

# Statistics for Me

Maria Mckinley

August 21, 2006

## 1 The PDF and the CDF

Figure 1 is an example of a distribution of the age of death for a week in some city. In Figure 2, we see how the histogram looks more like a normal density as the number of data points increases. The normal curve is plotted over the histogram. If we normalize the curve so the area under the curve is one (by changing the y axis), we have the probability density function (pdf).

The area under the entire curve of a pdf, or the total probability of having *any* observed value, is one.

$$\int_{-\infty}^{\infty} p(x) dx = 1; \quad p(x) \geq 0 \quad (1)$$

To find the probability of picking a person at random, who died, for instance in their 70's, find the area under the curve between 70 and 79. In general terms, the probability of finding a particular random variable  $x$  between  $x_1$  and  $x_1 + \delta x$  is given by finding the area under the curve, between  $x_1$  and  $x_1 + \delta x$ :

$$P(x_1 < x < x_1 + \delta x) = \int_{x_1}^{x_1 + \delta x} p(x) dx \quad (2)$$

As the interval between  $x_1$  and  $x_1 + \delta x$  gets smaller and smaller, the probability can be estimated by the value of  $p$  at  $x_1$  multiplied by  $\delta x$

$$\lim_{\delta x \rightarrow 0} P(x_1 < x < x_1 + \delta x) = p(x_1) \delta x_1 \quad (3)$$

Note, however, that because this is a continuous distribution, the probability of a random variable being exactly equal to  $x$  is zero, because the area under a single number is zero.

The following relationship follows from the above equation, and will be useful later on.

$$p(x_1) = \frac{\partial P(x_0 < x < x_1)}{\partial x_1} \quad (4)$$

If a scientist were to give a measured value of, for instance, the density of water, they give the estimated error as well. If the value given is  $1.1\text{g/cm}^2 \pm$

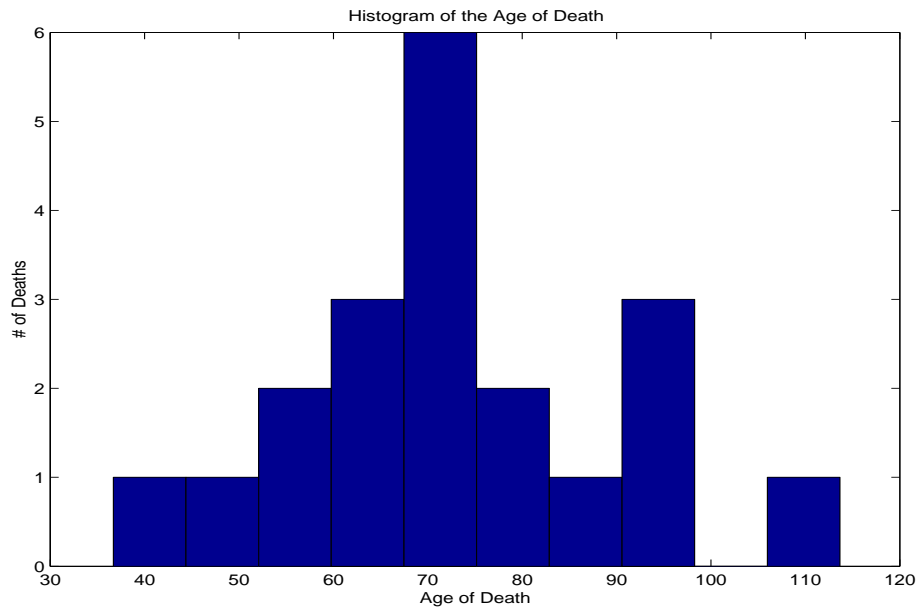


Figure 1: Distribution of numbers representing the age of death for a week in some city.

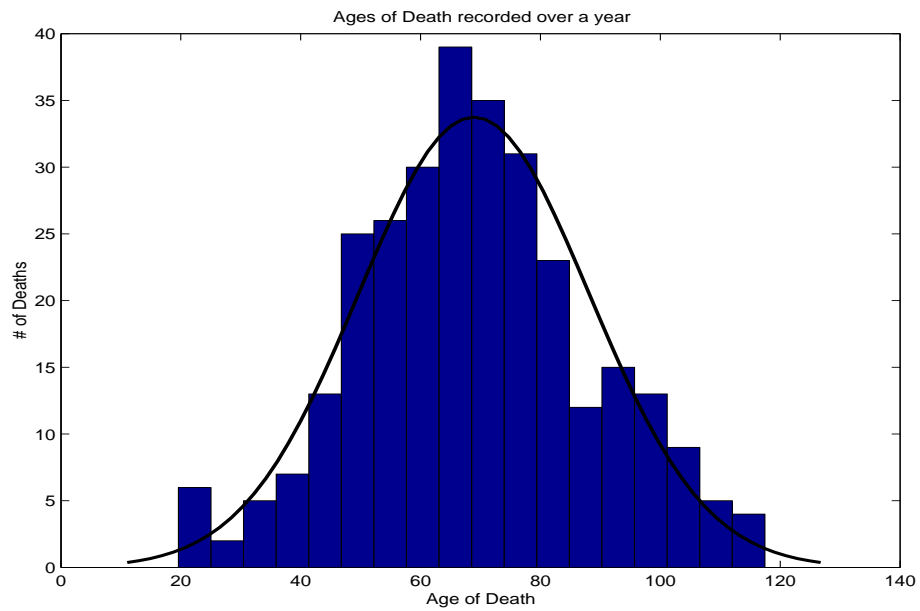


Figure 2: Distribution of probabilities for the ages of death shown in the previous figure.

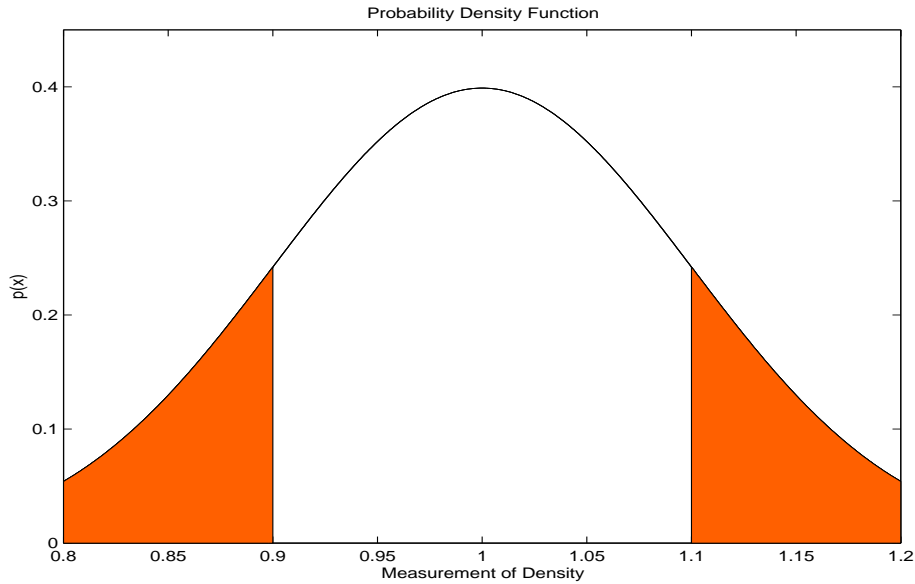


Figure 3: Probability Density Function for a Scientist's Measurement of Density

0.1g/cm<sup>2</sup>, how accurate is this measurement? In other words, what is the probability that they would have measured a value that was larger than 1.1g/cm<sup>2</sup> or smaller than 0.9g/cm<sup>2</sup>? This is the area in the shaded portions of Figure 3.

$$P = 2 \int_{-\infty}^{0.9} p(x) dx \quad (5)$$

When the probability histogram follows the normal curve, it can be summarized by the expected value and standard error. This means that if 1.1 and 0.9 are one standard deviation from the mean, then the probability is approximately 1/3 that the actual value falls in that range.

A related function to the pdf is the cumulative distribution function (cdf). In our first example, if we were interested in finding out the probability of finding someone who died before they turned 60, we would find the area under the curve from 60 to negative infinity. This area is highlighted in Figure 4. As we choose greater ages, or move along the x axis towards infinity, the probability goes from zero, at age zero, and to one at age infinity(? -Everyone will die before they are infinitely old!). This is a description of the cdf.

$$\lim_{x_2 \rightarrow \infty} P(x < x_2) = 1; \quad \lim_{x_2 \rightarrow -\infty, 0} = 0 \quad (6)$$

Figure 5 shows the cdf for the example of the ages of death. The y axis is the probability a random x will be less than,  $x_2$ , whereas in the pdf the area

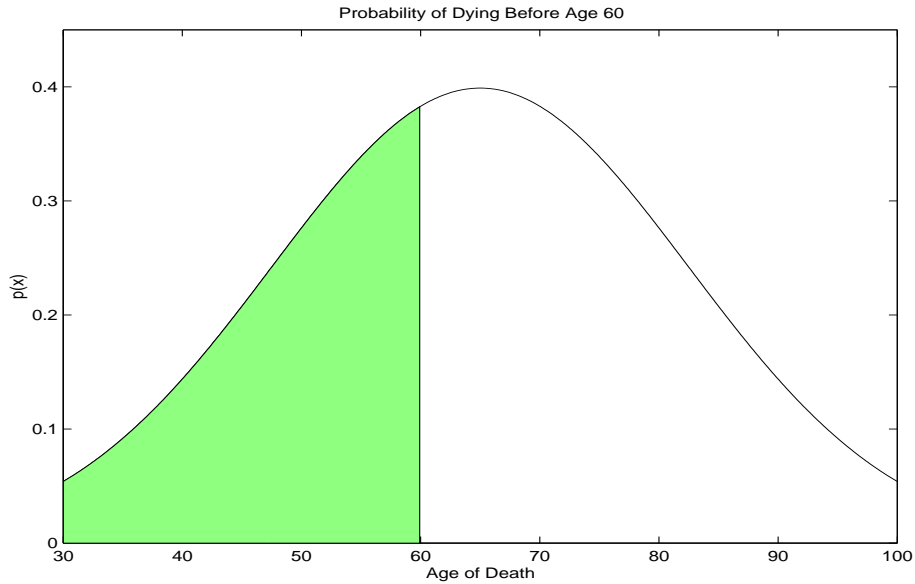


Figure 4: Pdf for Age of Death, probability of dying before age 60 highlighted.

under the curve between  $x$  and  $x_2$  was the probability  $x$  would be between  $x$  and  $x_2$ .

The following equations describe the relationship between the pdf and the cdf:

$$P(x) = \int_{-\infty}^x p(x) dx \quad p(x) = \frac{\partial P}{\partial x} \quad (7)$$

Another distribution we can look at to demonstrate this relationship is the exponential. The pdf is shown in Figure 6. Once again, it follows the definition of the area under the curve being one.

$$\int_0^{\infty} e^{-x} dx = 1 \quad (8)$$

The equation of the cdf is found by integrating  $p(x)$ .

$$P(x) = \int_0^x e^{-\bar{x}} d\bar{x} \quad (9)$$

$$\begin{aligned} &= \int_0^{-x} e^{\bar{u}} d\bar{u}; \quad \text{where } u = -\bar{x} \quad \text{and} \quad d\bar{u} = -d\bar{x} \\ &= -[e^{-x} - e^0] \\ &= 1 - e^{-x} \end{aligned} \quad (10)$$

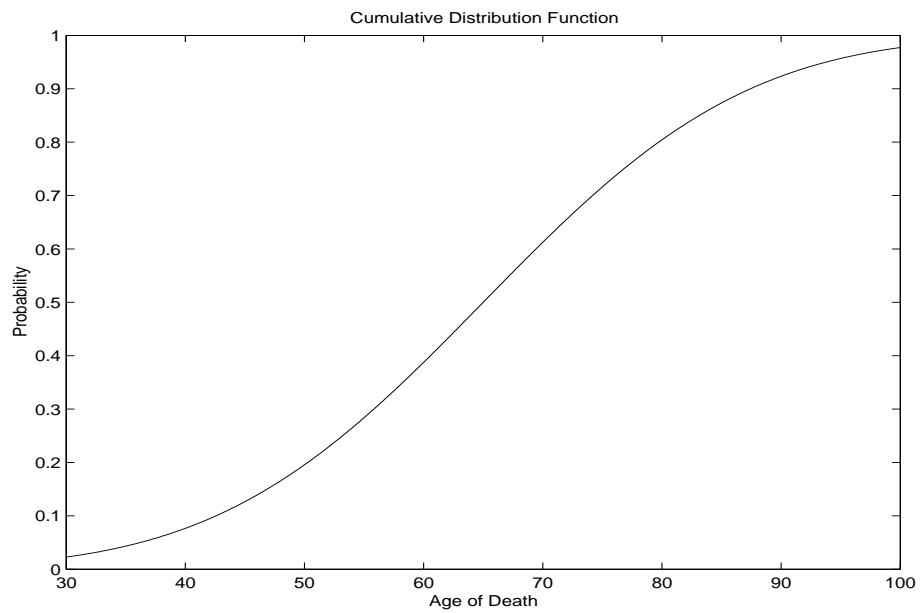


Figure 5: Cdf for Age of Death, probability of  $x$  being equal to, or less than,  $x_2$

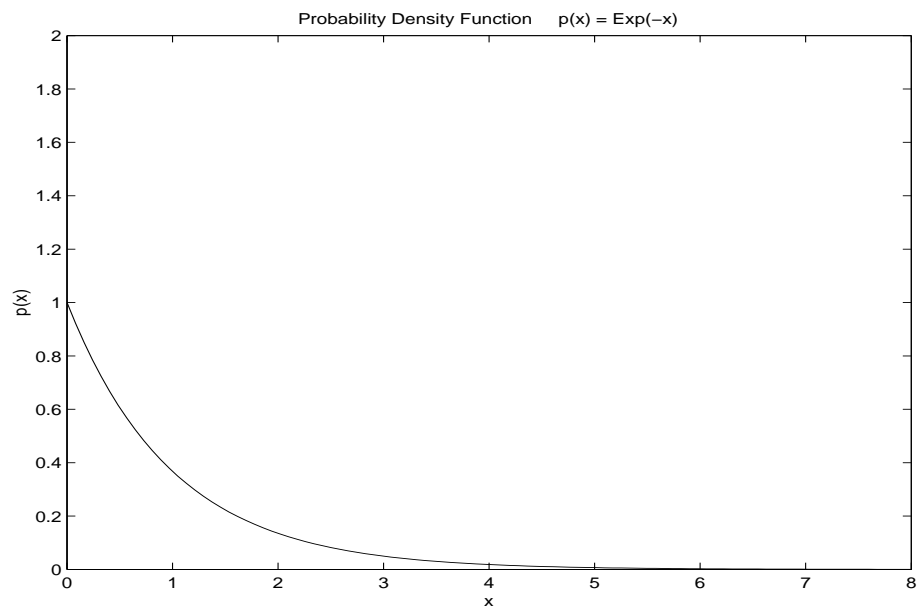


Figure 6: Pdf for the Exponential Distribution

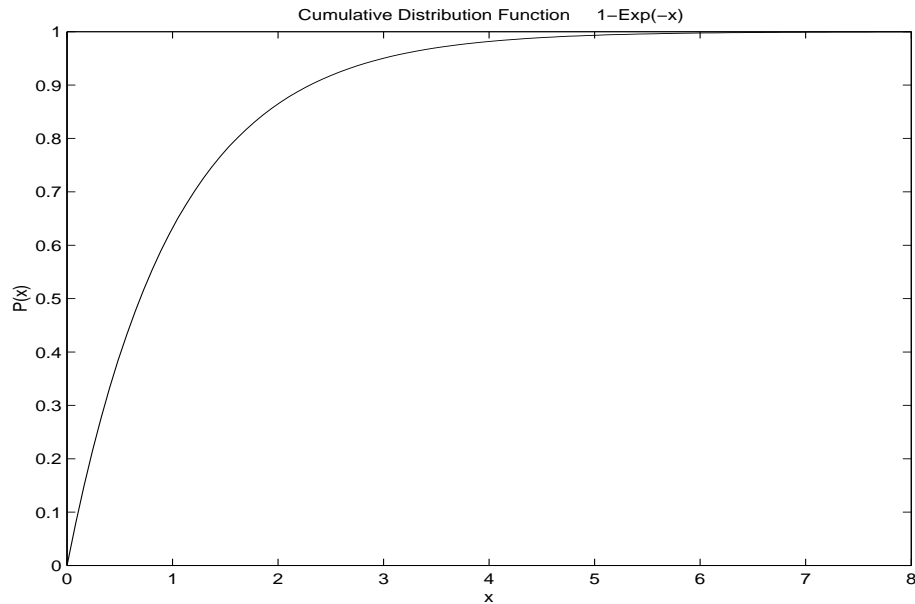


Figure 7: Cdf for the Exponential Distribution

So the equation for the cdf, in this case, is  $u = 1 - e^{-x}$ . The cdf is plotted in Figure 7. If we solve this equation for  $x$ , we get  $x = -\ln(1 - u)$ . We can use this equation to generate a random distribution, and we will do this in section 4.

Let's determine the mean of the exponential distribution. It is possible to calculate the mean without knowing the actual numbers of each value, it is enough to know the proportion of the population at each value. For example, we can calculate the average weight of a group of children, by knowing that 1/3 weigh 40 lbs. and 2/3 weigh 45 lbs.

$$40 \times 1/3 + 45 \times 2/3 = 41.7$$

So the mean can be expressed as:

$$\bar{x} = \sum (x, \text{proportion}) \quad (11)$$

Of course, if we break down our pdf into infinitesimal proportions, the summation becomes an integral, so the population mean,  $\langle x \rangle$  is defined by:

$$\langle x \rangle = \int x p(x) dx \quad (12)$$

We use the two symbols for mean interchangeably, in an attempt to make the equations more readable,  $\bar{x} = \langle x \rangle$ .

Suppose we now wanted to determine how far most values are from this average. It seems that we would determine how far each point was from the

mean, and then take the average of these numbers. But, if we did exactly this, we would get zero! This is because for every number that is above the mean, there is a number that is below the mean that will cancel it out.

$$\begin{aligned}
 \langle (x - \bar{x}) \rangle &= \int p(x) \cdot (x - \bar{x}) dx & (13) \\
 &= \int xp(x) dx - \bar{x} \int p(x) dx \\
 &= \bar{x} - \bar{x} \\
 &= 0
 \end{aligned}$$

(14)

To get around this, we need to get rid of the negative sign somehow. One way to do this is to solve for the variance instead.

$$\begin{aligned}
 \langle (x - \bar{x})^2 \rangle &= \int (x - \bar{x})^2 dx & (15) \\
 &= \int (x^2 - 2x\bar{x} + \bar{x}^2)p(x) dx \\
 &= \int x^2 p(x) dx - 2\bar{x} \int xp(x) dx + \bar{x}^2 \int p(x) dx \\
 &= \langle x^2 \rangle - 2\langle x \rangle^2 + \langle x \rangle^2 \\
 &= (\langle x^2 \rangle - \langle x \rangle^2)
 \end{aligned}$$

(16)

(17)

To fix the units problem (squared terms), we use the standard deviation, which is just the square root of the variance.

A helpful aside: Approximately 68 deviation of the mean, and approximately 95 deviations of the average. This is not always true, but a good rule of thumb.

The mean is also known as the first moment. In general, the moments are given by:

$$M_n = \langle x^n \rangle = \int x^n p(x) dx \quad (18)$$

The variance is known as the 2nd moment about the mean, where a moment about the mean is defined by:

$$C_n = \langle (x - \bar{x})^n \rangle = \int (x - \bar{x})^n p(x) dx \quad (19)$$

Moments	Moments about the mean
$M_0 = 1$	
$M_1 = \langle x \rangle$	$C_1 = 0$
$M_2 = \langle x^2 \rangle$	$C_2 = M_2 - M_1^2$
$M_3 = \langle x^3 \rangle$	$C_3 = M_3 - 3M_2M_1 + 2M_1^3$

We began by looking at a distribution of ages of death, and then saw how this distribution can be related to the pdf. The pdf told us about confidence and probability, and we used this information to determine the relation of the cdf to the pdf. From here we went over how to get the mean of a distribution, and eventually how to get all of the moments and the moments about the mean. We should now look at the distribution of the time of night this writing was done, and think about how this might be correlated to the probability of it being coherent!

## 2 Reporting Error

Let's think about the scientist who is measuring density again. If we assume that the mean that she measured is the correct one, what is the probability that someone else measuring the density would get the same result? Notice this isn't quite the same question we asked before when looking at the pdf. The pdf we are now discussing is actually made up of probabilities, rather than data that has been fitted with a pdf. IF IT ISN'T THE SAME PDF, WHERE DOES THIS ONE COME FROM? It is this pdf that scientists use to determine error, but there are quite a few different ways to report it. We will now go through the various ways of reporting error, starting with the most informative to the least informative.

1. Show the pdf. This is especially helpful if the pdf is not gaussian.
2. Report a confidence interval. For example, 90% confidence that the value lies between 2 and 3. So, where does this confidence interval come from? Figure 8 shows the mean, the mode, and the median values for a non-gaussian distribution, and we can see these are not the same. The median describes the point on which the area is the same on both sides (ie. 50%), shown as the line through the distribution. The mean, in this case, lies just to the left of the line dividing the area. If you were to imagine a fulcrum balancing the distribution, the fulcrum would be placed at the mean, which is why it is known as the first moment. The most probable value is the mode or the maximum. For determining confidence intervals, ie. what is the most likely value range, it is the most probable value that interests us. But, should the confidence interval be centered on the mode? Figure 9 shows 3 different ways of showing the same area. The top graph shows a 68% area centered on the mode, and we can see this is not quite right. The lowest x value in the area shown has a higher probability than the highest x value. In other words, there are values that are more likely outside of the area than some of the values within the area chosen. A better choice is the middle graph, which shows an area not centered on the mode, but one in which the two extremes of the area have the same probability. This explains why confidence intervals are not always symmetric, e.g. the most probable value in this case is  $7 + 1.7, -2.5$ . In this case it is obvious that the distribution is not gaussian. However, in

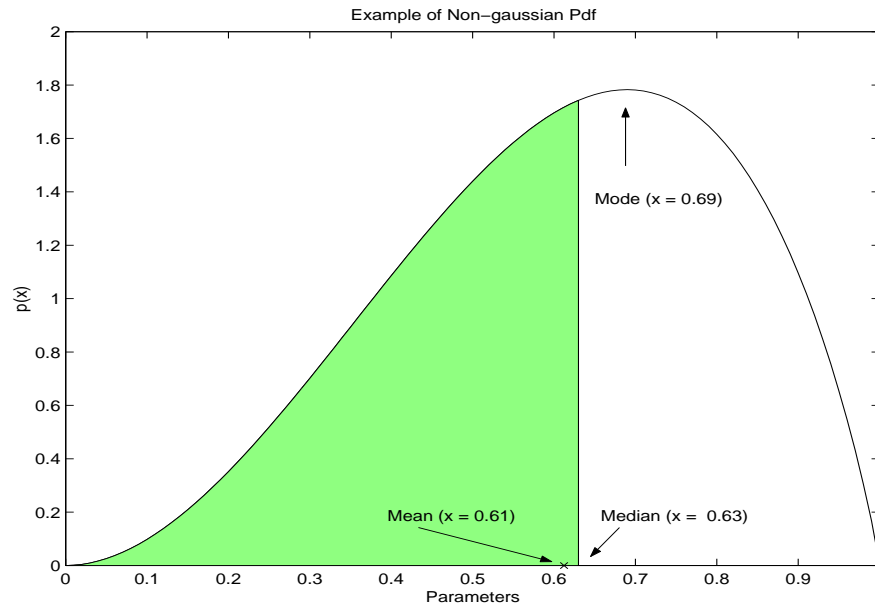


Figure 8: Example of a non-gaussian pdf that should be shown when reporting errors

the case of a bimodal distribution this is not enough, and the pdf should be disclosed.

3. The most common way people choose to report errors, ie. to characterize the pdf, is to report the standard deviation, for example,  $5 \pm .2$ . If the distribution is gaussian, this is fine, but for other distributions the standard deviation and 68% may be different numbers. Since it is not obvious from the notation, it should be reported which number is being used.
4. The better than nothing alternative to the previous methods is to guess, ie. not base it on the pdf at all.
5. The last and least desirable alternative is to not report an error estimate.

### 3 Propagating Error

So you have measured  $x$ , and calculated the error estimate as  $\bar{x} \pm \delta x$ . What if you now want to report the error on a function of  $x$ ,  $\bar{f} \pm \delta f$ . For example, one can imagine measuring two sides of a triangle, and solving for the length of the third side (see Figure 10). But what happens to the error?

Let's look at the even simpler relationship:

$$f = x + y \quad \text{where} \quad x = x \pm \delta x \quad \text{and} \quad y = y \pm \delta y$$

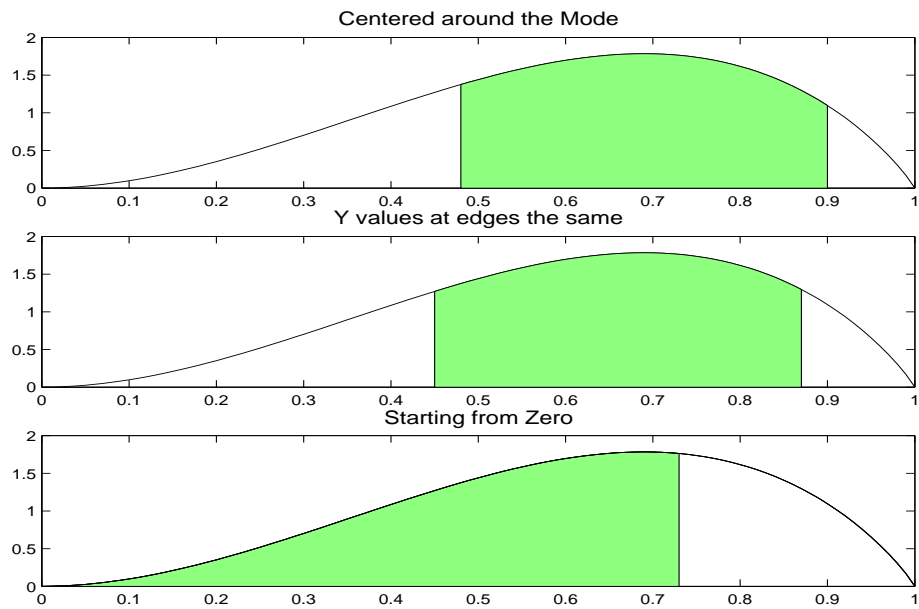


Figure 9: 3 graphs showing 68% area

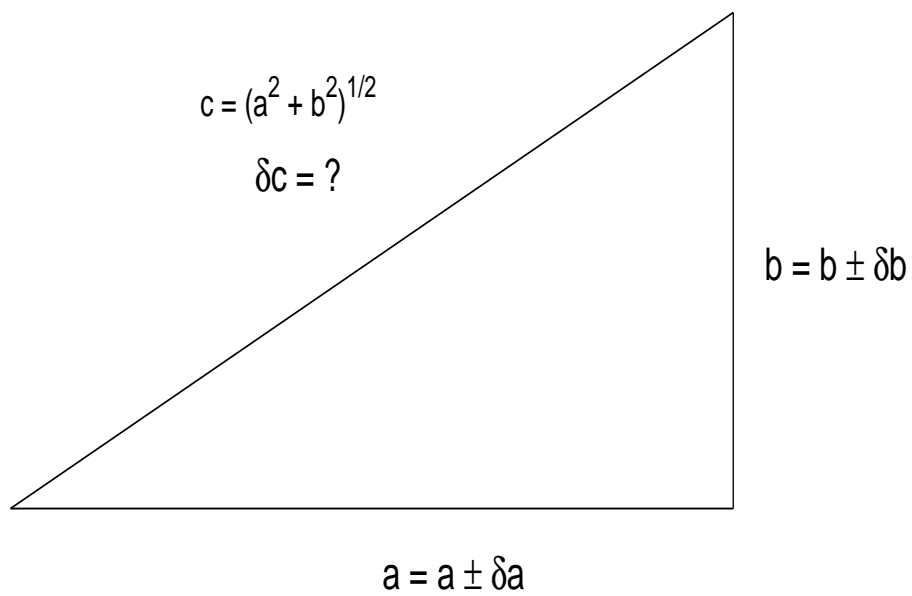


Figure 10: Measuring the sides of a triangle

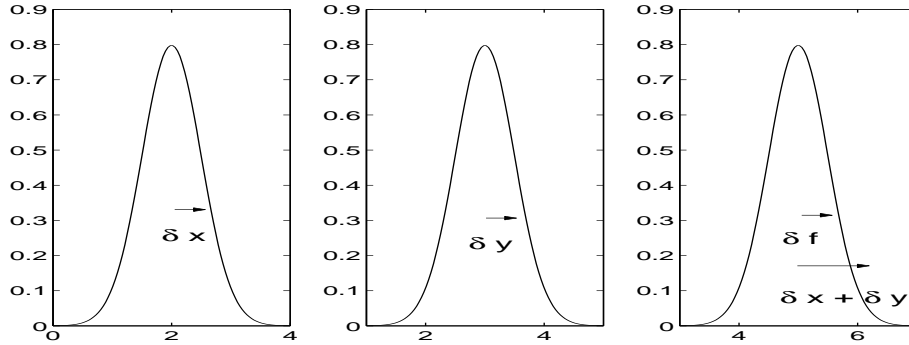


Figure 11:  $f = x + y$  and  $\partial f = ?$

One might assume that the error can also be added:

$$f = (\bar{x} + \bar{y}) \pm (\delta x + \delta y)$$

But this is wrong! To see this we should look at the distributions of  $\bar{x}$ ,  $\bar{y}$ , and  $\bar{f}$ . Figure 11 shows these distributions,  $x = 2 \pm 0.5$ ,  $y = 3 \pm 0.5$ , and  $f \pm \delta f = (x \pm \delta x) + (y \pm \delta y) = 7 \pm ?$ . It should be obvious that even though the variables add in a straightforward fashion, the error does not add in the same way.

The error in  $x$  and  $y$  are the standard deviations of those distributions.

$$(\delta x)^2 = \frac{1}{N} \sum_i^N (x_i - \bar{x})^2 \quad \text{where} \quad \bar{x} = \frac{1}{N} \sum_i^N x_i \quad (20)$$

$$(\delta f)^2 = \frac{1}{N} \sum_i^N (f_i - \bar{f})^2 \quad \text{where} \quad f_i = x_i + y_i \quad \text{and} \quad \bar{f} = \bar{x} + \bar{y}^* \quad (21)$$

$$= \frac{1}{N} \sum_i^N (x_i + y_i - \bar{x} - \bar{y})^2 \quad (22)$$

$$= \frac{1}{N} \sum_i^N [(x_i - \bar{x}) + (y_i - \bar{y})]^2 \quad (23)$$

$$= \frac{1}{N} \sum_i^N (x_i - \bar{x})^2 + \frac{1}{N} \sum_i^N 2(x_i - \bar{x})(y_i - \bar{y}) + \frac{1}{N} \sum_i^N (y_i - \bar{y})^2 \quad (24)$$

---

\*proof for  $\bar{f} = \bar{x} + \bar{y}$ :

$$\bar{f} = \frac{1}{N} \sum_i^N f_i$$

which gives us:

$$(\delta f)^2 = (\delta x)^2 + 2 \text{cov}(x, y) + (\delta y)^2$$

where  $\text{cov}(x, y)$  is the covariance matrix. In the case where  $x_i$  and  $y_i$  are uncorrelated, the covariance matrix is equal to zero, and the middle term drops out, leaving:

$$(\delta f)^2 = (\delta x)^2 + (\delta y)^2$$

We now return to looking at how error propagates in a function. Let's look at the more complicated function.  $f(x) = x^2$ . If  $x$  is  $2.0 \pm 0.2$ , what do we expect  $f$  to be? We begin with the same equations, but now we remain in the more general notation  $f(x)$ :

$$(\delta x)^2 = \frac{1}{N} \sum_i^N (x_i - \bar{x})^2 \quad \text{where} \quad \bar{x} = \frac{1}{N} \sum_i^N x_i \quad (25)$$

$$(\delta f)^2 = \frac{1}{N} \sum_i^N (f_i - \bar{f})^2 \quad \text{where} \quad \bar{f} = f(\bar{x}) \quad \dagger \quad (26)$$

$$= \frac{1}{N} \sum_i^N (f(x_i) - \bar{f})^2 \quad (27)$$

$$= \frac{1}{N} \sum_i^N [f(\bar{x}) + (x_i - \bar{x})f'(\bar{x}) \dots - \bar{f}]^2 \quad (28)$$

$$= \frac{1}{N} \left[ \sum_i^N (x_i - \bar{x})^2 \right] [f'(\bar{x})]^2 \quad (29)$$

$$= (\delta x)^2 \left| \frac{\delta f}{\delta x} \Big|_{\bar{x}} \right|^2 \quad (30)$$

$$\text{therefore}(\delta f) = (\delta x) \left| \frac{\delta f}{\delta x} \Big|_{\bar{x}} \right| \quad (31)$$

---


$$\begin{aligned} &= \frac{1}{N} \sum_i^N (x_i + y_i) \\ &= \frac{1}{N} \sum_i^N x_i + \frac{1}{N} \sum_i^N y_i \\ &= \bar{x} + \bar{y} \end{aligned}$$

<sup>†</sup>general proof for  $\bar{f} = f(\bar{x})$ :

$$\bar{f} = \frac{1}{N} \sum_i^N f_i$$

In this case,  $f(x) = \bar{x}^2$ , so  $\frac{\partial f}{\partial x} = 2\bar{x}$ , and the propagated error is  $\delta f = 2\bar{x} \delta x$ . We measured  $x = 2$ , so  $f(x) = 4$ , and the original error,  $\delta x$ , was 0.2, so  $\delta f = 0.8$ .

For the more general case,  $\bar{f} = \bar{x}^n$ , we have

$$\begin{aligned} \delta f &= n\bar{x}^{n-1}(\delta x) \\ &= n\frac{\bar{x}^n}{\bar{x}}(\delta x) \\ &= n\frac{\bar{f}}{\bar{x}}(\delta x) \quad (\text{using } \bar{f} = \bar{x}^n) \\ \left(\frac{\delta f}{\bar{f}}\right) &= n\left(\frac{\delta x}{\bar{x}}\right) \end{aligned} \tag{32}$$

$$\text{fractional error in } f \uparrow = \quad \uparrow \text{ fractional error in } x$$

It is worthwhile to examine fractional errors for a moment at this point. If we add  $N$  things with error  $\sim (\delta x)$ , where all  $x$ 's are approximately  $\bar{x}$  ( $x_i \sim \bar{x}$ , so  $f \sim N\bar{x}$ ) and all error is approximately  $\delta x$  ( $(\delta x_i) \sim (\delta x)$ ), then we have:

---


$$\begin{aligned} &= \frac{1}{N} \sum_i^N f(x_i) \\ &= \frac{1}{N} \sum_i^N [f(\bar{x}) + (x_i - \bar{x})f'(\bar{x}) + \dots] \\ &= \frac{1}{N} \sum_i^N 1 + \frac{1}{N} f'(\bar{x}) \sum_i^N (x_i - \bar{x}) + \dots \\ &= f(\bar{x}) + \dots \end{aligned}$$

$$\begin{aligned}
f &= \sum_i^N x_i \\
(\delta f)^2 &= N(\delta x)^2 \\
(\delta f) &= \sqrt{N}(\delta x) \\
\left(\frac{\delta f}{f}\right) &\sim \frac{\sqrt{N}}{N} \frac{\delta x}{x} \sim \frac{1}{\sqrt{N}} \frac{\delta x}{x}
\end{aligned}$$

Which is why we say fractional errors decrease like, or noise scales with,  $N^{1/2}$ .

Two other examples, demonstrating how differently error can propagate, are given in the following examples:

$$\begin{array}{ll}
f = \log x & f = e^x \\
\delta f = \frac{1}{x} \delta x & \delta f = e^x \delta x \\
\delta f = \frac{\delta x}{x} & \frac{\delta f}{f} = \delta x
\end{array}$$

Notice how in the first example directly above, the error in  $f$  is equal to the fractional error in  $x$ , and in the second example, the fractional error in  $f$  is equal to the error in  $x$ .

We are now ready to talk about error propagation in a function of two variables, each with their own error.

## 4 Fitting Data to a Model or Fitting a Model to Data

Let us begin by constructing a data set. We want our data to fit along a line, and be distributed normally around this line. So we pick a line,  $y = ax + b$ , and set up a vector of  $x$ , let's begin with just 5 data pts,  $x = 1, 2, 3, 4, 5$ . We choose an  $a$  and a  $b$ , for our example  $a = -2$  and  $b = 5$ . Now, if we just wanted a line, we would solve for  $y$ , but we want something that looks like a data set that could have been recorded from a real experiment, so we will choose random numbers which have  $y$  for the mean, and some standard deviation, in our case, 0.5. Figure 8 shows the data graphed, along with error bars showing the standard deviation, and the line used to create the data. What if we wanted to create a model that fit this data, how would we do it?

We want to find the model,  $\vec{a}$ , that is most likely to give this distribution of data points. In this case, it looks like a linear distribution, so we have the simple case of  $\vec{a}$  having only two parameters,  $a$  and  $b$ . Let's say we have already come up with a model; how would we figure out how likely that model is to fit the data? The distribution of the data points around the mean  $y(x_i)$  should

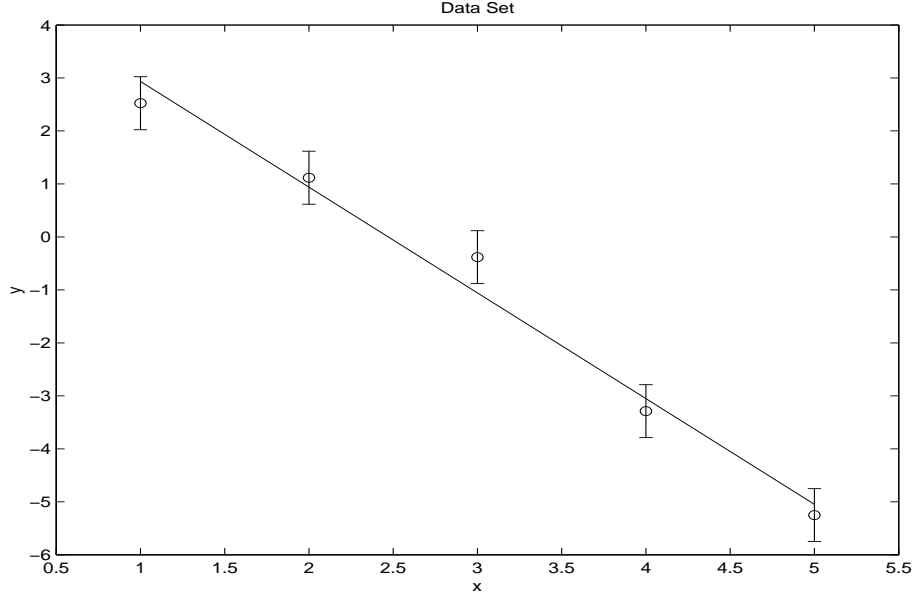


Figure 12: Data distributed around a line

follow a gaussian distribution, where  $\delta y_i$  = “uncertainty”. The equation of this pdf for each data point is given by:

$$p(y_i) = \frac{1}{\sqrt{2\pi(\delta y_i)}} \exp \left[ -\frac{1}{2} \left( \frac{y_i - y(x_i)}{\delta y_i} \right)^2 \right] \quad (33)$$

and the probability that this data set could have occurred, given these parameters, is the product of the probabilities of each point. We assume each data point has the same standard deviation.

$$P(y_1, y_2, \dots, y_N) = \prod_i (2\pi\delta y_i)^{-1/2} \exp \left[ -\frac{1}{2} \left( \frac{y_i - y(x_i)}{\delta y_i} \right)^2 \right] \quad (34)$$

$$= \prod_i (2\pi\delta y_i)^{-1/2} \exp \left[ -\frac{\chi^2}{2} \right] \quad (35)$$

where

$$\chi^2 = \sum_i \left[ \frac{y_i - y(x_i)}{\delta y_i} \right]^2 \quad (36)$$

is the only thing that depends on the parameters  $\vec{a}$ . According to maximum likelihood, we should choose  $\vec{a}$  to maximize P (the probability that each data

point could have occurred), and therefore we need to minimize  $\chi^2$ . For our simple case with two parameters:

$$\chi^2 = \sum_i^N \left[ \frac{y_i - ax_i - b}{\delta y_i} \right]^2 \quad (37)$$

at the minimum the derivatives of  $\chi^2(a, b)$  with respect to a and b vanish:

$$\frac{\partial \chi^2}{\partial a} = \sum_i^N \frac{y_i - ax_i - b}{(\delta y_i)^2} = 0 \quad (38)$$

$$\frac{\partial \chi^2}{\partial b} = \sum_i^N \frac{y_i - ax_i - b}{(\delta y_i)^2} = 0 \quad (39)$$

We make the following substitutions:

$$S = \sum_i^N \frac{1}{(\delta y_i)^2} \quad S_x = \sum_i^N \frac{x_i}{(\delta y_i)^2} \quad S_y = \sum_i^N \frac{y_i}{(\delta y_i)^2} \quad (40)$$

$$S_{xy} = \sum_i^N \frac{x_i y_i}{(\delta y_i)^2} \quad S_{xx} = \sum_i^N \frac{x_i x_i}{(\delta y_i)^2} \quad S_{yy} = \sum_i^N \frac{y_i y_i}{(\delta y_i)^2} \quad (41)$$

To get the following equations:

$$S_{xy} - aS_{xx} - bS_x = 0 \quad (42)$$

$$S_y - aS_x - bS = 0 \quad (43)$$

We can either use substitution or a matrix equation.

$$\begin{pmatrix} S_{xx} & S_x \\ S_x & S \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} S_{xy} \\ S_y \end{pmatrix} \quad (44)$$

where the determinate is  $SS_{xx} - S_x^2$ . Solve for a and b to get:

$$a = \frac{S_x S_y - S_{xy} S}{S_x^2 - S_{xx} S} \quad (45)$$

$$b = \frac{S_x S_{xy} - S_y S_{xx}}{S_x^2 - S_{xx} S} \quad (46)$$

We have now minimized  $\chi^2$  to get an a and b that will maximize the probability of getting a given data set. An assumption we have made is that errors are Gaussian and independent from data point to data point.

An interesting exercise we can do to verify a and b, is to make sure our units are consistent. We can see that the units of S must be equal to the units of  $y^{-2}$ ,  $[S_y] = [y]^{-1}$ ,  $[S_x] = [x][y]^{-2}$ , and so on. If we then write a in terms of the units of x and y we have

$$\begin{aligned} [a] &= \frac{[x][y]^{-3}}{[x]^2[y]^{-4}} \\ &= [y/x] \end{aligned} \tag{47}$$

which, fortunately, is the units of a slope.

Looking back to equation 23, we see that  $\chi^2$  is going to be approximately N. More precisely, the expectation value of  $\chi^2$  is the number of data points minus the number of parameters, otherwise known as  $\nu$ , or the degrees of freedom in the fit. The standard deviation of  $\chi^2$  is  $\sqrt{2\nu}$ .

Let's return to our data set. If we do this business of minimizing  $\chi^2$  to get an a and a b, here is what we get: a = -1.9954, and b = 4.9298. Assuming this is correct, how good is the answer? When we plug this a and b and our data points into our equation for  $\chi^2$ , we get  $\chi^2 = 3.01$ . This answer makes sense, since we have 5 data points and 2 parameters. How much do we have to change a or b to get a  $\chi^2$  that is outside the standard deviation? By trying several different a's and b's we determine that  $a = -2.00 \pm 0.08$  and  $b = 4.93 \pm 0.27$  are the values within one standard deviation. Can we increase the accuracy by increasing the number of data points? What happens if we use 1000 data pts. instead of 5? After minimizing  $\chi^2$ , in this case we get a = -1.9999 and b = 4.9474.  $\chi^2$  is 1,029, and the variance in a and b are  $-1.99985 \pm 0.00015$  and  $4.947 \pm 0.090$ . This seems better, but we still don't really know how good a fit we have. The question we should now be thinking about is, what is the probability of getting a value of  $\chi^2$  that is as poor or worse as the value we have, by chance?

Figure 9 highlights the probability of getting a  $\chi^2$  as bad or worse than what we got ( $\chi^2 = 3.01$ ); if the area ( $\epsilon = 1 - P(\chi^2)$ ) is much smaller than one, than the model is wrong (with  $1 - \epsilon$  confidence). If  $\epsilon$  is of order one, then it is a typical data set. If  $1 - \epsilon$  is much smaller than one, then the data is "too good", and the error bars were probably underestimated. In this case,  $\epsilon = 1 - 0.6099 = .3901$ , so it is a typical data set.

We can repeat this procedure of creating data sets of 5 points, and then minimizing  $\chi^2$  to get an a and b. If we then solving for the corresponding values of  $\chi^2$ , we get a distribution of  $\chi^2$ s. This distribution, in this case for 1,000 data sets of 5 points, along with the pdf for the same degrees of freedom superimposed, is plotted in Figure 10.

**REGRESSION** To determine the correct way, we start with the simple example of a function of one variable,  $f(x) = x$ . This would be the case, for example if you had several people measure the length of something. If the error for each measurement was the same, ie. each measurement was correct within one millimeter, then we could assume that x is the mean of these measure-

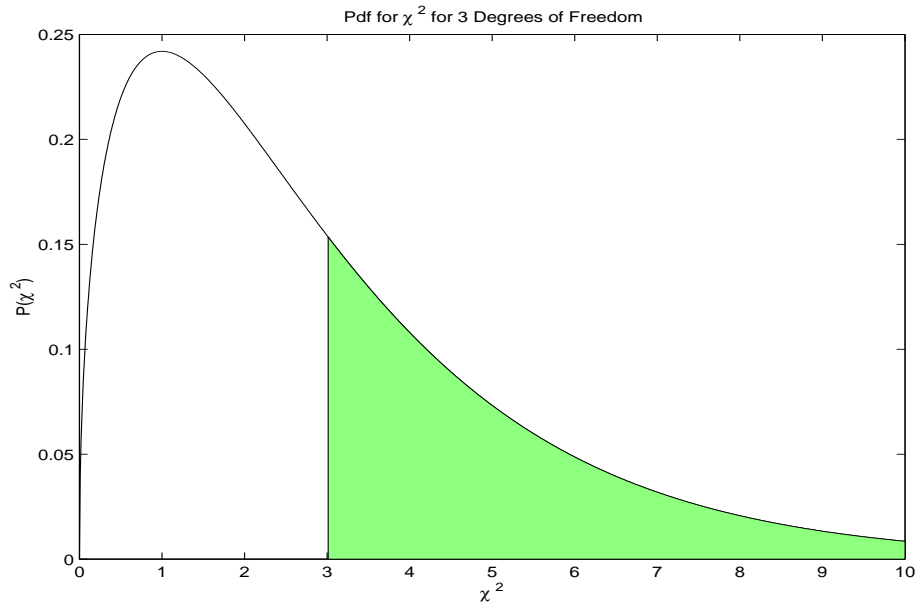


Figure 13: Probability of getting a  $\chi^2$  as bad or worse than what we have.

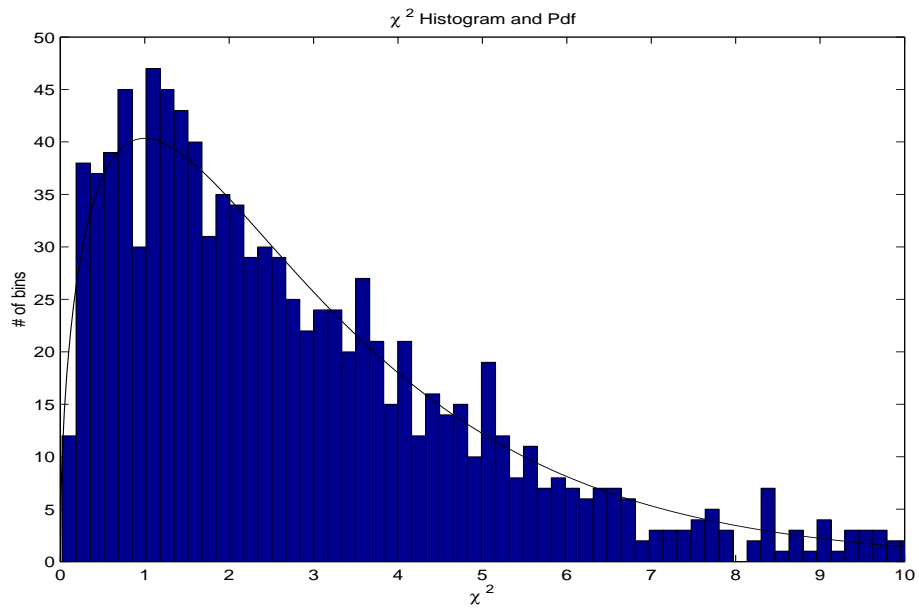


Figure 14: Histogram of 1,000  $\chi^2$ s, with pdf for  $\chi^2$  for the appropriate degree of freedom (3) superimposed.

ments,  $\pm$  the standard deviation. If some people were more accurate, however, we might want to give their measurement more consideration. To do this we minimize  $\chi^2$  and solve for  $x$ . The equation for  $\chi^2$  is,

$$\chi^2 = \sum_i^N \left( \frac{x_i - x}{\delta x_i} \right)^2$$

Minimize:

$$\begin{aligned} \frac{\partial \chi^2}{\partial x} &= \sum_i^N \frac{-2(x_i - x)}{(\delta x_i)^2} = 0 \\ &= \sum_i^N \frac{-x_i}{(\delta x_i)^2} + \sum_i^N \frac{x}{(\delta x_i)^2} \end{aligned}$$

Define some variables:

$$S = \sum_i^N \frac{1}{(\delta x_i)^2} \quad S_x = \sum_i^N \frac{x_i}{(\delta x_i)^2}$$

Substitute and solve for  $x$ :

$$\begin{aligned} -S_x + x * S &= 0 \\ x &= S_x / S \end{aligned} \tag{48}$$

We can put some real numbers to this example. Let's say we made 3 measurements,  $9.7 \pm 0.1$ ,  $9.8 \pm 0.2$ , and  $9.9 \pm 0.3$ . Plugging the values into equation 20, we get 9.7347 as our value of  $x$ . Notice that this is different from the average of the three numbers, 9.8, because it has taken into account the error in each number. This is also known as the weighted average. What about the error? In the same way that  $x$  was different from the average, the error is also different from the standard deviation. For reasons we will prove momentarily, the error adds in quadrature.

$$(\delta x)^2 = \sum_i^N \left( \frac{\partial x}{\partial x_i} \right)^2 (\delta x_i)^2 \tag{49}$$

Our equation for  $x$  was:

$$x = S_x / S \tag{50}$$

Now differentiate both sides:

$$\frac{\partial x}{\partial x_i} = \frac{1}{S} \frac{\partial S_x}{\partial x_i} - \frac{S_x}{S^2} \frac{\partial S}{\partial x_i} \tag{51}$$

The derivative of  $S_x$  with respect to  $x_i$  is  $\frac{1}{(\delta x_i)^2}$ , and the derivative of  $S$  with respect to  $x_i$  is zero. We plug these in to Equation 23 to get:

$$\frac{\partial x}{\partial x_i} = \frac{1}{S} \frac{1}{(\delta x_i)^2} - 0$$

now plug in to Equation 21 to get:

$$\begin{aligned}(\delta x)^2 &= \sum_i^N \frac{1}{S^2} \frac{1}{(\delta x_i)^4} (\delta x_i)^2 \\ &= \frac{1}{X^2} \sum_i^N \frac{1}{(\delta x_i)^2} \\ &= \frac{1}{S^2} S \\ &= \frac{1}{S} \\ \delta x &= \frac{1}{\sqrt{S}}\end{aligned}$$

So, for our measuring example, the error is 0.0857, which is much less than the error on any given measurement.