

The Statistics of Experimental Uncertainties

Michael Mulhearn

May 11, 2018

1 Statistics of Experiments

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

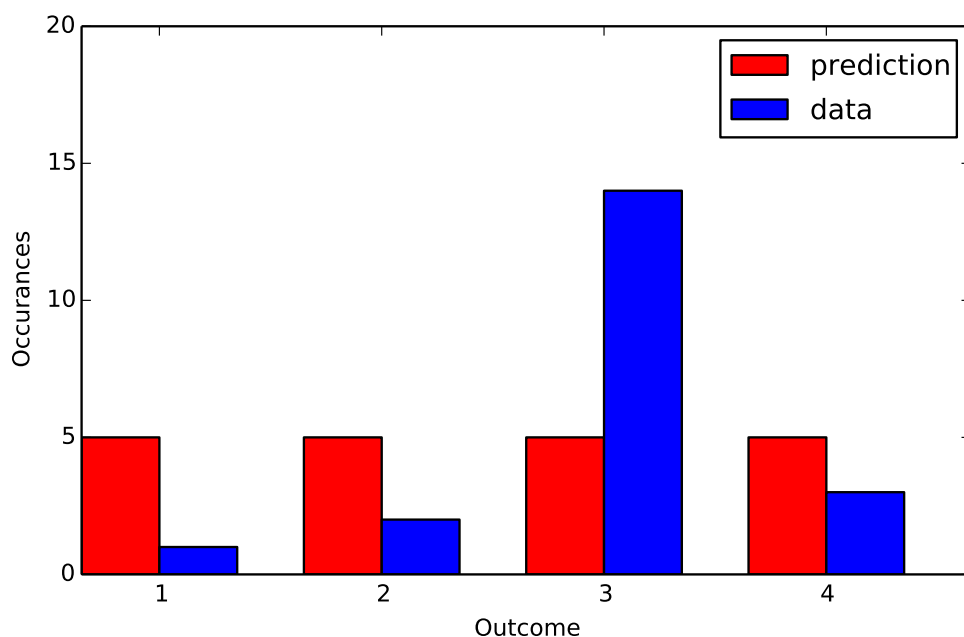


Figure 1: Comparison of experimental results with a prediction.

Scientists almost never display data this way (in bar graphs) 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.

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.

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:

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

where the binomial coefficients are defined by

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

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:

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

which is called the Binomial Distribution.

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:

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

For a continuous probability distribution, we would integrate instead:

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

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:

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

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:

$$\sigma^2 \equiv \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 \quad (7)$$

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.

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:

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

So the mean value is given by:

$$\bar{m} = n\epsilon \quad (8)$$

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:

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

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

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 effect 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 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 $\epsilon = \lambda/n$:

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

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 :

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

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

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:

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

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:

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

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

7 The Gaussian Distribution

Next, we consider the Poisson Distribution in the limit $\lambda \rightarrow \infty$. 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 value 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:

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

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:

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

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 14 yields:

$$\begin{aligned}
P(\delta) &= \frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}} \cdot \frac{1}{\exp \left(\lambda\delta + \lambda\frac{\delta^2}{2} \right)} \\
&= \frac{1}{\sqrt{2\pi\lambda}} \cdot \exp \left(-\lambda\frac{\delta^2}{2} \right)
\end{aligned} \tag{15}$$

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.

8 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. 2.

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>]

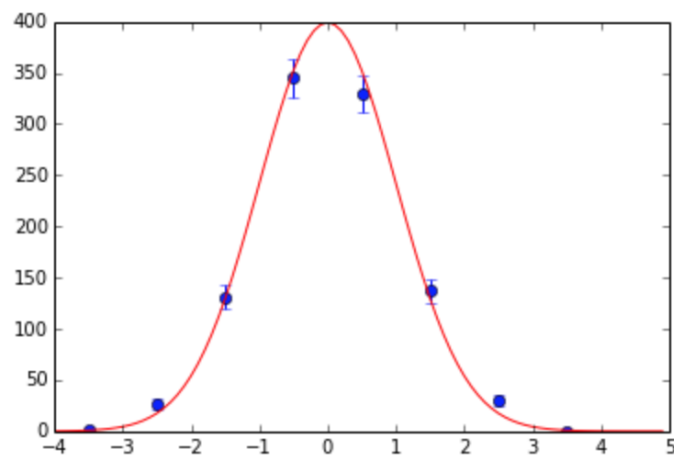


Figure 2: Histogram.

9 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:

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

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 $y = x - \lambda$.)

10 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.

11 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 $\sigma = 1/\text{sqrt}2$. 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 1.

12 Mean and Variance of a Sum

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

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

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

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

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

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

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:

$$\langle f(x) \cdot g(y) \rangle = \langle f(x) \rangle \cdot \langle g(y) \rangle \quad (17)$$

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:

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

13 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

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

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

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

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:

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

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.

14 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:

$$\sigma_y = k\sigma_x$$

15 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:

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

16 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, as has been conclusively demonstrated by many years of running Physics 116. 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 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.

newpage

17 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. 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.

18 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:

$$\phi(t) = \exp\left(\frac{1}{2}t^2\right) \tag{21}$$

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:

$$\langle f(z) \rangle_z = \int \left(\prod_i dx_i p(x_i) \right) f\left(\frac{1}{\sqrt{n}} \sum x_i\right) \tag{22}$$

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:

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

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

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

which is the characteristic function from the Gaussian distribution.