

Analysis of Experimental Data

Michael Mulhearn

March 6, 2019

Chapter 1

Statistical Distributions

1.1 Statistics of Experiments

A primary purpose of science is to predict the results of experiments. Consider a simple experiment with five possible outcomes which we repeat ten times. If our theoretical prediction is that each of these five outcomes is equally probable, then our prediction for a typical series of ten experiments would be for each outcome to occur two times. Now suppose we perform the experiment ten times and present the results like this:

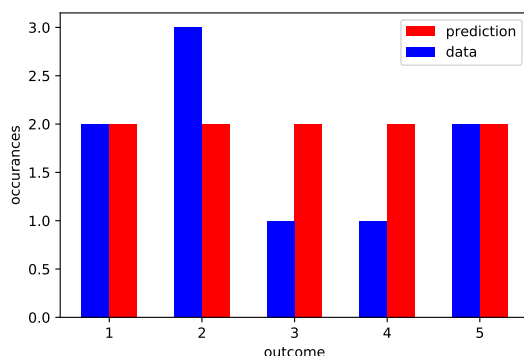


Figure 1.1: Comparison of experimental results with a prediction.

Scientists almost never display experimental data this way (as a bar graph) because it is nearly impossible to answer the crucial question *is this data consistent with this prediction?* Even if every outcome has an equal probability, the results of individual experiments experience statistical fluctuations. So even if the theory is correct, we will seldom reproduce exactly the theory prediction.

To interpret scientific experiments, it isn't enough to have a single prediction for the outcome of an experiment, instead, you need a prediction for the statistical distribution of outcomes: a probability distribution function. We'll start this discussion, therefore, by deriving three of the most frequently encountered probability distributions: the Binomial Distribution, the Poisson Distribution, and the Gaussian Distribution.

1.2 The Binomial Distribution

The Binomial Distribution is the most general of the distributions we'll consider, but it is a bit cumbersome to use in practice. The more familiar Poisson and Gaussian distributions are limiting cases of this distribution.

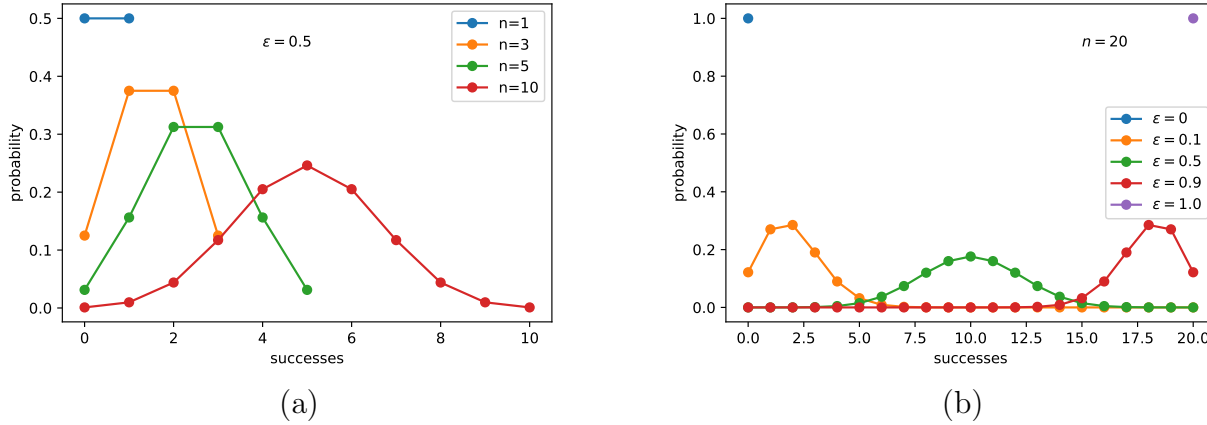


Figure 1.2: The binomial distribution for several different values of the parameters (a) n and (b) ϵ .

Suppose you repeat a particular process n times, and each time you have the same probability ϵ of a particular outcome, which, without losing generality, we'll call “success”. The probability of having exactly m successes after n trials is simply given by:

$$P = \sum_i p_i$$

where i runs over all specific outcomes with m successes and p_i is the probability of each specific outcome. However, as these specific outcomes all contain exactly m successes, they share the same probability, namely:

$$p_i = \epsilon^m (1 - \epsilon)^{n-m}$$

and so we are left to consider simply the total number of specific outcomes containing m successes.

The quantity we need is provided by the binomial theorem from mathematics, which states that:

$$(p + q)^n = \sum_{m=0}^n \binom{n}{m} p^m q^{n-m} \quad (1.1)$$

where the binomial coefficients are defined by

$$\binom{n}{m} = \frac{n!}{m! (n-m)!} \quad (1.2)$$

and are also often referred to in other contexts as n -choose- m . The binomial coefficient simply tells us how many times we can choose m instances of p instead of q , from n factors, and so it is precisely the combinatoric factor that we need.

The probability of obtaining m successes after n trials with probability ϵ is therefore given by:

$$P(m; n, \epsilon) = \binom{n}{m} \epsilon^m (1 - \epsilon)^{n-m} \quad (1.3)$$

which is called the Binomial Distribution.

1.3 Mean and Variance

Given a probability distribution, the most urgent questions are generally “what is the mean value we can expect from this distribution?” and “how close to the mean value are most of the outcomes?” The first answer localizes the distribution while the second answer describes its width.

To calculate the mean value in answer to the first question, we simply calculate a weighted average:

$$\langle m \rangle \equiv \bar{m} \equiv \sum_m m P(m) \quad (1.4)$$

For a continuous probability distribution, we would integrate instead:

$$\langle x \rangle \equiv \bar{x} \equiv \int x P(x) dx \quad (1.5)$$

We usually answer the second question in terms of the variance, σ^2 , of the distribution:

$$\sigma^2 \equiv \langle (x - \bar{x})^2 \rangle$$

Other answers have problems, e.g. $\langle x - \bar{x} \rangle$ can be zero or nearly so, even for wide distributions, as long as it is symmetric. You could fix this by calculating $\langle |x - \bar{x}| \rangle$ but this is generally much harder to calculate, and less useful, than the variance. For instance, it is left as an exercise to show that:

$$\langle (x - \bar{x})^2 \rangle = \langle x^2 \rangle - \bar{x}^2 \quad (1.6)$$

using the fact that \bar{x} is simply a number, and so $\langle \bar{x} \rangle = \bar{x}$. We often write this result equivalently as:

$$\sigma^2 \equiv \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 \quad (1.7)$$

Which shows explicitly that we need only calculate $\langle x \rangle$ and $\langle x^2 \rangle$ in order to determine the variance of a distribution.

1.4 Mean and Variance of the Binomial Distribution

The mean value of Binomial Distribution is given by:

$$\begin{aligned} \bar{m} &= \sum_{m=0}^n m P(m) \\ &= \sum_{m=0}^n m \binom{n}{m} \epsilon^m (1 - \epsilon)^{n-m} \end{aligned}$$

which looks rather daunting! The trick is to use the Binomial Theorem (1.1) and define a function of two independent variables p and q given by:

$$f(p, q) = (p + q)^n = \sum_{m=0}^n \binom{n}{m} p^m q^{n-m}$$

We then calculate:

$$\frac{\partial f}{\partial p} = n(p + q)^{n-1} = \sum_{m=0}^n m \binom{n}{m} p^{m-1} q^{n-m}$$

and multiplying by p we have:

$$np(p+q)^{n+1} = \sum_{m=0}^n m \binom{n}{m} p^m q^{n-m}$$

which is true for any p and q . We now substitute the particular values $p = \epsilon$ and $q = 1 - \epsilon$ and find that:

$$n\epsilon = \sum_{m=0}^n m \binom{n}{m} \epsilon^m (1 - \epsilon)^{n-m} \equiv \sum_{m=0}^n m P(m) = \bar{m}$$

So the mean value is given by:

$$\bar{m} = n\epsilon \tag{1.8}$$

or the total number of trials times the probability of success for each trial, a wholly plausible answer.

For the variance, we use a variation of the same trick, this time using the second partial derivative:

$$p^2 \cdot \frac{\partial^2 f}{\partial p^2} = n(n-1)p^2(p+q)^{n-2} = \sum_{m=0}^n m(m-1) \binom{n}{m} p^m q^{n-m}$$

and again putting $p = \epsilon$ and $q = 1 - \epsilon$ to find that:

$$\begin{aligned} n(n-1)\epsilon^2 &= \sum_{m=0}^n (m^2 - m) \binom{n}{m} p^m q^{n-m} \\ &= \sum_{m=0}^n (m^2 - m) P(m) \\ &= \langle m^2 - m \rangle = \langle m^2 \rangle - \langle m \rangle \end{aligned}$$

and as $\langle m \rangle = n\epsilon$ we have:

$$\langle m^2 \rangle = n(n-1)\epsilon^2 + n\epsilon$$

And so:

$$\sigma^2 = \langle m^2 \rangle - \langle m \rangle^2 = n(n-1)\epsilon^2 + n\epsilon - n^2\epsilon^2$$

or simply:

$$\sigma^2 = n\epsilon(1 - \epsilon) \tag{1.9}$$

Note that if $\epsilon = 0$ or $\epsilon = 1$, there is only one outcome (all failures or all success) and so the variation is zero.

1.5 The Poisson Distribution

Suppose we have some time interval over which we expect to observe a mean number of events λ . The events must be independent of one another: an event occurring at a particular time cannot affect the time at which the next event occurs. We divide the time interval over which the λ events are expected to occur into n sub-intervals, each with an equal probability to contain an event. These intervals will be all the same size if the events are uniformly distributed in time, but if the events are not uniformly distributed, the sub-intervals are simply chosen to ensure the probability

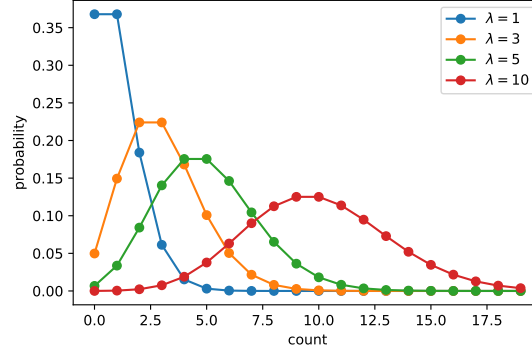


Figure 1.3: The Poisson distribution for several values of parameter λ .

is the same in each interval. Once cast this way, we can interpret this as a binomial distribution, with probability to contain an event, by construction, given by $\epsilon = \lambda/n$:

$$\begin{aligned}
 P(m) &= \binom{n}{m} \epsilon^m (1 - \epsilon)^{n-m} \\
 &= \frac{n!}{m! (n-m)!} \left(\frac{\lambda}{n}\right)^m \left(1 - \frac{\lambda}{n}\right)^{n-m} \\
 &= \left(\frac{\lambda^m}{m!}\right) \left(1 - \frac{\lambda}{n}\right)^n \left[\frac{n!}{(n-m)!} \cdot \frac{1}{n^m}\right]_1 \left[\left(1 - \frac{\lambda}{n}\right)^{-m}\right]_2
 \end{aligned}$$

We obtain the Poisson distribution by considering the limit that $n \rightarrow \infty$. It is left as an exercise to show that both $[\dots]_1 \rightarrow 1$ and $[\dots]_2 \rightarrow 1$ as $n \rightarrow \infty$. Recalling that

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

we obtain the Poisson distribution, the probability for observing m events for a mean of λ :

$$P(m; \lambda) = \frac{\lambda^m}{m!} e^{-\lambda} \tag{1.10}$$

Notice that there is no longer a parameter n , since we took $n \rightarrow \infty$, and so m now ranges from 0 to ∞ . Note also that λ is a real number, not necessarily an integer, even though it is the mean of integer values. For example, the mean of the integers 1 and 2 is the real value 1.5.

1.6 Mean and Variance of The Poisson Distribution

The mean of the Poisson distribution is given by:

$$\begin{aligned}
 \bar{m} &= \sum_{m \geq 0} m P(m) \\
 &= \sum_{m \geq 0} m \frac{\lambda^m}{m!} e^{-\lambda}
 \end{aligned}$$

Since the first term ($m = 0$) is zero, we have:

$$\begin{aligned}
 \bar{m} &= e^{-\lambda} \sum_{m \geq 1} \frac{\lambda^m}{(m-1)!} \\
 &= \lambda e^{-\lambda} \sum_{m \geq 1} \frac{\lambda^{m-1}}{(m-1)!} \\
 &= \lambda e^{-\lambda} \sum_{n \geq 0} \frac{\lambda^n}{n!} \\
 &= \lambda e^{-\lambda} e^{\lambda} \\
 \bar{m} &= \lambda
 \end{aligned} \tag{1.11}$$

which should come as no surprise, as the assumption in the derivation was the that mean number of events was λ .

For the variance, we use a similar manipulation to calculate:

$$\begin{aligned}
 \langle m^2 \rangle &= \sum_{m \geq 0} m^2 P(m) \\
 &= \sum_{m \geq 0} m^2 \frac{\lambda^m}{m!} e^{-\lambda} \\
 &= \lambda \sum_{m \geq 1} m \frac{\lambda^{m-1}}{(m-1)!} e^{-\lambda} \\
 &= \lambda \sum_{n \geq 0} (n+1) \frac{\lambda^n}{(n)!} e^{-\lambda} \\
 &= \lambda \langle m+1 \rangle = \lambda (\lambda + 1)
 \end{aligned}$$

And so:

$$\begin{aligned}
 \sigma^2 &= \langle m^2 \rangle - \langle m \rangle^2 \\
 &= \lambda (\lambda + 1) - \lambda^2 \\
 \sigma^2 &= \lambda.
 \end{aligned} \tag{1.12}$$

That is, the variance of a Poisson distribution is simply the mean.

1.7 The Gaussian Distribution

The last and most important probability distribution we will consider is the Gaussian distribution:

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x-\lambda)^2}{2\sigma^2}\right) \tag{1.13}$$

which is plotted in Fig. 1.4. The Gaussian distribution (also sometimes called the normal distribution) appears frequently as a consequence of the central limit theorem, which states that the sum of independent random variables, when appropriately average (such as by taking the mean value) approaches a Gaussian distribution. As one example of the CLT, we'll show in the next section that the Gaussian distribution emerges from the Poisson distribution in the limit that $\lambda \rightarrow \infty$.

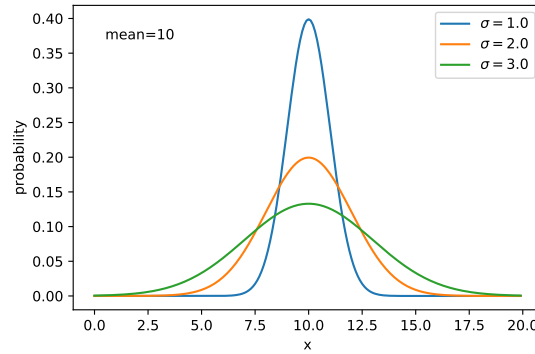


Figure 1.4: The Gaussian distribution for a mean of 10 and several values of parameter σ .

The Binomial and Poisson distributions are discrete functions: they describe the probability of outcomes which are integer quantities. The Poisson probability for having 2.31 events is either undefined or taken to be zero. The value of the Binomial and Poisson distribution function at a particular integer value is simply the probability of that particular outcome. To determine the probability that an outcome is within a range of integers, say m_1 to m_2 the probability distribution function is simply added:

$$P = \sum_{m=m_1}^{m_2} P(m).$$

The Gaussian distribution is a continuous function that describes the probability *density* at each position x . The probability of any particular outcome, say $x = 1.24323$, is vanishing small. But the probability that the value lies within a range of values is non-zero, and is determined by integrating:

$$P = \int_{x_{\min}}^{x_{\max}} P(x; \lambda, \sigma) dx$$

The Gaussian distribution is called a probability density function (PDF). A discrete probability distribution is sometimes referred to as probability mass function (PMF), to draw upon an analogy with physical mass and density: a PDF is probability per unit volume while a PMF is simply a probability.

The Gaussian distribution is normalized so that:

$$\int_{-\infty}^{+\infty} P(x) dx = 1$$

To determine the mean value of a function $f(x)$ we likewise integrate:

$$\langle f(x) \rangle \equiv \int_{-\infty}^{+\infty} f(x) P(x) dx = 1$$

In particular the mean value of x is:

$$\langle x \rangle \equiv \int_{-\infty}^{+\infty} x P(x) dx = 1$$

and:

$$\langle x^2 \rangle \equiv \int_{-\infty}^{+\infty} x^2 P(x) dx = 1$$

We have defined the Gaussian distribution with the suggestive parameters λ and σ . It is left as an exercise to show that indeed, the mean is λ and the variance is σ^2 .

1.8 The Gaussian Distribution as a Limiting Case

As expected from the Central Limit Theorem, the Poisson Distribution in the limit $\lambda \rightarrow \infty$ approaches the Gaussian distribution. In this case, we can apply the Stirling Approximation:

$$\lim_{n \rightarrow \infty} n! = \sqrt{2\pi n} e^{-n} n^n$$

to the Poisson distribution as follows:

$$\begin{aligned} P(m) &= \frac{\lambda^m}{m!} e^{-\lambda} \\ &\rightarrow \frac{\lambda^m e^{-\lambda}}{\sqrt{2\pi m} e^{-m} m^m} \\ &= \frac{e^{m-\lambda}}{\sqrt{2\pi \lambda}} \left(\frac{\lambda}{m} \right)^{m+\frac{1}{2}} \end{aligned}$$

Now we consider a new variable δ , defined by

$$\delta \equiv \frac{m - \lambda}{\lambda}$$

which measures the difference between the observed number of events m and the mean of the distribution, as a fraction of the mean. Intuitively, the function is getting very narrow, and so we expect this to be a small quantity, but let's check this. First we have:

$$\langle \delta \rangle = \frac{\langle m \rangle - \lambda}{\lambda} = \frac{\lambda - \lambda}{\lambda} = 0$$

but also:

$$\langle \delta^2 \rangle = \frac{\langle (m - \lambda)^2 \rangle}{\lambda^2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

where we have used the fact that the variance is given by $\langle (m - \lambda)^2 \rangle = \lambda$, and so as $\lambda \rightarrow \infty$ we have

$$\langle \delta^2 \rangle \rightarrow 0$$

So we can write:

$$m = \lambda(1 + \delta) \tag{1.14}$$

where we expect the approximation $\delta \rightarrow 0$ to hold as long as we require $\lambda \rightarrow \infty$. So now we can write the distribution in terms of the small quantity δ and the large quantity λ as:

$$\begin{aligned} P(\delta) &= \frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}} \left(\frac{\lambda}{\lambda(1+\delta)} \right)^{\lambda(1+\delta)+\frac{1}{2}} \\ &= \frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}} \cdot \frac{1}{X} \end{aligned} \tag{1.15}$$

where we define the quantity:

$$X = (1 + \delta)^{\lambda(1+\delta)+\frac{1}{2}}$$

which can be approximated as follows:

$$\begin{aligned}
 \ln X &= \left(\lambda(1 + \delta) + \frac{1}{2} \right) \cdot \ln(1 + \delta) \\
 &= \left(\lambda(1 + \delta) + \frac{1}{2} \right) \cdot \left(\delta - \frac{\delta^2}{2} + \mathcal{O}(\delta^3) \right) \\
 \frac{\ln X}{\lambda} &= \left(1 + \delta + \frac{1}{2\lambda} \right) \cdot \left(\delta - \frac{\delta^2}{2} + \mathcal{O}(\delta^3) \right) \\
 &= (1 + \delta + \mathcal{O}(\delta^2)) \cdot \left(\delta - \frac{\delta^2}{2} + \mathcal{O}(\delta^3) \right) \\
 &= \delta + \frac{\delta^2}{2} + \mathcal{O}(\delta^3),
 \end{aligned}$$

where in the second to last step we used $\mathcal{O}(\frac{1}{\lambda}) \sim \mathcal{O}(\delta^2)$. Neglecting the small quantities, we can approximate

$$X = \exp \left(\lambda\delta + \lambda\frac{\delta^2}{2} \right)$$

which, when plugged back into Equation 1.15 yields:

$$\begin{aligned}
 P(\delta) &= \frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}} \cdot \frac{1}{\exp(\lambda\delta + \lambda\frac{\delta^2}{2})} \\
 &= \frac{1}{\sqrt{2\pi\lambda}} \cdot \exp \left(-\lambda\frac{\delta^2}{2} \right)
 \end{aligned} \tag{1.16}$$

Recalling that:

$$\begin{aligned}
 \delta &\equiv \frac{m - \lambda}{\lambda} \\
 P(x) &= \frac{1}{\sqrt{2\pi\lambda}} \cdot \exp \left(-\frac{(m - \lambda)^2}{2\lambda} \right)
 \end{aligned} \tag{1.17}$$

Which we can identify as a special case of the Gaussian distribution of Equation 1.13 with $\sigma^2 = \lambda$ and $x = m$:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma}} \cdot \exp \left(-\frac{(x - \lambda)^2}{2\sigma^2} \right)$$

It may trouble you that x is continuous while m here was an integer. But note that in the limit $\lambda \rightarrow \infty$, the difference between m/λ and $(m + 1)/\lambda$ is infinitesimal, and so this is effectively a continuous function, consistent with the Gaussian distribution.

1.9 Histograms in Scientific Python

Suppose a particular variable x is measured 1000 times. One way to visualize the collected data is shown in Fig. 1.5a, which simply plots each measurement value above the measurement number (from 0 to 1000). In this example, the number of measurements that occur within the range from $x = 100$ to $x = 105$ is 181. This count is plotted as the red data point in Fig. 1.5b, 181 entries located above $x = 102.5$, the center of the range. If we repeat this exercise across a number or

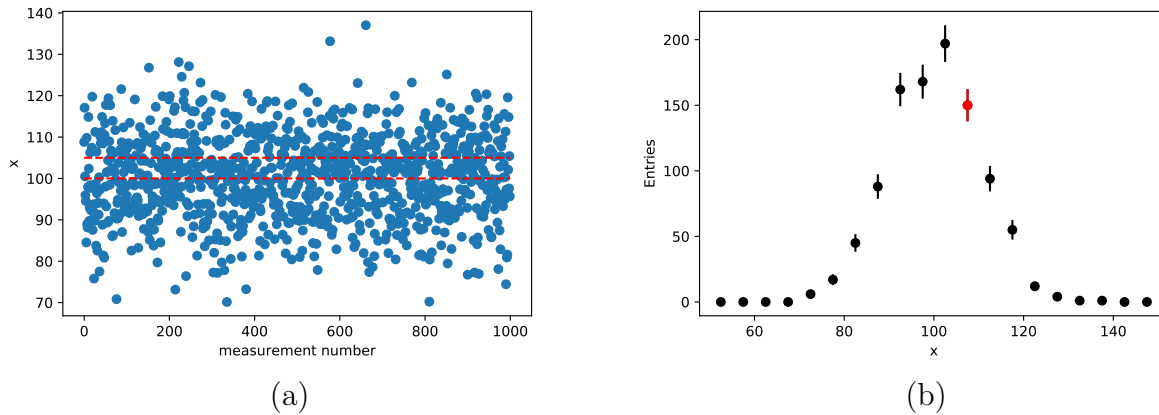


Figure 1.5: The 1000 measurements of variable x in (a) are used to produce the histogram in (b). The red data point in (b) is the count of the number of entries in range indicated by the red dashed lines in (a).

ranges, the resulting plot in Fig. 1.5b is called a histogram. A histogram reports the number of entries that occur within each of a sequence of consecutive ranges of x -values. Each range considered is called a histogram bin, and the choice of which bins to use is at the discretion of the analyzer.

The content of each bin is a single number, a count, and is therefore subject to the Poisson distribution. We can estimate the mean of the Poisson distribution by the measured value, so that $\lambda = N$. For the Poisson distribution, the variance $\sigma^2 = \lambda$, and so $\sigma = \sqrt{N}$. It is customary to draw a line of size \sqrt{N} when plotting a histogram value N . This is an example of an error bar, which indicates how well our measurement has determined a particular value.

An example producing a histogram in Scientific Python is shown in Fig. 1.6. The data to plot is simply a sequence of 1000 values randomly and uniformly chosen in the range $[0, 20]$:

```
x = np.random.uniform(high=20.0,size=1000)
```

To create a histogram from these 1000 values, we use the `np.histogram` function:

```
counts,edges = np.histogram(x,bins=10,range=(0,20))
```

where we have specified 10 bins, uniformly covering the range from 0 to 20. The function returns to arrays, which we save as `counts` and `edges`. The `counts` array contains the bin contents, the count of the number of values in each bin:

```
counts: [ 92  82 123  96  85 106 105  99  99 113]
```

The `edges` array contains the edges of the bins:

```
edges: [ 0.  2.  4.  6.  8. 10. 12. 14. 16. 18. 20.]
```

You'll notice that 10 consecutive bins have 11 edges. For plotting continuous data, one choice is to plot the contents at the center of each bin:

```
cbins = (edges[:-1] + edges[1:])/2.0
```

the two slices `edges[:-1]` and `edges[1:]` are all but the last and all but the first. The average of the two is the center of each bin:

```

x = np.random.uniform(high=20.0,size=1000)
counts,edges = np.histogram(x,bins=10,range=(0,20))
print("counts: ", counts)
print("edges:   ", edges)
cbins = (edges[:-1] + edges[1:])/2.0
err    = counts**0.5
print("cbins:   ", cbins)
plt.errorbar(cbins,counts,yerr=err,fmt="ko")
plt.xlabel("x")
plt.ylabel("Entries")

```

```

counts:  [ 81  96 104 108 104  99 102 110  98  98]
edges:   [ 0.  2.  4.  6.  8. 10. 12. 14. 16. 18. 20.]
cbins:   [ 1.  3.  5.  7.  9. 11. 13. 15. 17. 19.]

```

```
Text(0,0.5,'Entries')
```

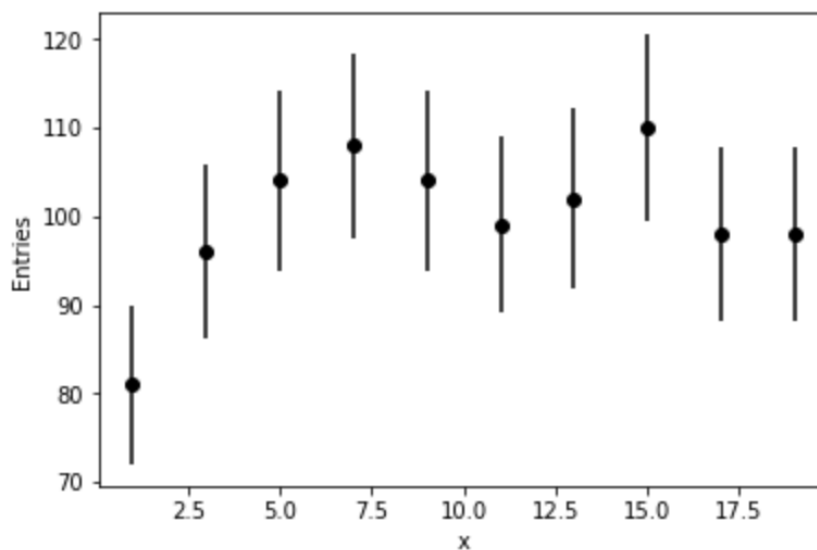


Figure 1.6: Example producing a histogram in Scientific Python.

```
cbins:    [ 1.  3.  5.  7.  9. 11. 13. 15. 17. 19.]
```

The error bar values are chosen as the square root of the bin values:

```
err  = counts**0.5
```

The histogram is plotted using the `plt.errorbar` function:

```
plt.errorbar(cbins,counts,yerr=err,fmt="ko")
```

which plots the bin contents `counts` at the bin center values `cbins` using the square root error bars in the array `err`, using the format "ko" for black circles.

1.10 Comparing a Histogram to a Probability Distribution Function

Our theoretical models often predict a PDF for some observable variable x . As experimentalists, we are often therefore concerned with the question as to whether our collected data for an observable x is consistent with the theoretical PDF. A visual approach to answering this question is to plot the data in a histogram, and to draw the PDF as a curve normalized to the histogram.

To predict the number of events in a bin with edges x_{lower} and x_{upper} , in principle we need to integrate the PDF and normalize to the number of experiments:

$$N_{\text{pred}} = N_{\text{meas}} \int_{x_{\text{lower}}}^{x_{\text{upper}}} p(x) dx$$

In practice, we generally choose the bin sizes small enough that the PDF is approximately constant during the entire bin, and in this case, the prediction can be taken as:

$$N_{\text{pred}} = N_{\text{meas}} \Delta x p(x)$$

where Δx is the width of each bin. This scale factor $N_{\text{meas}} \Delta x$ allows us to compare a continuous function to data collected in discrete bins, as shown in Fig. 1.7.

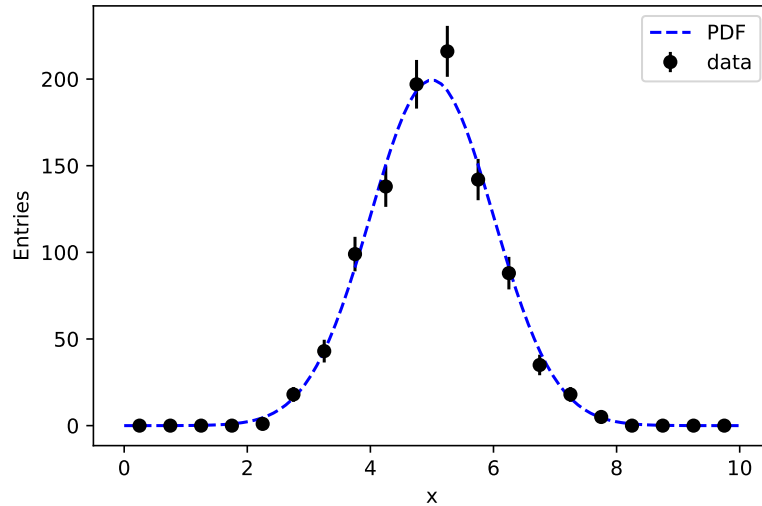


Figure 1.7: The Gaussian PDF scaled to compare to data from a Gaussian distribution. In this case, there are 1000 total entries in the histogram and the bin size is 0.5, for a scale factor of 500.

1.11 Homework Exercises for Distributions

Problem 1: Show that the Binomial distribution, $P(m)$, in Equation 1.3 is properly normalized:

$$\sum_{m=0}^n P(m) = 1$$

as a consequence of the Binomial Theorem (Equation 1.1).

Problem 2: Show that Equation 1.6 is correct.

Problem 3: Show that the Poisson distribution, $P(m)$, in Equation 1.10 is properly normalized:

$$\sum_{m \geq 0} P(m) = 1.$$

Hint: recall the Taylor series expansion for e^λ .

Problem 4: Show that the Gaussian distribution, $P(x)$, in Equation 1.13 is properly normalized:

$$\int_{-\infty}^{\infty} P(x) dx = 1.$$

Problem 5: Show that the mean of the Gaussian distribution has been correctly identified in Equation 1.13. That is, show explicitly that:

$$\int_{-\infty}^{\infty} x P(x) dx = \lambda$$

Problem 6: Show that the variance of the Gaussian distribution has been correctly identified in Equation 1.13. That is, show explicitly that:

$$\int_{-\infty}^{\infty} x^2 P(x) dx = \sigma^2$$

when we take $\lambda = 0$ (which is equivalent to simply changing variables $y = x - \lambda$.)

Chapter 2

Statistical Uncertainties

2.1 Reporting Experimental Uncertainties

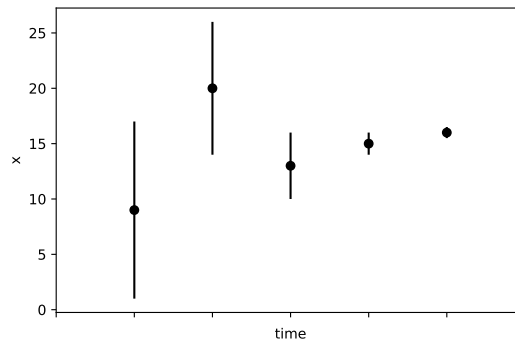


Figure 2.1: The evolution of a measurement of a quantity x over time.

A scientific measurement is meaningless without an associated uncertainty. For instance, we might report a measured distance as:

$$x = 1.2 \pm 0.3 \text{ m}$$

and a time as:

$$t = 4.76 \pm 0.13 \mu\text{s}.$$

When reporting a number with an associated uncertainty, we round both the value and the uncertainty to one significant digit in the uncertainty, so:

$$x = 1.245 \pm 0.313 \text{ m} \rightarrow x = 1.2 \pm 0.3 \text{ m}$$

If the first nonzero digit in the uncertainty is a zero or one, we keep two digits instead:

$$x = 1.245 \pm 0.113 \text{ m} \rightarrow x = 1.25 \pm 0.11 \text{ m}.$$

We'll define this uncertainty quite precisely, but let's start with a conceptual definition. Each particular measurement of a quantity x makes a claim about where the true value of x is most likely to lie. If we assume that measurements improve with time and better reproduce the true value, we

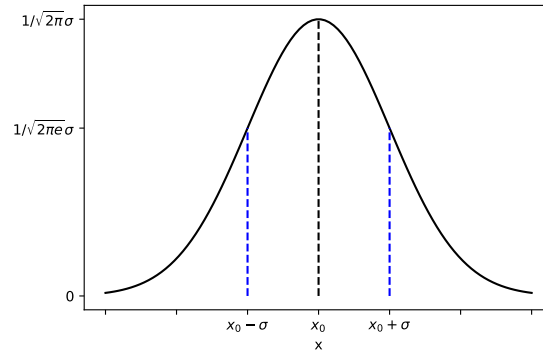


Figure 2.2: The interpretation of experimental uncertainties as the parameters of a Gaussian distribution for the true value.

expect a series of measurements to proceed like in Fig. 2.1, with future measurements shown to be consistent with the uncertainties reported by previous experiments.

We can therefore think of a measurement and its uncertainty as describing a PDF for the outcome of experiments in the distant future with uncertainties so small as to effectively measure the true value of the quantity. Because most measured quantities have many independent sources of uncertainties, the Central Limit Theorem implies that the true value of the measured quantity is best described by a Gaussian Distribution. We define the measured value x as the mean of this distribution, and the uncertainty as the σ of the Gaussian.

2.2 The Error Function

Suppose we measure the speed of light in vacuum to be:

$$3.3 \pm 0.2 \times 10^8 \text{ m/s}^2$$

How consistent is this with the generally accepted value:

$$3.0 \times 10^8 \text{ m/s}^2$$

where we have omitted the uncertainty here because it is much smaller than our uncertainty! The interpretation of the uncertainty as the σ of a Gaussian distribution allows us to give quite precise answers. To begin we might say that it is consistent within:

$$\frac{3.3 - 3.0}{0.2} = 1.5 \sigma.$$

but we can also ask, what is the probability enclosed in 1.5σ . To answer this question, we use the error function:

$$\text{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x \exp(-t^2) dt$$

which is simply the integral from $-x$ to x of a Gaussian with $\sigma = 1/\sqrt{2}$. So if we want to calculate the probability contained in n sigma we would calculate:

$$\text{erf}\left(n \cdot \frac{1}{\sqrt{2}}\right)$$

which is tabulated in Table 2.1.

interval	error function	integrated probability
$\pm 1\sigma$	$\operatorname{erf}\left(\frac{1}{\sqrt{2}}\right)$	68.3%
$\pm 2\sigma$	$\operatorname{erf}\left(\frac{2}{\sqrt{2}}\right)$	95.4%
$\pm 3\sigma$	$\operatorname{erf}\left(\frac{3}{\sqrt{2}}\right)$	98.5%
$\pm 4\sigma$	$\operatorname{erf}\left(\frac{4}{\sqrt{2}}\right)$	99.5%
$\pm 5\sigma$	$\operatorname{erf}\left(\frac{5}{\sqrt{2}}\right)$	99.8%

Table 2.1: The integrated probability for a Gaussian distribution within the stated bounds.

2.3 Mean and Variance of a Sum

So far we've been calculating expectation values of functions of one random variable:

$$\langle f(x) \rangle = \int f(x) P(x) dx$$

but this easily generalizes to two or more random variables:

$$\langle f(x, y) \rangle = \int \int f(x, y) P(x, y) dx dy$$

If the random variables are independent, then we have $P(x, y) = P(x) P(y)$ and so

$$\langle f(x, y) \rangle = \int \int f(x, y) P(x) P(y) dx dy$$

And also we have

$$\begin{aligned}
 \langle f(x) + g(y) \rangle &= \int \int dx dy P(x) P(y) (f(x) + g(y)) \\
 &= \left(\int P(y) dy \right) \cdot \left(\int P(x) f(x) dx \right) + \left(\int P(x) dx \right) \cdot \left(\int P(y) f(y) dy \right) \\
 &= 1 \cdot \int P(x) f(x) dx + 1 \cdot \int P(y) f(y) dy
 \end{aligned}$$

which we can write much more simply as:

$$\langle f(x) + g(y) \rangle = \langle f(x) \rangle + \langle g(y) \rangle$$

revealing the power of the expectation value notation for tackling independent random variables. In the homework you will show that similarly:

$$\langle f(x) \cdot g(y) \rangle = \langle f(x) \rangle \cdot \langle g(y) \rangle \quad (2.1)$$

We can therefore conclude that the mean of a sum of two random variables:

$$\langle x + y \rangle = \langle x \rangle + \langle y \rangle$$

is just the sum of the mean.

For simplicity let's assume that $\langle x \rangle = \langle y \rangle = 0$, and so:

$$\begin{aligned}\langle (x + y)^2 \rangle &= \langle x^2 + 2xy + y^2 \rangle \\ &= \langle x^2 \rangle + 2 \langle xy \rangle + \langle y^2 \rangle \\ &= \langle x^2 \rangle + 2 \langle x \rangle \langle y \rangle + \langle y^2 \rangle \\ &= \langle x^2 \rangle + \langle y^2 \rangle\end{aligned}$$

from which it follows that:

$$\begin{aligned}\sigma^2(x + y) &= \langle (x + y)^2 \rangle - (\langle x + y \rangle)^2 \\ &= \langle (x + y)^2 \rangle \\ &= \langle x^2 \rangle + \langle y^2 \rangle \\ &= \sigma_x^2 + \sigma_y^2\end{aligned}$$

In the exercises you will show that even if $\langle x \rangle$ and $\langle y \rangle$ are non-zero, we still obtain:

$$\sigma^2(x + y) = \sigma_x^2 + \sigma_y^2$$

that is, the uncertainties in a sum add in quadrature.

2.4 Uncertainties Add in Quadrature

In the previous section, we discussed that the variance of the sum of two random variables drawn from any distribution is just the sum of their variances:

$$\sigma^2(x + y) = \sigma_x^2 + \sigma_y^2$$

This is a quite general result. In the context of experimental uncertainties, we consider experimental measurements to be drawn from a Gaussian distribution and the square-root of the variance (σ) is called the uncertainty.

If we calculate the sum s of two independent measurements x and y with uncertainties σ_x and σ_y , it's clear that the variance of the sum will be simply $\sigma_x^2 + \sigma_y^2$. If the resulting distribution of $x + y$ is a Gaussian distribution, as seems probable due to the Central Limit Theorem, then the uncertainties simply add in quadrature:

$$\sigma_{x+y} = \sqrt{\sigma_x^2 + \sigma_y^2}.$$

Let's show explicitly that this is the case. If we wish to know the probability that two random variables x and y add to some particular value $u = x + y$, we simply integrate the total probability of x and y subject to the requirement $u = x + y$:

$$\begin{aligned}P(u) &= \int dx \int dy P_x(x) P_y(y) \delta(u - (x + y)) \\ &= \int dx P_x(x) P_y(u - x)\end{aligned}$$

If we make the (very often valid) assumption that x and y are Gaussian distributed, and, for simplicity, assume that the mean values are zero (or simply change coordinates), so that we have:

$$\begin{aligned} P_x(x) &= \frac{1}{\sqrt{2\pi}a} \exp\left(-\frac{x^2}{2a^2}\right) \\ P_y(y) &= \frac{1}{\sqrt{2\pi}b} \exp\left(-\frac{y^2}{2b^2}\right) \end{aligned}$$

And so the mean value probability distribution function for u is now:

$$\begin{aligned} P(u) &= \frac{1}{2\pi ab} \int dx \exp\left(-\frac{x^2}{2a^2}\right) \exp\left(-\frac{(u-x)^2}{2b^2}\right) \\ &= \frac{1}{2\pi ab} \int dx \exp\left(-\frac{a^2+b^2}{2a^2b^2} \left\{x^2 - \frac{2a^2}{a^2+b^2}ux + \frac{a^2}{a^2+b^2}u^2\right\}_1\right) \end{aligned}$$

We deal with the term in brackets ($\{\}_1$) by completing the square. Simply note that

$$\left(x - \frac{a^2}{a^2+b^2}u\right)^2 = x^2 - \frac{2a^2}{a^2+b^2}ux + \frac{a^4}{(a^2+b^2)^2}u^2 \quad (2.2)$$

reproduces the first and second terms, so we can replace:

$$x^2 - \frac{2a^2}{a^2+b^2}ux = \left(x - \frac{a^2}{a^2+b^2}u\right)^2 - \frac{a^4}{(a^2+b^2)^2}u^2 \quad (2.3)$$

to obtain:

$$\begin{aligned} \{\}_1 &= \left(x - \frac{a^2}{a^2+b^2}u\right)^2 - \frac{a^4}{(a^2+b^2)^2}u^2 + \frac{a^2}{a^2+b^2}u^2 \\ &= \left(x - \frac{a^2}{a^2+b^2}u\right)^2 + \frac{a^2b^2}{(a^2+b^2)^2}u^2 \end{aligned}$$

and substituting back into the original expression we obtain:

$$P(u) = \frac{1}{2\pi ab} \left\{ \int_{-\infty}^{+\infty} dx \exp\left(-\frac{a^2+b^2}{2a^2b^2} \left(x - \frac{a^2}{a^2+b^2}u\right)^2\right) \right\}_2 \exp\left(-\frac{1}{2} \frac{u^2}{a^2+b^2}\right)$$

making a substitution of variables (u is constant during the integration):

$$y = x - \frac{a^2}{a^2+b^2}u$$

the definite integral in brackets yields:

$$\{\}_2 = \sqrt{(2\pi)} \frac{ab}{\sqrt{a^2+b^2}} \quad (2.4)$$

and we have at last:

$$P(u) = \frac{1}{\sqrt{2\pi}\sqrt{a^2+b^2}} \exp\left(-\frac{1}{2} \frac{u^2}{a^2+b^2}\right)$$

which is a Gaussian distribution with variance:

$$\sigma^2 = a^2 + b^2,$$

that is, uncertainties add in quadrature.

2.5 Handling Constants

When adding a known constant to a measured quantity:

$$y = x + C$$

we can consider the constant to have uncertainty $\sigma = 0$. And therefore, from the addition in quadrature rule:

$$\sigma_y = \sigma_x$$

If we multiply a measured quantity by a known constant

$$y = kx$$

we consider that:

$$dx \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = d(kx) \frac{1}{\sqrt{2\pi}k\sigma} \exp\left(-\frac{(kx-k\mu)^2}{2(k\sigma)^2}\right)$$

and conclude that:

$$\sigma_y = k\sigma_x$$

2.6 General Propagation of Uncertainties

We are now ready to handle propagation of uncertainties for a general function. Suppose we measure $x = x_0 \pm \sigma_x$ and $y = y_0 \pm \sigma_y$ and we wish to know the resulting uncertainty on the calculated quantity $f(x, y)$.

Now we Taylor expand the function about the measured values x_0 and y_0 :

$$f(x, y) = f(x_0 + \Delta x, y_0 + \Delta y) \sim f(x_0, y_0) + \left. \frac{df}{dx} \right|_{x_0, y_0} \Delta x + \left. \frac{df}{dy} \right|_{x_0, y_0} \Delta y$$

We note that

$$\begin{aligned} \Delta x &\equiv x - x_0 \\ \Delta y &\equiv y - y_0 \end{aligned}$$

where we can consider x_0 and y_0 to be precisely known constants that happen to coincide with our best measured value of x and y . In this case, our measurement amounts to:

$$\begin{aligned} \Delta x &= 0 \pm \sigma_x \\ \Delta y &= 0 \pm \sigma_y \end{aligned}$$

and our uncertainties are associated with Δx and Δy , not the “constants” x_0 and y_0 .

Turning back to the Taylor expansion, we see now that:

$$f(x, y) \sim A + B\Delta x + C\Delta y$$

where the constants are:

$$\begin{aligned} A &= f(x_0, y_0) \\ B &= \left. \frac{df}{dx} \right|_{x_0, y_0} \\ C &= \left. \frac{df}{dy} \right|_{x_0, y_0} \end{aligned}$$

We know how to propagate uncertainties in this case, which involves scaling, adding in quadrature, and adding a constant:

$$\sigma_f^2 = B^2 \sigma_x^2 + C^2 \sigma_y^2$$

And plugging in the constants:

$$\sigma_f^2 = \left(\left. \frac{df}{dx} \right|_{x_0, y_0} \right)^2 \sigma_x^2 + \left(\left. \frac{df}{dy} \right|_{x_0, y_0} \right)^2 \sigma_y^2$$

2.7 Variance of the Mean

Suppose we have an apparatus that can measure a quantity x with uncertainty σ , and we make a series of N measurements $x_1, x_2, x_3, \dots, x_N$.

We can calculate the mean value:

$$X = \frac{x_1 + x_2 + x_3 + \dots + x_N}{N}$$

The uncertainty on the mean value can be calculated by propagating the uncertainties in the usual way:

$$\begin{aligned} \sigma_X &= \sqrt{\left(\frac{dX}{dx_1} \sigma \right)^2 + \left(\frac{dX}{dx_2} \sigma \right)^2 + \dots + \left(\frac{dX}{dx_N} \sigma \right)^2} \\ &= \sqrt{N \left(\frac{\sigma}{N} \right)^2} \\ &= \frac{\sigma}{\sqrt{N}} \end{aligned}$$

If we assume that the mean value is approaching the true value of the quantity x , then repeating the measurement N times reduces the uncertainty by a factor \sqrt{N} .

It would seem, therefore, that by simply repeating the same experiment many times, we could reach very high precision even with very lousy equipment. But this is, unfortunately, not generally the case. The statistical uncertainties which we have considered here do indeed diminish with more data. But there are also systematic uncertainties, which are the same for each measurement. The mean of many measurements tends toward the true value plus some constant (and generally unknown) offset known as the systematic uncertainty.

Generally in the early stages, experiments are statistics limited, and taking more data helps reduce the overall uncertainty. But at a later stage, the experiment becomes systematics limited. More data does not help unless the apparatus is improved. Often, in complicated experiments, increased statistics also gives the experimenter a better handle on the systematic uncertainties, allowing for improvements to both the statistical and systematic uncertainties.

2.8 Exercises for Uncertainties

Problem 1: Suppose you measure an RMS voltage as $V = 43.2145$ mV and you calculate the uncertainty on this measurement to be 0.471 mV. How should you report this measurement? How about if the uncertainty were 1.07 mV?

Problem 2: Show that Equation 2.1 is valid.

Problem 3: Show that for independent random variables x and y :

$$\langle (x + y)^2 \rangle - \langle x + y \rangle^2 = \langle x^2 \rangle - \langle x \rangle^2 + \langle y^2 \rangle - \langle y \rangle^2.$$

What does this imply about the uncertainty on the sum of two variables, assuming all of the PDFs are Gaussian?

Problem 4: For a measurement $x_0 \pm \sigma$ we define the fractional uncertainty as σ/x_0 . Use the general formula for propagating uncertainties to show that for products and ratios, the *fractional* uncertainties add in quadrature.

Problem 5: Suppose you measure the system gain for your new ionization chamber to be 12 mV per collected electron. Suppose you measure a pulse with a height of 6.2 V. First estimate the uncertainty on this measured value due to statistical fluctuations from the number of electrons. You can neglect any uncertainties on the system gain measurement and the voltage reading itself as these are very small compared to the uncertainty due to electron statistics. Next, suppose your pre-amplifier adds about 300 mV of noise to this measurement. What is the total uncertainty for this measurement?

Problem 6: The rare decay of the Higgs boson ($m \sim 125$ GeV in units where $c = 1$) into pairs of photons played a crucial role in its discovery at the Large Hadron Collider. The invariant mass of a particle that decays into two massless particles p_1 and p_2 can be calculated from:

$$m = \sqrt{2p_1p_2(1 - \cos \theta)}$$

Suppose you were sitting in the control and saw a nice looking two photon event with $p_1 = 53 \pm 2$ GeV, $p_2 = 75 \pm 3$ GeV and $\theta = 2.8 \pm 0.1$. Calculate the invariant mass of these two photons (answer in GeV) and the uncertainty.

Problem 7: A perennial problem with drawing data in histograms is what to do about bins that have zero events in them. Those crazy people that insist that data has no statistical uncertainty (only predictions have statistical uncertainty!) laugh maniacally when we struggle with what to do in these bins. This is where our highly useful fiction that data has an associated uncertainty manifestly falls apart!

The canonical way to draw uncertainties on histogram data is to take the uncertainty in a bin as \sqrt{n} when n is the number of entries in the bin. But for $n = 0$, this implies the uncertainty is zero! If you take this too literally during the statistical analysis of your data, it will cause serious problems (as we'll see when we get to curve fitting.) But for histogram plotting purposes, it is

standard practice to put zero uncertainty on bins with zero events: anything else tends to clutter a plot without adding much useful information.

Another proposal one sees occasionally is to draw the data point at 0 with an error bar going from 0 to 1. I must admit, I never understand this proposal, as there is a far better choice that we will work out here. When we draw typical error bars corresponding to the uncertainty $\pm\sigma$ we are integrating 68.3% of the Gaussian PDF. At zero events, the Gaussian distribution is not the right distribution, and we should instead use the Poisson distribution suitable for small numbers of events. We observed 0 events, so we'll start our range at $\lambda = 0$. We want to find the largest value of lambda that keeps the probability p of observing 0 events (as we have) at $p \geq 1 - 0.683$. Find this value of lambda.

Problem 8: Those crazy people that think only predictions should be drawn with uncertainties argue like this: the Poisson distribution has one parameter λ which is both the mean and the variance. We expect to see λ events on average with fluctuations of about $\sigma = \sqrt{\lambda}$, leading to some observed value n . When we plot uncertainties of \sqrt{n} we are plotting the wrong uncertainty, which should be $\sqrt{\lambda}$. Calculate how much these two estimates vary for a typical one sigma fluctuation, up and down, for $\lambda = 10$ and $\lambda = 100$. Do the crazy people have a point?

Problem 9: In particle physics, we always need to understand the efficiency of our detectors. For instance, if n muons are produced by collisions, we'd like to know how many are actually recorded successfully by our muon detectors. If we measure m muons that are actually detected, we know that our efficiency is $\epsilon = m/n$. No problem here!

You will often see young particle physicists (and occasionally old!) run into problems estimating the uncertainty of this estimate. They reason like this: n is just a constant, with no uncertainty. I could pick to study exactly 1000 events, for instance. The only number which I actually measure is m , and that should fluctuate by $\sigma_m = \sqrt{m}$. So my uncertainty on the efficiency is just $\sigma_\epsilon = \sigma_m/n = \sqrt{m}/n$. Problems arise because we often build good detectors with high efficiency. So consider, for instance $n = 100$ and $m = 95$. This leads to a measurement of the efficiency as 0.95 ± 0.10 which seems to include the impossible value $\epsilon = 1.05$.

One way out of this embarrassment is to realize that this problem is more appropriate for the binomial distribution which has two parameters, the number of trials n and the success rate ϵ which is precisely the efficiency we are attempting to measure, and for which we all agree our best estimate is $\epsilon = m/n$. Unlike our incorrect assumption that $\sigma_m = \sqrt{m}$, for the Binomial distribution $\sigma_m = \sqrt{n\epsilon(1-\epsilon)}$. Calculate the resulting uncertainty on the efficiency ϵ and check that there is no longer a problem for $n = 100$ and $m = 95$.

There's another way to solve this problem, and that is to realize that m and n are not independent variables, because $m > n$ is impossible. If we instead think of this in terms of the number of events p that pass, and the number of events f that fail, these variable are independent. Calculate the uncertainty on the efficiency:

$$\epsilon = \frac{p}{p+f}$$

using standard propagation of uncertainties and the fact that $\sigma_p = \sqrt{p}$ and $\sigma_f = \sqrt{f}$. You should be able to reproduce the same answer as above.

Chapter 3

Statistical Analysis

3.1 Likelihood and χ^2

Suppose we make a series of measurements:

$$\{x_1 \pm \sigma_1, x_2 \pm \sigma_2, \dots, x_n \pm \sigma_n\} \equiv x_i \pm \sigma_i$$

and we would like to quantify how likely this outcome is to have occurred as the result of a corresponding theoretical prediction for each measurement:

$$\{X_1, X_2, \dots, X_n\} \equiv X_i$$

Assuming the uncertainties on each x are Gaussian, the probability of one measurement is:

$$P_i = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(X_i - x_i)^2}{2\sigma_i^2}\right)$$

And the probability for the complete set of measurements, called the Likelihood, is the product of these probabilities for each measurement:

$$\mathcal{L} = \prod_i P_i = \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(X_i - x_i)^2}{2\sigma_i^2}\right)$$

Now being physicists, we hate products and prefer sums, so we apply a logarithm, and there is an annoying factor of $\frac{1}{2}$ in the exponential, so we multiply by 2. We'd also like to construct a quantity that tells us about how far apart a prediction is from our measurement, and so we multiply by -1 , to obtain:

$$-2 \log \mathcal{L} = \sum_i \frac{(X_i - x_i)^2}{\sigma_i^2} + 2 \log(\sqrt{2\pi}\sigma_i) \quad (3.1)$$

Assuming the experimental uncertainties, σ_i , are known, the second term is simply a constant of the experiment setup. The first term is what we call the χ^2 metric:

$$\chi^2 \equiv \sum_i \frac{(X_i - x_i)^2}{\sigma_i^2} \quad (3.2)$$

A small value of χ^2 means that the result is close to the theoretical prediction and a large value means that the result is unlikely to have occurred as a result of the prediction. If the uncertainties

and prediction are all correct, we would expect each x_i to differ from the prediction X_i by about σ_i . So in this case we would expect:

$$\chi^2 = \sum_i \frac{(X_i - x_i)^2}{\sigma_i^2} \sim \sum_i 1 = N \quad (3.3)$$

This assumes that the prediction X_i is simply given to us. We'll revisit this assumption when we discuss degrees of freedom.

3.2 Maximal Likelihood Method

Often, we are interested to know which particular parameters of a model maximize the likelihood of the data we have collected. For instance, suppose we made N measurements:

$$\{x_1 \pm \sigma, x_2 \pm \sigma, \dots, x_N \pm \sigma\}$$

of the same quantity, and we would like to find out which single value X is most consistent with these n measurements. Our χ^2 for this simple model is then:

$$\chi^2 = \sum_i \frac{(X - x_i)^2}{\sigma^2}.$$

We can find the particular value of X that maximizes the likelihood by finding the value that minimizes the χ^2 , because $\chi^2 \sim -2 \log \mathcal{L}$. We find the minimum from the condition:

$$\frac{d\chi^2}{dX} = 0,$$

which amounts to:

$$\begin{aligned} 0 &= 2 \sum_i \frac{(X - x_i)}{\sigma^2} \\ 0 &= \left(X \sum_i 1 \right) - \left(\sum_i x_i \right) \\ X &= \frac{1}{N} \sum_i x_i \end{aligned} \quad (3.4)$$

which of course is just the mean of the measurements, as expected.

The χ^2 formalism is quite general. Suppose at each position x_i we make a measurement y_i which has a corresponding theoretical prediction $f(x_i; a, b)$ where a and b are parameters of the theory. In this case, the χ^2 is given:

$$\chi^2 = \sum_i \frac{(f(x_i; a, b) - y_i)^2}{\sigma_i^2}$$

and to determine the best fit values for a and b we would require:

$$\begin{aligned} \frac{\partial \chi^2}{\partial a} &= 0 \\ \frac{\partial \chi^2}{\partial b} &= 0 \end{aligned}$$

We can accommodate any number of theory parameters in this fashion. And the interpretations of x and y are endless. We can imagine taking measurements of the voltage at particular times, measuring the electric field strength at particular radii, the number of cars produced in a factory each month, and so on.

3.3 Interpretation of $\Delta\chi^2$

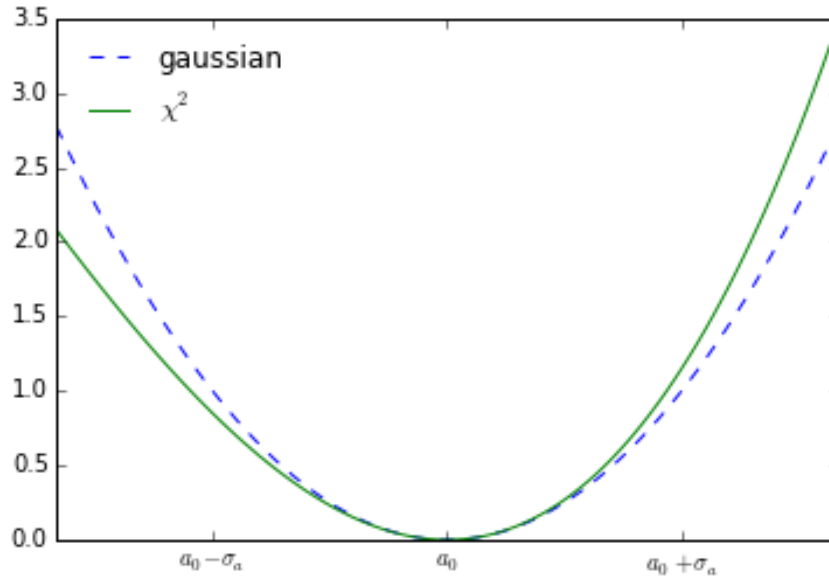


Figure 3.1: Approximation of χ^2 near minimum.

Suppose at particular points $\{x_i\}$ we have made measurements $\{y_i \pm \sigma_i\}$ with corresponding theoretical predictions $f(x_i, a)$. Using the χ^2 formalism we can determine the best fit value a_0 for the parameter a . But this is of very little use unless we can also determine the corresponding uncertainty σ_a associated with the best fit value a_0 .

One approach is to simply apply propagation of uncertainties to the formula determined by minimizing the χ^2 , for instance, Equation 3.4. But this approach can be tedious or even unusable in cases where the χ^2 is minimized numerically and no closed form solution is available. It is well worthwhile, therefore, to consider an alternative approach, that determines these uncertainties directly from the Likelihood and its corresponding χ^2 distribution.

Consider the Likelihood associated with this series of measurements:

$$\mathcal{L}(a) = \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2} \frac{(f(x_i, a) - y_i)^2}{\sigma_i^2}\right)$$

This complicated function in the end is just the PDF for the true value of the parameter a , as determined by our data. We are well justified, therefore, to assume this PDF is equivalently a simple Gaussian distribution:

$$\mathcal{L}(a) = \frac{1}{\sqrt{2\pi}\sigma_a} \exp\left(-\frac{(a - a_0)^2}{2\sigma_a^2}\right)$$

explicitly in terms of the best fit value a_0 and the uncertainty σ_a . The χ^2 is then simply:

$$\chi^2(a; a_0, \sigma_a) = \frac{(a - a_0)^2}{\sigma_a^2} \quad (3.5)$$

We see that:

$$\frac{d\chi^2}{da} = \frac{2(a - a_0)}{\sigma_a^2} = 0$$

when $a = a_0$ exactly as expected. We also see that:

$$\frac{d^2\chi^2}{da^2} = \frac{2}{\sigma_a^2}$$

As shown in Fig. 3.1 the χ^2 for a Gaussian likelihood is a parabola, and so the second derivative *evaluated anywhere* yields a term containing the uncertainty σ_a . However, the region near the minimum of any χ^2 can distribution can be approximated as a parabola about the minimum, and so for a general χ^2 distribution, we have:

$$\sigma_a^2 = \frac{2}{\left. \frac{d^2\chi^2}{da^2} \right|_{a_0}} \quad (3.6)$$

The uncertainty on any parameter can be estimated from the curvature of the χ^2 at the minimum. Another approach, sometimes used in numerical calculations, is to note that χ^2 changes by a factor of one when you move a distance σ_a from the minimum.

3.4 Uncertainty on the Mean

Earlier we used the χ^2 formalism to show that the best fit constant value describing a set of repeated measurements of the same quantity:

$$\{x_1 \pm \sigma, x_2 \pm \sigma, \dots, x_N \pm \sigma\} \equiv x_i \pm \sigma_i$$

is, not surprisingly, the mean value of the data:

$$X_0 = \frac{1}{N} \sum_i x_i.$$

We are now in a position to determine the uncertainty on this mean value, by first taking the second derivative of the χ^2 function:

$$\begin{aligned} \chi^2 &= \sum_i \frac{(X - x_i)^2}{\sigma^2} \\ \frac{d\chi^2}{dX} &= \sum_i \frac{2(X - x_i)}{\sigma^2} \\ \frac{d^2\chi^2}{dX^2} &= \sum_i \frac{2}{\sigma^2} \\ &= \frac{2N}{\sigma^2} \end{aligned}$$

From which we conclude that the uncertainty on the mean is given by:

$$\sigma_X^2 = \frac{2}{\left. \frac{d^2\chi^2}{dX^2} \right|_{X_0}} = \sigma^2/N$$

a result we obtained before by simply propagating uncertainties.

3.5 Weighted Mean

Suppose we make a series of measurements of the same quantity as before, but now each measurement has a different uncertainty:

$$\{x_1 \pm \sigma_1, x_2 \pm \sigma_2, \dots, x_N \pm \sigma_N\} \equiv x_i \pm \sigma_i$$

Our χ^2 formalism allows us to determine the single value X that best matches this data. The χ^2 is simply:

$$\chi^2 = \sum_i \frac{(X - x_i)^2}{\sigma_i^2}$$

And the minimum value of χ^2 occurs at:

$$\frac{d\chi^2}{dX} = 0$$

and so

$$\begin{aligned} \frac{d\chi^2}{dX} &= \sum_i \frac{2(X - x_i)}{\sigma_i^2} = 0 \\ 0 &= \sum_i \frac{x_i}{\sigma_i^2} - X \sum_i \frac{1}{\sigma_i^2} \\ X &= \sum_i w_i x_i \end{aligned}$$

where

$$w_i = \frac{1/\sigma_i^2}{\sum_j 1/\sigma_j^2}. \quad (3.7)$$

The uncertainty is determined from the second derivative of the χ^2 function:

$$\frac{d^2\chi^2}{dX^2} = \sum_i \frac{2}{\sigma_i^2}$$

From which we determine that the uncertainty on the mean is:

$$\sigma_X^2 = \frac{2}{\left. \frac{d^2\chi^2}{dX^2} \right|_{X_0}} = \left(\sum_i \frac{1}{\sigma_i^2} \right)^{-1}$$

3.6 Degrees of Freedom

We noted before that we expect the χ^2 function for a series of N measurements to have a value of approximately N , because each term in the χ^2 sum is approximately one. We need to revisit this estimate when we have we are comparing our data to a best fit function.

It's instructive to consider some extreme cases. Consider the case that we make a single measurement $x_1 \pm \sigma_1$. In this case, the mean value is simply $X = x_1$ and the χ^2 is zero. Suppose we make two measurements. In this case, the best fit line $y = ax + b$ will pass through the two points, and again we will have $\chi^2 = 0$. Quite in general, we can always fit N data points exactly with N parameters. When we discuss the typical size of χ^2 , the important quantity is the number of degrees of freedom (NDF), which is the number of measurements minus the number of fit parameters. The χ^2 for a function that has been *fitted* to N data points should have:

$$\chi^2/\text{NDF} \sim 1$$

3.7 The Best Estimate for σ

So far we have presumed that we know the uncertainty associated with each measurement, but suppose we don't know this.

Minimizing χ^2 is no help here, because we can make σ as large as we want to minimize χ^2 . This is because we treated the uncertainties as constants! Return to:

$$-2 \log \mathcal{L} = \sum_i \left(\frac{(X_i - x_i)^2}{\sigma^2} + 2 \log(\sqrt{2\pi}\sigma) \right) \quad (3.8)$$

Differentiating wrt σ and setting to zero:

$$0 = \left(-2 \sum_i \frac{(X_i - x_i)^2}{\sigma^3} \right) + \frac{2\sqrt{2\pi}}{\sqrt{2\pi}\sigma} \sum_i \quad (3.9)$$

$$\sigma^2 = \frac{\sum_i (X_i - x_i)^2}{N} \quad (3.10)$$

This analysis assumes we have not found the best fit values for the predictions X_i . In this case, we simply replace N with the degrees of freedom:

$$\sigma^2 = \frac{\sum_i (X_i^{\text{fit}} - x_i)^2}{\text{NDF}}$$

but this quantity is simply the χ^2 per degree of freedom for $\sigma_i = 1$:

$$\sigma^2 = \frac{\chi^2(\sigma_i = 1)}{\text{NDF}}.$$

So if we do not know the uncertainties, we can set $\sigma_i = 1$, and then simply scale the squares of the uncertainties by the χ^2 per degree of freedom at the best fit values for the parameter.