Analysis of Experimental Data

Michael Mulhearn

June 2, 2021

Chapter 1

Statistical Distributions

1.1 Statistics of Experiments

A primary purpose of science is to predict the results of experiments. Consider a simple experiment with five possible outcomes which we repeat ten times. If our theoretical prediction is that each of these five outcomes is equally probable, 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 as in Fig. 1.1. Scientist almost never display experimental data in this

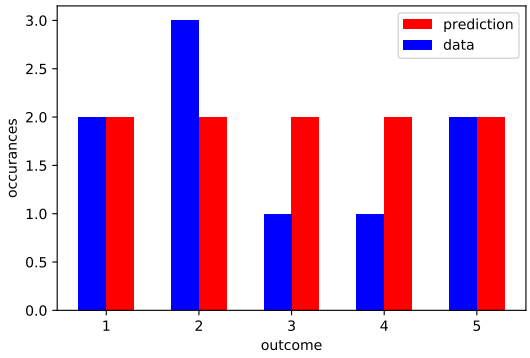


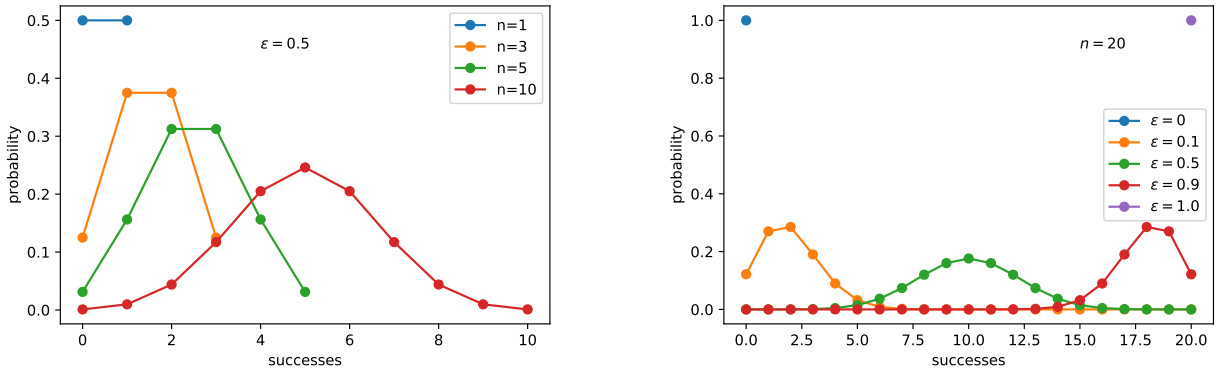
Figure 1.1: Comparison of experimental results with a prediction.

format (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 measured results of individual experiments will vary from 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. The mathematical framework for providing this prediction is called a probability distribution, which reports the probability of each possible outcome for a random variable. 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.



(a) (b)

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

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

$$P = \sum_i p_i$$

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

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

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

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

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

where the binomial coefficients are defined by

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

where the symbol on the left is read as "n choose m". The binomial coefficient 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:

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

which is called the Binomial Distribution.

1.3 The Mean and Variance of a Probability Distribution

Given a probability distribution describing the outcomes of an experiment, our most urgent questions are generally "what is a typical outcome?" and "how widely will outcomes typically vary from one another?".

There are a number of ways to quantify the answer to first question, but generally the most useful answer is the mean value of the distribution, μ, which we calculate as a weighted average across all possible outcomes. For the Binomial distribution, which describes probabilities for outcomes which are natural numbers up to n, we calculate the sum:

$$\mu = \sum_{m=0}^{n} m\, P(m) \tag{1.4}$$

For a continuous probability distribution, where outcomes can be any real value, we integrate instead:

$$\mu = \int_{-\infty}^{+\infty} x\, P(x)\, dx \tag{1.5}$$

across a range appropriate to the situation, typically $[-\infty, \infty]$.

A mode of a distribution is an outcome for which the distribution function takes on a maximal value (in the case of continuous distributions, the requirement is often relaxed to require only a local maximum). A median of a distribution is an outcome at which at most 50% of the probablity is contained above the value and at most 50% of the probablity is contained below the value. Both of these quantities provide information about typical values for a distribution, but they suffer the significant defect that they are not necessarily unique for a given distribution.

The mean is an example of an expectation value. In general, the epectation value of a function $f(x)$ of a random variable x drawn from a probability distribution function $P(x)$ is:

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

For a discrete random variable, the integral becomes a sum as in Eqn. 1.4. Using this notation, we can define the mean as the expectation value of the random variable:

$$\mu \equiv \langle x \rangle$$

To answer the second question, we define an expectation value which characterizes the variation of outcomes from the mean value, which we call the variance (σ^2) of the distribution:

$$\sigma^2 \equiv \langle (x-\mu)^2 \rangle\,.$$

The square root of the variance is referred to as the standard deviation (σ).

There are other ways to quantify the variation from outcome to outcome, but they are generally inferior to the variance in some way. For example $\langle x - \mu \rangle$ can be zero, even for wide distributions, as long as $P(x)$ is symmetric. The quantity $\langle |x-\mu| \rangle$ (called the average deviation) avoids this pitfall but is generally much harder to calculate, due to the absolute value.

The variance, on the other hand, is quite convenient to calculate. For example, in the homework you will show that:

$$\langle (x-\mu)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2\,. \tag{1.6}$$

We need only calculate $\langle x \rangle$ and $\langle x^2 \rangle$ in order to determine both the mean and variance of a distribution.

1.4 Mean and Variance of the Binomial Distribution

The mean value of Binomial Distribution is given by:

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

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

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

We then calculate:

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

and multiplying by p we have:

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

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

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

So the mean value is given by:

$$\mu = n\epsilon \tag{1.7}$$

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

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

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

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

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

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

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

And so:

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

or simply:

$$\sigma^2 = n\,\epsilon\,(1-\epsilon) \tag{1.8}$$

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

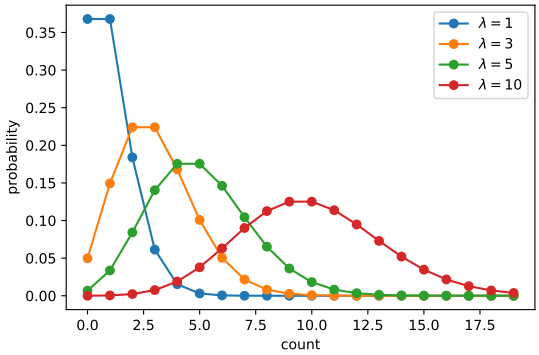


Figure 1.3: The Poisson distribution for several values of parameter λ.

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 is the same in each interval. The probability of an event occuring in each subinterval is $\epsilon = \lambda/n$. If we make n large enough, the probability of observing two events in an interval is vanishing small: $\epsilon^2 = \lambda^2/n^2 \ll \epsilon$. Once cast this way, we can interpret the outcome of the counting experiment as drawn from a Binomial distribution of n trials each with $\epsilon = \lambda/n$:

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

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

$$P(m\ ;\ \lambda) = \frac{\lambda^m}{m!}\, e^{-\lambda} \tag{1.9}$$

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

1.6 Mean and Variance of The Poisson Distribution

The mean of the Poisson distribution is given by:

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

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

$$\begin{aligned} \mu &= e^{-\lambda} \sum_{m\geq 1} \frac{\lambda^m}{(m-1)!} \\ &= \lambda e^{-\lambda} \sum_{m\geq 1} \frac{\lambda^{m-1}}{(m-1)!} \\ &= \lambda e^{-\lambda} \sum_{n\geq 0} \frac{\lambda^n}{n!} \\ &= \lambda e^{-\lambda} e^{\lambda} \\ \mu &= \lambda \end{aligned} \tag{1.10}$$

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

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

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

And so:

$$\begin{aligned} \sigma^2 &= \langle m^2 \rangle - \langle m \rangle^2 \\ &= \lambda\, (\lambda + 1) - \lambda^2 \\ \sigma^2 &= \lambda. \end{aligned} \tag{1.11}$$

That is, the variance of a Poisson distribution is simply the mean. This is probably the single most practical result from statistics. Whenever a measurement amounts to a counting experiment, simply from knowing the outcome of the experiment, the count N, we can estimate that mean value $\mu \sim N$ and also the variance $\sigma^2 = N$. Typical flutuations in a count N are characterized by $\sigma = \sqrt{N}$.

1.7 The Gaussian Distribution

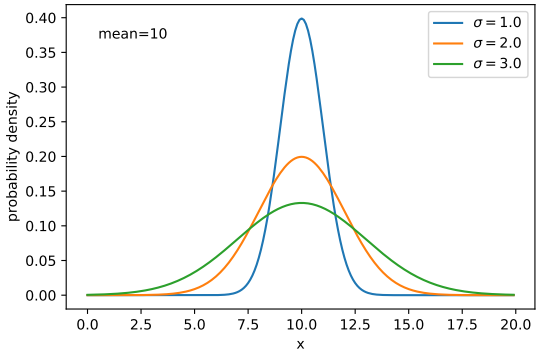


Figure 1.4: The Gaussian distribution for a mean of 10 and several values of parameter σ.

The next distribution we will consider is the Gaussian distribution:

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{1.12}$$

which is plotted in Fig. 1.4. The Gaussian distribution (often called the normal distribution) is the most important and useful distribution as a consequence of the central limit theorem (CLT). The CLT states that when independent random variables drawn from any distribution of finite variance are added together and appropriately normalized (such as by taking their average) the distribution of the resulting random variable approaches the Gaussian distribution as the number of entries in the sum increases to infinity. A proof of the CLT is in the Appendix.

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

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

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

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

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

The Gaussian distribution is normalized so that:

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

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

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

In particular the mean value of x can be calculated by:

$$\langle x\rangle \equiv \int_{-\infty}^{+\infty} xP(x)\,dx$$

and:

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

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

1.8 The Gaussian Distribution as a Limiting Case

As expected from the Central Limit Theorem, the Poisson Distribution in the limit $\lambda \to \infty$ approaches the Gaussian distribution. 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:

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

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:

$$m = \lambda(1+\delta) \tag{1.13}$$

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:

$$\begin{aligned} P(\delta) &= \frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}} \left(\frac{\lambda}{\lambda(1+\delta)} \right)^{\lambda(1+\delta)+\frac{1}{2}} \\ &= \frac{e^{\lambda\delta}}{\sqrt{2\pi\lambda}} \cdot \frac{1}{X} \end{aligned} \tag{1.14}$$

where we define the quantity:

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

which can be approximated as follows:

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

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 1.14 yields:

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

Recalling that:

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

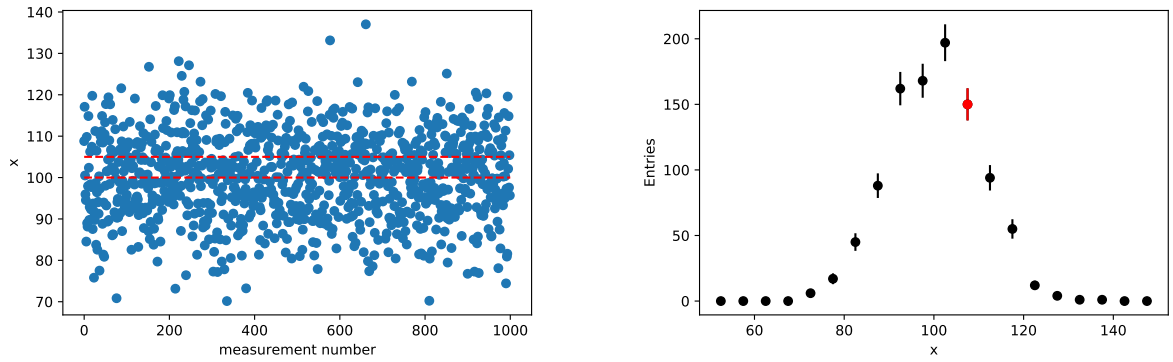
$$P(x) = \frac{1}{\sqrt{2\pi\lambda}} \cdot \exp\left(-\frac{(m-\lambda)^2}{2\lambda}\right) \tag{1.16}$$

Which we can identify as a special case of the Gaussian distribution of Equation 1.12 with $\sigma^2 = \mu = \lambda$ and $x = m$:

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

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

1.9 Histograms in Scientific Python



(a) (b)

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

Suppose a particular variable x is measured 1000 times. One way to visualize the collected data is shown in Fig. 1.5a, which simply plots each measurement value above the measurement number (from 0 to 1000). In this example, the number of measurements that occur within the range from $x = 100$ to $x = 105$ is 181. This count is plotted as the red data point in Fig. 1.5b, 181 entries located above $x = 102.5$, the center of the range. If we repeat this exercise across a number or ranges, the resulting plot in Fig. 1.5b is called a histogram. A histogram reports the number of entries that occur within each of a sequence of consecutive ranges of x-values. Each range considered is called a histogram bin, and the choice of which bins to use is at the discretion of the analyzer.

The content of each bin is a single number, a count. In Section 1.6, an appropriate estimate for typical fluctuation of a count N is shown to be $\sqrt{N}$. It is customary to indicate the size of these typical fluctuations with an "error bar", a line through the data point at N with length $\sqrt{N}$. The error bars provide a visual cue to how much we expect that a repeated identical measurement would vary with respect to the presented data.

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

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

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

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

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

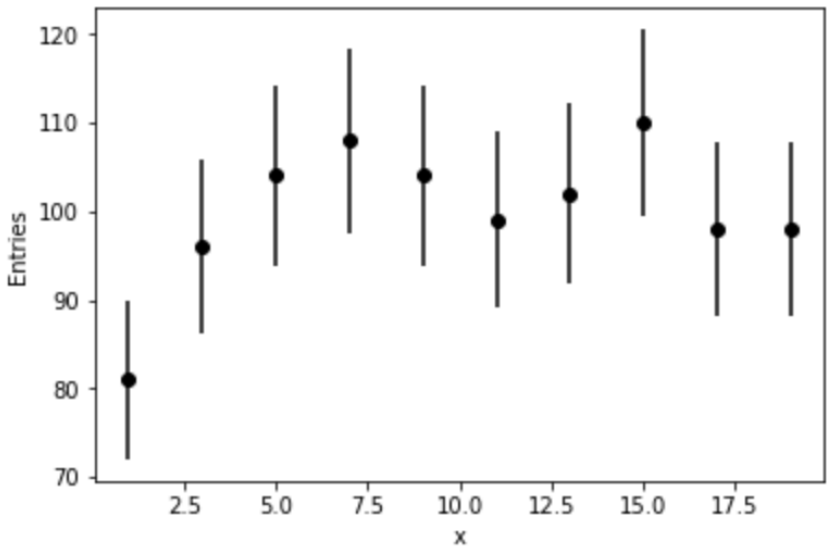


Figure 1.6: Example producing a histogram in Scientific Python.

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

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

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

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

The edges array contains the edges of the bins:

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

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

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

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

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

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

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

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

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

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

1.10 Comparing a Histogram to a Probability Distribution Function

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

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

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

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

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

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

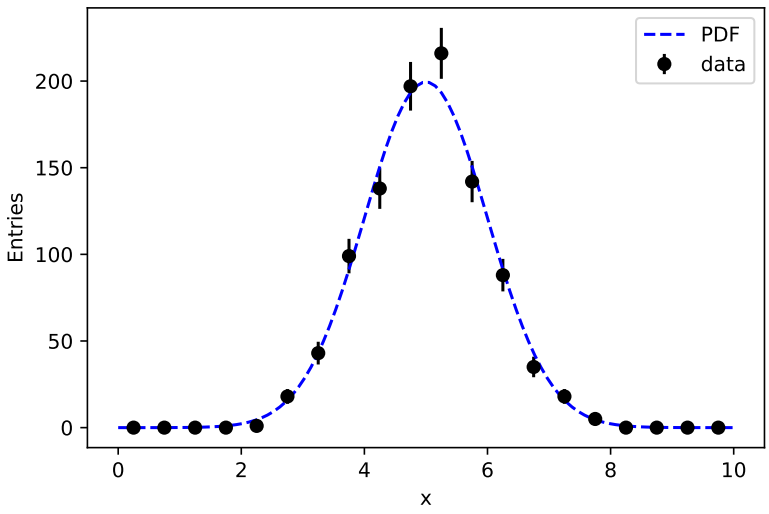


Figure 1.7: The Gaussian PDF scaled to compare to data from a Gaussian distribution. In this case, there are 1000 total entries in the histogram and the bin size is 0.5, for a scale factor of 500.

1.11 Homework Exercises for Distributions

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

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

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

Problem 2: Show that Equation 1.6 is correct.

Problem 3: Show that the Poisson distribution, $P(m)$, in Equation 1.9 is properly normalized:

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

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

Problem 4: Show that the Gaussian distribution, $P(x)$, in Equation 1.12 is properly normalized:

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

Problem 5: Show that the mean of the Gaussian distribution has been correctly identified in Equation 1.12. That is, show explicitly that:

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

Problem 6: Show that the variance of the Gaussian distribution has been correctly identified in Equation 1.12. 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 $y = x - \lambda$.)

Problem 7: Suppose that 300 students are taking introductory physics from a particularly boring professor. Each student has a 50% chance of falling asleep at some point during lecture and staying asleep until the end of class. What is the mean and variance of the number of students asleep at the end of class?

Problem 8: Suppose that in your spare time, you build a muon detector to impress your undergraduate research advisor. On your own, you find that you detect, on average, one cosmic ray muon per minute. What is the probability of observing zero muons when your advisor finally gives you three minutes of their attention? What is the probability of dazzling your advisor by observing six muons during three minutes?

Problem 9: Consider a special probability distribution with the following properties: (A) it is only defined for outcomes 0,1,2,3 and 4 (B) it has two modes at outcomes 0 and 4, (C) all five outcomes are medians. Sketch the probability distribution and determine its mean.

Problem 10: Suppose that two independent experiments provide counts a and b, which we can assume are drawn from a Poisson distribution. Now consider the counting experiment that results from simply adding these counts to obtain $a + b$. Show that that the variance of this sum is given by:

$$\sigma^2(a + b) = \sigma_a^2 + \sigma_b^2$$

or equivalently:

$$\sigma(a + b) = \sqrt{\sigma_a^2 + \sigma_b^2}$$

Hint: use the property of the Poisson distribution that $\sigma^2 = \lambda = \mu$. We'll show in the next chapter that, quite generally, uncertainties (σ) add in quadrature in this way.

Chapter 2

Experimental Uncertainties

2.1 Reporting Experimental Uncertainties

A scientific measurement is meaningless without an associated uncertainty. Without an associated uncertainty, we can't make any meaningful prediction about the results that someone reproducing our measurement would obtain. We also cannot quantitatively compare our measurement to previous measurements or predictions.

We report experimental uncertainties along with measured values. For instance, we might report a measured distance as:

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

and a time as:

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

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

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

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

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

This is because the effect of rounding to one digit can be a 50% effect in this case. Sometimes this rule is extended to the case when the first nonzero digit is a two. For this class, you may assume that we use only one significant digit of uncertainty unless that digit is one, in which case we use two significant digits.

2.2 Interpreting Experimental Uncertainties

We will define experimental uncertainty precisely, but let's start with a conceptual definition. A typical sequence of experimental measurements is shown in Fig. 2.1. Over time, technology improves, science advances, and the uncertainty on the measured value of x improves. The uncertainty on a measurement amounts to a claim about the likely outcome of future experiments with arbitrarily small uncertainty.

We can therefore think of a measurement and it's uncertainty as describing a probability distribution for the future outcome of an experiment in the future with uncertainties that are negligible on the scale of our present uncertainty.

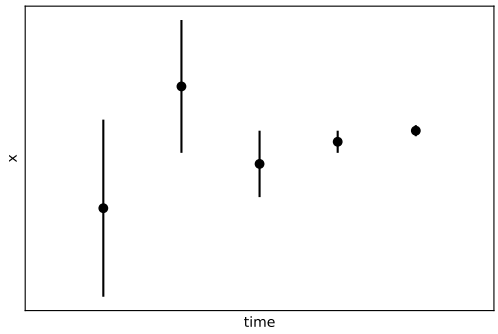


Figure 2.1: The evolution of a measurement of a quantity x over time.

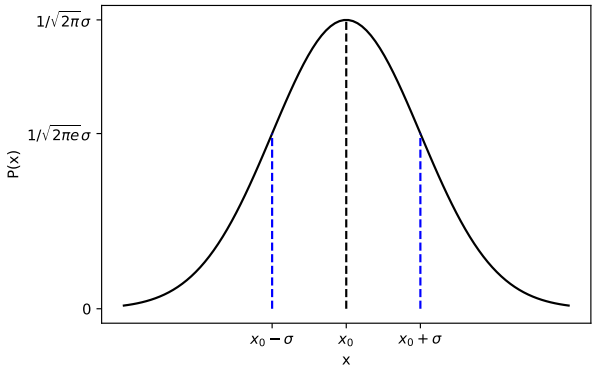


Figure 2.2: The interpretation of experimental uncertainties as the parameters of a Gaussian distribution for the true value.

But which distribution should we use? The Central Limit Theorem informs us that any quantity that results from a suitable average of random uniform variables approaches the Gaussian Distribution. Most measurements, even relatively simple ones, effectively average many different real world factors which randomly influence the result, and so the Central Limit Theorom implies that the Guassian Distribution is often the best distribution to describe an experimental outcome. We define the measured value x as the mean and the experimental uncertainty as the σ of the Gaussian.

2.3 Probabilistic Interpretation of Uncertainty

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:

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

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 $\sigma = 1/\sqrt{2}$. 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 2.1.

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%
$\pm 3\sigma$	$\text{erf}\left(\frac{3}{\sqrt{2}}\right)$	99.7%
$\pm 4\sigma$	$\text{erf}\left(\frac{4}{\sqrt{2}}\right)$	99.993%
$\pm 5\sigma$	$\text{erf}\left(\frac{5}{\sqrt{2}}\right)$	99.99994%

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

In practice, scientist often compare measurement in terms of σ. Measured values that differs from a prediction by one or even two σ are considered to be consistent, because nearly 5% of

our measurements will be off by two σ or more. A three sigma discrepancy, which occurs with a probabilty of 0.3% is starting to get interesting: either someone has likely made a mistake (the theorist or the experimentalist) or something new and exciting is happening, or both. In my field of particle physics, we call three σ effects "initial evidence", but hold off on claiming discovery of a new particle until five σ. This is because if you collect lots of data and do lots of different analyses, it starts to become likely that you will encounter an occassional three σ statistical fluctuation. It would be quite an embarrasment to give away a Nobel prize for a "discovery" that turns out to be merely a statistical fluctuation!

2.4 Independent Random Variables

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 $P(x,y) = P(x)\,Q(y)$ and so

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

And also we have

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

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:

$$\langle f(x) \cdot g(y) \rangle = \langle f(x) \rangle \cdot \langle g(y) \rangle \tag{2.1}$$

Even in two-dimensions, it is easy to see that expectation value of a constant is simply the constant:

$$\begin{aligned}
\langle C \rangle &= \int\int dx\, dy\ C\, P(x)\, Q(y) \\
&= C \int\int dx\, dy\ P(x)\, Q(y) \\
&= C
\end{aligned}$$

Combined with the previous results, we can see that constants move through expectation values intuitively:

$$\langle f(x,y) + C \rangle = \langle f(x,y) \rangle + C \tag{2.2}$$

and:

$$\langle C f(x,y) \rangle = C \langle f(x,y) \rangle \tag{2.3}$$

In the homework, you will show that for a constant C:

$$\sigma^2(Cx) = C^2 \sigma^2(x) \tag{2.4}$$

or equivalently:

$$\sigma(Cx) = C \sigma(x).$$

We can also 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 means. For simplicity let's assume that $\langle x \rangle = \langle y \rangle = 0$, and so:

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

from which it follows that:

$$\begin{aligned} \sigma^2(x+y) &= \langle (x+y)^2 \rangle - (\langle x+y \rangle)^2 \\ &= \langle (x+y)^2 \rangle \\ &= \langle x^2 \rangle + \langle y^2 \rangle \\ \sigma^2(x+y) &= \sigma_x^2 + \sigma_y^2 \end{aligned} \tag{2.5}$$

In the exercises you will show that this result persists even when $\langle x \rangle$ and $\langle y \rangle$ are non-zero. We say that independent uncertainties add in quadrature, that is:

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

2.5 Repeated Experiments

Suppose that we have an apparatus which can measure a quantity x with uncertainty σ_x, and that we repeat a series of N measurements $x_1, x_2, x_3, \ldots, x_N$. Each x_i is a random variable drawn from a distribution with variance σ_x and mean μ.

A useful quantity to calculate from our N measurements is the sample mean:

$$\bar{x} \equiv \frac{1}{N} \sum_{i=1}^{N} x_i = \frac{x_1 + x_2 + x_3 + \ldots + x_N}{N}$$

Just as each of our individual measurements have mean value:

$$\langle x_i \rangle = \mu$$

the sample mean also has mean value:

$$\langle \bar{x} \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle x_i \rangle = \frac{1}{N} \sum_{i=1}^{N} \mu = \mu$$

The uncertainty on the sample mean can be calculated using the results of the previous sections. Using Equations 2.5 and 2.4 we see that:

$$\begin{aligned} \sigma^2(\bar{x}) &= \sigma^2 \left(\frac{1}{N} \sum_{i=1}^{N} x_i \right) \\ &= \sum_{i=1}^{N} \sigma^2 \left(\frac{x_i}{N} \right) \\ &= \frac{1}{N^2} \sum_{i=1}^{N} \sigma^2 (x_i) \end{aligned}$$

Now we need only note that $\sigma^2(x_i) = \sigma_x^2$ by definition, and complete the sum:

ext, we move the constant $1/N$ outside the sum:

$$\begin{aligned} \sigma^2(\bar{x}) &= \frac{1}{N^2} \sum_{i=1}^{N} \sigma_x^2 \\ &= \frac{\sigma_x^2}{N} \\ \sigma(\bar{x}) &= \frac{\sigma_x}{\sqrt{N}} \end{aligned}$$

This shows that the sample mean measures the same mean value μ but with an uncertainty that is reduced by a factor $1/\sqrt{N}$. Taking additional data reduces the uncertainty of a measurement.

2.6 General Propagation of Uncertainties

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

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

$$f(x, y) \sim f(x_0, y_0) + \left. \frac{\partial f}{\partial x} \right|_{x_0, y_0} (x - x_0) + \left. \frac{\partial f}{dy} \right|_{x_0, y_0} (y - y_0)$$

which we can write as:

$$f(x, y) \sim Ax + By + C$$

where the constants are:

$$\begin{aligned} A &= \left.\frac{\partial f}{\partial x}\right|_{x_0,y_0} \\ B &= \left.\frac{\partial f}{dy}\right|_{x_0,y_0} \\ C &= f(x_0,y_0) - Ax_0 - By_0 \end{aligned}$$

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

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

And plugging in the constants:

$$\sigma_f^2 = \left(\left.\frac{\partial f}{\partial x}\right|_{x_0y_0}\right)^2 \sigma_x^2 + \left(\left.\frac{\partial f}{dy}\right|_{x_0y_0}\right)^2 \sigma_y^2 \tag{2.6}$$

This is a plausible result. As x moves by a typical amount σ_x from the value x_0, we expect f to change by approximately:

$$\Delta f = \left.\frac{\partial f}{\partial x}\right|_{x_0y_0} \sigma_x$$

The result we obtain is this amount of variation added in quadrature with a similar variation due to variable y.

This result was obtained by the Taylor series approximation, and is therefore only appropriate when the uncertainties are small enough that the function f is not changing dramatically as x and y vary by σ_x and σ_y. This is usually the case for a well designed experiment. A poorly designed experiment that determines a quantity f which is varying dramatically, from poorly measured quantities, would likley benefit from a redesign, for instance by measuring the dramatically varying quantity f more directly! A typical case where this approximation is problematic is when calculating recipocals of quantities which are within a few sigma of zero.

2.7 Systematic Uncertainties

If by repeating an experiment N times we can reduce the statistical uncertainty by a factor of $\sqrt{N}$, it might seem that we can reach any desired level of experimental uncertainty simply by repeating an experiment many times. Unfortunately, statistical uncertainty is just one component of experimental uncertainty. There is also systematic uncertainty.

Our derivation of the Binomial, Poisson, and Gaussian distributions all shared a common assumption: that each outcome is independent of previous outcomes. But real measurements inevitably contain flaws which introduce a **constant** unknown bias relative to future measurements. We describe the expected size of this bias as an additional uncertainty called the systematic uncertainty. Since this bias is constant, and so not subject to statistical fluctuations, repeating an experiment does nothing to reduce systematic uncertainties.

The only way to reduce systematic uncertainties is to design a better experiment. Generally in the early stages, experiments are statistically limited, and taking more data helps reduce the overall uncertainty. But at a later stage, the experiment becomes systematically 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. Determining and minimizing systematic uncertainties is one of the greatest challenges that an experimenter faces.

Often scientific results report a measurement with both the statistical and a systematic uncertainty, e.g.:

$$x = 1.21 \pm 0.03(\text{stat}) \pm 0.05(\text{syst})$$

Systematic uncertainties and statistical uncertainties are considered to be independent, so the total experimental uncertainty is their sum in quadrature:

$$\sigma_{\text{total}} = \sqrt{\sigma_{\text{syst}}^2 + \sigma_{\text{stat}}^2}.$$

Often, systematic uncertainties are the result of something outside the experimenters direct control. One common example is when a measurement depends on a measured value determined by another experiment. For example, suppose you determine a position x from your measurement of a time t by the relation $x = vt$. But you do not measure the velocity v yourself and instead rely on a previous measurement of v with total uncertainty σ_v. In this case, by standard propagation of uncertainties, we find that:

$$\frac{\sigma_x}{x} = \frac{\sigma_v}{v}.$$

So a 10% uncertainty on the value v would lead to a 10% systematic uncertainty on our measured value x.

Another common source of systematic uncertainties is when a measured value depends on something outside the experimenters direct control. These types of uncertainties are usually estimated by determining how much the uncontrolled quantity is likely to vary, and the effect this variation has on the measured value. For example, suppose a measurement was sensitive to the amount of light that leaks into the apparatus, which is different depending on the weather and time of day. In this case, you might estimate the systematic by comparing results of measurements made at noon to similar measurements made at midnight.

Apart from building a better experiment, the most common way to improve systematics is by calibration. You could eliminate the systematic due to the velocity v in the first example by measuring the relationship between x and t yourself. In the light leakage example, you could record the time each measurement was made, and correct for the effect of light leakage.

2.8 Exercises for Uncertainties

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

Problem 2: Show that Equation 2.1 is valid.

Problem 3: Show that Equation 2.4 is valid.

Problem 4: Show that Equation 2.5 is generally valid by showing:

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

Problem 5: 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 6: Suppose you measure the system gain (i.e. the increase in output voltage for each electron) 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 7: 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 room 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 8: 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 crazy people 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 and associating it to the wrong quantity! What we should be doing is comparing our measured value n to a prediction λ with $\sigma = \sqrt{\lambda}$.

Let's see how much these two different estimates of the uncertainty typically vary. For $\lambda = 10$, calculate typical values for one sigma variation, that is, calculate $n_1 = \lambda + \sqrt{\lambda}$ and $n_2 = \lambda - \sqrt{\lambda}$. It's OK that these will be non-integer values, but you can round them off to the nearest integer if this bothers you. Next calculate the uncertainty you would estimate from the measured values n_1 and n_2, that is $\sqrt{n_1}$ and $\sqrt{n_2}$ and compare to $\sqrt{\lambda}$. Then repeat for $\lambda = 100$.

Our estimate of the uncertainty as $\sqrt{n}$ is indeed biased with respect $\sqrt{\lambda}$. However, this approach to determining the uncertainty of our data avoids a much more important bias, because it can be determined directy from the collected data, independently of any theoretical prediction.

Problem 9: In particle physics, we always need to understand the efficiency (E) 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 $E = m/n$. 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 $\sigma_E = \sigma_m/n = \sqrt{m}/n$. 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 we can interpret as a claim that there is 16% chance of a future measurement being above 1.05, which is clearly impossible!

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 binomial success rate $\epsilon = E = m/n$. For $n = 100$ and $m = 95$, calculate the uncertainty σ on the value E from the variance of the Binomial distribution. Check that the efficiency E plus one sigma is now safely below one. This gives a nice intuitive explanation for the factors ϵ and $1 - \epsilon$ in the variance of the binomial distribution: they cause the variance to varnish as ϵ aproaches either zero or one.

Problem 10: (Optional and Challenging) There's another way to solve the previous 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 $\sigma_f = \sqrt{f}$. With some work, you should be able to reproduce the same result for the uncertainty as that obtained by assuming a binomial distribution.

Chapter 3

Statistical Analysis

3.1 Sample Mean and Sample Variance

Let's return to the case of N repeated measurements of the same quantity x. We can interpret our experiment as drawing random variables from a distribution with mean μ and variance σ^2, with our experimental uncertainty identified as σ, the square root of the variance.

The quantities μ and σ^2 are often referred to as the *population mean* and *population variance*, because they are associated with the entire (possibly infinite) population, from which our finite and relatively small data sample is drawn. Throughout these lecture notes, I have mostly used the alternative terms mean of the distribution and variance of the distribution.

One goal of the experimenter is to determine the best estimate for the value μ using our limited sample of N values of x. We've seen that we can estimate μ using the sample mean:

$$\bar{x} = \frac{1}{N}\sum_i x_i$$

which we can think of as an alternative approach to the computing integral:

$$\mu = \langle x \rangle = \int_{-\infty}^{+\infty} xP(x)dx$$

The experimental outcomes are drawn from P(x), so the average of these outcomes approximates the integral.

In general, the experiment can estimate any

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

as

$$\bar{f} = \frac{1}{N}\sum_i f(x_i).$$

An essential experimental technique is the ability to estimate uncertainty from repeated measurements. Even if you know the uncertainty of a measurement, this technique is valuable as a means of validating your experiment. We can estimate the variance of the distribution

$$\sigma^2 = \langle (x-\mu)^2 \rangle \sim \langle (x-\bar{x})^2 \rangle$$

from our limited sample of N values of x as:

$$s_N^2 = \frac{1}{N}\sum_i (x_i - \bar{x})^2.$$

The quantity s_n^2 is called the sample variance. You will show in the exercises that a better estimate for σ^2 is provided by the *unbiased* sample variance:

$$s^2 = \frac{1}{N-1}\sum_i (x_i - \bar{x})^2. \tag{3.1}$$

This form corrects for the bias introduced by using the estimate $\bar{x}$ in place of μ. The quantity s_n is called the uncorrected sample standard deviation and the quantity s is called the corrected sample standard deviation. Note that while s^2 is an unbiased estimate for σ^2, the value s remains a biased estimate for σ, despite the correction.

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

$$\{a_1, a_2, \ldots, a_n\} \equiv a_i$$

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

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

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

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

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 to get rid of it. We obtain:

$$-2\ln\mathcal{L} = \sum_i \frac{(a_i - x_i)^2}{\sigma_i^2} + 2\ln(\sqrt{2\pi}\sigma_i) \tag{3.2}$$

Assuming the experimental uncertainties, σ_i, are known, the second term is simply a constant of the experiment setup. The first term is referred to as the χ^2 test, or even just χ^2 ("chi-squared"):

$$\chi^2 \equiv \sum_i \frac{(a_i - x_i)^2}{\sigma_i^2} \tag{3.3}$$

A small value of χ^2 means that the result is close to the theoretical prediction and a large value means that the result is unlikely to have occurred as a result of the prediction. If the uncertainties and prediction are all correct, then we expect each outcome x_i to be a random variable drawn from

a distribution with mean value a_i and variance σ_i^2. We can calculate the expected value for the χ^2 as:

$$\begin{aligned}\langle\chi^2\rangle &= \langle\sum_i \frac{(a_i - x_i)^2}{\sigma_i^2}\rangle \\ &= \sum_i \frac{\langle(a_i - x_i)^2\rangle}{\sigma_i^2} \\ &= \sum_i \frac{\sigma_i^2}{\sigma_i^2} \\ &= \sum_i 1 \\ &= N\end{aligned}$$

This interpretation assumes that the prediction a_i is exactly the mean of the distribution, and is not biased by our experimental results. That is, we have not modified the prediction in any way to fit our results. We will revisit this assumption later.

3.3 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:

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

of the same quantity x, each with the same uncertainty σ_x, and we would like to find out which value is most consistent with these N measurements. Our χ^2 test for this simple model is then:

$$\chi^2 = \sum_i \frac{(a - x_i)^2}{\sigma_x^2}.$$

We can find the particular value of a, called a_0, that maximizes the likelihood by finding the minimum of the χ^2, because $\chi^2 \sim -2\ln\mathcal{L}$. We find the minimum from the condition:

$$\left.\frac{\partial\chi^2}{\partial a}\right|_{a_0} = 0,$$

which amounts to:

$$\begin{aligned}0 &= 2\sum_i \frac{(a_0 - x_i)}{\sigma_x^2} \\ 0 &= \left(a_0\sum_i\right) - \left(\sum_i x_i\right) \\ a_0 &= \frac{1}{N}\sum_i x_i = \bar{x} \qquad (3.4)\end{aligned}$$

which is the sample mean $\bar{x}$ for our N measurements. This exercise shows that the sample mean is in fact our best estimate for the mean of the distribution which the measurements x_i are drawn from.

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

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

and to determine the best fit values a_0 and b_0 we would require:

$$\left.\frac{\partial \chi^2}{\partial a}\right|_{a_0, b_0} = 0$$
$$\left.\frac{\partial \chi^2}{\partial b}\right|_{a_0, b_0} = 0$$

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

3.4 Interpretation of $\Delta\chi^2$

Suppose that 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.

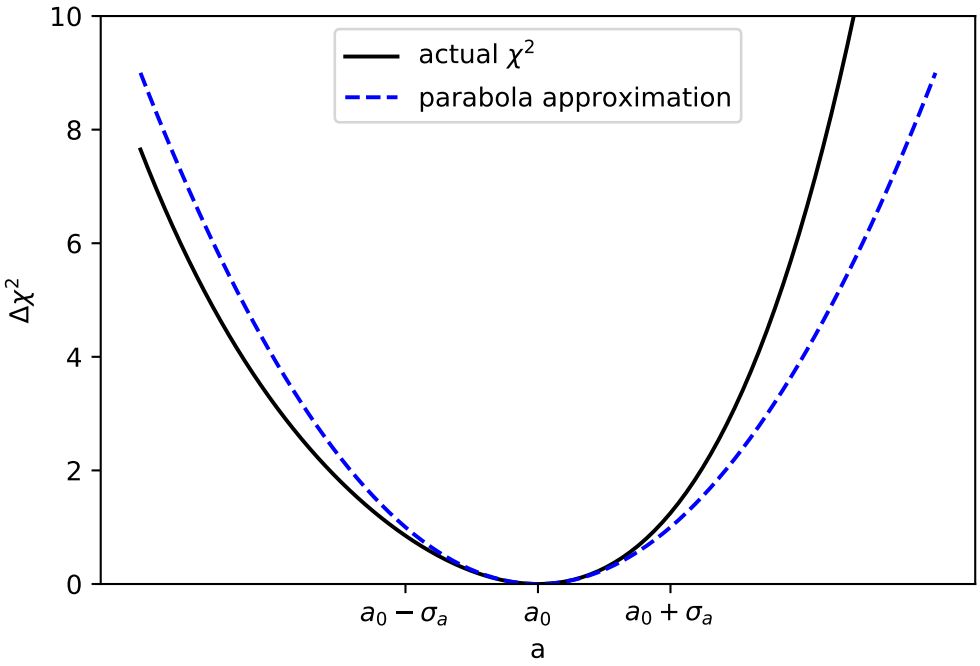


Figure 3.1: Approximation of a χ^2 by a parabola near the minimum.

One approach is to simply apply propagation of uncertainties to the formula determined by minimizing the χ^2, for instance, Equation 3.5. But this approach can be tedious or even unusable

in cases where the χ^2 is minimized numerically and no closed form solution is available. It is well worthwhile, therefore, to consider an alternative approach, that determines these uncertainties directly from the Likelihood and its corresponding χ^2 distribution.

Consider the Likelihood associated with this series of measurements:

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

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

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

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

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

We see that:

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

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

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

The χ^2 for a simple Guassian likelihood is a parabola, and so the second derivative *evaluated anywhere* yields a term containing the uncertainty σ_a. However, as illustrated in Fig. 3.1, the region near the minimum of an any χ^2 distribution can be approximated as a parabola about the minimum, with a second derivative set equal to that of the χ^2. And so for any χ^2 distribution, we have:

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

The uncertainty on any parameter can be estimated from the curvature of the χ^2 at the minimum.

3.5 Uncertainty on the Sample Mean

Let's consider again a set of repeated measurements of the same quantity each with the same uncertainty σ_x:

$$\{x_1 \pm \sigma_x, x_2 \pm \sigma_x, ..., x_N \pm \sigma_x\} \equiv x_i \pm \sigma_x$$

Earlier we used the χ^2 test to determine the best fit constant value a_0 for these measurements is the sample mean:

$$a_0 = \bar{x} = \frac{1}{N}\sum_i x_i.$$

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

$$\begin{aligned} \chi^2 &= \sum_i \frac{(a - x_i)^2}{\sigma_x^2} \\ \frac{d^2\chi^2}{da^2} &= \sum_i \frac{2}{\sigma_x^2} \\ &= \frac{2N}{\sigma_x^2} \end{aligned}$$

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

$$\sigma_{\bar{x}}^2 = \sigma_a^2 = \frac{2}{\frac{d^2\chi^2}{dX^2}\Big|_{a_0}} = \sigma_x^2/N$$

a result we obtained earlier by propagating uncertainties.

3.6 Weighted Mean

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

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

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

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

And the minimum value of χ^2 occurs at:

$$\frac{d\chi^2}{da}\Big|_{a_0} = 0$$

and so

$$\begin{aligned} 0 &= \sum_i \frac{2\,(a_0 - x_i)}{\sigma_i^2} = 0 \\ 0 &= \sum_i \frac{x_i}{\sigma_i^2} - a_0 \sum_i \frac{1}{\sigma_i^2} \\ a_0 &= \sum_i w_i\, x_i \end{aligned}$$

where

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

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

$$\left.\frac{d^2\chi^2}{da^2}\right|_{a_0} = \sum_i \frac{2}{\sigma_i^2}$$

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

$$\sigma_a^2 = \frac{2}{\left.\frac{d^2\chi^2}{da^2}\right|_{a_0}} = \left(\sum_i \frac{1}{\sigma_i^2}\right)^{-1}$$

3.7 Degrees of Freedom

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

It's instructive to consider some extreme cases. Consider the case that we make a single measurement $x_1 \pm \sigma_i$. In this case, the best fit is simply $a_0 = x_1$ and the χ^2 is zero. Suppose we make two measurements. In this case, the best fit line $y = a\,x + b$ will pass through the two points, and again we will have $\chi^2 = 0$. Quite in general, we can always fit N data points exactly with N parameters. When we discuss the typical size of chi^2, the important quanity is the number of degrees of freedom (NDF), which is the number of measurements minus the number of fit parameters. The χ^2 for a function that has been *fitted* to N data points should have:

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

3.8 The Best Estimate for σ

So far we have presumed that we know the uncertainty associated with our experimental measurements. Suppose we do not know out experimental uncertainty, and wish to estimate it from the data, just as we do for parameters. Minimizing χ^2 appears to be of no help here, because we can make σ as large as we want to minimize χ^2.

This is because the χ^2 term is just one term in the log likelihood:

$$-2\ln\mathcal{L} = \sum_i \left(\frac{(a_i - x_i)^2}{\sigma^2} + 2\ln(\sqrt{2\pi}\sigma)\right)$$

We've been able to ignore the second term so far under the assumption that the experimental uncertainties are constant terms in the likelihood, not parameters to minimize!

To find the values of σ which maximizes the likelihood, we need differenentiate the entire expression with respect to σ, and set it to zero:

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

and solving for σ:

$$\sigma^2 = \frac{\sum_i (a_i - x_i)^2}{N} \tag{3.8}$$

This analysis assumes we have not found the best fit values for the predictions a_i. To adapt this result to the case of fitted parameters, we simply replace N with the degrees of freedom:

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

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

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

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

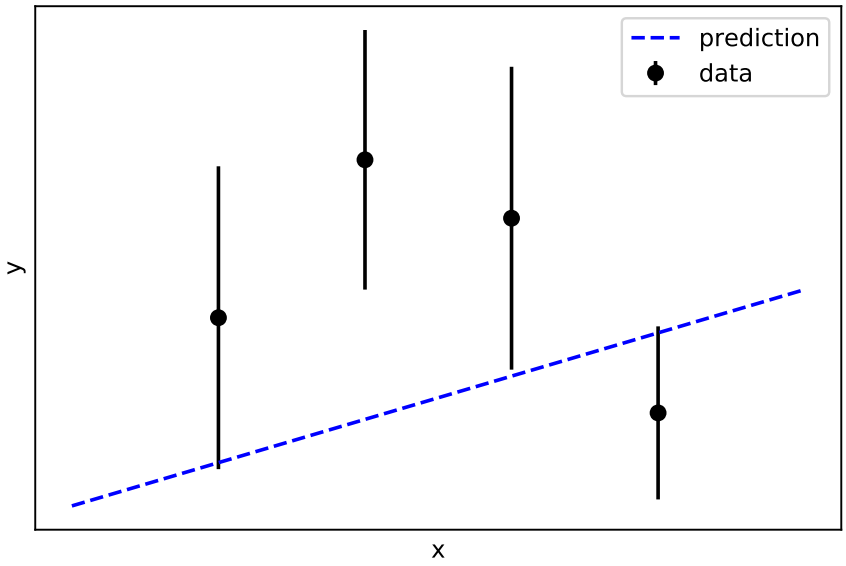
3.9 Homework Exercises for Statistical Analysis

Problem 1: Suppose you collect N measurements of $y_i \pm \sigma$ each at angle θ_i. You would like to fit your data to the model:

$$y = A \sin \theta$$

where A is an unknown parameter. Use the χ^2 technique to derive a formula for the best fit value of A and its uncertainty.

Problem 2: Estimate the χ^2 for the following data with respect to the prediction:



Problem 3: Sketch your own prediction for a relationship between two variables x and y. Make it somewhat complicated. Draw typical experimental data which might result from an experiment which is consistent with the prediction. (No calculation needed!)

Problem 4: Repeat Problem 3, but for an experiment that has a large discrepancy with the prediction.

Problem 5: Suppose that you measure the same quantity x a total N times, with the same statistical experimental uncertainty σ for each measurement. Suppose you do not know the value of the experimental uncertainty σ. First show that the best estimate for the mean value of x is the sample mean $\bar{x}$, independent of the value of σ. Next, use the procedure of Section 3.8 to show that the best estimate for σ^2 is given by Eqn. 3.1. Last, calculate the statistical uncertainty of the quanitity $\bar{x}$.

Problem 6: (Optional and Challenging) The sample variance S_N^2 provides a biased estimate for

the variance σ^2 of the distribution. Show explicitly that:

$$\langle S_N^2 \rangle = \frac{N-1}{N}\,\sigma^2.$$

One way to approach this problem is to show that

$$\langle S_N^2 \rangle = \langle x^2 \rangle - \langle \bar{x}^2 \rangle$$

in contrast to what you might have been tempted to assume:

$$\langle S_N^2 \rangle = \langle x^2 \rangle - \langle \bar{x} \rangle^2 \quad \text{(incorrect)}$$

From there, it follows pretty easily from

$$\langle \bar{x}^2 \rangle = \sigma^2/N + \mu^2$$

which itself follows from our result for repeated measurements:

$$\sigma(\bar{x}^2) = \sigma^2/N.$$

I wish it were true that this exceptionally clever correction was also extremely useful... but the reality is that we usually just make sure N is large enough that it hardly matters!

Appendix A

Practice Problems

A.1 Experimental Uncertainties

1. You make four independent measurements of the same quantity x:

$$\{x_i\} = \{5, 6, 4, 5\}$$

what is your best estimate for the value of x? If the uncertainty on each measurement of x was $\sigma = 1$, what is the uncertainty on your best estimate of x?

2. You make two independent measurements of the same quantity x. The first measurement was $x = 5 \pm \sqrt{2}$ and the second was $x = 8 \pm 1$. What is your best estimate for the value of x?

3. You make 100 independent measurements of the quantity x and find that your collected data has a mean of 23.2341 and a variance of 25. What is your best estimate for x and the uncertainty on this estimate?

4. You measure the speed of light v by measuring a time t and distance x according to the formula:

$$v = x/t$$

Derive an expression for fractional uncertainty $\sigma(v)/v$ in terms of the fractional uncertainties of x and t.

5. The relationship between the kinetic energy of a particle K and a measured voltage V has been calibrated as:

$$K = aV + b.$$

Derive an expression for the uncertainty on the value K from the uncertainties on a, b, and V.

6. Suppose you measure the energy U from a voltage V and a capacitance C from the expression:

$$U = \frac{1}{2}CV^2$$

(a) If the fractional uncertainty on C is 3% and the factional uncertainty on V is 2%, what is the fractional uncertainty on the energy U? (b) Suppose you repeat the measurement of V many times, using the same capacitor C. What is the best uncertainty you could hope to achieve on the energy U?

7. The funding agencies give you one million dollars based on your celebrated measurement of the quantity x which had 12% statistical uncertainty and a 3% systematic uncertainty. You can spend the money on either: (a) repeating the experiment an additional eight times (for a total of 9), or (b) reduce the systematic uncertainty to 1% without repeating the experiment. Which option is the best use of the money?

8. The funding agencies give you ten million dollars based on your controversial measurement of the quantity y which had 1% statistical uncertainty and a 10% systematic uncertainty. You can spend the money on either: (a) repeating the experiment one million times, or (b) reducing the systematic uncertainty from 10% to 8% without repeating the experiment. Which option is the best use of the money?

9. In the US, blood pressure varies by about $\sim 10\%$. A cereal manufacturer conducts a study of $N = 100$ participants who ate their cereal daily for one year. The study found that the average blood pressure of this group was 1 % lower than the average for the US population. Does their claim that "Super Choco Bloxs lowers blood pressure!" have scientific merit?

10. Typical IQs are in the range from $100 \pm 10\%$ in the total US population. A study of 100 college students that regularly drank beer found that the average IQ of this group was 110. Is this result statistically significant? Is there a likely source of systematic uncertainty?

A.2 Answers

1. 5.0 ± 0.5
2. 7
3. 23.2 ± 0.5
4. $$\frac{\sigma_v}{v} = \sqrt{\left(\frac{\sigma_x}{x}\right)^2 + \left(\frac{\sigma_t}{t}\right)^2}$$
5. $\sigma_K = \sqrt{V^2\sigma_a^2 + a^2\sigma_V^2 + \sigma_b^2}$
6. 5%
7. option a
8. option b
9. No.
10. Statistically significant, but major systematic due to comparison of college students to general population.

Appendix B

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:

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

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:

$$\langle f(z) \rangle_z = \int \left(\prod_i dx_i \; p(x_i) \right) \; f\left(\frac{1}{\sqrt{n}} \sum x_i \right) \tag{B.2}$$

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.

Appendix C

Independent 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:

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

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

$$\left(x - \frac{a^2}{a^2+b^2}u\right)^2 = x^2 - \frac{2a^2}{a^2+b^2}ux + \frac{a^4}{(a^2+b^2)^2}u^2 \tag{C.1}$$

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

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

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:

$$\{\}_2 = \sqrt{(2\pi)}\frac{ab}{\sqrt{a^2+b^2}} \tag{C.3}$$

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.

Appendix D

Notation

We write a sum as:

$$\sum_{i=1}^{N} x_i = x_1 + x_2 + ... + x_N$$

when they are clear from context, we leave off the limits and write:

$$\sum_{i} x_i = x_1 + x_2 + ... + x_N.$$

Similarly, we write a product as:

$$\prod_{i=1}^{N} x_i = x_1 \cdot x_2 \cdot ... \cdot x_N$$

Although we avoid doing so in these lecture notes, if you encounter an integral

$$\int f(x)\, dx$$

it most likely refers to a definite integral over the whole range, e.g. e.g. from $-\infty$ to $+\infty$, and not an indefinite integral.

We write the binomial coefficients

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

which is pronounced "n choose m".

For a function of multiple variables $f(x, y)$ we write the partial derivative:

$$\frac{\partial f}{\partial y}$$

which is the derivative of f with respect to y while holding x constant. To indicate a particular value for the derivative at the location x_0, y_0 we write:

$$\left.\frac{\partial f}{\partial y}\right|_{x_0,y_0}$$

which we pronounce as the partial derivative of f with respect to y evaluated at (x_0, y_0).

We write a probability distribution P of a random variable x with parameters a and b as:

$$P(x; a, b)$$

We can ommit the parameters and write $P(x)$ if we do not wish to make them explicit. Although these notes attempt to avoid this, sometimes $P(x)$ and $P(y)$ are considered to be two different functions, one for variable x and one for variable y.

We define the expectation value of a quantity $f(x)$ for a random variable x by the integral:

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

We write the mean of a distribution as:

$$\mu \equiv \langle x \rangle$$

We write the variance (σ^2) of a random variable x as:

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

The square root of the variance is called the standard deviation σ. When it is not clear from context, we can explicitly indicate the random variable in the variation, for example σ_y^2 or $sigma^2(y)$. For example:

$$\sigma^2(f(x)) \equiv \langle (f(x) - \langle f(x) \rangle)^2 \rangle$$

An experimental measurement of a quanity x with value x_0 and uncertainty σ is written as:

$$x_0 \pm \sigma$$

Often, we refer to the measured value x_0 as simply x.

We indicate a set of N experimental measurements of a quantity x by:

$$\{x_i\} = \{x_1, x_2, \ldots, x_N\}$$

We include their uncertainties as:

$$\{x_i \pm \sigma_i\} = \{x_1 \pm \sigma_1, x_2 \pm \sigma_2, \ldots, x_N \pm \sigma_N\}$$

Experimental estimates of a quantity $f(x)$ from N experimental measurements of a quantity x are denoted with the bar notation:

$$\bar{f} = \frac{1}{N} \sum_{i=0}^{N} f(x_i)$$

The sample mean is:

$$\bar{x} = \frac{1}{N} \sum_{i=0}^{N} x_i$$

The sample variance is defined as:

$$s_N^2 = \frac{1}{N} \sum_i (x_i - \bar{x})^2.$$

The unbiased sample variance is defined as:

$$s^2 = \frac{1}{N-1} \sum_i (x_i - \bar{x})^2.$$

The quantity s_n is called the sample standard deviation or the uncorrected sample standard deviation and the quantity s is called the corrected sample standard deviation.