Analysis of Experimental Data

Michael Mulhearn

December 22, 2018

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, than 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:
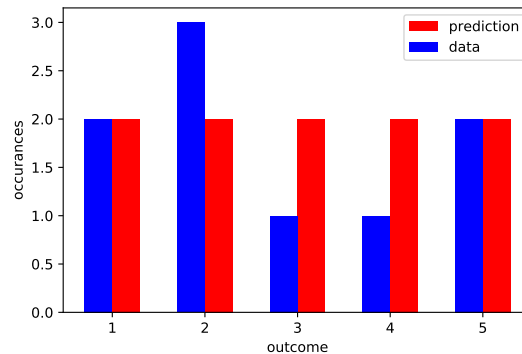

Figure 1: Comparison of experimental results with a prediction.

Scientist 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 well 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
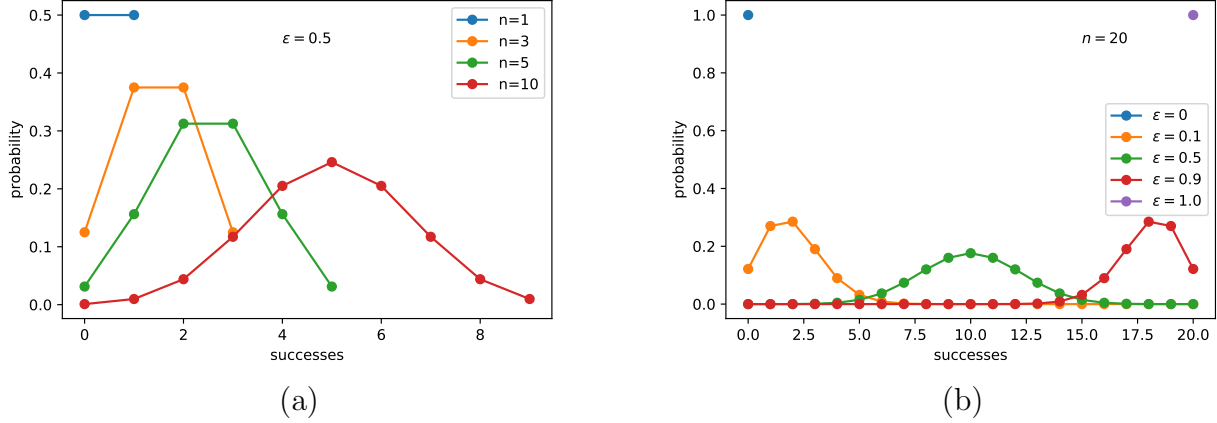
cases of this distribution.

Figure 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:

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

where the binomial coefficients are defined by

$$\binom{n}{m} = \frac{n!}{m! \, (n - m)!} \tag{2}$$

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:

$$P(m;\, n, \epsilon) = \binom{n}{m} \epsilon^m \, (1 - \epsilon)^{n-m} \tag{3}$$

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:

$$\langle m \rangle \equiv \bar{m} \equiv \sum_m m\, P(m) \tag{4}$$

For a continuous probability distribution, we would integrate instead:

$$\langle x \rangle \equiv \bar{x} \equiv \int x\, P(x)\, dx \tag{5}$$

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:

$$\langle (x - \bar{x})^2 \rangle = \langle x^2 \rangle - \bar{x}^2 \tag{6}$$

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:

$$\sigma^2 \equiv \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 \tag{7}$$

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) 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:

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

So the mean value is given by:

$$\bar{m} = n\epsilon \tag{8}$$

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:

$$n(n-1)\epsilon^2 = \sum_{m=0}^{n} (m^2 - m) \binom{n}{m} p^m q^{n-m}$$

$$= \sum_{m=0}^{n} (m^2 - m)P(m)$$

$$= \langle m^2 - m \rangle = \langle m^2 \rangle - \langle m \rangle$$

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:

$$\sigma^2 = n\,\epsilon\,(1 - \epsilon) \tag{9}$$

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 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
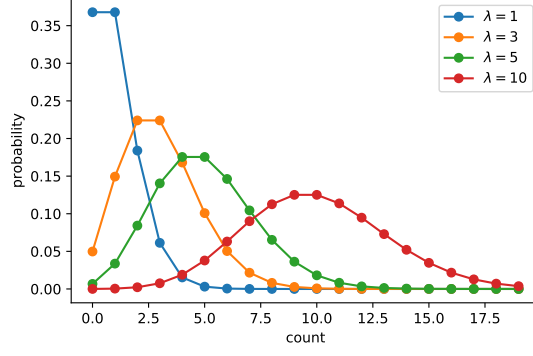
Figure 3: The Poisson distribution for several values of parameter λ.

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 $\epsilon = \lambda/n$:

$$
\begin{aligned}
P(m) &= \binom{n}{m} \epsilon^m (1-\epsilon)^{n-m} \\[2mm]
&= \frac{n!}{m!\,(n-m)!} \left(\frac{\lambda}{n}\right)^m \left(1-\frac{\lambda}{n}\right)^{n-m} \\[2mm]
&= \left(\frac{\lambda^m}{m!}\right) \left(1-\frac{\lambda}{n}\right)^n \left[\frac{n!}{(n-m)!}\cdot\frac{1}{n^m}\right]_1 \left[\left(1-\frac{\lambda}{n}\right)^{-m}\right]_2
\end{aligned}
$$

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

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

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

$$
P(m\,;\,\lambda) = \frac{\lambda^m}{m!} e^{-\lambda} \tag{10}
$$

Notice that there is no longer a parameter n, since we took $n \to \infty$, and so m now ranges from 0 to ∞.

1.6   Mean and Variance of The Poisson Distribution

The mean of the Poisson distribution is given by:

$$
\begin{aligned}
\bar{m} &= \sum_{m\geq 0} m\,P(m) \\[2mm]
&= \sum_{m\geq 0} m\,\frac{\lambda^m}{m!} e^{-\lambda}
\end{aligned}
$$

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

$$
\begin{aligned}
\bar{m} &= e^{-\lambda} \sum_{m \geq 1} \frac{\lambda^m}{(m-1)!} \\
&= \lambda e^{-\lambda} \sum_{m \geq 1} \frac{\lambda^{m-1}}{(m-1)!} \\
&= \lambda e^{-\lambda} \sum_{n \geq 0} \frac{\lambda^n}{n!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
\bar{m} &= \lambda
\end{aligned}
\tag{11}
$$

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:

$$
\begin{aligned}
\sigma^2 &= \langle m^2 \rangle - \langle m \rangle^2 \\
&= \lambda\,(\lambda+1) - \lambda^2 \\
\sigma^2 &= \lambda.
\end{aligned}
\tag{12}
$$

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

1.7   The Gaussian Distribution

Next, we consider the Poisson Distribution in the limit $\lambda \to \infty$. In this case, we can apply the Stirling Approximation:

$$
\lim_{n \to \infty} n! = \sqrt{2\pi n}\; e^{-n}\, n^n
$$

to the Poisson distribution as follows:

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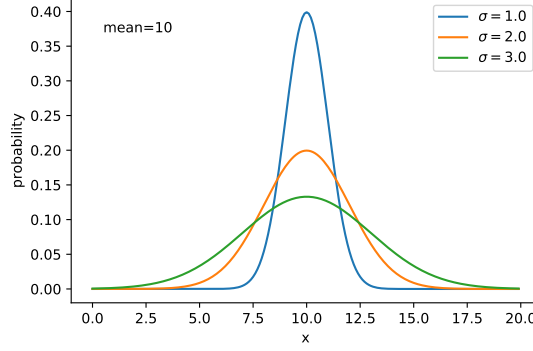\end{aligned}
$$

Figure 4: The Gaussian distribution for a mean of 10 and several values of parameter σ.

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 $\lambda \to \infty$ we have

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

So we can write:

$$m = \lambda(1 + \delta) \tag{13}$$

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

$$
\begin{aligned}
P(\delta) &= \frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}} \left(\frac{\lambda}{\lambda(1+\delta)} \right)^{\lambda(1+\delta)+\frac{1}{2}} \\
&= \frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}} \cdot \frac{1}{X}
\end{aligned}
\tag{14}
$$

where we define the quantity:

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

which can be approximated as follows:

$$\ln X = \left(\lambda(1+\delta) + \frac{1}{2}\right) \cdot \ln(1+\delta)$$

$$= \left(\lambda(1+\delta) + \frac{1}{2}\right) \cdot \left(\delta - \frac{\delta^2}{2} + \mathcal{O}(\delta^3)\right)$$

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

$$= \left(1 + \delta + \mathcal{O}\left(\delta^2\right)\right) \cdot \left(\delta - \frac{\delta^2}{2} + \mathcal{O}(\delta^3)\right)$$

$$= \delta + \frac{\delta^2}{2} + \mathcal{O}(\delta^3),$$

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

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

which, when plugged backed into Equation 14 yields:

$$P(\delta) = \frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}} \cdot \frac{1}{\exp\left(\lambda\delta + \lambda\frac{\delta^2}{2}\right)}$$

$$= \frac{1}{\sqrt{2\pi\lambda}} \cdot \exp\left(-\lambda\frac{\delta^2}{2}\right) \tag{15}$$

Now notice that Equation 13 implies that m is quite large, and so may now be treated as a continuous variable, which we will rename x (since m looks like an integer value), hence we have:

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

and likewise we know that the variance of the original Poisson distribution is given by $\sigma^2 = \lambda$, and so we can rewrite Equation 15 in the (hopefully) more familiar form:

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x-\lambda)^2}{2\sigma^2}\right) \tag{16}$$

Which is a Gaussian distribution with mean value λ and variance σ^2. The proof that these quantities are indeed the mean and the variance is left as an exercise.

1.8 Continuous versus Discrete Distributions

. . .

1.9 Histograms and Distributions in Scientific Python

The experimental analog to a theoretical probability distribution function is a histogram: a graphical representation of the distribution of numerical data. Suppose we measure a continuous quantity x and wish to compare the outcome of many measurements of x with our theoretical prediction.

The probability of any particular outcome, say x=1.24567392874, is vanishing small. Which is why our prediction is in the form of a probability distribution function $p(x)$, where $p(x)dx$ gives the infinitesimal probability for an outcome between x and $x + dx$.

To display our data in a histogram, we divide the range of x into several bins, and simply tabulate how many experiments yielded a result in each bin.

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. 5.

To aid in the comparison between the data and theoretical prediction, we put error bars on each data point. In this case, we have used one-sigma uncertainties from the poisson variance $\sigma^2 = N$. That is, we have taken the square root of the total number of events in each bin as the statistical uncertainty. We'll see in the next section that this is consistent with the standard interpretation of uncertainties.

```
In [23]: samples = np.random.normal(size=1000)
         bins = np.arange(-4,5) # integer binning from -5 to 5
         cbins = 0.5*(bins[:-1] + bins[1:])  # bin centers, for plotting
         counts,dummy = np.histogram(samples,bins)
         errs = np.sqrt(counts)
         #plot(cbins,hist) # <- this version is for amateur hour!

         errs = np.sqrt(counts)
         plt.errorbar(cbins, counts, yerr=errs, fmt='o')


         fbins = np.arange(-4,5,0.1)
         y = mlab.normpdf( fbins , 0.0, 1.0)
         y = y*1000
         plt.plot(fbins,y,'r-')
```
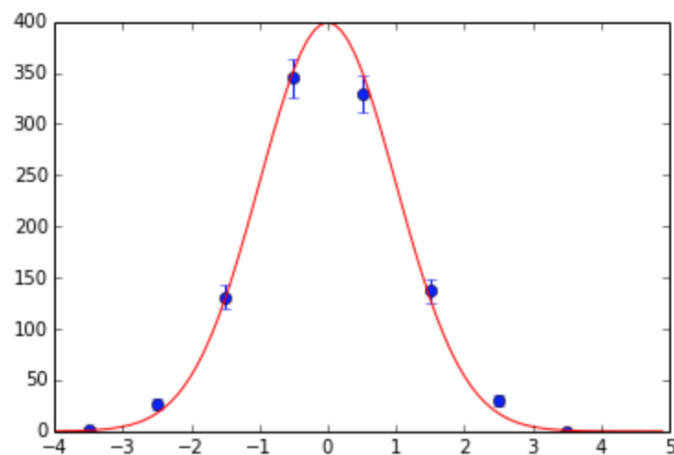
Out[23]: [<matplotlib.lines.Line2D at 0x10f5dbd10>]

Figure 5: Histogram.

1.10 Homework Exercises for Distributions

Problem 1: Show that the Binomial distribution, $P(m)$, in Equation 3 is properly normalized:

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

as a consequence of the Binomial Theorem (Equation 1).

Problem 2: Show that Equation 6 is correct.

Problem 3: Show that the Poisson distribution, $P(m)$, in Equation 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 16 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 16. That is, show explicitly that:

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

Problem 6: Show that the variance of the Gaussian distribution has been correctly identified in Equation 16. 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 $y = x - \lambda$.)

1.11 Scientific Python Exercises for Distributions

Problem 1: (Basic plotting of distributions)

Problem 2: (Show the limit of sum of many flat distribution results in Gaussian)

$\cdots$

2 Statistical Uncertainties

2.1 Precision versus Accuracy

. . .

2.2 Importance of the Gaussian Distribution

2.3 Reporting Experimental Uncertainties

A scientific measurement is meaningless without an associated uncertainty. The convention is to report only one significant digit worth of uncertainty (or two if the first digit is a 1 or 2). So we don't report $x = 1.234 \pm 0.303$ m, but instead would report $x = 1.2 \pm 0.3$ m or $t = 4.76 \pm 0.13$ s.

The uncertainty is intended to indicate how close the true value for measured quantities is to the best estimate. The Central Limit Theorem informs us that the sum of a large number of random variables approaches a Gaussian distribution. Since experimental measurements are often repeated many times, and the sources of noise causing the measurement to differ from the true value are numerous, most often the experimental uncertainties can be treated as coming from a Gaussian distribution. That is to say that the true value of the measured quantity is assumed to be described by a Gaussian distribution with mean value at the best estimated value and with σ given by the reported uncertainty.

2.4 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:

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

which is simply the integral from $-x$ to x of a Gaussian with $\sigma = 1/sqrt2$. So if we want to calculate the probability contained in n sigma we would calculate:

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

which is tabulated in Table 1.

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

Table 1: The integrated probability for a Gaussian distribution within the stated bounds.

2.5 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

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

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:

$$\langle f(x) \cdot g(y) \rangle = \langle f(x) \rangle \cdot \langle g(y) \rangle \tag{17}$$

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. Supposing that $\langle x \rangle = \langle y \rangle = 0$ we also have:

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

In the exercises you will show using expectation values that the variance of the sum of two independent random variables is just the sum of their variances:

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

2.6 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 $u = x + y$:

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

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:

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

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

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

15

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

$$\left(x - \frac{a^2}{a^2+b^2}u\right)^2 = x^2 - \frac{2a^2}{a^2+b^2}ux + \frac{a^4}{(a^2+b^2)^2}u^2 \tag{18}$$

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

$$x^2 - \frac{2a^2}{a^2+b^2}ux = \left(x - \frac{a^2}{a^2+b^2}u\right)^2 - \frac{a^4}{(a^2+b^2)^2}u^2 \tag{19}$$

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:

$$\{\}_2 = \sqrt{(2\pi)}\frac{ab}{\sqrt{a^2+b^2}} \tag{20}$$

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.

2.7 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:

$$\sigma_y = \sigma_x$$

16

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:

$$\sigma_y = k\sigma_x$$

2.8   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:

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

We note that

$$\Delta x \equiv x - x_0$$
$$\Delta y \equiv y - y_0$$

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:

$$\Delta x = 0 \pm \sigma_x$$
$$\Delta y = 0 \pm \sigma_y$$

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:

$$A = f(x_0, y_0)$$
$$B = \frac{df}{dx}\bigg|_{x_0,y_0}$$
$$C = \frac{df}{dy}\bigg|_{x_0,y_0}$$

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:

$$\sigma_f^2 = \left(\frac{df}{dx}\bigg|_{x_0y_0}\right)^2\sigma_x^2 + \left(\frac{df}{dy}\bigg|_{x_0y_0}\right)^2\sigma_y^2$$

2.9 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, ..., x_N.

We can calculate the mean value:

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

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

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

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, as has been conclusively demonstrated by many years of running Physics 116. The statistical uncertainties which we have considered here due 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 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.

2.10 Uncertainties in Scientific Python

2.11 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 17 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:

$$m = \sqrt{2p_1 p_2 (1 - \cos\theta)}$$

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 vari-

ance. 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.

3   Statistical Analysis

3.1   Likelihood and χ^2

Suppose we make a series of measurements:

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

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

$$\{y_1, y_2, \ldots, y_n\} \equiv y_i$$

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

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

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

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

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 multiple by 2, and, generally being pessimists, we prefer to minimize instead of maximize, so we multiple by -1, so that at last we calculate:

$$-2\log\mathcal{L} = \sum_i \frac{(y_i - x_i)^2}{\sigma_i^2} - \log(\sqrt{2\pi}\sigma_i) \tag{21}$$

Assuming the experimental uncertainties, σ_i, are known, the second term is simply a constant. To maximize the likelihood, we therefore minimize:

$$\chi^2 \equiv \sum_i \frac{(y_i - x_i)^2}{\sigma_i^2} \tag{22}$$

A small value of χ^2 means that the result is very 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 y_i by about σ_i. So in this case we would expect:

$$\chi^2 = \sum_i \frac{(y_i - x_i)^2}{\sigma_i^2} \sim \sum_i 1 = N \tag{23}$$

3.2   Minimizing χ^2

If we wish to determine the theoretical prediction that best describes our data, we simply minimize the χ^2 (which amounts to maximizing the likelihood).

Consider, for example, the case where we measure a particular quantity x a total of n times, and so obtain measurements $x_i \pm \sigma_i$. We wish to extract from our data our best estimate for the

true value of x. In this case, the prediction y_i will have the same value for every every i, and we construct χ^2 as:

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

Where m is the parameter we wish to extract from the data. The minimum value of χ^2 occurs at:

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

and so

$$\frac{d\chi^2}{dm} = \sum_i \frac{2(m - x_i)}{\sigma_i^2} = 0$$

$$0 = m \sum_i \frac{1}{\sigma_i^2} - \sum_i \frac{x_i}{\sigma_i^2}$$

$$m = \sum_i w_i x_i$$

where

$$w_i = \frac{1/\sigma_i^2}{\sum_j 1/\sigma_j^2} \tag{24}$$

In the case that all the uncertainties are the same $\sigma_i = \sigma$ the weighting factor is simply $w_i = 1/N$ and we see that the best estimate for m is just the mean value of our measurements.

3.3   Extracting Uncertainties

The example measurement in the preceding section showed us that the best estimate for the true value of x from a series of measurement is a weighted mean. Since the mean is a sum of measured values, we know how to determine the uncertainty on m by propagating the uncertainty from the individual measurements:

$$\sigma_m^2 = \sum_i \left(\frac{dm}{dx_i}\right)^2 \sigma_i^2$$

$$\sigma_m^2 = \sum_i (w_i \sigma_i)^2$$

$$\sigma_m^2 = \sum_i \left(\frac{1}{\sum_j 1/\sigma_j^2}\right)^2$$

$$1/\sigma_m^2 = \frac{1}{N}\left(\sum_j 1/\sigma_j^2\right)^2$$

In case the uncertainties are the same, this reduces to:

$$\sigma_m^2 = \sigma^2/N$$

But now let's consider a much more powerful approach. Let's imagine that we have two parameters a and b that we have determined (using the results of the previous section) to have best fit values a_0 and b_0. We would like to determine the uncertainty on these extracted parameters. The χ^2

22

sums over the experimental measurements and uncertainties, so in the end it is simply a function of the parameters we seek to determine. We can therefore think of our experiment as having an equivalent χ^2 where each parameter was simply directly measured with it's (currently unknown) experimental uncertainty:

$$\chi^2 = \frac{(a-a_0)^2}{\sigma_a^2} + \frac{(b-b_0)^2}{\sigma_b^2}$$

In this case:

$$\frac{1}{2}\frac{d^2\chi^2}{da^2} = \frac{1}{\sigma_a^2}$$

So we see that we can extract the uncertainties on our fit parameters from the second derivative of the χ^2. So the minimum of the χ^2 function is the best fit value, and the curvature at that point is related to the experimental uncertainty on the extracted parameters.

In our example, this leads to:

$$\sigma_m^2 = \sigma^2/N$$

3.4   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:

$$-2\log\mathcal{L} = \sum_i \left(\frac{(m-x_i)^2}{\sigma^2} - \log(\sqrt{2\pi}\sigma) \right) \tag{25}$$

$$-2\log\mathcal{L} = \sum_i \frac{(m-x_i)^2}{\sigma^2} - \log(\sqrt{2\pi}\sigma) \tag{26}$$

Differenentiating wrt σ and setting to zero:

$$0 = \left(\sum_i \frac{(m-x_i)^2}{\sigma^3} \right) - \frac{N}{\sigma} \tag{27}$$

$$\sigma^2 = \frac{\sum_i (m-x_i)^2}{N} \tag{28}$$

4   Systematic Uncertainties

4.1   The Difference Between Statistical and Systematic Uncertainties

The Poisson distribution has the widely appreciated feature that the mean equals the variance, or equivalently, that $\sigma \sim \frac{1}{\sqrt{n}}$. We interpret this uncertainty as resulting from the statistical fluctuations of n events, and we call this uncertainty σ the *statistical* uncertainty. A crucial assumption in the derivation of the Poisson distribution was that the events are independent. ldots

4.2   The χ^2 Metric

Presented with a series of N measurements:

$$\{x_1 \pm \sigma_1, x_2 \pm \sigma_2, ..., x_N \pm \sigma_N\}$$

J7CFFRP[H0909986JGM we are often concerned with how consistent these results are with a corresponding theoretical prediction:

$$\{X_1, X_2, ..., X_N\}$$

Assuming Gaussian uncertainties for each measurement x, then the probability of one particular measurement is just a Gaussian PDF:

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

measurements is called the *Likelihood* and it is just the product of the probabilities of each measurement:

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

Since sums are generally easier than products, we consider:

$$\log \mathcal{L}$$

And since physicist think about minimization rather than maximization (pessimists? gravity?) we take:

$$-\log \mathcal{L}$$

And since there is an annoying factor of $1/2$ in the exponent, we finally take

$$-2\log \mathcal{L}$$

which we can calculate as:

$$-2\log \mathcal{L} = -2\sum_i \log \left(\frac{1}{2\pi\sigma_i} \right) + \sum_i \frac{(x_i - X_i)^2}{\sigma_i^2}$$

The first term is a constant that does not depend on the measured and predicted values, only on the precision of the experiment. The second term, is what we call the χ^2 metric:

$$\chi^2 = \sum_i \frac{(x_i - X_i)^2}{\sigma_i^2} \tag{29}$$

When χ^2 is small, the measured values $\{x_i\}$ are relatively close to the predicted values $\{X_i\}$ as defined by the experimental uncertainties. When χ^2 is large, the predicted values are well outside the experimental uncertainties of the measurements.

4.3 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:

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

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:

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

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:

$$
\begin{aligned}
0 &= 2 \sum_i \frac{(x_i - X)}{\sigma^2} \\
0 &= \sum_i x_i - X \sum_i \\
X &= \frac{1}{N} \sum_i x_i
\end{aligned}
\tag{30}
$$

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:

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

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

$$
\begin{aligned}
\frac{d\chi^2}{da} &= 0 \\
\frac{d\chi^2}{db} &= 0
\end{aligned}
$$

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.

Exercise Calculate a general solution for the best fit parameters of the straight line fit $f = ax+b$.

4.4 Interpretation of $\Delta\chi^2$

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 30. 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 it's corresponding χ^2 distribution.

Let's consider the Likelihood associated with this series of measurements, and interpret it as a probability distribution function for the random variable a:

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

Assume for the moment that this PDF is a Gaussian distribution, and so can written alternatively as:

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

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

$$\chi^2(a; a_0, \sigma_a) = \frac{(a - a_0)^2}{\sigma_a^2} \tag{31}$$

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. 6 the χ^2 for a Guassian 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:

$$\sigma_a^2 = \frac{2}{\left.\frac{d^2\chi^2}{da^2}\right|_{a_0}} \tag{32}$$

Exercise Show that at the minimum, varying a parameter by it's uncertainty changes the χ^2 by 1.

4.5   Numerical Least Squares Fitting

The χ^2 formalism lends itself easily to numerical solutions, since the problem amounts to the optimization problem of finding the minimum of a function. A popular approach is the gradient descent approach. The minimum is approached in steps, with the gradient calculated at each step, and used to point the direction of the next step. The size of the step can be chosen so that the gradient is approximately constant during the step, such as by estimating the second derivative. Once at the minimum, the uncertainties can be estimated from the second derivative according to Equation 32.
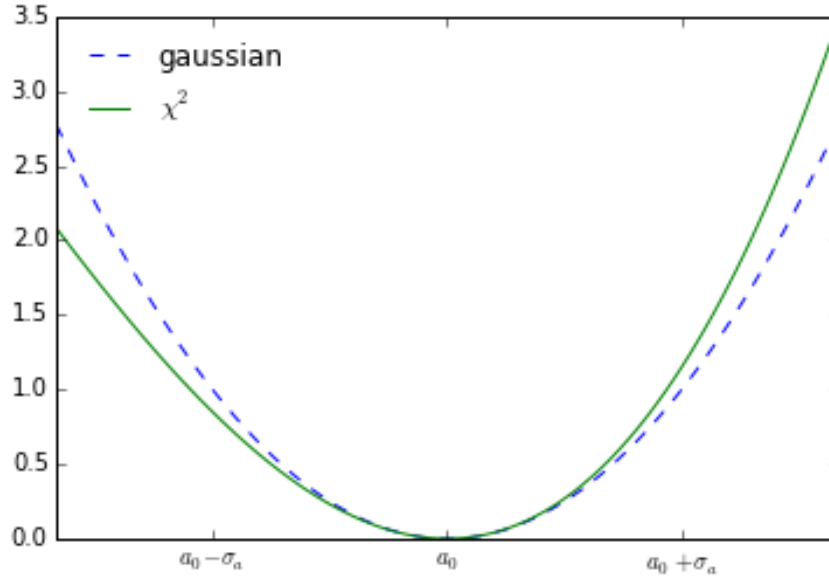
Figure 6: Approximation of χ^2 near minimum.

The major limitation of the steepest descent approach is that it will generally find only the nearest *local* minimum. There is no guarantee that the global minimum has been reached. This limitation is generally avoided by making an initial guess which is known to be near the actual minimum. Also, the results of fits should always be evaluated to make certain they are sensible... local minimums that are not the absolute minimum almost always look ridiculous when plotted against the data!

A complete example implementing a straight line fit in scientific python is included below. Note a few snippets in particular:

```
def func(x, a, b):
    return a*x + b
popt, pcov = curve_fit(func, x, y, sigma=yunc, p0=[1,1],absolute_sigma=True)
```

The fit function is the user defined function "func", which in this case, takes two parameters (a and b). The call to curve_fit is where the actual fit is performed. Notice that the uncertainties are provided explicitly, the initial guesses are provided, and the uncertainties are specified as absolute.

The default behavior of curve_fit is rather unscientific and intended for the general public. It simply assumes all uncertainties are equal (if sigma is unspecified) and scales the uncertainties by an amount that gives a reasonable χ^2 at the minimum. This is seldom what a physicists wants.

You obtain the fitted parameters and their uncertainties from the return values of the functions:

```
# report the fitted values:
print "fitted values:  ", popt


# report the uncertainties, determined from the diagonals of the covariance matrix:
perr = np.sqrt(np.diag(pcov))
print "statistical uncertainties:  ", perr
```

The non-diagonal elements are beyond the scope of this course, but generally provide information about how the fit parameters are correlated. Generally speaking, highly correlated parameters should be avoided by choosing a more appropriate parameterization.

The full example is here:

```
#
# fiteg.py
#
# An example program for fitting a straight line to data with uncertainties:
#

import matplotlib.pyplot as plt
import numpy as np
from scipy.optimize import curve_fit

# generate simulated data y = a*x + b according to:
a = 3;
b = 5;
x = np.arange(0,10)
y = a*x + b
yunc = 0.5 + y*0.03
# smear data by the uncertainties yunc
y = y + np.random.normal(scale=yunc)

# plot the simulated data with uncertainties:
fig, ax = plt.subplots(1,1)
ax.set_xlabel('x')
ax.set_ylabel('y')
ax.errorbar(x,y,yunc,fmt="ko")

# fit the data to the straight line implemented in "func":
# note we provide explicitly:
# - the uncertanities on y (sigma = yunc)
# - the starting guess for our parameters (p0=[1,1])
# - we specify the sigmas are absolute, not just relative, when determining statistical
#   uncertainties (absolute_sigma=True)  A good test is to set this False and make
#   certain it doesn't dramatically change your results, which would be evidence that
#   either your uncertainties are inaccurate or the model is poor.

def func(x, a, b):
    return a*x + b
popt, pcov = curve_fit(func, x, y, sigma=yunc, p0=[1,1],absolute_sigma=True)

# report the fitted values:
print "fitted values:  ", popt

# report the uncertainties, determined from the diagonals of the covariance matrix:
perr = np.sqrt(np.diag(pcov))
print "statistical uncertainties:  ", perr
```

5 Appendix: The Central Limit Theorem

First we'll need to define the characteristic functions of a probability distribution function (PDF), which is simply the expectation value of the complex exponential:

$$\phi(t) \equiv \langle \exp(itx) \rangle$$

this is, equivalently, just the Fourier Transform of the PDF:

$$\langle \exp(itx) \rangle = \int_{-\infty}^{+\infty} dx \, p(x) \exp(itx)$$

The characteristic function of a Gaussian PDF with $\sigma = 1$ and mean value $\mu = 0$ is then:

$$
\begin{aligned}
\phi(t) &= \langle \exp(itx) \rangle \\
&= \int_{-\infty}^{+\infty} dx \, \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \exp(itx) \\
&= \int_{-\infty}^{+\infty} dx \, \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2 - i2tx}{2}\right)
\end{aligned}
$$

And now completing the square by noting:

$$(x - it)^2 = x^2 - i2tx + (it)^2$$

so that the characteristic function is:

$$
\begin{aligned}
\phi(t) &= \int_{-\infty}^{+\infty} dx \, \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x - it)^2 + (it)^2}{2}\right) \\
&= \exp\left(\frac{1}{2}t^2\right) \int_{-\infty}^{+\infty} dx \, \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x - it)^2}{2}\right)
\end{aligned}
$$

where the integral is now simply the integral of a Gaussian distribution (albeit one with a complex mean value) which integrates to simply 1. And so the characteristic function of the Gaussian distribution is:

$$\phi(t) = \exp\left(\frac{1}{2}t^2\right) \tag{33}$$

The proof of the Central Limit Theorem amounts to showing that the characteristic function of a sum of random variables, from any PDF, with finite variance converges to the characteristic function for the Gaussian distribution. Due to Levy's Continuity Theorem, the proof of which we'll leave to the mathematicians, this means that the PDF for the sum of random variables converges to the Gaussian distribution.

We can assume, without losing generality, that we have n independent random variables x_i identically distributed with mean $\mu = 0$ and $\sigma = 1$. We consider the random variable z which is the sum of these random variables:

$$z = \frac{1}{\sqrt{n}} \sum x_i$$

The expectation value for a function of z is therefore an integral over all of the x_i variables each weighted by the PDF:

$$\langle f(z) \rangle_z = \int \left(\prod_i dx_i \, p(x_i) \right) f\left(\frac{1}{\sqrt{n}} \sum x_i\right) \tag{34}$$

The characteristic function of the random variable z is therefore:

$$
\begin{aligned}
\phi(t) &= \langle \exp(itz) \rangle_z \\
&= \int \left(\prod_i dx_i \, p(x_i) \right) \exp \left(\frac{it}{\sqrt{n}} \sum_i x_i \right) \\
&= \int \left(\prod_i dx_i \, p(x_i) \right) \left(\prod_i \exp \left(\frac{itx_i}{\sqrt{n}} \right) \right) \\
&= \prod_i \int dx_i \, p(x_i) \exp \left(\frac{itx_i}{\sqrt{n}} \right) \\
&= \prod_i \langle \exp \left(\frac{itx_i}{\sqrt{n}} \right) \rangle \\
&= \left[\langle \exp \left(\frac{itx}{\sqrt{n}} \right) \rangle \right]^n
\end{aligned}
$$

where in the last step we have used the fact that the x_i are identical independent random variables. Now expanding the exponential as a Taylor series:

$$
\begin{aligned}
\phi(t) &= \left[\langle 1 + \frac{it}{\sqrt{n}} x + \frac{(it)^2}{2n} x^2 + \ldots \rangle \right]^n \\
&= \left[1 + \frac{it}{\sqrt{n}} \langle x \rangle + \frac{(it)^2}{2n} \langle x^2 \rangle \right]^n \\
&= \left[1 - \frac{t^2}{2n} \right]^n
\end{aligned}
$$

Now in the limit $n \to \infty$ this becomes:

$$
\begin{aligned}
\phi(t) &= \lim_{n \to \infty} \left[1 - \frac{t^2}{2n} \right]^n \\
&= \exp \left(\frac{1}{2} t^2 \right)
\end{aligned}
$$

which is the characteristic function from the Gaussian distribution.