DISTRIBUTION FUNCTIONS

HOMEWORK
Read Taylor, "An Introduction to Error Analysis", Chapters 4, 5, 10, 11. Do problems 4.5, 4.6, 5.21, 5.36, 11.10, 11.20.

PURPOSE
To understand the nature of the Gaussian and Poisson distribution functions and their relationship to error analysis.

INTRODUCTION
Often, repeated measurements of a physical quantity will not each give the same value. Rather, there will be a distribution of measured values about some central value. Here we will consider two of the important types of distributions. First, there may be a single true value for the quantity being measured, but random measuring errors due to the apparatus or techniques will cause the result to vary. In principle, these random errors can be reduced by improving the apparatus or techniques, such that the width of the distribution of measured values will decrease. The measured values in this case usually form a continuous distribution, but in this lab, we will work with discrete values for the measured variable.

The second type of distribution arises when the measurement process has inherent statistical fluctuations that cannot be reduced by improving techniques. For example, in the decay of a radioactive sample, we can only give the probability that each nucleus will decay within a given time interval. Even if the counter works perfectly, the number counted for the same time interval will vary. The measurements for this second case have discrete values, e.g., counts of decays.

For the first case, the distribution of measured values about the mean is frequently described by a bell-shaped curve called the Normal or Gaussian distribution which is characterized by two parameters – the mean that gives the location of the peak of the distribution, and the standard deviation that gives the width of the distribution. For the second case of statistical fluctuations, such as nuclear decay, the distribution about the mean is given by the Poisson distribution, characterized by only one parameter – the mean. If the mean number of decays per chosen counting time interval is large compared to the size of the fluctuations, the Poisson distribution reduces to a Gaussian distribution. Interestingly, even for a mean as small as 10 the Poisson distribution is very well approximated by a Gaussian distribution.

In this experiment you will study the radioactive decay of nuclei for the case where the counting interval is very short and Poisson statistics apply and for longer counting intervals where the distribution can be approximated by a Gaussian. For both cases, you will evaluate the standard deviation and study its relationship to the uncertainty in your measurements.

THEORY

A. Statistical Definitions

Your textbook gives a thorough discussion of distribution functions. Here we will only give a brief summary of some important points. The mean and variance of a parent distribution in random variable x are defined by:

$$\mu = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} x_i \qquad \sigma^2 = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$$

From a limited set of measurements (n not infinite) we can estimate μ from the average value of the x_i. Likewise, the variance is estimated from the "sample" of measurements and also uses the average value of the x_i.

$$\mu \approx \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \sigma^2 \approx \sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

The factor n-1 gives a better approximation of the variance than n because we are using only an estimate of μ rather than its actual value. The square root of the variance is called the standard deviation.

For a Gaussian distribution there is a 68% chance that a single measurement will fall between $\mu-\sigma$ and $\mu+\sigma$. There is a 95.4% chance that it will fall within 2σ of μ and a 99.7% chance that it will fall within 3σ of μ.

Next, we would like to know how close our estimate of μ is to the true value, i.e., how close $\bar{x}$ is to μ. This is not given by σ_x, which is characteristic of the parent distribution (x), but rather by

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

where n measurements form the sample; this is characteristic of the distribution of $\bar{x}$. Thus, as more measurements are made the uncertainty on the estimate of the mean decreases, but only as the square root of the number of measurements.

B. Distribution Functions

A distribution function P(x) gives the probability that a single measurement of a quantity whose true value is μ will give a specific value of x; more specifically, when x is a continuous variable, P(x)dx is the probability the measured value of x will lie between x and x+dx. The area under the curve, $\int P(x)dx$, equals 1, which simply means that there is 100% probability of getting *some* value of x in any measurement.

The most commonly encountered distribution is the *Normal* or *Gaussian* distribution:

$$P_G(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (1)$$

This curve is peaked at x = μ and is symmetric about μ. For a set of measurements, $\bar{x}$ is a good estimate of μ. The reason $P_G(x)$ is encountered so often is given by the central limit theorem which (very roughly) states that if the fluctuations in the measured value of x are caused by several (K) independent factors each with its own (not necessarily normal) distribution, then in the limit that K becomes large the overall distribution approaches a normal distribution.

The second common distribution function (applicable to integer variables) is the *Poisson* distribution in N:

$$P_P(N) = \frac{\mu^N}{N!} e^{-\mu} \qquad (2)$$

where N is a non-negative integer, and μ, the expected value of N, is not necessarily an integer. $P_P(N)$ is an asymmetric, peaked function, whose maximum value does not coincide with μ. For a set of measurements of N, the average $\overline{N}$ is again a good estimate of μ.

Notice that while Gaussian $P_G(N)$ appears to involve two parameters μ and σ, Poisson $P_P(N)$ involves only μ, and both the peak position and width of Poisson $P_P(N)$ are solely determined by μ. The standard deviation of Poisson $P_P(N)$ is $\sigma = \sqrt{\mu}$. For large μ, Poisson $P_P(N)$ becomes almost symmetric and can be accurately approximated by Gaussian $P_G(N)$ with $\sigma = \sqrt{\mu}$. We will use this property to study the Gaussian distribution.

APPARATUS

Radioactive source (5 mC Cs 137)
Geiger tube
Electronic counter
LabPro Interface
Computer

EXPERIMENT

In this experiment you will measure the activity of a radioactive ^{137}Cs source. The half-life of the ^{137}Cs (30 yr.) is long compared to the length of the experiment so that its activity can be considered constant. You will use a computer to acquire data on the sample's activity and then compare the data to the theoretical Gaussian and Poisson distributions in order to understand the role that distribution functions play in your data.

The nuclear decay is detected by a Geiger tube. Each time a gamma or beta ray from a decaying nucleus enters the Geiger tube a current pulse is generated. The tube is connected to a counter circuit which converts it into a 5 V, 1 ms wide voltage pulse which is available at the phone jack on the back of the counter. The counter also counts the pulses and displays the count on the front panel. However, you will not use this feature. Instead, the voltage pulse from the counter is fed into a LabPro interface. The LabPro interface transmits the decay data to the computer through its input port. Using the program Logger Pro, you will repeatedly record the number of counts in a time interval (whose length you have chosen) and make a plot of the distribution of counts about the average value. You will then export the data to MATLAB to analyze them, compare to the theoretical distribution functions, and plot your results.

PROCEDURE

This is a long experiment. To complete it in time, one must fully understand the theory before beginning the experiment and analyzing the results.

A. The Geiger tube should be connected to the BNC connector on the back of the counter, and the DIG/SONIC1 input of the LabPro interface to the jack on the back of the counter. The LabPro interface is connected to the serial port of the computer.

B. Make sure the counter is turned on (the LabPro interface is always on). Place the radioactive source under the Geiger tube.

Use the program Logger Pro to record the count (decay) rate of the source. To set up the Logger Pro program, see the appendix page at the end of the manual. Choose suitable values for Collection LENGTH and count interval (SAMPLING RATE). Click button COLLECT to collect data. To save your data, choose FILE: EXPORT DATA AS TEXT … (not SAVE).

The count rate in a given interval is the number of counts divided by the count interval. The rate you will observe depends on how close the source is to the tube. Adjust the distance so that the count rate is less than but near 100 counts/second (cps). *[The reason for this is that the counter generates 1 ms wide pulses which the computer counts. If two nuclei decay within 1 ms of each other and the two gamma or beta rays enter into the detector, the computer will only record one. If the count rate approaches 1000 cps, you will lose many counts due to this dead time and your data will be distorted.]*

You will now do a series of "runs" each of which measures the decay counts N in a set of n consecutive intervals making up the total collection time of the run. The collection of decay counts N in a run constitute measurements of the unknown true count N_0, and decay rates $R = N/\Delta t$ constitute measurements of the unknown true rate R_0 (Δt is the time interval of one measurement of N).

C. We start by studying the distribution of N in a run. Be sure not to move the Cs source relative to the Geiger counter between measurements. Using Logger Pro, set the measuring interval to 1/2 second and measure N for a run time of 30 seconds. This gives you n = 60 measurements of N. Record $\overline{N}$, the average value of N, and the standard deviation, σ_N, which are both calculated by Logger Pro (click the STAT button). You are trying to determine $R_0 = N_0/\Delta t$, the true value of the decay rate. Your best estimate of R_0 is $\overline{N}/\Delta t$, with the uncertainty given by 'standard error' $(\sigma_N/\Delta t)/\sqrt{n}$. Report this best estimate and uncertainty.

The distribution of N in the run is expected to be Gaussian. To test this prediction, plot a histogram of the obtained values of N using a bin width of 1 using MATLAB (see the MATLAB hints in the appendix at the end). Does the histogram resemble a Gaussian curve centered on $\overline{N}$?

For this relatively small number of samplings (60 in this case), the histogram may not look too close to a Gaussian. For this reason, let's collect data again for a much longer run (5 min) with the same sampling interval of ½ second. This should produce 600 data points. Analyze these for $\overline{N}$ and σ_N and plot the histogram. Does it remind a Gaussian better?

For both measurements, plot a smooth (analytic) curve of a Gaussian function, properly normalized, on top of your histograms (on the same plot), so that a comparison can be made.

The significance of σ_N is that if **one more 1/2 s measurement** is made, there is a 68% probability that the value of N obtained will lie between $N_o - \sigma$ and $N_o + \sigma$, where $N_o \approx \overline{N}$ and $\sigma \approx \sigma_N$. To test his prediction, plot a very "coarse grained" histogram, with a large bin width of 2σ centered on $\overline{N}$. The number of counts in the central bin should thus be approximately 68% of the total number of counts. Does the histogram verify the prediction?

D. Now we study the distribution of N in a large sample, comparing it to the expected distribution. For a 10-minute run, make a histogram of N with bin width of 1. Then calculate the expected distribution of events for a Gaussian distribution using the values of $\overline{N}$ and σ_N that you observe for this data set. Multiply your theoretical distribution by the total number of samples, such that it is normalized to the data histogram. Plot the data and the theoretical curve in the same graph and compare.

E. Again we study the distribution of N in a large sample, i.e., for a long run. For the same 10-minute run, prepare a table giving the percentage of intervals where N falls within $\pm \sigma_N$, $\pm 2\sigma_N$, and $\pm 3\sigma_N$ of the mean. In your table, compare these numbers with the theoretically expected percentages. Use MATLAB to devise a histogram to give you the numbers for this table. Print the histogram and explain it in your report.

F. So far, the distribution of measurements of N about $\overline{N}$ has been well described by a Gaussian distribution. We now want to study the Poisson distribution which will describe the data when $\overline{N}$ becomes smaller. To reach this limit it is not necessary to move the source relative to the counter; instead, you can simply change the measuring interval to 1/20 s. Since $\overline{N}$ was about 50 counts/interval for a 1/2 s measuring interval, it will now be less than 5 counts/interval and the distribution will no longer be Gaussian. Using Logger Pro, record data for a run time of 1 minute. Record the distribution of events and calculate the expected distribution of events for a Poisson distribution using the value of $\overline{N}$ calculated for this run. Present your results in a single graph comparing theory and experiment.

G. With the measuring interval set to 1/10 s and a run time of 20 s, record $\overline{N}$ and σ_N for a series of runs as the source is moved progressively further away from the counter so that $\overline{N}$ takes on a few values between 0.5 and 5. For each run observe how the shape of the distribution changes and note the location of the peak value as compared to $\overline{N}$. [The two no longer coincide since the distribution is not symmetric.] Make a plot of $\sqrt{\overline{N}}$ vs σ_N to verify the theoretical prediction that $\sigma_N = \sqrt{\overline{N}}$. Include the distributions of N in your report (use command subplot() to put all distributions on one page).

If the mean number of counts is still > 1 even for the farthest possible position of the source from the Geiger tube, you can reduce it more by flipping the sample (after flipping, the gamma rays will have to pass through a thin aluminum plate supporting the sample, which will additionally reduce the count rate). Repeat measurement as above one more time, but with the source flipped. Plot its histogram. What is the mean value? Does it resemble a Poisson distribution closer now?

I. In your lab report, write up each section (C-G) separately in a manner complete enough so that your results can be followed by someone who does not have a copy of this lab handout. Give a description of the purpose of the section, describe the procedure, list the data in a table, describe the analysis, and give conclusions. Graphs and tables should be numbered and referred to in the text (i.e., "see Figure 4" or “see Table 3”, not "see one of the attached plots".) **Make sure to answer all questions.**

APPENDICES

1. HINTS FOR USING MATLAB

Read the help information on commands given in these hints.

To import a Logger Pro text file of exported data (first method):

Select the file in the Matlab directory listing window,
Right click on the file

- Import data …
- Next>
- Create vectors from each column using column names
- Finish

To import a Logger Pro text file of exported data (second method):

Use command importdata(filename).
Look at help importdata for example use.
This method is useful in a '.m' file that has "clear all".

To histogram Counts in bins of 1 between 25 and 60:

- x = 25:1:60;
- hist(Counts, x);
- Read the help on "hist" to obtain the counts in, and centers of, bins.

Use "normpdf" function to compute Normal distribution function values

- y = normpdf (x, mu, sigma); % 'See the help page with 'doc normpdf'

Use "plot" function for xy-plot with data x and y (type 'doc plot' in the command line for the usage detail)

Use "hold on" command to overlay multiple plots on the same graph (type 'doc hold' in the command line for the usage detail)

To histogram Counts in bins of 2*sigma centered on Nmean

- x = (Nmean – 6*sigma): (2*sigma): (Nmean + 6*sigma);
- hist(Counts, x);

For semilog plots of data with errorbars, use the errorbar() and log10() functions, e.g.

- X = log10(n)
- E = sigma ./ sqrt(n)
- errorbar(X, Nbar, E, 'o')
- Unfortunately, the semilogx() function does not plot errorbars, and the errorbar function cannot do log axes. A combination of both using hold almost works but not quite.

For plotting a straight line on the errorbar() plot, note this convenient construction for making an array with all values the same

- 45.3*ones(size(Nbar));

Use "poisspdf" function to compute Poisson distribution function values

- y = poisspdf (x, lambda); % 'See the help page with 'doc poisspdf' command

If you attempt to manually compute Poisson distribution, the factorial function, factorial(n), can be used in calculating the expected Poisson distribution. But it does not allow for n to be an array. So, a "for" loop must be used (see help for command "for"), e.g.,

```
for i = 1:length(n)
   poisson(i) = mu^n(i) * exp(mu) / factorial(n(i));
end
```

2. HINTS FOR USING LOGGER PRO

To setup the program, go to:

Menu EXPERIMENT
- CONNECT INTERFACE
- CONNECT ON PORT: select COM1

Button LABPRO
- Drag RADIATION RM-BTD to DIG/SONIC1 box
- CLOSE

Menu EXPERIMENT
- DATA COLLECTION …
- Tab COLLECTION
- LENGTH 30 seconds
- Sample at Time Zero: off
- SAMPLING RATE 0.5 secs/sample
- DONE

To collect data, click button COLLECT.

To export data, go to:

Menu FILE
- EXPORT DATA AS TEXT.