Chapter 4

Useful distributions

This chapter introduces four important distributions that can be used to describe a variety of situations. The first distribution encountered is that of the binomial distribution (section 4.2). This is used to understand problems where the possible outcomes are binary, and usually categorised in terms of success and failure. For example one can consider the situation of either detecting of failing to detect a particle passing through some apparatus as a binary event. The detection efficiency[1] in this particular problem is the parameter p of the binomial distribution. Typically one finds that $p \sim 1$ when working with efficient detectors. The Poisson distribution (section 4.3) can be used to understand rare events where the total number of trials is not necessarily known, and the distribution depends on only the number of observed events and a single parameter λ that is both the mean and variance of the distribution. For example the Poisson distribution can be used to describe the uncertainties on the content of each bin in Figure **??**, which is a topic discussed in more detail in chapter 6. The third distribution discussed here is the Gaussian distribution. This plays a significant role in describing the uncertainties on measurements where the number of data are large. Finally the χ^2 distribution is introduced in section 4.5. This distribution is typically encountered less frequently than the others, but still plays an important role in statistical data analysis. One of the uses of this distribution is to quantify the so-called goodness of fit between a model and a set of data. The probability determined for a given χ^2 and number of degrees of freedom can be a useful factor in determining if a fit result is valid, a topic encountered in chapter 8.

These distributions are related to each other: in the limit of infinite data the binomial and χ^2 distributions tend to a Gaussian distribution. The Poisson distribution tends to a Gaussian distribution in the limit that $\lambda \to \infty$, and the binomial distribution is related to the Poisson distribution in the limit $p \to 0$. Additional distributions are often encountered and a number of potentially useful ones are described in appendix B. Section 4.1 introduces the formalism required to determine the expectation values of discrete and continuous distributions, which will be of use in the remainder of this chapter. It is worth noting that ones ability to correctly manipulate these distributions may vary depending on the algorithm used (see section 4.6).

4.1 Expectation values of probability density functions

The notion of a probability density function (PDF) was introduced in section 2.5. For some PDF denoted by $P(x)$ describing a continuous distribution (or $P(x_i)$ for a discrete distribution), we can compute the ***expectation value*** (the average value) of some quantity as the integral (or sum) over the quantity multiplied by the PDF. For example the expectation value of the variable x, distributed according to the PDF $P(x)$ in

[1] The efficiency of detecting a particle is the ratio of the number of detected particles divided by the total number of particles. This topic is discussed in section 5.3.

the domain $-\infty < x < \infty$, is

$$\langle x \rangle = \int_{-\infty}^{+\infty} x P(x) dx, \tag{4.1.1}$$

in analogy with the discussion in section 3.2. If we replace the variable x in Eq. (4.1.1) by a more complicated expression then we can compute the expectation values for other quantities. For example the mean value of $V = (x - \overline{x})^2$ is given by

$$\langle V \rangle = \int_{-\infty}^{+\infty} (x - \overline{x})^2 P(x) dx. \tag{4.1.2}$$

The equivalent equations for a discrete distribution $P(x_i)$ are

$$\langle x \rangle = \sum_i x_i P(x_i), \tag{4.1.3}$$

and

$$\langle V \rangle = \sum_i (x_i - \overline{x})^2 P(x_i), \tag{4.1.4}$$

where the sum is over all bins. These results will be useful throughout the remainder of this chapter.

4.2 Binomial distribution

Consider the case of a flipping an unbiased coin as described in section 2.7.1. There are two possible outcomes; H and T. If we choose H, and flip a coin, the probability of a success p is 0.5, and the probability of a failure $(1 - p) = q$ is also 0.5. We can try flipping the coin a number of times n. For each coin flip there are two possible outcomes: success and failure, thus there are 2^n possible permutations of flipping the coin. It is possible to compute the number of combinations of obtaining r successes from n trials as

$$^nC_r = \frac{n!}{r!(n-r)!}. \tag{4.2.1}$$

The probability of success and failure are equal for flipping an unbiased coin, and we previously assigned them values of p and $1 - p$, respectively. We can multiply the number of possible permutations as given by ^nC_r by the probability of r successes and $n - r$ failures in order to obtain

$$P(r; p, n) = p^r (1 - p)^{n-r} \frac{n!}{r!(n-r)!}. \tag{4.2.2}$$

The result $P(r; p, n)$ is the probability of obtaining r successes from n experiments where the probability of a successful outcome of an experiment is given by p. We have seen this result before in the context of flipping a coin (see section 2). Now we have obtained a generalised solution.

The ***binomial distribution*** is the distribution corresponding to the probabilities computed using Eq. (B.1.3) and is a function of r, p, and n. In general this can be used to describe the probability of the number of possible outcomes when repeating an experiment with binary output a number of times. Figure 4.1 shows the binomial distribution expected for several different cases of n and p. Note that when $p = 0.5$, and $q = 0.5$, the distribution obtained is symmetric about the mean value. The mean and variance of a binomial probability distribution are given by (See section 4.2.1 for the proof of these results)

$$\langle r \rangle = np, \quad (4.2.3)$$
$$V(r) = np(1-p). \quad (4.2.4)$$

Hence the standard deviation of a binomial distribution is $\sigma = \sqrt{np(1-p)}$.

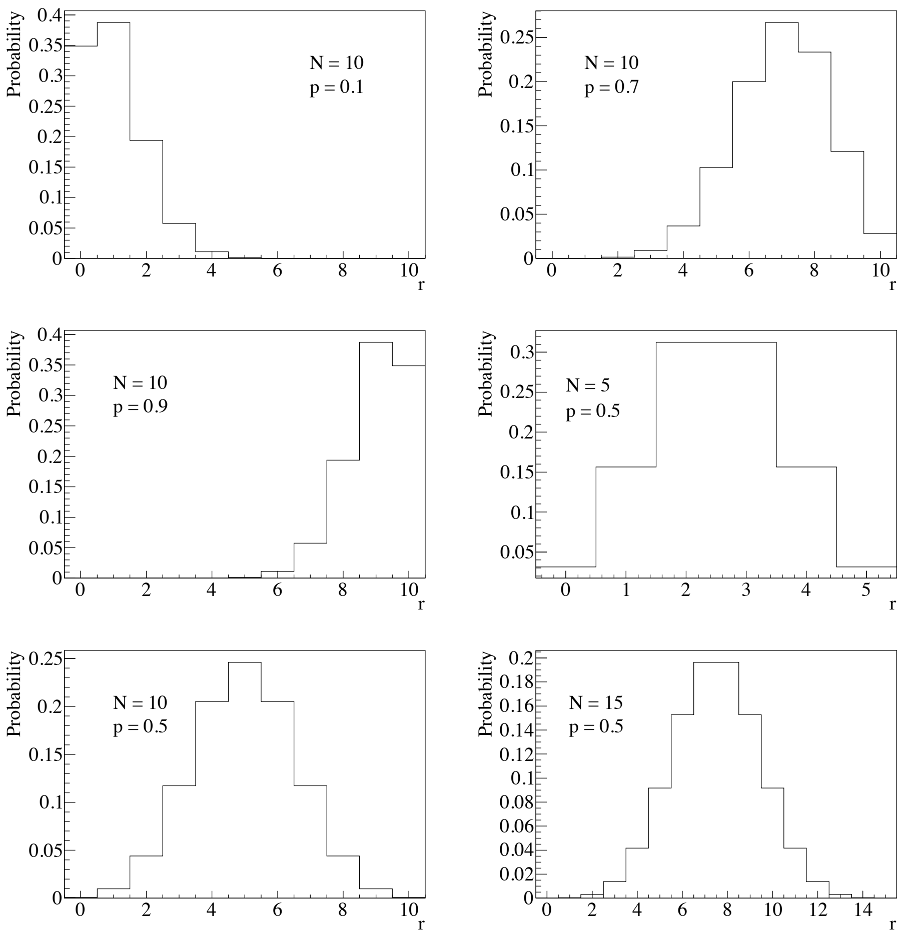

Figure 4.1: Binomial distributions for different values of n and p.

Example: Consider two coins - one being unbiased with $p = 0.5$, and the other being biased toward heads with $p = 0.7$. If we flip both coins ten times, what is the probability that we obtain equal numbers of heads (H) and tails (T) for each coin?

Firstly, let's consider how many combinations of H and T there are in order to obtain this result. This is given by Eq. (4.2.1) where $n = 10$ and $r = 5$, so the number of combinations we are interested in is

$$^{10}C_5 = \frac{10!}{5!5!}. \quad (4.2.5)$$

This can be used to compute the probability of obtaining 5H5T with a fair coin, using Eq. (B.1.3) which is

$$P(5; 0.5, 10) = 0.5^5(1 - 0.5)^5 \frac{10!}{5!5!}. \quad (4.2.6)$$

$$= 0.246. \quad (4.2.7)$$

If we use the second coin, then as $p = 0.7$ we obtain a probability of $P(5; 0.7, 10) = 0.103$. There is a factor of 2.4 difference in the probability of obtaining equal numbers of heads compared to tails when using the biased coin instead of the fair coin. The mean number of heads obtained with the fair coin is five, however if we use the biased coin we would find that the mean number of heads obtained would be seven. The distributions of possible outcomes as a function of r (the number of heads obtained) for these experiments are shown in Figure 4.1. The top-right figure is for the biased coin, and the bottom-left one is for the unbiased coin. Tables of binomial probabilities can be found in appendix D.1.

Another way of viewing the binomial distribution can be illustrated in terms of the concept of a testing for a biased coin. Instead of considering absolute probabilities one can study the likelihood distribution and compare possible outcomes. If we perform an experiment where we flip a coin that we suspect is biased some number of times n, then we can use the number of observed heads r to compute the probability $P(p; n, r) \propto p^r(1 - p)^{n-r}$. The likelihood distribution $L(p; n, r)$ is obtained when the maximum value of $P(p; n, r)$ is set to unity. Figure 4.2 shows the likelihood of p obtained for r heads for five and fifteen trials. If you flipped a coin five times and got zero heads, you might start to become suspicious that the coin could be biased, but you would not be able to rule out the possibility that the coin was really unbiased. One can however compare the likelihood of an unbiased coin against one with a given bias, and compute a likelihood ratio of the two possible outcomes. If on the other hand you flip a coin 15 times and only obtain tails, then you would be left with two possibilities. Either the coin is biased, or your 15 trials led to an extremely unlucky outcome. The former conclusion is the most likely one given the available data, however it should be noted that while in this case the likelihood that the coin is unbiased is very small, it is not zero. Hence it is possible to conclude that the coin is biased only if one is prepared to accept the possibility that there is a chance that the conclusion may be incorrect. This issue exemplifies the need for hypothesis testing which is discussed in chapter 7.

4.2.1 Derivation of the mean and variance of the binomial distribution

The binomial distribution is given by Eq. (B.1.3). For this to be valid, the total probability of anything to happen should be unity. In other words, it should be certain that something happens for a given value of p, and a given number of trials n. So we can sum Eq. (B.1.3) over r to verify this property

$$\sum_{r=0}^{r=n} P(r; p, n) = \sum_{r=0}^{r=n} p^r(1 - p)^{n-r} \frac{n!}{r!(n - r)!}, \quad (4.2.8)$$

$$= (1 - p)^n + np(1 - p)^{n-1} + \frac{n(n - 1)}{2!} p^2(1 - p)^2 + \ldots$$
$$+np^{n-1}(1 - p) + p^n. \quad (4.2.9)$$

One can prove by induction (for $n = 0, 1, 2, 3, \ldots$, thus for arbitrary n) that the sum of this binomial expansion is always unity.

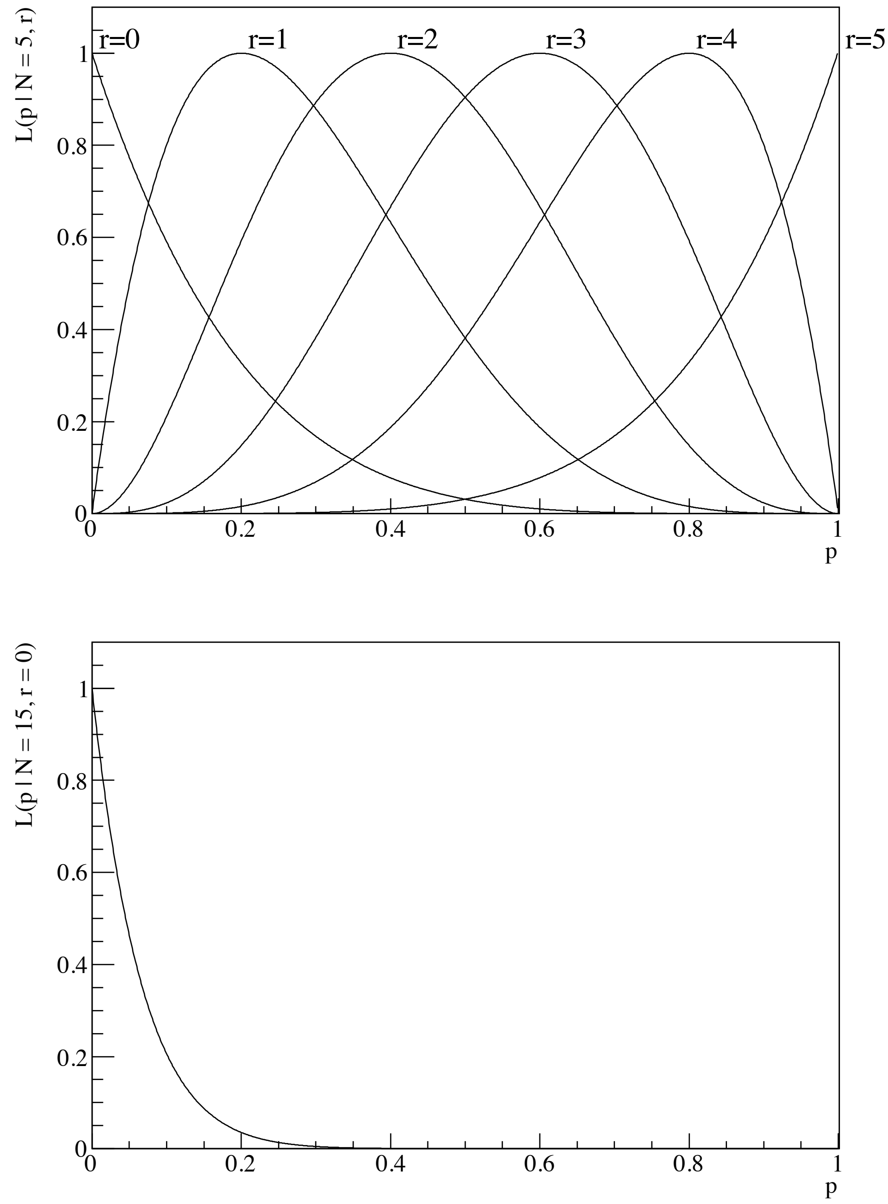

Figure 4.2: The likelihood of p obtained for (top) a coin tossed 5 times, obtaining r heads, and (bottom) a coin tossed 15 times, obtaining no heads.

Following on from Eq. (4.1.3), the mean value of r is given by

$$\begin{aligned}
\langle r \rangle &= \sum_{r=0}^{r=n} rP(r;p,n), && (4.2.10)\\
&= \sum_{r=0}^{r=n} rp^r(1-p)^{n-r}\frac{n!}{r!(n-r)!}, && (4.2.11)\\
&= np\sum_{r=0}^{r=n} p^{r-1}(1-p)^{n-r}\frac{(n-1)!}{(r-1)!(n-r)!}, && (4.2.12)\\
&= np\sum_{r=0}^{r=n} P(r;p-1,n-1), && (4.2.13)\\
&= np, && (4.2.14)
\end{aligned}$$

as the sum over all possible outcomes of the binomial distribution is unity. In order to compute the variance on r, one needs to use the above result. The starting point to determine $V(r)$ is

$$V(r) = \sum_{r=0}^{r=n}(r-np)^2 P(r;p,n), \quad (4.2.15)$$
$$= \langle r^2\rangle - \langle r\rangle^2, \quad (4.2.16)$$
$$= \langle r^2\rangle - n^2p^2. \quad (4.2.17)$$

In order to compute $\langle r^2\rangle$, in analogy with the derivation above for $\langle r\rangle$, one needs absorb the factor of r^2 into the $r!$ term in the binomial series. It is not possible to do this directly, however one can absorb a factor of $r(r-1)$ into that sum, and hence compute $V(r)$ via

$$V(r) = \langle r(r-1)\rangle + \langle r\rangle - \langle r\rangle^2. \quad (4.2.18)$$

As anticipated, the right hand side of the previous equation reduces to the desired result of $np(1-p)$.

4.3 Poisson distribution

The ***Poisson distribution*** is given by

$$P(r,\lambda) = \frac{\lambda^r e^{-\lambda}}{r!}, \quad (4.3.1)$$

which is a function of the number of observed events r, and λ. The parameter λ is the mean and variance of the distribution (as shown in section 4.3.1). Figure 4.3 shows the Poisson probability distribution for several different values of λ. For small λ the distribution is asymmetric and skewed to the right. As λ increases the Poisson distribution becomes more symmetric.

This is an important distribution often used in science when studying the occurrence of rare events in a continually running experiment. For example, the Poisson distribution can be used to describe radioactive decay (see section **??**) or particle interactions where we have no idea how many decays occur in total, but can study events over a finite time to understand the underlying behaviour. Tables of Poisson probabilities can be found in appendix D.2.

If one considers the situation where r events have been observed, one can ask the question, what is the most likely value of λ corresponding to this observation. Figure 4.4 shows the corresponding likelihood distribution for $r = 0, 1, 2, 3, 4$, and 5 events. One can see that for $r > 0$ the Poisson distribution has a definite non-zero maximum to the likelihood distribution.

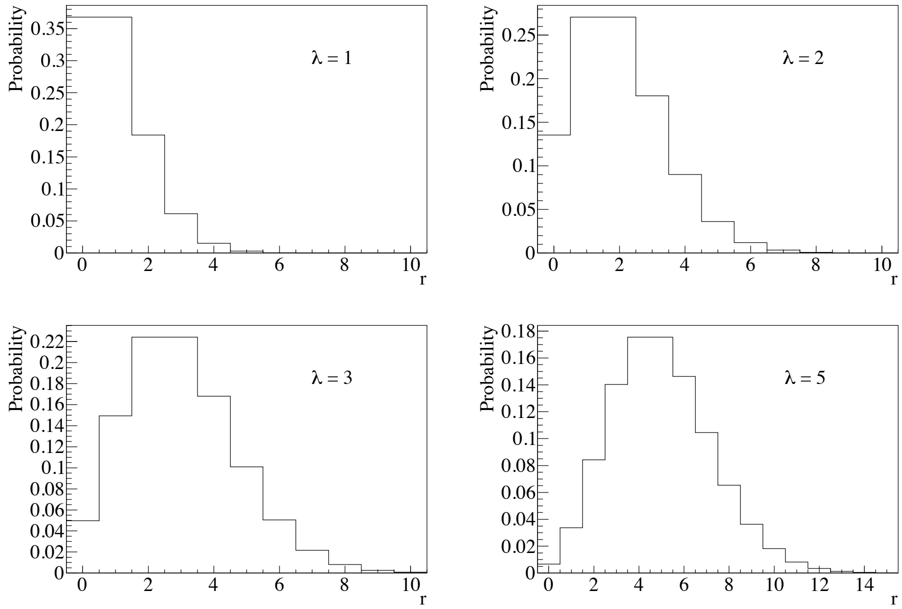

Figure 4.3: Poisson probability distributions for different values of λ.

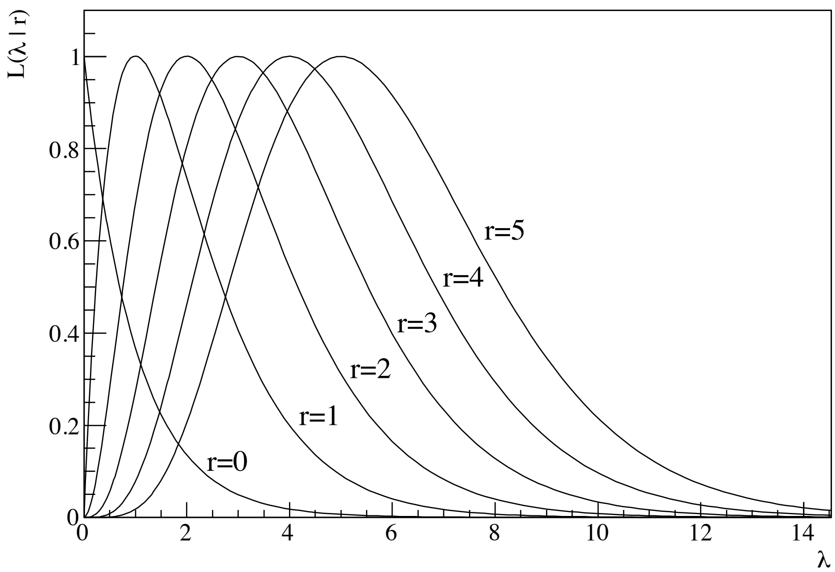

Figure 4.4: Likelihood distributions corresponding to Poisson probability distributions for different values of r as a function of λ.

4.3.1 Derivation of the mean and variance of the Poisson probability distribution

If one considers the Poisson probability distribution as given by Eq. (B.1.13), then the sum of this distribution over all values of r is unity. This can be shown as follows

$$P(r,\lambda) = \frac{\lambda^r e^{-\lambda}}{r!}, \tag{4.3.2}$$

$$\sum_{r=0}^{r=\infty} P(r,\lambda) = e^{-\lambda}\left[1+\lambda+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\frac{\lambda^4}{4!}+\ldots\right], \tag{4.3.3}$$

$$= e^{-\lambda}e^{\lambda}, \tag{4.3.4}$$

as the term in the square brackets is the Maclaurin series expansion for e^λ. As required the Poisson probability distribution is normalised to unity.

Following on from section 4.1, the mean value of r is given by

$$\begin{aligned}
\langle r \rangle &= \sum_{r=0}^{r=\infty} rP(r,\lambda), && (4.3.6)\\
&= \sum_{r=0}^{r=\infty} r\frac{\lambda^r e^{-\lambda}}{r!}, && (4.3.7)\\
&= 0 + \lambda e^{-\lambda} + \frac{2\lambda^2 e^{-\lambda}}{2!} + \frac{3\lambda^3 e^{-\lambda}}{3!}, && (4.3.8)\\
&= \lambda e^{-\lambda}\left[1 + \lambda + \frac{\lambda^2}{2!} + \ldots\right], && (4.3.9)\\
&= \lambda. && (4.3.10)
\end{aligned}$$

Similarly the variance of r is given by

$$\begin{aligned}
V(r) &= \sum_{r=0}^{r=\infty} (r-\lambda)^2 P(r,\lambda), && (4.3.11)\\
&= \sum_{r=0}^{r=\infty} (r-\lambda)^2 \frac{\lambda^r e^{-\lambda}}{r!}, && (4.3.12)\\
&= \lambda e^{-\lambda}\left[\lambda + (1-\lambda)^2 + \frac{(2-\lambda)^2\lambda}{2!} + \frac{(3-\lambda)^2\lambda^2}{3!} + \ldots\right]. && (4.3.13)
\end{aligned}$$

After gathering together the terms in the square bracket, these simplify to the Maclaurin series expansion for e^λ and we find

$$\begin{aligned}
V(r) &= \lambda e^{-\lambda} e^{\lambda}, && (4.3.14)\\
&= \lambda. && (4.3.15)
\end{aligned}$$

Thus we obtain the expected results that the PDF is normalised to a total probability of one, and that the distribution has a common mean and variance given by λ.

4.4 Gaussian distribution

The ***Gaussian distribution***, also known as the ***normal distribution***, with a mean value μ and standard deviation σ as a function of some variable x is given by

$$P(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}. \quad (4.4.1)$$

This distribution will be discussed in more detail in chapter 5 when considering the nature of statistical errors. Figure 4.5 shows a Gaussian distribution with $\mu = 0$ and $\sigma = 1$. Often it is useful to transform data from the x space to a corresponding z space which has a mean value of zero, and a standard deviation of one. This transformation is a one to one mapping given by

$$z = \frac{x-\mu}{\sigma}, \quad (4.4.2)$$

for any given Gaussian distribution $G(x, \mu, \sigma)$. Hence the Gaussian distribution in terms of z is

$$P(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}. \tag{4.4.3}$$

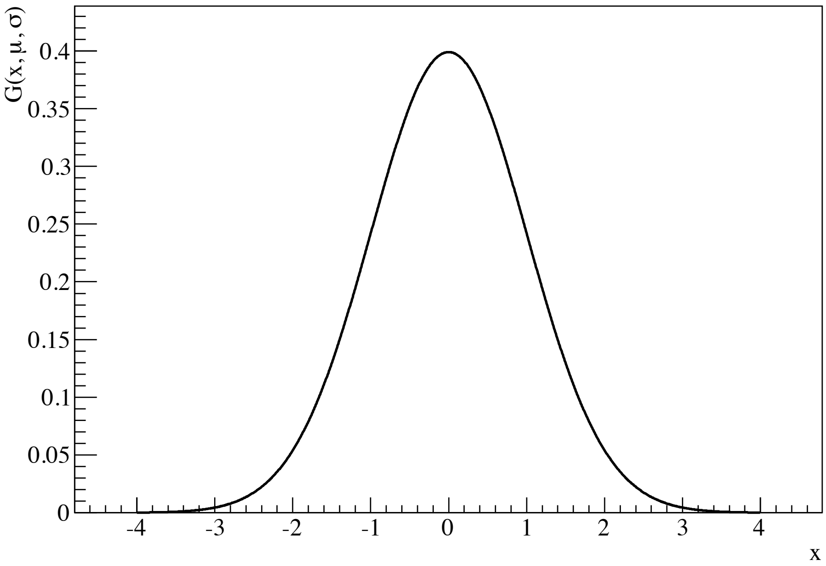

Figure 4.5: Gaussian distribution with a mean value $\mu = 0$ and width $\sigma = 1$. This is equivalent to $P(z)$ given in Eq. (4.4.3).

There are a number of useful results associated with integrals of the Gaussian distribution. These are discussed in the following, and can be used in order to determine the confidence intervals discussed in chapter 6. Firstly the normalisation condition for a Gaussian distribution is given by

$$I = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} e^{-(x-\mu)^2/2\sigma^2} dx, \tag{4.4.4}$$

$$= 1. \tag{4.4.5}$$

While it is possible to analytically integrate the Gaussian distribution, it is often convenient to perform a numerical integration instead (see appendix C for a short introduction to this subject). Tables D.8 and D.9 summarise the integral of a Gaussian function from $-\infty$ to a given number of standard deviations above the mean, and within a given number of standard deviations above and below the mean, respectively.

One can extend the Gaussian distribution to encompass more than one dimension, resulting in the so-called multi-normal distribution, or multivariate Gaussian distribution. This distribution is discussed in section 6.8.1 in the context of confidence intervals corresponding to an *error ellipse* resulting from two correlated parameters.

4.5 χ^2 distribution

The $\boldsymbol{\chi^2}$ ***distribution*** is given by

$$P(\chi^2, \nu) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)}(\chi^2)^{(\nu/2-1)}e^{-\chi^2/2}, \tag{4.5.1}$$

as a function of χ^2, and the number of ***degrees of freedom*** ν. The χ^2 sum is over squared differences between data x_i and a hypothesis $\widehat{x}$, normalised by the uncertainty on the data $\sigma(x_i)$. Hence

$$\chi^2 = \sum_{i=data}\left[\frac{x_i - \widehat{x}}{\sigma(x_i)}\right]^2. \tag{4.5.2}$$

The quantity ν is given by the number of entries in the data sample less any constraints imposed or adjusted to compute the χ^2. The quantity $\Gamma(\nu/2)$ is the ***Gamma distribution*** which has the form

$$\Gamma(\nu/2) = (\nu/2 - 1)!, \tag{4.5.3}$$

and is valid for positive integer values of ν. The Gamma function for $\nu = 1$ is $\Gamma(1/2) = \sqrt{\pi}$.

For example, if we have a sample of n data points, then there are $n - 1$ degrees of freedom (as the total number of data points is itself a constraint), so $\nu = n-1$. Given some model, we can then compute the value of χ^2 using Eq. (4.5.2), and thus $P(\chi^2, \nu)$ using Eq. (4.5.1). Figure 4.6 shows examples of the χ^2 distribution for different numbers of degrees of freedom. The χ^2 distribution can be used to validate the agreement between some model and a set of data. As a general rule of thumb a reasonable agreement between a set of data $\Omega(x)$ and some model of the data $f(x)$ is obtained for $\chi^2/\nu \sim 1$. The χ^2 probability distribution enables one to go beyond applying a simple rule of thumb and allows the experimenter to compute a probability associated with the agreement between data and model. More generally this type of comparison is often referred to as ***goodness of fit*** (GOF) test.

4.6 Computational issues

When a distribution involves the computation of very large or very small numbers, then it is usually necessary to take care in the calculation. A classic example of this is the computation of n factorial (n!). If we use a computer to calculate n! for us with a 4-bit integer representation, then we will find that we can calculate 13!=1,932,053,504, but we obtain a solution for 14! that is smaller than 13!, which is clearly not correct. The reason that we compute the wrong answer is that we run out of memory allocated for the representation of the integer result. We can use a longer integer representation for this calculation, however we will run into the same problem once again for some larger value of n – so we may only temporarily postpone the problem. When a situation like this is encountered, a computer will not give us a warning, and will merrily continue to do what it is told, so it is up to the person doing the calculation to make sure that the result is sensible.

A second illustration of computational problems that can be encountered is given by the binomial distribution

$$P(r; p, n) = p^r(1-p)^{n-r}\frac{n!}{r!(n-r)!}. \tag{4.6.1}$$

In addition to having to compute factorials that we know from above can be problematic, we also have to compute powers of numbers less than one. As either r or $n - r$ become large, then the powers of p and q we need to compute may quickly become small, and here we run into another problem. If we represent a

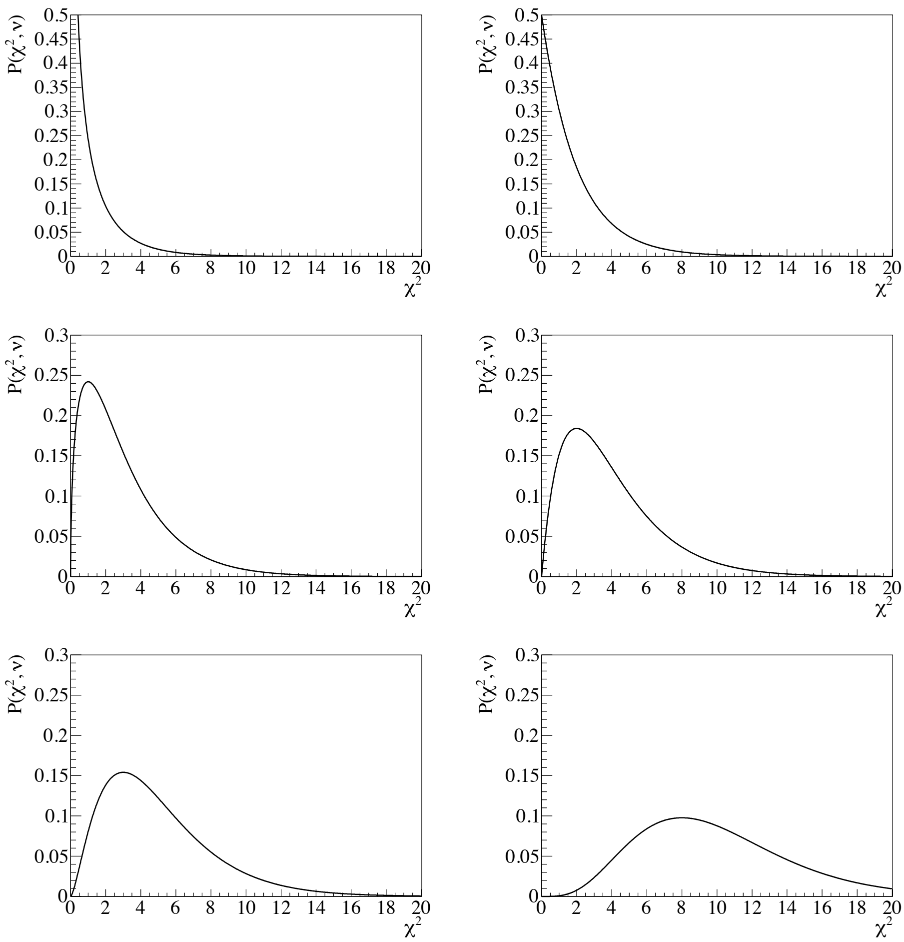

Figure 4.6: Examples of the χ^2 probability distribution for $\nu = 1, 2, 3, 4, 5$, and 10 (left to right, top to bottom). Note that the vertical scales are not the same for all plots.

real number with some floating point form in a computer, then that too has a finite resolution. We may be able to compute p^r and q^{n-r}, but find that the product of the two small but finite numbers is zero. In such instances, it can be beneficial to multiply the small numbers by a large factor to compute the probability, then divide out those large factors at the end of the calculation so that the total probability for anything to happen remains normalised to unity. Again it is up to the person doing the calculation to ensure that the result obtained is correct, and any scale factors introduced into a computation are sufficient to compute the result to the desired level of accuracy.

For such cases it would be desirable if we could use a more robust algorithm, rather than postpone problems resulting from the precision of number representation in a computer by buying a new computer or compiler with a larger bit representation for floating point or integer numbers. Indeed often it is possible to find a more robust algorithm to compute a quantity with a little thought: take for example the case of the Poisson probability distribution

$$P(r, \lambda) = \frac{\lambda^r e^{-\lambda}}{r!}. \tag{4.6.2}$$

If we compute $P(0, \lambda)$, then we can obtain $P(1, \lambda)$ by multiplying our result for $P(0, \lambda)$ by λ and dividing by the incremented value of r (in this case by one). So in general we can use the following rule to compute successive terms of a Poisson probability distribution

$$P(r+1,\lambda) = \frac{\lambda P(r,\lambda)}{r+1}. \tag{4.6.3}$$

Using this algorithm we never have to compute the factorial of a large number, and at each step in the calculation we find that we obtain sensible results, even for large r. Common numerical issues such as the ones presented here are discussed in literature on computer algorithms, e.g. see volume one of Knuth (1998).