8
25	9
18	10
14	11
7	12
2	13

Find the mean number in a litter, and the standard deviation.

12. A lump of Shakespeareum (a fictitious radioactive element) contains 10^{21} nuclei. The probability that any one will decay in 10 sec is found to be 2×10^{-21}. Find the probability that in a given 10-sec period *no* decays will occur. Also one decay, two, three, etc. Find the number of decays per 10 sec such that the probability of *more* decays than this number is less than 0.1%. The answer to this part will determine what is meant by "etc." in the first part.

13. A group of underfed chickens were observed for 50 consecutive days and found to lay the following numbers of eggs:

Eggs laid	*No. of days*
0	10
1	13
2	13
3	8
4	4
5	2

Show that this is approximately a Poisson distribution. Calculate the mean and standard deviation directly from the data. Compare with standard deviation predicted by the Poisson distribution.

14. Derive Eq. (8.6) for the *mean* of the Poisson distribution directly from Eqs. (6.2) and (8.5). To evaluate the sum, insert Eq. (8.5) into Eq. (8.7) and differentiate the result with respect to a. This procedure is similar to that used in Appendix B for the binomial distribution.

15. Derive Eq. (8.9) for the variance of the Poisson distribution directly from Eqs. (6.4) and (8.5) by the same procedure suggested in Prob. 14.

16. During a summer shower which lasted 10 min, 10^6 raindrops fell on a square area 10 m on a side. The top of a convertible was actuated by a rain-sensing element 1 cm square; so the interior of the car was protected in case of rain.

a. Find the probability that at least one raindrop landed on the element.

b. In such a shower how much time must elapse after the shower begins, on the average, before the top closes itself?

17. It has been observed in human reproduction that twins occur approximately once in 100 births. If the number of babies in a birth follows a Poisson distribution, calculate the probability of the birth of quintuplets. Do you think it likely that octuplets have ever been born in the history of man?

18. A coin is tossed 10,000 times; the results are 5176 heads and 4824 tails. Is this a reasonable result for a symmetric coin, or is it fairly conclusive evidence that the coin is asymmetric? (*Hint:* Calculate the total probability for *more* than 5176 heads in 10,000 tosses. To do this, approximate

the binomial distribution by a Gauss distribution, as discussed in the last paragraph of Appendix C.)

19. Calculate the mean deviation for the Gauss distribution. Express your result as a multiple of σ.

20. If a set of measurements is distributed according to the Gauss distribution, find the probability that any single measurement will fall between $(m - \frac{1}{2}\sigma)$ and $(m + \frac{1}{2}\sigma)$.

21. The "probable error" of a distribution is defined as the error such that the probability of occurrence of an error whose absolute value is less than this value is ½. Find the probable error for the normal (Gauss) distribution, and express it as a multiple of σ. Is this *the most probable error?* If not, what is?

22. Show that the graph representing the results of Prob. 1 can be approximated by a normal distribution curve. Find the appropriate mean and standard deviation for this curve.

23. Consider the data of Prob. 27. If these are normally distributed, and if two additional measurements are made, find:

a. The probability that *both* will be in the interval 54.98 to 55.02 cm.

b. The probability that *neither* will be in this interval.

24. The measurements (x) in a certain experiment are distributed according to the function

$$F(x) = A/[(x - m)^2 + b^2].$$

a. Sketch the function.

b. Find the value of A needed to normalize the function.

c. What is the mean of the distribution?

d. Discuss the standard deviation of the distribution.

25. Suppose that the function of Prob. 24 were "cut off" at $x = m \pm b$. That is, $F(x)$ is the given function in the inter-

val $m - b < x < m + b$, but $F(x) = 0$ for values outside this interval. Answer the questions of Prob. 24.

26. An object undergoes simple harmonic motion with amplitude A and frequency f according to the equation $x = A \sin 2\pi ft$, where x represents the displacement from equilibrium. Calculate the mean and standard deviation of the position and of the speed of the object.

27. The height of a mercury column in a manometer was measured using a cathetometer. The following measurements were obtained:

cm	*cm*
55.06	54.99
54.92	55.02
55.01	55.03
55.00	55.02
54.98	54.97

Test these data using Chauvenet's criterion to determine which should be discarded. After discarding the appropriate data, recompute the mean. By how much does it differ from the original mean? Compare this difference with σ for the set of data.

28. Apply the χ^2 test to the data of Prob. 13.

29. Discuss how the χ^2 test might be applied in Prob. 18.