Statistical Analysis

Michael Mulhearn

April 1, 2018

1 Likelihood and χ^2

Suppose we make a series of measurements:

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

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

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

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

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

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

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

Now being physicists, we hate products and prefer sums, so we apply a logarithm , and there is an annoying factor of $\frac{1}{2}$ in the exponential, so we multiple by 2, and, generally being pessimists, we prefer to minimize instead of maximize, so we multiple by -1, so that at last we calculate:

$$-2\log\mathcal{L} = \sum_i \frac{(y_i - x_i)^2}{\sigma_i^2} - \log(\sqrt{2\pi}\sigma_i) \tag{1}$$

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

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

A small value of χ^2 means that the result is very close to the theoretical prediction and a large value means that the result is unlikely to have occurred as a result of the prediction. If the uncertainties and prediction are all correct, we would expect each x_i to differ from the prediction y_i by about σ_i. So in this case we would expect:

$$\chi^2 = \sum_i \frac{(y_i - x_i)^2}{\sigma_i^2} \sim \sum_i 1 = N \tag{3}$$

2 Minimizing χ^2

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

Consider, for example, the case where we measure a particular quantity x a total of n times, and so obtain measurements $x_i \pm \sigma_i$. We wish to extract from our data our best estimate for the true value of x. In this case, the prediction y_i will have the same value for every every i, and we construct χ^2 as:

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

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

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

and so

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

where

$$w_i = \frac{1/\sigma_i^2}{\sum_j 1/\sigma_j^2} \tag{4}$$

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

3 Extracting Uncertainties

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

$$\begin{aligned}
\sigma_m^2 &= \sum_i \left(\frac{dm}{dx_i}\right)^2 \sigma_i^2 \\
\sigma_m^2 &= \sum_i (w_i\sigma_i)^2 \\
\sigma_m^2 &= \sum_i \left(\frac{1}{\sum_j 1/\sigma_j^2}\right)^2 \\
1/\sigma_m^2 &= \frac{1}{N}\left(\sum_j 1/\sigma_j^2\right)^2
\end{aligned}$$

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

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

But now let's consider a much more powerful approach. Let's imagine that we have two parameters a and b that we have determined (using the results of the previous section) to have best fit values a_0 and b_0. We would like to determine the uncertainty on these extracted parameters. Theχ^2 sums over the experimental measurements and uncertainties, so in the end it is simply a function of the parameters we seek to determine. We can therefore think of our experiment as having an equivalent χ^2 where each parameter was simply directly measured with it's (currently unknown) experimental uncertainty:

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

In this case:

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

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

In our example, this leads to:

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

4 The Best Estimate for σ

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

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

$$-2\log\mathcal{L} = \sum_i \left(\frac{(m-x_i)^2}{\sigma^2} - \log(\sqrt{2\pi}\sigma)\right) \tag{5}$$

$$-2\log\mathcal{L} = \sum_i \frac{(m-x_i)^2}{\sigma^2} - \log(\sqrt{2\pi}\sigma) \tag{6}$$

Differenentiating wrt σ and setting to zero:

$$0 = \left(\sum_i \frac{(m-x_i)^2}{\sigma^3}\right) - \frac{N}{\sigma} \tag{7}$$

$$\sigma^2 = \frac{\sum_i (m-x_i)^2}{N} \tag{8}$$