Error Analysis II

1 Introduction

This lab has two goals. First we will measure the decay rate of a radioactive nucleus to study Poisson statistics. Second we will study the process of fitting data by the least-squares method, and by the χ^2 minimization method.

2 Poisson Statistics

The Poisson distribution is the correct one for "counting" experiments. So if one is counting events which occur at a definite rate, but where the basic process is statistical in nature, one uses Poisson statistics to understand one's data. A good example is radioactive decays, where the nuclei of the radioactive element decay at random times, but the number of total nuclei determines the average observed decay rate. A counter example is counting clock ticks where there is a definite rate, but the ticks are evenly spaced in time and there is no random component to the times.

For the reserve books in the Physics Library for this course, see Taylor, chapter 11, or Bevington, section 3-2, for discussions of Poisson statistics.

The main result of a study of Poisson statistics is that, after counting for some interval, t, and measuring a number of events, n,

- The best estimate of the true number of counts, μ, is n; i.e., $\mu = n$.
- The best estimate of the counting rate is n/t.
- The best estimate of the standard deviation is the square root of n; i.e.., $\sigma = \sqrt{n}$.
- The error in the mean is equal to the standard deviation.
- If one repeats the experiment many times one would find a distribution of n values that follows the Poisson distribution, $P(n, \mu) = e^{-\mu}\mu^n/n!$

Recall that for the Gaussian distribution there were three quantities that were important: the mean, μ, the standard deviation, σ, and the error of the mean, $\overline{\sigma}$. For the Poisson distribution there are only two important quantities, the mean and the standard deviation, because the error in the mean equals the standard deviation.

The Poisson distribution, for values of μ below about 10, has a tail on the right. For large values of μ the distribution becomes symmetric. Figure 1 shows two Poisson distributions, for $\mu = 5$ and $\mu = 50$. Part (b) of Figure 1 also has a fit to a Gaussian function superimposed on the histogram. It fits very well. This illustrates the fact that the Poisson distribution becomes very close to Gaussian for large μ.

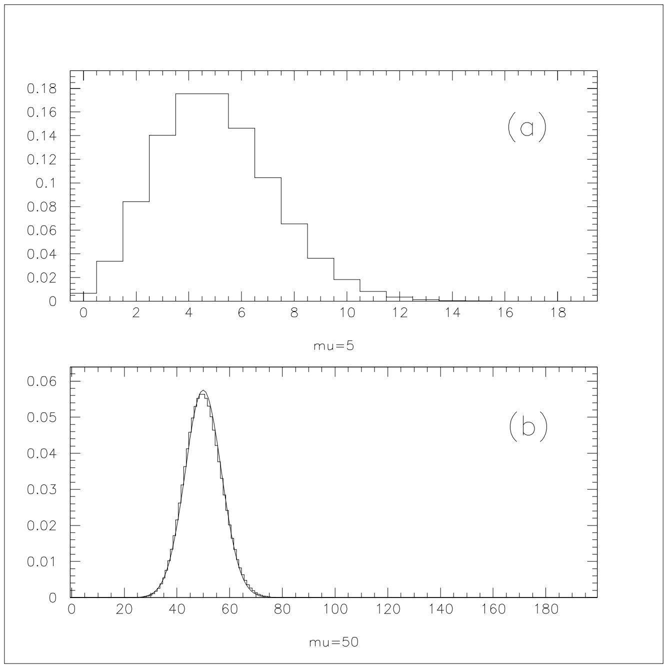

Figure 1: Poisson distributions. (a): $\mu = 5$. (b): $\mu = 50$.

3 Fitting Data by the Least Squares Method

The three Physics 275 books on reserve in the Physics Library have sections on fitting data using the Least Squares method. See Baird section 6-7, Taylor chapter 8, and Bevington chapter 6.

The Least Squares method is appropriate for fitting data to a straight line if the uncertainties in all the data points are equal. This is the method used in Excel. Basically the method minimizes the deviations of the data points from the fit.

If we have a set of data points (x_i, y_i) where x_i represents the independent variable and y_i is the dependent variable (i.e., we set our experimental situation to value x_i and measure the value of y_i), and we want to fit those data to the function $f(x)$, then in this method we minimize $X^2 = \Sigma(y_i - f(x_i))^2$. We use the square of the deviations to avoid subtracting deviations. If $f(x)$ is a straight line then the answers for the slope, m, and intercept, b, can be found algebraically.

4 Fitting Data by the χ^2 Minimization Method

Two of the Physics 275 books on reserve in the Physics Library have sections on fitting data using the χ^2 minimization method. See Taylor chapter 12, and Bevington chapter 11.

If the uncertainties in all the data points are not equal (which happens much more often than not), then we will want to count those data points with small uncertainty more highly in fitting a function to the data. If $f(x)$ is not a straight line, we will not be able to find the solution to the problem in closed form. In either of these cases the method of Least Squares will not work. Here we must perform a more complicated procedure to find the best values of the function's parameters. One such method (the simplest) is called χ^2 minimization.

Again, if we have a set of data points (x_i, y_i, σ_i), where the uncertainty in each y_i is σ_i, and we want to fit the data to the function $f(x)$, we minimize $\chi^2 = \Sigma(y_i - f(x_i))^2/\sigma_i^2$. To "minimize" χ^2 means to adjust the parameters of $f(x)$ to make χ^2 the smallest possible. Here we have weighted each deviation by the uncertainty in the measurement.

After finding the values of the parameters that minimize χ^2, the value that χ^2 takes is important. Since each data point's deviation is divided by the expected variation between it and the function, each term in the sum should contribute a value to χ^2 of about unity. So if there are n data points one would expect $\chi^2 \sim n$. But there is one complication one must deal with. If there are p fitting parameters, each one will reduce the expected value of χ^2 by approximately unity. So we really expect that a good fit will have $\chi^2 \sim n - p$. The quantity $n - p$ is called the "number of degrees of freedom" of the fit. If χ^2 is much less than, or much more than $n - p$, then there will be something wrong with the fit. Either the uncertainties, σ_i, are wrong, or the minimizing procedure did not converge correctly.

The Experiment

Today's lab has two activities: first measuring the rate of radioactive decays, and testing Poisson statistics as we do so; and second fitting a file of data by two methods: least squares and χ^2 minimization.

We will measure the decay rate of a radioactive source using a geiger counter which is hooked up to LoggerPro. Here are instructions for this apparatus: click on LoggerPro; drag the Radiation RM-BTD icon into the DIG/SONIC1 box; set it to 20 samples/second and length 10 seconds. Then do the experiment with 100 samples, and perform the rest of the enumerated points below.

Your lab report should consist of ALL the graphs mentioned in this section, all with titles and axes labeled. Include any calculations by writing them on the graph in what you think is an appropriate place.

For the radioactive decay and Poisson statistics section of the lab, follow this procedure:

1. Record, in Excel, 200 measurements of the rate of radioactive decays from LoggerPro. Find the mean of all the measurements, and the standard deviation. Calculate the mean rate of radioactive decays (number of particles detected per second). Include uncertainties in your answer. Use all of your data for this.

2. Make a histogram of the 200 measurements. Draw a vertical line on the histogram at the position of the mean. Calculate the standard deviation graphically and compare it to your calculated value. For a Gaussian distribution (also approximately true for a Poisson distribution), the Full Width at Half Max (FWHM) is equal to 2.35 times the standard deviation. So find the standard deviation by the equation, $\sigma = \text{FWHM}/2.35$.

3. Combine the 200 measurements into 20 groups, each of 10 measurements; i.e., group together measurements 1-10, 11-20, etc. Calculate the mean and standard deviation of each group. Histogram the means. Calculate the mean and standard deviation of the group means.

4. Change the histograms of events in parts 2 and 3 into histograms of the rate of nuclear decays by dividing the number of events by the live time of the detector and re-histograming the result. Plot the two histograms on the same graph. Why are they so different?

For the fitting part of this lab, you will find a file called simdata.asc by clicking on "my computer", disc C:, "program files", and "paw". Open this file and copy the contents into Excel, plot it, and fit it to a straight line. Excel uses least squares to perform this fit. Include the plot in your lab report.

Next read the same data into PAW. Again plot it and fit it to a straight line. PAW uses χ^2 minimization for its fits. Include the plot in your lab report. State the result of the fit including the uncertainty. Compare the answers you got for m and b by the two methods. Which method gives the more accurate answer?

Here are the PAW commands you will need.

```
vector/read x,y,sigma simdata.asc
histogram/create/1d 1 'simdata' 8 1.3 2.9
```

```
histogram/put/contents 1 y
histogram/put/errors 1 sigma
histogram/fit 1 p1
fortran/file 66 fit1.ps
meta 66 -111
histogram/fit 1 p1
close 66
```