It is a capital mistake to theorise before one has data. Insensibly one begins to twist the facts to suit theories, instead of theories to suit facts

—Sir Arthur Conan Doyle

CHAPTER 

Theoretical Distributions

Science is supposed to explain to us what is actually happening, and indeed what will happen, in the world. Unfortunately as soon as you try and do something useful with it, sordid arithmetical numbers start getting in the way and messing up the basic scientific laws. An unbiased coin may perfectly well come down heads uppermost 55 times out of 100. A decaying radioactive source may give 95 counts on a Geiger counter in one minute, and 110 counts in the next. A 10 volt power supply across a resistor marked 100 Ω may give a reading of 103 mA on your ammeter. Predictions from basic laws are modified by statistical distributions, arising from the finite size of the data sample, the experimental accuracy, and similar causes. This chapter deals with the basic ideas of distributions, and especially with the three fundamental statistical distributions: the binomial, the Poisson, and the Gaussian. Only by understanding the ways the distributions give rise to the data can one go on to use the particular behaviour of the data to produce general statements about the processes that produced them in the first place—or, as Holmes puts it, to twist your theory to suit your observed facts.

3.1 GENERAL PROPERTIES OF DISTRIBUTIONS

3.1.1 A Simple Distribution

Suppose you toss four coins. This is a simple example, and I will not pretend it is of any intrinsic interest; nevertheless we will go through it in detail as it introduces concepts that will be needed later for real problems.

For each coin the probability of the head landing uppermost is $\frac{1}{2}$, and so is the probability for the tail. We want to discuss the various possible outcomes for the four coins, and their probabilities.

1. The four coins could all land head uppermost. The probability of the first coin giving a head is $\frac{1}{2}$; so are those for the second, third, and fourth. To find the combined probability of all four giving a head we multiply the individual probabilities together, so the probability of four heads is $(\frac{1}{2})^4$. Call this $P(4)$; then $P(4) = \frac{1}{16}$.
2. Suppose the first three coins land head upwards, the fourth tail upwards. The combined probability for this is again the product of the individual ones, which gives $\frac{1}{8}$ for the first three and $\frac{1}{2}$ for the fourth, as the probability of a tail is also $\frac{1}{2}$, again giving $\frac{1}{16}$. However, if we ask for three heads and one tail, without specifying which coin gives the tail, there are four choices, namely HHHT, HHTH, HTHH, and THHH, each with the same probability of $\frac{1}{16}$, so the total probability $P(3)$ for three heads and a tail is: $P(3) = 4 \times \frac{1}{16} = \frac{1}{4}$.
3. For two heads and two tails there are six permutations of coins—HHTT, HTHT, HTTH, TTHH, THTH, and THHT—each of probability $\frac{1}{16}$, so the probability $P(2)$ of getting two heads and two tails is $\frac{3}{8}$.
4. For one head and three tails the probability is the same as one tail and three heads, so we can write down at once $P(1) = P(3) = \frac{1}{4}$.
5. Likewise, for no heads and four tails, $P(0) = P(4) = \frac{1}{16}$.

A quick check can be done by making sure that the total probability of *something happening is 1:*

$$\sum_r P(r) = P(0) + P(1) + P(2) + P(3) + P(4) = \frac{16}{16} = 1.$$

So if r is the number of heads ($r = 0, 1, 2, 3, 4$), we have a collection of probabilities $P(r) = (\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16})$, giving the probability that a toss of four unbiased coins will give r heads. This is a simple example of a *probability distribution.*

3.1.2 The Law of Large Numbers

Having all these numbers, let us try and do something with them. The probabilities are in the ratio 1:4:6:4:1, i.e. if one tosses four coins sixteen

times, there should be one result with four heads, four with three heads and a tail, etc. Four coins were accordingly tossed sixteen times, and the results are shown in the following table:

Number of heads	$r=4$	$r=3$	$r=2$	$r=1$	$r=0$
Theory predicts	1	4	6	4	1
Data	2	7	2	4	1

They do not agree. There is certainly a similarity in the pattern, but the numbers do not match perfectly. Indeed, it would have been surprising if they had. With such a small number of tosses (only sixteen) statistical fluctuations are substantial. To give the numbers a chance, the experiment was repeated with a 160, 1600, and 16000 trials:†

Number of heads	$r=4$	$r=3$	$r=2$	$r=1$	$r=0$
160 tosses					
Theory predicts	10	40	60	40	10
Data	10	40	61	38	11
1600 tosses					
Theory predicts	100	400	600	400	100
Data	125	403	567	409	96
16000 tosses					
Theory predicts	1000	4000	6000	4000	1000
Data	1009	3946	5992	4047	1006

The agreement becomes better and better as the number of trials increases and random effects are smoothed out.

The theory predicts a set of probabilities. The observed data frequencies do not quite agree with them. However, as the size of the data sample N increases the fluctuations cancel out, and the frequencies tend to the probabilities as N tends to infinity. This is the *law of large numbers.*

3.1.3 Expectation Values

If you know the probability distribution for some number r—often, in an attempt to add excitement to the subject, called the number of 'successes'—

†Simulated on a computer, of course. You are urged to try some cointossing experiments of your own, to appreciate the way in which the experimental distributions never quite agree with

one thing you can easily compute is the average number of 'successes' you would expect. This is called the *expectation value* of r and is written $\langle r \rangle$, or sometimes $E(r)$.† It is given by

$$\langle r \rangle = \sum_r rP(r). \tag{3.1}$$

For example, with four coins, as discussed in section 3.1.1, the average number of heads is given by

$$0 \times \frac{1}{16} + 1 \times \frac{1}{4} + 2 \times \frac{3}{8} + 3 \times \frac{1}{4} + 4 \times \frac{1}{16} = 2$$

which is an obvious result, but shows how the formula works.

Note that $\langle r \rangle$ is not necessarily the most probable result, although it is in this example. For five coins, $\langle r \rangle = 2.5$.

Any function of r also has its expectation value, defined in the same way:

$$\langle f \rangle = \sum_r f(r)P(r). \tag{3.2}$$

One useful way to think of the expectation value is in terms of gambling; suppose there is a random process (like a fruit machine) with various possible outcomes r, each of which has probability $P(r)$, and pays out an amount $f(r)$. Then the expectation value $\langle f \rangle$ is what you would expect, on average, to win, and would be an exactly fair fee to pay the organiser of the game for taking part.

There is an obvious parallel between an expectation value and the mean of a data sample (as described in the previous chapter). The former is a sum over a theoretical probability distribution and the latter is a (similar) sum over a real data sample. The law of large numbers ensures that if a data sample is described by a theoretical distribution, then as N, the size of the data sample, goes to infinity,

$$\bar{f} \to \langle f \rangle. \tag{3.3}$$

Note that expectation values add

$$\langle f+g \rangle = \sum (f+g)P(r) = \sum fP(r) + \sum gP(r) = \langle f \rangle + \langle g \rangle$$

but they *do not* multiply. In general, $\langle fg \rangle \neq \langle f \rangle \langle g \rangle$ unless f and g are *independent*.

3.1.4 Probability Density Distributions

Continuous variables need treating slightly differently from discrete variables. Suppose you are measuring the lengths of a large number of pieces

†The expectation value of r itself is also often denoted by the symbol μ.

of string, randomly distributed between 10 cm and 12 cm. Somebody asks you how many are 11 cm long. The answer has to be none. There will presumably be some between 10.5 and 11.5, probably a few between 10.9 and 11.1, maybe a couple between 10.99 and 11.01, but it's unlikely there will be any in the narrow range between 10.99999 and 11.00001, and if you insist that the value has to be exactly 11.00000000000... the range is so small that the probability vanishes.

However, the probability that x will lie within a specified range—like 10.9 to 11.1 cm—is a finite and perfectly sensible thing to talk about, and this is described by the *probability density distribution*, $P(x)$, defined by

$$\text{Probability (result lies between } x_1 \text{ and } x_2) = \int_{x_1}^{x_2} P(x)\,dx$$

or equivalently

$$P(x) = \text{limit}_{\delta x \to 0} \frac{\text{Probability (result lies between } x \text{ and } x + \delta x)}{\delta x}.$$

Probabilities are pure numbers. Probability densities, on the other hand, have dimensions, the inverse of those of the variable x to which they apply.

For expectation values the same ideas apply as for the earlier probability functions, except that you get integrals instead of summations:

$$\langle x \rangle = \int_{-\infty}^{\infty} xP(x)\,dx \tag{3.4}$$

$$\langle f \rangle = \int_{-\infty}^{\infty} f(x)P(x)\,dx. \tag{3.5}$$

If you have done some quantum mechanics you may have met expressions like $\langle x \rangle = \int \psi^*(x)x\psi(x)\,dx$. The meaning of the symbol is exactly the same; it is the expected average value of the result. It is tempting to go further and equate the quantity $\psi^*(x)\psi(x) = |\psi(x)|^2$ with the probability density $P(x)$, but this is wrong, as it does not work for expectation values of quantities (like momentum) that involve differential operators.

3.2 THE BINOMIAL DISTRIBUTION

The binomial distribution describes processes with a given number of identical trials, with two possible outcomes. Examples are tossing coins (heads or tails), quality checks of components (pass or fail), treatment of patients (kill or cure), and many similar. We call the two outcomes, without prejudice, 'success' and 'failure', and denote the probability of a success as p, and that

of failure therefore $1-p$.† This basic process is repeated n times—n is called the number of *trials*—and the distribution gives the probability of r successes (and thus $n-r$ failures) out of these n trials, each of which has an individual probability of success p.

3.2.1 The Binomial Probability Distribution Formula

The probability of r successes from n trials is a generalisation of the particular case considered in detail in section 3.1.1. It is made up of two factors. Firstly, there are 2^n possible permutations of success and failure, of which the number with r successes is the number of ways of selecting r from n:

$$ {}_nC_r = \frac{n!}{r!(n-r!)}. $$

Secondly, as there are r successes of probability p, and likewise $n-r$ failures of probability $1-p$, the combined result has a probability obtained by multiplying all these together, namely $p^r(1-p)^{n-r}$

Putting these two factors together gives the

Binomial probability distribution *The probability of r successes out of n tries, each of which has probability p of success, is*

$$ P(r;p,n) = p^r(1-p)^{n-r}\frac{n!}{r!(n-r)!}. \tag{3.6} $$

As this probability depends not only on r, the number of successes, but also on the intrinsic probability p and number of trials n, they are also shown as arguments of P, separated from r by a semicolon. This is a purely artistic device to show that usually one considers how P varies with r for a given n and p.

The ${}_nC_r$ are the binomial coefficients, so the total probability of *something* happening is the binomial expansion of $[p+(1-p)]^n$, and is therefore 1, as it has to be

$$ \sum_{r=0}^{n} p^r(1-p)^{n-r}\,{}_nC_r = [p+(1-p)]^n = 1^n = 1. \tag{3.7} $$

The important properties of the binomial distribution are (proofs, if desired, are given in section 3.2.2)

the mean number of success is $\quad \langle r \rangle = np$ (3.8)

†Some people define the probability of a failure as q. This makes formulae simpler, at the price of a new symbol and having to remember that q is always equal to $1-p$. Follow your

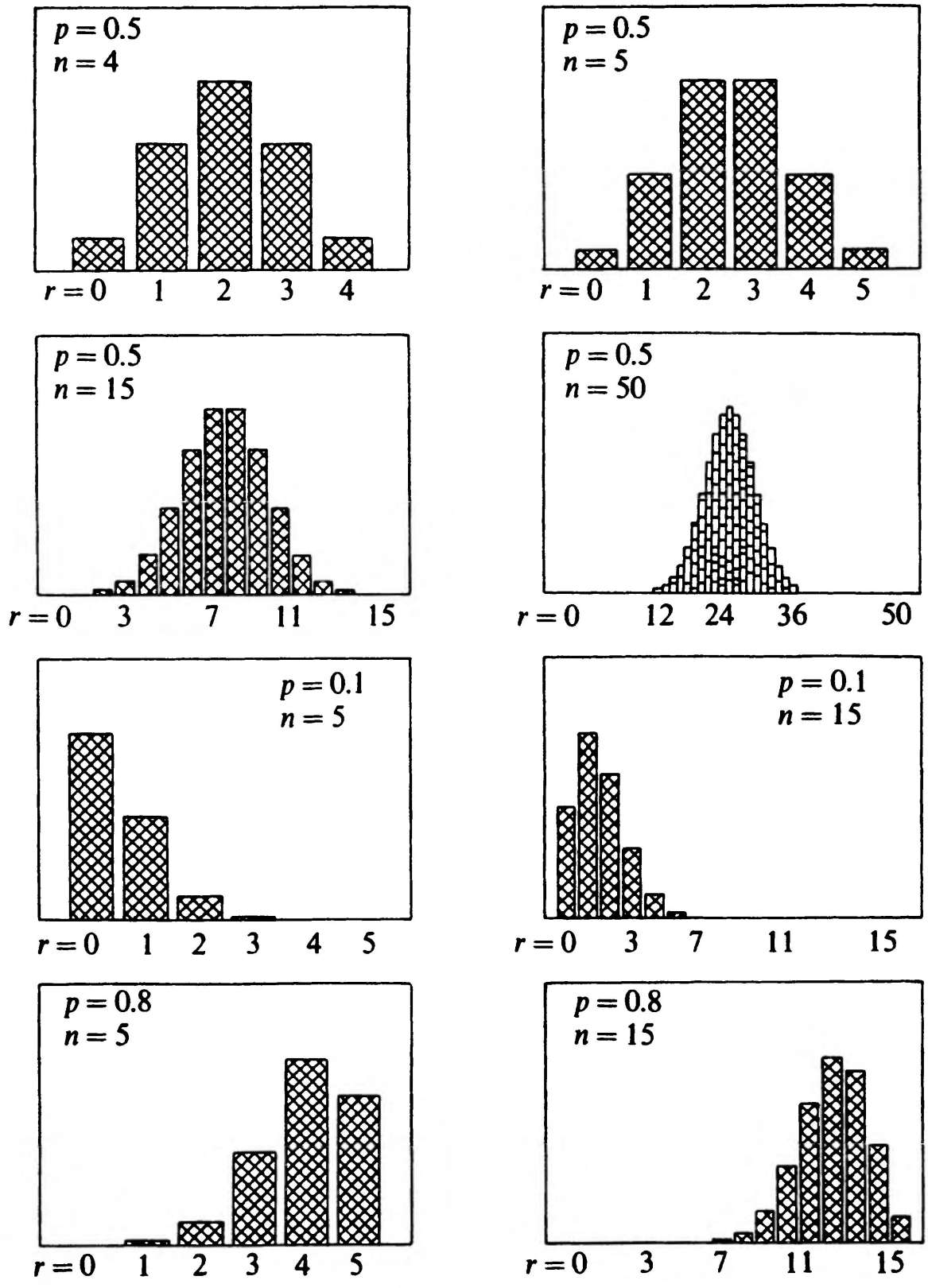


Fig. 3.1. Some binomial distributions, with various values of n and p. (The vertical scale is arbitrary).

the variance is $V(r) = np(1-p)$ (3.9)

and thus the standard deviation is $\sigma = \sqrt{np(1-p)}$. (3.10)

Some examples of binomial distributions are shown in Figure 3.1. They peak around the value np, as expected. As n increases, the peak, in proportion to the full range of n, becomes progressively narrower, albeit slowly. The relative width of the peak also depends on p, and (for the same n) peaks with p close to 0 or 1 are narrower than those with p near 0.5.

Example Detector efficiencies
You are trying to measure the tracks of cosmic ray particles using spark chambers, which are 95% efficient. You make the sensible decision that at least three points are needed to define a track. How efficient at detecting tracks would a stack of three chambers be? Would using four or five chambers give a significant improvement?

The probability of three hits from three chambers is

$$P(3; 0.95, 3) = 0.95^3 = 0.857$$

so this would be 85.7% efficient. For four chambers the probability of three or four hits is

$$P(3; 0.95, 4) + P(4; 0.95, 4) = 0.171 + 0.815 = 98.6\%.$$

For five chambers,

$$P(3; 0.95, 5) + P(4; 0.95, 5) + P(5; 0.95, 5) = 0.021 + 0.204 + 0.774 = 99.9\%.$$

Example Guessing cards
In an experiment into extrasensory perception, a subject guesses the symbol on a card. There are five different symbols so they have a 20% chance of guessing right by chance. If they guess six cards, what is the probability of getting more than half correct by chance?

The probability is

$$P(4; 0.2, 6) + P(5; 0.2, 6) + P(6; 0.2, 6) = 1.54\% + 0.154\% + 0.0064\% = 1.7\%.$$

★3.2.2 Proof of Properties of the Binomial Distribution

To prove equation 3.8, put the binomial formula (equation 3.6) in the expectation value (equation 3.1)

$$\langle r \rangle = \sum_{r=0}^{r=n} r p^r (1-p)^{n-r} \frac{n!}{r!(n-r)!}.$$

Take out a factor of np and drop the $r = 0$ term (which is zero):

$$\langle r \rangle = np \sum^{r=n} p^{r-1} (1-p)^{n-r} \frac{(n-1)!}{(r-1)!(n-r)!}.$$

Substituting $r' = r - 1$, $n' = n - 1$, this becomes

$$\langle r \rangle = np \sum_{r'=0}^{r'=n'} p^{r'}(1-p)^{n'-r'} \frac{n'!}{r'!(n'-r')!}.$$

The sum is the expansion of $[p + (1-p)]^{n'}$, and is just 1 (by equation 3.7). Therefore,

$$\langle r \rangle = np.$$

To find $V(r)$, start with the expression

$$\langle r(r-1) \rangle = \sum_{r=0}^{r=n} r(r-1) p^r (1-p)^{n-r} \frac{n!}{r!(n-r)!}.$$

Similar treatment (the first two terms are now zero) gives

$$\langle r(r-1) \rangle = p^2 n(n-1) \sum_{r'=0}^{r'=n'} p^{r'}(1-p)^{n'-r'} \frac{n'!}{r'!(n'-r')!}$$

where $r' = r - 2, n' = n - 2$. The sum is again 1, so

$$\langle r^2 - r \rangle = n(n-1)p^2$$

and using $\langle r \rangle = np$

$$\langle r^2 \rangle - \langle r \rangle^2 = n(n-1)p^2 + np - (np)^2$$

$$V(r) = np(1-p)$$

which is equation 3.9.

3.3 THE POISSON DISTRIBUTION

The binomial distribution describes cases where particular outcomes occur in a certain number of trials, n. The Poisson distribution describes cases where there are still particular outcomes but no idea of the number of trials; instead these are *sharp events occurring in a continuum.* For example, during a thunderstorm there will be a definite number of flashes of lightning (sharp events), but it is meaningless to ask how often there was not a flash. A Geiger counter placed near a radioactive source will produce definite clicks, but not definite non-clicks.

If in such an experiment one knows that the average number of events is, say, ten a minute, then in a particular minute one expects on average ten events, though intuitively one feels that nine or eleven would be unremarkable... but suppose there were five or fifteen? Is that compatible, or has something changed? We need to know the probability of obtaining a particular number of events, given the average number. This can be analysed by taking the limit of the binomial distribution, in which the number of tries

n, becomes large while at the same time the probability p becomes small, with their product constant.

3.3.1 The Poisson Probability Formula

Suppose that on average λ events would be expected to occur in some interval. Split the interval up into n very small equal sections, so small that the chance of getting two events in one section can be discounted. Then the probability that a given section contains an event is $p = \lambda/n$.

The probability that there will be r events in the n sections of the interval is given by the binomial formula (equation 3.6)

$$P(r; \lambda/n, n) = \frac{\lambda^r}{n^r}\left(1 - \frac{\lambda}{n}\right)^{n-r} \frac{n!}{r!(n-r)!}.$$

As $n \to \infty$ with r finite the factorials give a power of n:

$$\frac{n!}{(n-r)!} = n(n-1)(n-2)\cdots(n-r+1) \to n^r$$

and an exponential appears:

$$\left(1 - \frac{\lambda}{n}\right)^{n-r} \to \left(1 - \frac{\lambda}{n}\right)^{n} \to e^{-\lambda}$$

(This limit is actually a definition of e^x; alternatively it can be seen by taking logarithms of both sides and using $\ln(1+\delta) \approx \delta$.)

Inserting these two limits in the binomial formula above gives the

Poisson probability formula *The probability of obtaining r events if the mean expected number is λ is*

$$P(r; \lambda) = \frac{e^{-\lambda}\lambda^r}{r!} \tag{3.11}$$

In calculating a series of Poisson probabilities it is often convenient to start with $P(0)$, which is just $e^{-\lambda}$, and then successively multiply by λ and divide by $1, 2, 3, 4, \ldots$ to get $P(1)$, $P(2)$, $P(3)$, $P(4), \ldots$.

Example Fatal horse kicks

The classic example of Poisson statistics is the set of figures on the numbers of Prussian soldiers kicked to death by horses. In ten different army corps, over twenty years (in the last century), there were 122 deaths, so that λ, the mean number of deaths in one corps in one year, is $\frac{122}{200} = 0.610$. The probability of no deaths occurring, in a given corps for a given year, is $P(0; 0.61) = e^{-0.61}0.61^0/0! = 0.5434$; to get the prediction for the number of cases where no fatality occurred we just multiply by the number of cases considered (200) to get 108.7. Actually there were 109, so the agreement is virtually perfect. The full data show similar excellent agreement.

Number of deaths in 1 corps in 1 year	Actual number of such cases	Poisson prediction
0	109	108.7
1	65	66.3
2	22	20.2
3	3	4.1
4	1	0.6

Example Supernova neutrinos
Here are the numbers of neutrino events detected in 10 second intervals by the Irvine–Michigan–Brookhaven experiment on 23 February 1987—around which time the supernova S1987a was first seen by astronomers:

No. of events	0	1	2	3	4	5	6	7	8	9
No. of intervals	1042	860	307	78	15	3	0	0	0	1
Prediction	1064	823	318	82	16	2	0.3	0.03	0.003	0.0003

Ignoring the interval with nine events (for a strict justification of this see problem 8.2) the mean $\bar{r}$ is

$$\frac{860 + 307 \times 2 + 78 \times 3 + 15 \times 4 + 3 \times 5}{1042 + 860 + 307 + 78 + 15 + 3} = 0.77.$$

The Poisson predictions this gives are shown, and agree well with the data, except for the interval with nine events. This shows that the background due to random events is Poisson, and well understood, and the nine events are not a fluctuation on background, and came from the supernova.

Looking at the formula, or at the distributions shown in Figure 3.2, you can see that for λ below 1.0, the most probable result is zero. For higher values a peak develops, but note that this is below λ—although λ is the mean, it is not the mode. Indeed the formula shows that, for λ integer, $r = \lambda$ and $r = \lambda - 1$ are equally probable.

The Poisson distribution is always broader than a binomial distribution with the same mean. The Poisson variance is equal to the mean λ, whereas the binomial variance $np(1-p)$ is always smaller than the mean np. This is understandable, as the number of binomial success does have an upper limit (as r cannot exceed n) whereas the Poisson distribution can have a long tail. This upper tail is a characteristic of the Poisson distribution.

The important properties of the Poisson distribution (proofs, if desired, are in section 3.3.2) are

the total probability is 1 $$\sum_{r=0}^{\infty} P(r;\lambda) = 1 \qquad (3.12)$$

the mean number of events is $\langle r \rangle = \lambda$ (3.13)

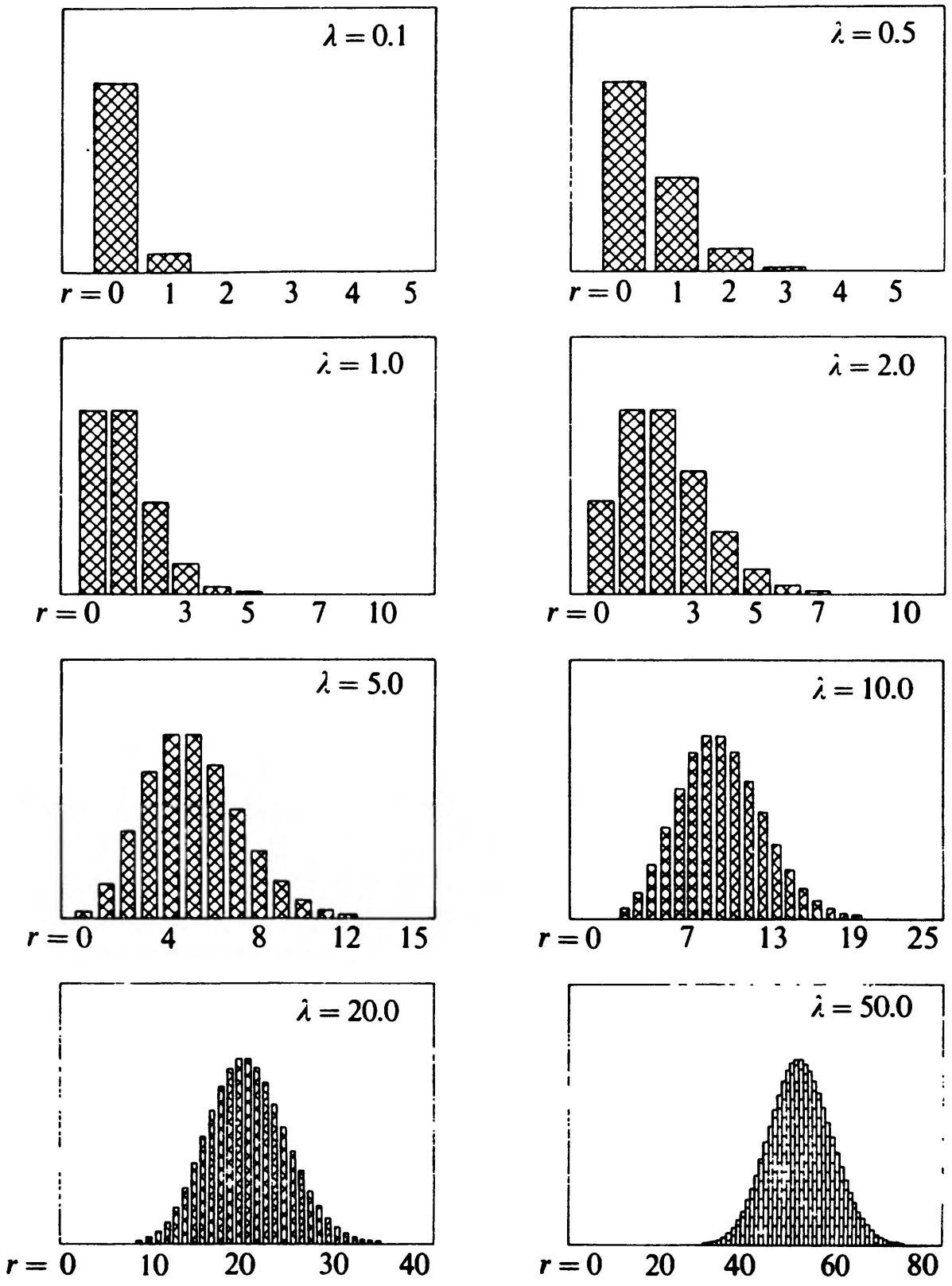


Fig. 3.2. Some Poisson distributions, with various values of λ. The vertical scale is arbitrary.

with variance $V(r) = \lambda$ (3.14)

and thus a standard deviation $\sigma = \sqrt{\lambda}$ (3.15)

and this last is overwhelmingly the most important thing to remember: for a Poisson distribution, the standard deviation is just the square root of the mean number of events.

Example More horse kicks
In the previous example of the Prussian horsemen, the mean was found to be 0.610. The variance is 0.608—almost identical.

The Poisson can make a useful approximation to the binomial distribution in cases where the number of trials, n, is large, and/or the probability p is small—it is easier to calculate as it does not involve messy factorials.

Example Poisson approximation of a binomial
If there are 100 trials, with individual probability of success of 2%, then the binomial probabilities for the numbers of successes are

r	0	1	2	3	4	5	6
P(binomial)	13.3%	27.1%	27.3%	18.2%	9.0%	3.5%	1.1%

The Poisson distribution, for a mean of 2, gives the probabilities

P(Poisson)	13.5%	27.1%	27.1%	18.0%	9.0%	3.6%	1.2%.

Unless you are very demanding, this accuracy is presumably ample, and the computation is much easier—try them yourself and see.

★3.3.2 Proof of Properties of the Poisson Distribution

To show that the normalisation (equation 3.12) is correct is straightforward

$$\sum_{r=0}^{\infty} P(r;\lambda) = e^{-\lambda} \sum \frac{\lambda^r}{r!}$$

$$= e^{-\lambda} e^{\lambda} \quad \text{(as the sum is just the expansion of } e^{\lambda})$$

$$= 1.$$

$\langle r \rangle$ is given by

$$\langle r \rangle = \sum^{\infty} r e^{-\lambda} \frac{\lambda^r}{r!}.$$

Drop the $r = 0$ term and take out some factors:

$$\langle r \rangle = \lambda e^{-\lambda} \sum_{r=1}^{\infty} \frac{\lambda^{r-1}}{(r-1)!}.$$

Set $r' = r - 1$:

$$\langle r \rangle = \lambda e^{-\lambda} \sum_{r'=0}^{\infty} \frac{\lambda r'}{r'!}$$

and use equation 3.12 to get equation 3.13:

$$\langle r \rangle = \lambda$$

To find $V(r)$, start with

$$\langle r(r-1) \rangle = \sum_{r=0}^{\infty} r(r-1) e^{-\lambda} \frac{\lambda^r}{r!}.$$

As before, dropping the first two terms and putting $r' = r - 2$,

$$\langle r^2 - r \rangle = \lambda^2 e^{-\lambda} \sum_{r'=0}^{\infty} \frac{\lambda^{r'}}{r'!}$$

$$\langle r^2 \rangle - \langle r \rangle = \lambda^2$$

and then using equation 3.13 gives equation 3.14:

$$\langle r^2 \rangle - \langle r \rangle^2 = \lambda^2 + \lambda - \lambda^2$$

$$V(r) = \lambda.$$

★3.3.3 Two Poisson Distributions

If there are two separate types of events occurring according to Poisson statistics and we do not distinguish between the two (for example, a radioactive source containing two different unstable isotopes both giving identical clicks on a Geiger counter), then the probability of r events is also Poisson, with mean equal to the sum of the two means.

Suppose the two events types are called a and b, with individual means λ_a and λ_b, so we know the probability of observing r_a and r_b. A total of r events could be all of type b, or one of type a and the rest of type b, and so on. The total probability is given by

$$P(r) = \sum_{r_a=0}^{r} P(r_a; \lambda_a) P(r - r_a; \lambda_b)$$

$$= e^{-\lambda_a} e^{-\lambda_b} \sum \frac{\lambda_a^{r_a} \lambda_b^{r-r_a}}{r_a!(r-r_a)!}$$

$$= e^{-(\lambda_a + \lambda_b)} \frac{(\lambda_a + \lambda_b)^r}{r!} \sum_{r_a=0}^{r} \frac{r!}{r_a!(r-r_a)!} \left(\frac{\lambda_a}{\lambda_a + \lambda_b} \right)^{r_a} \left(\frac{\lambda_b}{\lambda_a + \lambda_b} \right)^{r-r_a}.$$

The summation, on closer inspection, is just the binomial expansion of $\left(\frac{\hat{\lambda}_a}{\hat{\lambda}_a+\hat{\lambda}_b}+\frac{\hat{\lambda}_b}{\hat{\lambda}_a+\hat{\lambda}_b}\right)^r$, which is just 1, so the result is

$$P(r)=e^{-(\hat{\lambda}_a+\hat{\lambda}_b)}\frac{(\hat{\lambda}_a+\hat{\lambda}_b)^r}{r!} \tag{3.16}$$

i.e. the sum of two Poisson processes is another Poisson process. This can be extended to any number of Poisson processes. The proof also shows (from the fact that the sum is a binomial expansion) that given r events, the distribution of events of type a is described by a binomial, $P\left(r_a;\frac{\hat{\lambda}_a}{\hat{\lambda}_a+\hat{\lambda}_b},r\right)$.

3.4 THE GAUSSIAN DISTRIBUTION

3.4.1 The Gaussian Probability Distribution Function

$$P(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}. \tag{3.17}$$

The *Gaussian* or *normal* is the most well known and useful of all distributions. It is a bell-shaped curve centred on, and symmetric about, $x=\mu$. The width is controlled by the parameter σ, which is also the standard deviation of the distribution (which will be shown in section 3.4.2). It is broad if σ is large, narrow if σ is small. At $x=\mu\pm\sigma$, $P(x)$ falls to 0.61 of its peak value—at a bit more than half. These are also the points of inflexion, where the second derivative is zero.

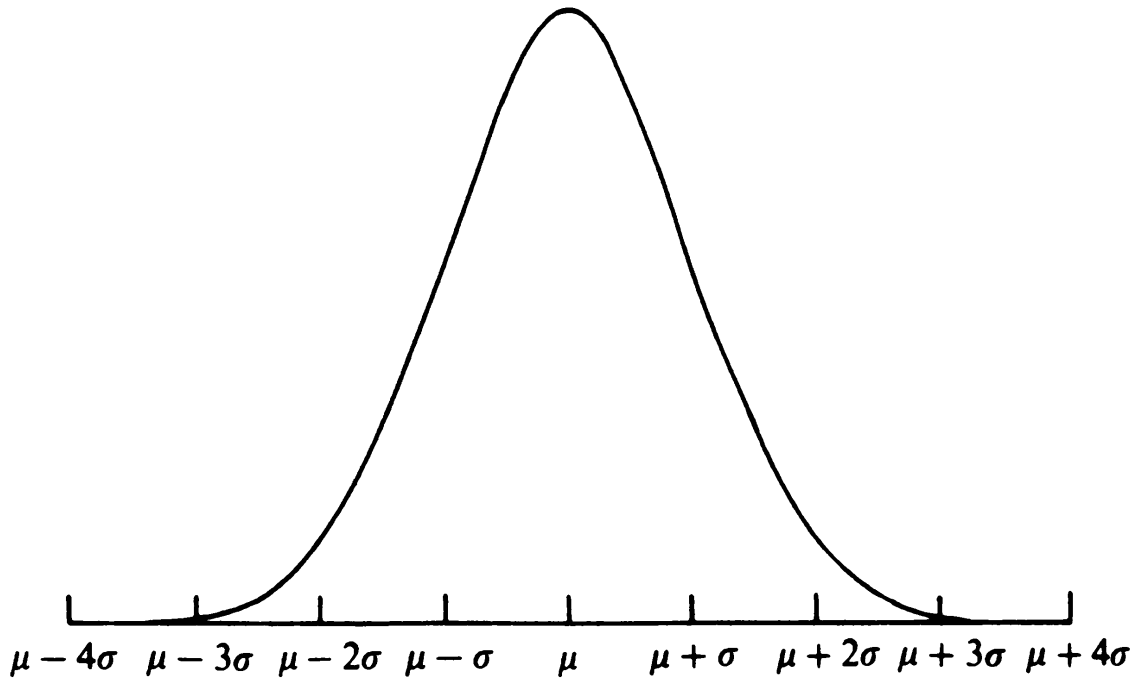


Fig. 3.3. The Gaussian distribution

Changing the value of μ shifts the distribution along the x axis without any change to its shape. Increasing or decreasing σ stretches or shrinks the curve about the central value. In this way all Gaussians are equivalent, in that a change of origin and scale reduces them to a standard form. This is why only one Gaussian is shown here, in contrast to the many pictures of different binomial and Poisson distributions. If you substitute $z = (x - \mu)/\sigma$ then the Gaussian becomes

$$\frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tag{3.18}$$

which is often called the *unit Gaussian* or *unit normal* distribution.

The important properties of the distribution (proofs, if required, are in section 3.4.2) are

it is normalized to 1:

$$\int_{-\infty}^{\infty} P(x; \mu, \sigma)\, dx = 1 \tag{3.19}$$

μ is the mean of the distribution:

$$\int_{-\infty}^{\infty} xP(x; \mu, \sigma)\, dx = \mu \tag{3.20}$$

(it is also the mode and the median.)

the standard deviation is σ, and variance σ^2:

$$\int_{-\infty}^{\infty} (x - \mu)^2 P(x; \mu, \sigma)\, dx = \sigma^2. \tag{3.21}$$

(This justifies our use of σ to represent both of the two quantities, the standard deviation of the distribution and the parameter in the Gaussian distribution formula, as they turn out to be the same.)

Although called after Gauss, the distribution was in fact discovered and investigated independently by many people. In France it is known as the *Laplacean*. The first recorded reference to it is by de Moivre (who was English) in 1733, in a work entitled *Approximatio ad summam terminorum binomii* $(a + b)^n$ *in seriem expansi.*

It is also often called the *normal* distribution. However, the use of this name implies a value-judgement (nobody, after all, would use an abnormal distribution) which is best avoided. It does indeed describe many different sorts of data, particularly in the field of measurement errors, but the reasons for this are complex and not to be glossed over by a bland label—this is the point of Lippman's famous remark (quoted by Poincaré): 'everybody believes in the law of errors, the experimenters because they think it is a

mathematical theorem, the mathematicians because they think it is an experimental fact.'

★3.4.2 Proof of Properties of the Gaussian

When working with Gaussians, it is usually simpler to shift the origin so that $\mu = 0$, but to leave in the scale factor of σ, as then the dimensions make sense. To prove the normalisation, we have to show that

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx = 1.$$

Setting $x' = x - \mu$ the expression becomes

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x'^2/2\sigma^2} dx'$$

and this integral is given in Table 3.1 (with $a = 1/2\sigma^2$), giving

$$\frac{1}{\sigma\sqrt{2\pi}} \sqrt{2\sigma^2\pi}$$

which is 1, as desired.

That μ is the mean of the distribution, which is also the expectation value $\langle x \rangle$, is obvious, but a proof can be spelt out if desired by writing

$$\langle x \rangle = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} x e^{-(x-\mu)^2/2\sigma^2} dx.$$

Putting $x = (x - \mu) + \mu$ and splitting the integral into two gives

$$\begin{aligned}\langle x \rangle &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} (x-\mu) e^{-(x-\mu)^2/2\sigma^2} dx + \mu \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx \\ &= 0 + \mu \times 1 \\ &= \mu.\end{aligned}$$

The variance is found from another standard integral from Table 3.1:

$$\begin{aligned}\langle (x-\mu)^2 \rangle &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} (x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \sigma^2.\end{aligned}$$

3.4.3 Definite Integrals

In working with the Gaussian function there are various standard integrals that occur frequently. Their derivation is usually straightforward, and can be found in any reputable mathematics textbook. They are collected here for convenience.

TABLE 3.1
USEFUL INTEGRALS

$$\int_{-\infty}^{\infty} e^{-ax^2}\,dx = \sqrt{\frac{\pi}{a}} \qquad \int_{-\infty}^{\infty} e^{-z^2/2}\,dz = \sqrt{2\pi}$$

$$\int_{0}^{\infty} x e^{-ax^2}\,dx = \frac{1}{2a} \qquad \int_{0}^{\infty} z e^{-z^2/2}\,dz = 1$$

$$\int_{-\infty}^{\infty} x^2 e^{-ax^2}\,dx = \frac{1}{2a}\sqrt{\frac{\pi}{a}} \qquad \int_{-\infty}^{\infty} z^2 e^{-z^2/2}\,dz = \sqrt{2\pi}.$$

Higher powers can be obtained by differentiating these with respect to a, giving

$$\int_{0}^{\infty} x^{2n+1} e^{-ax^2}\,dx = \frac{n!}{2a^{n+1}} \qquad \int_{0}^{\infty} z^{2n+1} e^{-z^2/2}\,dz = 2^n n!$$

$$\int_{-\infty}^{\infty} x^{2n} e^{-ax^2}\,dx = \frac{1.3.5...(2n-1)}{2^n a^n}\sqrt{\frac{\pi}{a}}$$

$$\int_{-\infty}^{\infty} z^{2n} e^{-z^2/2}\,dz = 1.3.5...(2n-1)\sqrt{2\pi}.$$

For any odd power, the symmetric integral vanishes:

$$\int_{-\infty}^{\infty} x^{2n+1} e^{-ax^2}\,dx = \int_{-\infty}^{\infty} z^{2n+1} e^{-z^2/2}\,dz = 0.$$

3.4.4 Indefinite Integrals

Unfortunately the indefinite integral of the Gaussian cannot be done analytically and written down as a nice expression. Instead you have to look it up in tables, or most reputable computers will provide a library function to evaluate it. Table 3.2 thus shows the value of the integrated Gaussian distribution, between the symmetric limits $-(x-\mu)/\sigma$ and $+(x-\mu)/\sigma$, i.e. the probability that, if an event is drawn from a Gaussian distribution, it will lie within some number of standard deviations of the mean. The probability that it will lie *outside* the range specified is, of course, just one minus the tabulated value.

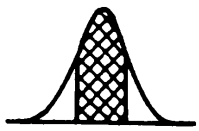
From Table 3.2 you can see that

68.27% of the area lies within σ of the mean,
95.45% lies within 2σ,
99.73% lies within 3σ.

If round numbers in the percentages are required, then

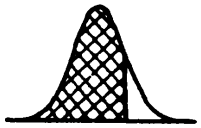
90% lie within 1.645σ,
95% lie within 1.960σ,
99% lie within 2.576σ,
99.9% lie within 3.290σ.

TABLE 3.2
TWO-TAILED GAUSSIAN INTEGRAL
Giving the percentage probability that a point lies within the given number of σ from the mean



$\frac{x-\mu}{\sigma}$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.00	0.80	1.60	2.39	3.19	3.99	4.78	5.58	6.38	7.17
0.10	7.97	8.76	9.55	10.34	11.13	11.92	12.71	13.50	14.28	15.07
0.20	15.85	16.63	17.41	18.19	18.97	19.74	20.51	21.28	22.05	22.82
0.30	23.58	24.34	25.10	25.86	26.61	27.37	28.12	28.86	29.61	30.35
0.40	31.08	31.82	32.55	33.28	34.01	34.73	35.45	36.16	36.88	37.59
0.50	38.29	38.99	39.69	40.39	41.08	41.77	42.45	43.13	43.81	44.48
0.60	45.15	45.81	46.47	47.13	47.78	48.43	49.07	49.71	50.35	50.98
0.70	51.61	52.23	52.85	53.46	54.07	54.67	55.27	55.87	56.46	57.05
0.80	57.63	58.21	58.78	59.35	59.91	60.47	61.02	61.57	62.11	62.65
0.90	63.19	63.72	64.24	64.76	65.28	65.79	66.29	66.80	67.29	67.78
1.00	68.27	68.75	69.23	69.70	70.17	70.63	71.09	71.54	71.99	72.43
1.10	72.87	73.30	73.73	74.15	74.57	74.99	75.40	75.80	76.20	76.60
1.20	76.99	77.37	77.75	78.13	78.50	78.87	79.23	79.59	79.95	80.29
1.30	80.64	80.98	81.32	81.65	81.98	82.30	82.62	82.93	83.24	83.55
1.40	83.85	84.15	84.44	84.73	85.01	85.29	85.57	85.84	86.11	86.38
1.50	86.64	86.90	87.15	87.40	87.64	87.89	88.12	88.36	88.59	88.82
1.60	89.04	89.26	89.48	89.69	89.90	90.11	90.31	90.51	90.70	90.90
1.70	91.09	91.27	91.46	91.64	91.81	91.99	92.16	92.33	92.49	92.65
1.80	92.81	92.97	93.12	93.28	93.42	93.57	97.71	93.85	93.99	94.12
1.90	94.26	94.39	94.51	94.64	94.76	94.88	95.00	95.12	95.23	95.34
2.00	95.45	95.56	95.66	95.76	95.86	95.96	96.06	96.15	96.25	96.34
2.10	96.43	96.51	96.60	96.68	96.76	96.84	96.92	97.00	97.07	97.15
2.20	97.22	97.29	97.36	97.43	97.49	97.56	97.62	97.68	97.74	97.80
2.30	97.86	97.91	97.97	98.02	98.07	98.12	98.17	98.22	98.27	98.32
2.40	98.36	98.40	98.45	98.49	98.53	98.57	98.61	98.65	98.69	98.72
2.50	98.76	98.79	98.83	98.86	98.89	98.92	98.95	98.98	99.01	99.04
2.60	99.07	99.09	99.12	99.15	99.17	99.20	99.22	99.24	99.26	99.29
2.70	99.31	99.33	99.35	99.37	99.39	99.40	99.42	99.44	99.46	99.47
2.80	99.49	99.50	99.52	99.53	99.55	99.56	99.58	99.59	99.60	99.61
2.90	99.63	99.64	99.65	99.66	99.67	99.68	99.69	99.70	99.71	99.72
3.00	99.73	99.74	99.75	99.76	99.76	99.77	99.78	99.79	99.79	99.80
3.10	99.81	99.81	99.82	99.83	99.83	99.84	99.84	99.85	99.85	99.86
3.20	99.86	99.87	99.87	99.88	99.88	99.88	99.89	99.89	99.90	99.90
3.30	99.90	99.91	99.91	99.91	99.92	99.92	99.92	99.92	99.93	99.93
3.40	99.93	99.94	99.94	99.94	99.94	99.94	99.94	99.95	99.95	99.95
3.50	99.95	99.96	99.96	99.96	99.96	99.96	99.96	99.96	99.97	99.97

TABLE 3.3
ONE-TAILED GAUSSIAN INTEGRAL
Giving the percentage probability that a point lies within the given number of σ to one side of the mean



$\frac{x-\mu}{\sigma}$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	50.00	50.40	50.80	51.20	51.60	51.99	52.39	52.79	53.19	53.59
0.10	53.98	54.38	54.78	55.17	55.57	55.96	56.36	56.75	57.14	57.53
0.20	57.93	58.32	58.71	59.10	59.48	59.87	60.26	60.64	61.03	61.41
0.30	61.79	62.17	62.55	62.93	63.31	63.68	64.06	64.43	64.80	65.17
0.40	65.54	65.91	66.28	66.64	67.00	67.36	67.72	68.08	68.44	68.79
0.50	69.15	69.50	69.85	70.19	70.54	70.88	71.23	71.57	71.90	72.24
0.60	72.57	72.91	73.24	73.57	73.89	74.22	74.54	74.86	75.17	75.49
0.70	75.80	76.11	76.42	76.73	77.04	77.34	77.64	77.94	78.23	78.52
0.80	78.81	79.10	79.39	79.67	79.95	80.23	80.51	80.78	81.06	81.33
0.90	81.59	81.86	82.12	82.38	82.64	82.89	83.15	83.40	83.65	83.89
1.00	84.13	84.38	84.61	84.85	85.08	85.31	85.54	85.77	85.99	86.21
1.10	86.43	86.65	86.86	87.08	87.29	87.49	87.70	87.90	88.10	88.30
1.20	88.49	88.69	88.88	89.07	89.25	89.44	89.62	89.80	89.97	90.15
1.30	90.32	90.49	90.66	90.82	90.99	91.15	91.31	91.47	91.62	91.77
1.40	91.92	92.07	92.22	92.36	92.51	92.65	92.79	92.92	93.06	93.19
1.50	93.32	93.45	93.57	93.70	93.82	93.94	94.06	94.18	94.29	94.41
1.60	94.52	94.63	94.74	94.84	94.95	95.05	95.15	95.25	95.35	95.45
1.70	95.54	95.64	95.73	95.82	95.91	95.99	96.08	96.16	96.25	96.33
1.80	96.41	96.49	96.56	96.64	96.71	96.78	96.86	96.93	96.99	97.06
1.90	97.13	97.19	97.26	97.32	97.38	97.44	97.50	97.56	97.61	97.67
2.00	97.72	97.78	97.83	97.88	97.93	97.98	98.03	98.08	98.12	98.17
2.10	98.21	98.26	98.30	98.34	98.38	98.42	98.46	98.50	98.54	98.57
2.20	98.61	98.64	98.68	98.71	98.75	98.78	98.81	98.84	98.87	98.90
2.30	98.93	98.96	98.98	99.01	99.04	99.06	99.09	99.11	99.13	99.16
2.40	99.18	99.20	99.22	99.25	99.27	99.29	99.31	99.32	99.34	99.36
2.50	99.38	99.40	99.41	99.43	99.45	99.46	99.48	99.49	99.51	99.52
2.60	99.53	99.55	99.56	99.57	99.59	99.60	99.61	99.62	99.63	99.64
2.70	99.65	99.66	99.67	99.68	99.69	99.70	99.71	99.72	99.73	99.74
2.80	99.74	99.75	99.76	99.77	99.77	99.78	99.79	99.79	99.80	99.81
2.90	99.81	99.82	99.82	99.83	99.84	99.84	99.85	99.85	99.86	99.86
3.00	99.87	99.87	99.87	99.88	99.88	99.89	99.89	99.89	99.90	99.90
3.10	99.90	99.91	99.91	99.91	99.92	99.92	99.92	99.92	99.93	99.93
3.20	99.93	99.93	99.94	99.94	99.94	99.94	99.94	99.95	99.95	99.95
3.30	99.95	99.95	99.95	99.96	99.96	99.96	99.96	99.96	99.96	99.97
3.40	99.97	99.97	99.97	99.97	99.97	99.97	99.97	99.97	99.97	99.98
3.50	99.98	99.98	99.98	99.98	99.98	99.98	99.98	99.98	99.98	99.98

The 2σ value is so close to 95% (and vice versa) that in practice the difference can often be ignored. From the 1σ value you obtain the useful rule of thumb that when a curve is shown going through a set of measured points with error bars, about one third of the error bars should miss the curve. Many people fail to realise this and overestimate their errors in an effort to make the curve go through all the points. It is thus a standard ploy in seminars, etc., when hapless speakers proudly present fitted data, to attack them for having too good a fit.

Sometimes you are interested in the probability of a value straying in one direction only—for example, you may want to be sure that some upper limit is not exceeded, but do not care how far it strays below the mean. For this you need the *one-tailed* probability, as shown in Table 3.3, as opposed to the *two-tailed* probability of Table 3.2.

Should you ever need to know the integrated Gaussian for any other (asymmetric) limits, it can be obtained from these tables by simple arithmetic. Indeed, Tables 3.2 and 3.3 can readily be obtained from each other, but both are given here for convenience of use.

3.4.5 Gaussian as Limit of the Poisson and Binomial

From the distributions shown Figure 3.2 it can be seen that for large λ, the Poisson distribution tends to a Gaussian shape, with $\mu = \lambda$, $\sigma = \sqrt{\lambda}$. In such cases the Gaussian may be used as a very convenient approximation to the Poisson. What is 'large' depends on how good an agreement you require. Some people put the requirement as low as $\lambda = 5$, but 10 is probably safer.

Proof: let $r = \lambda + x$, and use Stirling's approximation:

$$\ln r! \approx r \ln r - r + \ln \sqrt{(2\pi r)}.$$

Then, taking the logarithm of equation 3.11,

$$\ln P(r;\lambda) \approx -\lambda + r\ln\lambda - (r\ln r - r) - \ln\sqrt{2\pi r}$$

$$\approx -\lambda + r\left\{\ln\lambda - \ln\left[\lambda\left(1+\frac{x}{\lambda}\right)\right]\right\} + (\lambda + x) - \ln\sqrt{2\pi\lambda}.$$

Using the expansion $\ln(1+z) = z - z^2/2\ldots$,

$$\ln P(r;\lambda) \approx x - (\lambda + x)\left(\frac{x}{\lambda} - \frac{x^2}{2\lambda^2}\right) - \ln\sqrt{2\pi\lambda}$$

$$\approx -\frac{x^2}{2\lambda} - \ln\sqrt{2\pi\lambda}.$$

Thus, exponentiating,

$$P(x) = \frac{e^{-x^2/2\lambda}}{\sqrt{2\pi\lambda}}.$$

Example Poisson approximated by Gaussian
If λ is 5.3, then the probability of two events or less is 10.2%, using the Poisson formula. Approximating the histogram of the Poisson by the smooth Gaussian curve, the appropriate value for the Gaussian is halfway between the possible discrete values of 2 and 3, at 2.5 'events'. This is $(5.3-2.5)/\sqrt{5.3} = 1.22\sigma$ from the mean, and Table 3.3 gives this one-tailed probability as 11.1%.

Likewise the binomial tends to a Gaussian with $\mu = np$ and $\sigma = \sqrt{np(1-p)}$. (The proof is similar to that for the Poisson.) This happens first for $p \approx 0.5$; large or small values of p require a larger n. Indeed, almost everything tends to a Gaussian as the numbers become large—this is due to the *central limit theorem*, discussed in the next chapter.

★3.4.6 The Many-dimensional Gaussian

Consider a distribution in n variables, denoted by $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$—the notation is discussed in section 2.6.3. These can be written compactly as a vector $\mathbf{x}$, likewise the means, $\mu_{(1)}, \mu_{(2)}, \ldots, \mu_{(n)}$ can be written $\boldsymbol{\mu}$. The most general form of multi-dimensional Gaussian is an exponential of a quadratic form, which will contain terms in $x_{(i)}^2$, cross terms in $x_{(i)}x_{(j)}$, linear terms, and a constant, but nothing of higher power. This can be written:

$$P(\mathbf{x}) \propto \exp\left[-\frac{1}{2}(\tilde{\mathbf{x}} - \tilde{\boldsymbol{\mu}})\mathbf{A}(\mathbf{x} - \boldsymbol{\mu}) \right].$$

Even this contains some ambiguity, which can be resolved by insisting that $\mathbf{A}$ be symmetric:

$$A_{ij} = A_{ji}.$$

Henceforth suppose, without loss of generality, that all $\mu_{(i)}$ are zero.

It may be that $\mathbf{A}$ is diagonal, i.e. all the cross terms are zero. In that case $P(\mathbf{x})$ factorises into n independent Gaussians:

$$e^{-(A_{11}x_1^2 + A_{22}x_2^2 + A_{33}x_3^2 + \cdots)/2} = e^{-A_{11}x_1^2/2} e^{-A_{22}x_2^2/2} e^{-A_{33}x_3^2/2} \ldots$$

and the diagonal elements can be identified as

$$A_{ii} = \frac{1}{\sigma_i^2}.$$

As $\mathbf{A}$ is diagonal this can be written in the form

$$\mathbf{A} = \mathbf{V}^{-1}. \tag{3.22}$$

Now we go on to consider the general case and to show that the above equation is still true. Even if $\mathbf{A}$ is not diagonal, a unitary matrix $\mathbf{U}$ can always be found to diagonalise it; i.e.

$$\mathbf{U}\mathbf{A}\tilde{\mathbf{U}} = \mathbf{A}' \quad \text{where } \mathbf{A}' \text{ is diagonal}$$

Note: 'unitary' means that the transposed matrix is the same as the inverse: $\mathbf{U}^{-1} = \tilde{\mathbf{U}}$. The significance of this is that if one considers vectors $\mathbf{x}, \mathbf{y}, \ldots$ transformed by $\mathbf{U}$

$$\mathbf{x}' = \mathbf{U}\mathbf{x} \quad \mathbf{y}' = \mathbf{U}\mathbf{y} \quad \text{etc.}$$

then the transposes (denoted by a tilde, ~) are given by

$$\tilde{\mathbf{x}}' = \tilde{\mathbf{x}}\tilde{\mathbf{U}}$$

so the 'scalar product' of two vectors does not change under transformations by a unitary matrix,

$$\tilde{\mathbf{x}}'\mathbf{y}' = \tilde{\mathbf{x}}\tilde{\mathbf{U}}\mathbf{U}\mathbf{y} = \tilde{\mathbf{x}}\mathbf{U}^{-1}\mathbf{U}\mathbf{y} = \tilde{\mathbf{x}}\mathbf{y}$$

and they thus represent generalised rotations.

It is a basic fact of matrix algebra that for any symmetric matrix a unitary matrix can always be found for which $\mathbf{U}\mathbf{A}\tilde{\mathbf{U}}$ is diagonal.

The exponent $\tilde{\mathbf{x}}\mathbf{A}\mathbf{x}$ can be written

$$\tilde{\mathbf{x}}\tilde{\mathbf{U}}\mathbf{U}\mathbf{A}\tilde{\mathbf{U}}\mathbf{U}\mathbf{x}.$$

This is $\mathbf{x}'\mathbf{A}'\mathbf{x}'$, with $\mathbf{A}'$ diagonal. The variance matrix $\mathbf{V}'$, for the $\mathbf{x}'$, is thus diagonal with elements $(\mathbf{U}\mathbf{A}\tilde{\mathbf{U}})^{-1} = \mathbf{U}\mathbf{A}^{-1}\tilde{\mathbf{U}}$, by equation 3.22.

So we know the variance matrix for the $\mathbf{x}'$, and also that the $\mathbf{x}$ are related to these by $\mathbf{x} = \tilde{\mathbf{U}}\mathbf{x}'$. We now (anticipating a result from the next chapter) invoke the generalised combination of errors formula, equation 4.19, which gives the variance matrix for a set of variables which are a function of another set. (Incidentally, as in this case the relation is linear, the equation is exact and not an approximation.) The derivative matrix $\mathbf{G}$ in 4.19 is just $\mathbf{U}$, so

$$\begin{aligned} \mathbf{V} &= \tilde{\mathbf{U}}\mathbf{V}'\mathbf{U} = \tilde{\mathbf{U}}\mathbf{U}\mathbf{A}^{-1}\tilde{\mathbf{U}}\mathbf{U} \\ \mathbf{V} &= \mathbf{A}^{-1}. \end{aligned} \tag{3.23}$$

Result *The matrix in the exponent of the multidimensional Gaussian is the inverse of the covariance matrix.*

In full, with the normalisation (which can be found from the Jacobian of the $x \to x'$ transformation):

$$P(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{|\mathbf{V}|}} \exp\left[-\tfrac{1}{2}(\tilde{\mathbf{x}} - \tilde{\boldsymbol{\mu}})\mathbf{V}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right].$$

★3.4.7 The Binormal Distribution

For two dimensions (calling the variables x and y again, rather than x_1...

and $x_{(2)}$) the covariance matrix is

$$V = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

which has the inverse

$$\mathbf{V}^{-1} = \frac{1}{\sigma_x^2\sigma_y^2(1-\rho^2)} \begin{pmatrix} \sigma_y^2 & -\rho\sigma_x\sigma_y \\ -\rho\sigma_x\sigma_y & \sigma_x^2 \end{pmatrix}.$$

The full formula (including normalisation) for the binormal or two-dimensional Gaussian is thus

$$P(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{(1-\rho^2)}} \times \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) \right] \right\}. \tag{3.24}$$

This can be drawn on paper using contour lines. The contours of equal probability are curves for which the exponent in equation 3.24 is constant, and that is the equation of an ellipse. Manipulation of equation 3.24 shows that the ellipse for which the exponent is $-\frac{1}{2}$ has extreme x and y values at $\mu_x \pm \sigma_x$ and $\mu_y \pm \sigma_y$, i.e. it fits exactly into a rectangular box between these limits.

If you take a slice through the distribution, considering the distribution in y, say, for a fixed value of x, then, by inspection, equation 3.24 becomes a Gaussian distribution in y whose standard deviation is narrowed to $\sigma_y/\sqrt{1-\rho^2}$ and mean is $\mu_y + \rho(\sigma_y/\sigma_x)(x-\mu_x)$.

In two dimensions the unitary matrix U that diagonalises the exponent is the familiar rotation matrix

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

which rotates the (x, y) axes by some angle θ such that the major and minor axes of these ellipses coincide with the new axes: call them u and v. The three parameters of the binormal can thus be written σ_x, σ_y, ρ, as previously, or as σ_u, σ_v, θ, where u and v are uncorrelated and with standard deviations σ_u and σ_v, and the x, y system is given by rotating the u, v system through an angle θ.

A little algebra gives the relations between the two parameter sets:

$$\tan 2\theta = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2}$$

$$\sigma_u^2 = \frac{\cos^2\theta\sigma_x^2 - \sin^2\theta\sigma_y^2}{\cos^2\theta - \sin^2\theta} \qquad \sigma_v^2 = \frac{\cos^2\theta\sigma_y^2 - \sin^2\theta\sigma_x^2}{\cos^2\theta - \sin^2\theta}$$

$$\sigma_x^2 = \cos^2\theta\sigma_u^2 + \sin^2\theta\sigma_v^2 \quad \sigma_y^2 = \cos^2\theta\sigma_v^2 + \sin^2\theta\sigma_u^2$$

$$\rho = \sin\theta\cos\theta\frac{\sigma_u^2 - \sigma_v^2}{\sigma_x\sigma_y}.$$

Figure 3.4 shows the lines of constant probability for a two-dimensional distribution, where x and y are positively correlated. The ellipses of constant probability (at 90, 80,...,10% of the peak value) are shown. The parameters for this figure are

$$\sigma_u = 1.0 \quad \sigma_v = 0.5 \quad \theta = 45°$$

or, equivalently,

$$\sigma_x = \sqrt{\frac{5}{8}} \quad \sigma_y = \sqrt{\frac{5}{8}} \quad \rho = \frac{3}{5}.$$

★3.5 OTHER DISTRIBUTIONS

The Gaussian, Poisson, and binomial distributions are, in that order, far and away the most common and useful. However, they are not the only ones, and some others are described here; in addition the *χ^2 distribution*, *Student's t distribution*, and *Fisher's F distribution* will be discussed in later chapters.

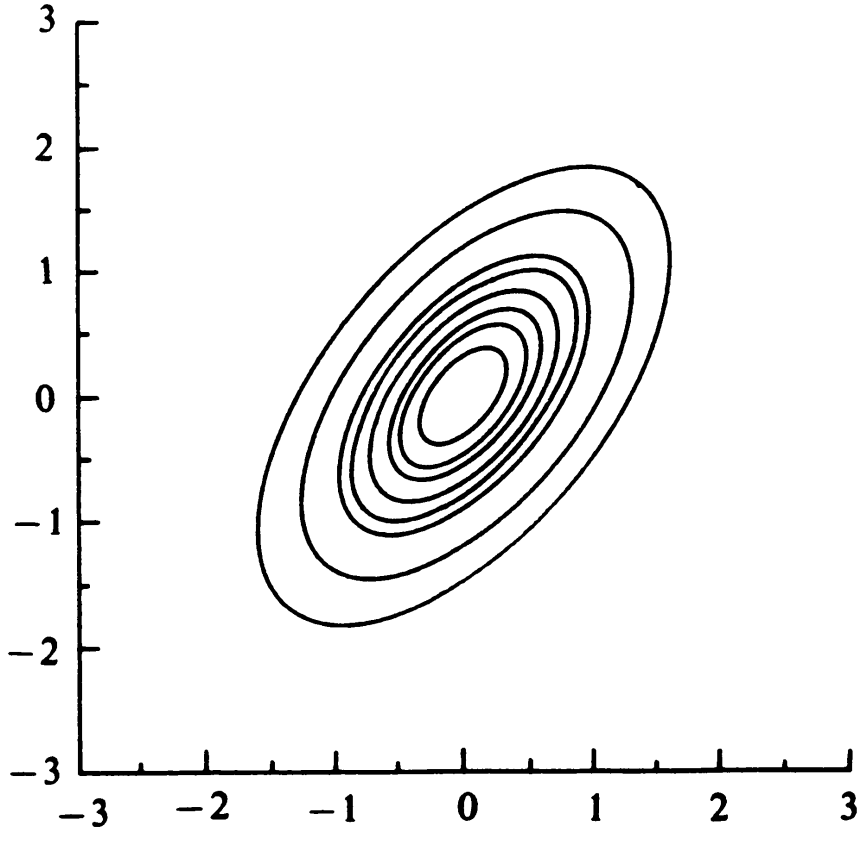


Fig. 3.4 The binormal distribution

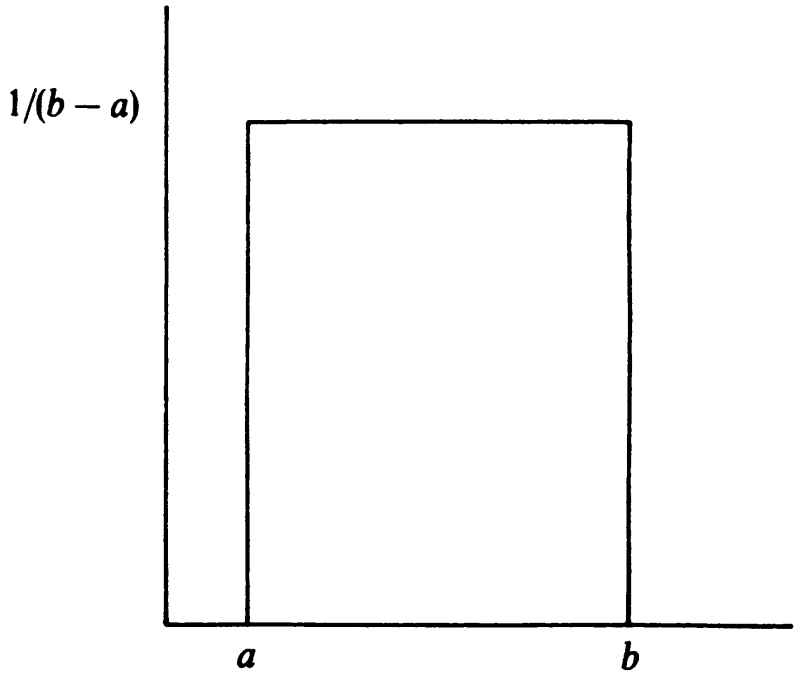


Fig. 3.5. The uniform distribution.

★3.5.1 The Uniform Distribution

Also known as the rectangular or top hat distribution, the uniform distribution (Figure 3.5) describes a probability which is constant over a certain range and zero outside it. If the range limits are a and b then

$$P(x) = \begin{cases} \dfrac{1}{b-a} & \text{for} \quad a \leqslant x \leqslant b \\ 0 & \text{elsewhere.} \end{cases}$$

The mean is obviously $(a+b)/2$. On doing the integral to obtain $\langle x^2 \rangle$ and thus the variance one gets

$$V(x) = \frac{(b-a)^2}{12} \tag{3.25}$$

i.e. the standard deviation for a uniform distribution is the width divided by $\sqrt{12}$.

★3.5.2 The Weibull Distribution

$$P(x; \alpha, \beta) = \alpha\beta(\alpha x)^{\beta-1} e^{-(\alpha x)^{\beta}}.$$

Originally invented to describe failure rates in ageing lightbulbs, the Weibull distribution (Figure 3.6) is useful for parametrising functions which rise as x increases from 0 and then fall again. α is just a scale factor. β expresses the sharpness of the peak. $\beta = 1$ gives the exponential function

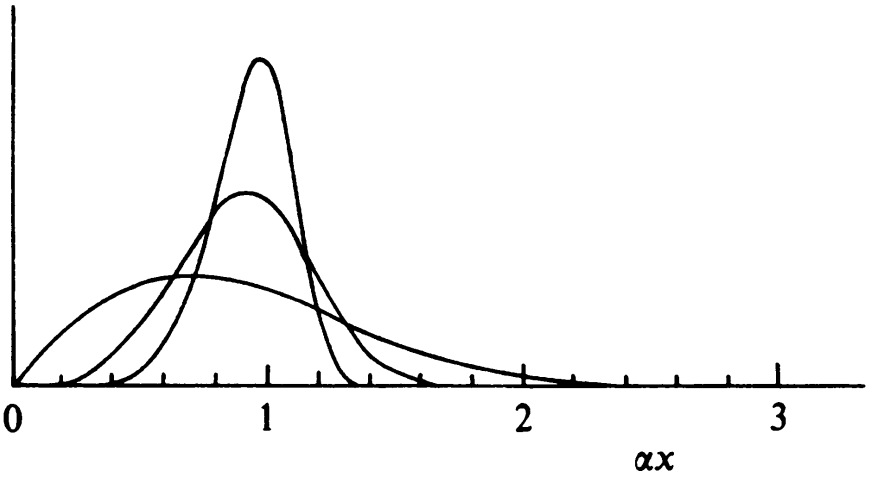


Fig. 3.6. Some Weibull functions. The successively sharper peaks are for $\beta = 2.0$, 4.0, and 7.0.

★3.5.3 The Breit–Wigner or Cauchy Distribution

$$F(m; M, \Gamma) = \frac{1}{2\pi} \frac{\Gamma}{(m - M)^2 + (\Gamma/2)^2}$$

$$F(z) = \frac{1}{\pi} \frac{1}{1 + z^2}.$$

The Breit–Wigner function, used by nuclear physicists to give the distribution of particles of mass m due to a resonance of mass M and width Γ, reduces to the Cauchy function $F(z)$ (Figure 3.7) by a change of origin and scale. Its chief feature is its unlovable mathematical behaviour. It does not have a variance as the integral $\int z^2 F(z)\,dz$ diverges.

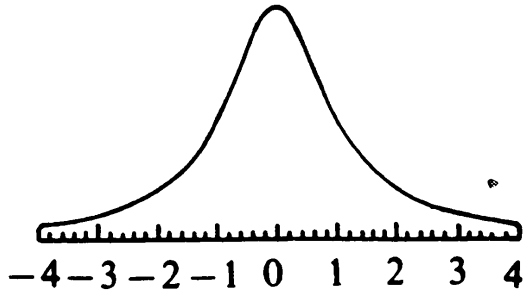


Fig. 3.7. The Cauchy function.

3.6 PROBLEMS

3.1

A defence system is 99.5% efficient in intercepting ballistic missiles. What is the probability that it will intercept all of 100 missiles launched against it? How many

missiles must an aggressor launch to have a better than evens chance of one or more penetrating the defences?

3.2
In the previous question, how many missiles would be needed to ensure a better than evens chance of more than two missiles evading the defences?

3.3
During a meteor shower, meteors fall at the rate 15.7 per hour. What is the probability of observing less than 5 in a given period of 30 minutes?

3.4
Repeat the previous problem, using the Gaussian approximation to the Poisson.

3.5
A student is trying to hitch a lift. Cars pass at random intervals, at an average rate of 1 per minute. The probability of a car giving a lift is 1%. What is the probability that the student will still be waiting:
(a) after 60 cars have passed?
(b) after 1 hour?

3.6
For a Gaussian distribution, using Tables 3.2 and 3.3:
(a) What is the probability of a value lying more than 1.23σ from the mean?
(b) What is the probability of a value lying more than 2.43σ above the mean?
(c) What is the probability of a value lying less than 1.09σ below the mean?
(d) What is the probability of a value lying above a point 0.45σ below the mean?
(e) What is the probability that a value lies more than 0.5σ but less than 1.5σ from the mean?
(f) What is the probability that a value lies above 1.2σ on the low side of the mean, and below 2.1σ on the high side?
(g) Within how many standard deviations does the probability of a value occurring equal 50%?
(h) How many standard deviations correspond to a one-tailed probability of 99%?

★*3.7*
Show that the skew and curtosis of a Gaussian are zero.