Simulated Radioactive Decay using Dice: Poisson Statistics

Objective

In this experiment, you will be simulating radioactive decay by rolling dice and taking a certain number away each roll based on probability and chance. Your goal will be to determine the decay rate for two different "species" and compare with theory.

Your successful analysis will include plotting and linearizing data, determining error from Poisson statistics, propagating error, and using a weighted least-squares curve fit to determine the decay rate and corresponding uncertainty in your measurement.

Theory and Background

In a radioactive decay process, the rate at which atoms are lost is proportional to the number of atoms

$$\frac{dN}{dt} = -\lambda N \,, \qquad (1)$$

where the decay constant λ is related to the probability of decay for that species of atom. We can solve this simple equation to obtain the typical exponential decay law for radioactivity

$$N(t) = N_o e^{-\lambda t} \,, \qquad (2)$$

for which N_o is the initial number of atoms.

By rolling a large set of dice again and again, and taking away the sixes for example after every roll (constant probability 1/6), we end up loosing dice at a rate proportional to the number of dice. The equation for the decay of dice is simply (2) where N_o now represents the initial number of dice, and t corresponds to the number of rolls.

The motivation for simulating radioactive decay for this experiment is to learn the rich analysis involved in counting experiments, without being bogged down by complex nuclear physics. Please pay attention to this lab. You will be doing it again with real particles (decaying muons) next term in PH351.

Note that counting experiments obey Poisson statistics. This is where events occur at random, but at a definite average rate. Examples in physics besides radioactive decay, are particle lifetime measurements, and photon counting. Like the Gaussian distribution, the Poisson distribution is a special case of the Binomial distribution (the one that determines probability of successes of coin flips). In fact, when the mean of a Poisson distribution becomes large, it approximates a Gaussian. The most important fact to remember about the Poisson distribution, however, is that the standard error is the square root of the mean

$$\sigma_x = \sqrt{< x >} \qquad (3)$$

Be careful not to confuse the statistical distribution with the trend of the data. In this case, Poisson statistics describes the distribution of **each** data point, not the trend line for the set. In this experiment, we expect an exponential decay for the trend whose individual data points are distributed about the mean (the one measurement we take for each point) by Poisson statistics.

The background reading necessary for the analysis can be found in Taylor Chapters 3, 8, and 11.

Procedure

Take a bucket of 200 dice an roll them in a place where they can be easily counted, picked up, and rolled again.

For your first data set you will be taking away the sixes after each roll until you have no dice left. Make sure to count the sixes you take away for each roll and mark them on the corresponding spreadsheet. If you find a convenient spot near one of the computers, you can just enter the data directly into the spreadsheet. If not, the old fashioned method of paper and pen will work fine too. You will find the dice entirely decay in about 30 rolls. Make sure that you set the dice to be eliminated far enough aside from the group to be rolled again so that they do not mix.

Based on the probability of obtaining a six, how many dice N do you expect to have remaining after $t = 1\ roll$? You can insert these values into (2) to determine the theoretical decay constant λ. With a slight bit of further manipulation, you can determine the half-life of the species in number of *rolls*. What is it?

For your second species, you will repeat the same procedure, but this time taking out both the twos and threes after each roll. What is the probability now for getting a two or a three per dice on any given roll? Again count the number for each roll, set these aside, and roll the remaining again. Keep rolling until you have taken away all the dice. You will find that this data set decays much faster. How much faster? Without using a calculator, use common sense and what you know about the new probability, compared to the first species, to guess the half-life of this species. What is the theoretical decay constant λ (feel free to use the calculator on this one)?

Similar to the Dartboard experiment, you will be using a spreadsheet to aid in the management of a large data set. After analyzing the data, you will determine an experimental value and uncertainty for the decay constant of both species of dice $\lambda \pm \sigma_\lambda$.

Data Analysis

The final product of your analysis will be one plot. On this graph you will show a linear representation of your raw data including appropriately calculated error bars on each data point for both species of dice decay. You will superimpose over each linearized data set (each species), the best-fit line calculated from a weighted least-squares fit. From the slope of these two lines you will be able to determine the decay constants λ for each respective species. Also from your fit routine, you will determine the corresponding uncertainties σ_λ in the slopes of these best-fit lines. As a check, you will compare the decay rates to those predicted by theory.

Part 1. Using the Spreadsheet to linearize data and propagate error

Using one of the PC's in lab, login as *phlab*, password *electron*. Click on the **Start** button, then **Lab Experiments**, **PH 350**, **Dice07**. Insert your data for the number of sixes you obtain after each roll into Column C of the spreadsheet. I have performed the simple math for you so that Column D shows the number of N dice remaining after each roll. Note that the total at the bottom of Column C should be 200 and that the last data point in Column D should be 0.

Using what you know about counting (Poissson) statistics, enter in the proper formula into Excel for the uncertainty so that Column E gives the values of uncertainty in each data point of the number of dice N remaining. This column of data corresponds to $N(t)$ in (2). Column F will give the corresponding fractional uncertainty. Are you surprised? As you can see, for low numbers of counts, the fractional uncertainty is pretty bad. Keep this in mind when plotting the data. A rule of thumb (though not hard and fast) is to ignore counts below 10. The fractional uncertainty here is just too large for the data points to be helpful. If you wanted to make the experiment 100 times more precise, how many dice would you need to start with?

You may plot the data in Column D against Column B, which corresponds to plotting $N(t)$ vs. t, to check by eye that it roughly looks exponential decay. For the data to be analyzed for your write up, you will linearize it and plot it.

Linearization allows a nice way to plot and analyze data that are described by complex functions. By linearization, you perform whatever math you need on an equation to turn the complex function into a simple line. This simplifies the analysis considerably for you. It also often increases the accessibility for your reader who can focus on the fact that you should be obtaining a plot of a line for results, should your model be correct.

To linearize (2) which describes the model here, we simply take the natural logarithm of both sides to obtain the line

$$y = \ln(N) = -\lambda t + \ln(N_o) \quad (4)$$

As you can see this is quite convenient since it allows us to determine the decay constant λ by simply finding the slope of the line we plot.

To graph your linearized data, you will need to plot $\ln(N)$ vs. t, so add a Column to Excel that gives the values $\ln(N)$. We are almost ready to plot, but need to consider the uncertainty. Earlier, you determined the uncertainty in the number of N dice remaining per roll using Poisson statistics. However, the error in $\ln(N)$ will not be the same as the error in N. In order to find the correct uncertainty, you need to propagate the error using the new relation between the data, $y = \ln(N)$. See Taylor Chapter 3 for information on propagation of errors. What is the new error σ_y? Does it look familiar? Add yet another column to Excel that represents the new uncertainties σ_y.

Part 2. Plotting your data

You will be using the program *Origin* 7.5 to produce high quality graphs of your results. **Graphs must be properly labeled and titled.**

Import your Excel data into Origin using the cut and paste function. For each species of dice, you will want three columns of data: number of rolls t, the natural log of the counts $y = \ln(N(t))$, and the uncertainties in the *y-data* σ_y. The values σ_y will be the error bars on your plot and will give the weighting for your least-squares fit.

Note that for this class you will need to include uncertainties on all measurements, as well as error bars on every data point in a graph. We will assume for this experiment (and generally most experiments) that there is negligible uncertainty in our dependent variable to be plotted on the *x*-axis.

To plot a graph in Origin that in includes error bars, go to **Plot**, **Special Line**, **y Error**. In the window prompt that pops up, you will be able to sort your columns by checking a box to denote what Column should be used for *x*-data, *y*-data, and the error bars (uncertainty). As mentioned before, eliminate data points whose raw count value N is less than 10. Do the data and error bars look as you expect? We will be quantitative about the analysis in the next section.

Part 3. Curve-Fitting

If you are familiar with Origin or other graphing/analysis software packages, it is quite tempting to fit your data with a couple of clicks of the mouse. However, it is very important that we know how these analysis packages work so we can decide which of the tools are appropriate for a particular set of data and which are not. For this part of the lab you will be making your own least-squares fitting algorithm. You may use any programming software you like including Excel (which can do the job quickly and just fine). You will need to include a hard copy of your code in the appendix of your write-up. If you use Excel, also send an electronic copy to the instructor since the cell equations do not print out.

Chapter 8 in Taylor for the details and equations regarding the method of least-squares. Because your error bars will vary in size from data point to data point, you will **need** to use the slightly more complicated **weighted least-squares fit.**

For a generic data set that can be modeled as a line

$$y_i = mx_i + b, \quad (5)$$

the method of least squares should give four calculated values: the best fit of the slope m, the corresponding error in the slope σ_m, the best fit of the *y*-intercept b, and the corresponding error in the intercept σ_b. Make sure to explain in words in your write-up the intuition of the method of least-squares? What gets minimized and why? What are the assumptions about the data when you use a least-squares fit? Do they conflict with any of our assumptions in this experiment?

Also, how did your values for the slope and error in the slope turn out? Did the theoretical value for the decay constant fall within the range you calculated , $\lambda \pm \sigma_\lambda$? If not, why? Give a common sense upper-bound estimate for the percent error in the decay constant. Do this based on the fractional uncertainty of the linearized data (note this is not the same as the fractional uncertainty we calculated earlier).

Once you have found the parameters from your fit, use them to superimpose a the best-fit line on top of your data. To do this, you will use the "function" option in Origin. Choose **File, New, Function, OK**. *Carefully* type in the equation for the best-fit line you wish to plot using the parameters for you least-squares algorithm. To overlay the function on your experimental data, use the "layer icon." This is the button is labeled "1" in the left-hand corner of your particular graphing window. Double-clicking this icon will allow you to combine elements of different plot windows on the same graph. Be sure to uncheck "rescale on ok" before plotting.

Now for the moment of truth...Let's see if we can trust Origin. Let's check Origin's built-in curve fit program to see if its linear fit uses the method of least-squares, and if it agrees with our calculation. Go to **Fitting**, **Tools**, **Linear Fit**. Under the "option" menu make sure to check the box that says "error as weight". Leave everything else unchecked. After hitting o.k., the best-fit line will be plotted, and the analysis from the fit will appear in the lower right-hand window. Does your best-fit line match Origin? Check the fit parameters just to be sure you agree. If so, you may use Origin's curve fit program for all other labs in this course. If not, why is it different? Who is right? In your plot, does the best-fit line stay within the error bars?

Use Origin's fit program or your own curve fit program to play around with adding or subtracting data points. What if you just use the first eight data points for the fit? We know that the higher the counts, the lower the uncertainty. So maybe we should use a small number of data points where the counts are the highest? However, we will be averaging over a smaller number of points. Does one overtake the other? Basically, would three really good points be better than three really good points and three more mediocre ones?

Make sure you do repeat this procedure for the second species of dice. Put both the data and curve-fits on the same single plot. Make sure to include your equation of your best-fit lines on the plot near the respective curves. Note you may want to combine data sets to one plot at the very end using the "layer icon" function in Origin.

Extra Credit

You will get extra credit if you say something about the quality of your fit. There are different ways to do this and not everyone agrees on a single method. Choose a way to comment on your fit and write a small program (like you did for the fit itself) to say something quantitative about the goodness-of-fit.