EAX - Error Analysis Exercise

All pages in this lab

I. Error Analysis Exercise

II. See Error Analysis Notes [1] for Optical Pumping

Contents

- 1 Note
- 2 References
- 3 Introduction
- 4 Problem Set
  - 4.1 Problem 1
  - 4.2 Problem 2
  - 4.3 Problem 3
  - 4.4 Problem 4
  - 4.5 Problem 5
  - 4.6 Problem 6
  - 4.7 Problem 7

Note

All new Physics 111 Advanced Lab students are required to complete this assignment at the beginning of the semester. It will be graded on 50 points basis; a late turn-in is allowed only with the instructor's approval before the due date. Don't jeopardize your grade on your first experiment by being late with this assignment. You need to know how to handle errors before you start a laboratory experiment.

Important: View the video introduction to error analysis [2] (you need to use your Berkeley email to access this). The Error Analysis Exercise due date is Advanced Lab Report Due Dates. [3]

References

[Books available online with UC Berkeley authentication at [4]]

1. P. Bevington, "Data Reduction and Error Analysis for the Physical Sciences" [5], McGraw-Hill. [An old standard that is pretty dry but straightforward. Chapter 5 is particularly important.]
In the 111-lab, the experiment does not end when you have finished collecting your data. In many labs, you will be required to perform a detailed analysis of the data you have acquired. The point of any scientific experiment is to make quantitative statements about the properties of the physical world. A common question is, are your measurements consistent with a particular theory or not? This question can only be answered by careful analysis, including both systematic uncertainties and statistical error.

The goals of this exercise are twofold. One is to familiarize students with the basics of error analysis. Ideally, this will serve as a guide during the acquisition and analysis of data throughout the advanced lab. The second goal is to introduce students to the Matlab numerical computing environment, which you will be using throughout the semester.

Before starting on EAX, please look over the Intro to Matlab section.

Problem Set

Problem 1

We want to measure the activity (number of decays per second) of a radioactive source so that we can use it to calibrate the equipment of the gamma-ray experiment. We use an electronic counter and a timer to measure the number of decays in a given time interval. In round numbers we measure 3000 decays in 2 minutes of observation. For how long should we measure (in seconds) in order to determine the activity with an uncertainty of 1.0%? Explain.

Problem 2

You are given two measurements of distance and their associated errors:

\[ A \pm \sigma_A \] and \[ B \pm \sigma_B \]. Calculate the error in the

(a) total distance
\[ A + B, \]

(b) difference
\[ A - B, \]

(c) the perimeter
\[ 2A + 2B, \]

(d) the product
\[ A \times B. \]

(e) the ratio
Problem 3

In this problem we will be generating and analyzing lists of normally distributed random numbers. The distribution we are sampling has true mean 0 and standard deviation 3.

1. If we sample this distribution N = 11 times, what do we expect the mean to be? How about the standard deviation? What is the uncertainty with which the mean is determined?

2. Using Matlab (or Python), generate a list of N = 11 normally distributed random numbers (the command randn(N,M) will generate M lists of length N with mean zero and standard deviation of 1). Calculate the mean, standard deviation and the error on the mean. Is this what you expected?

3. Now find the means, standard deviations, and uncertainties on the means for each of M = 2000 experiments, each with N = 11 measurements. Plot a histogram of the means from each experiment. How many experiments are compatible with the true mean of 0, i.e. the mean deviates less than the uncertainty of the mean? How many of the means are within 2 times the uncertainty? Is this what you expected? Compare the uncertainty on the means to the standard deviation of the means. This is why one often says that an estimate is within one or two sigma.

4. Now repeat questions 1-3 for N = 20, 40, 80, 800.

Problem 4

In this problem we will repeat the above process, but now using lists of exponentially distributed random numbers. The probability of selecting a random positive (x>0) number between x and x+dx is \( \propto e^{-x}dx \).

1. What do you expect the mean of the distribution to be? What do you expect the standard deviation to be? (Note: The standard deviation is defined exactly as it is for a normal distribution, but the "1 sigma = 68%" rule no longer applies to an exponential distribution). What do you expect the error on the mean for an experiment with N = 300 random samples to be? Given M = 600 experiments with N = 300 random numbers, what do you expect the distribution of means to look like? What is the uncertainty on the mean determined from the M=600 experiments?

2. Make a list of N = 300 exponentially distributed random numbers (Hint: this can be done starting with a uniform distribution of random numbers, or using commands in your programming language of choice). Calculate the mean and standard deviation.

3. Make M = 600 lists of N = 300 exponentially distributed random numbers. Make a histogram of the means. Does the distribution of means look as you thought? What is the standard deviation of the means. Does this agree with what you thought?

4. Repeat the previous steps for N = 3000 & 60000. Does the error on the mean scale as you thought? This is a demonstration of the Central Limit Theorem.

Problem 5

You are given a dataset (Peak.zip) from a gamma-ray experiment consisting of ~1000 hits. For each hit, the energy of the gamma-ray is recorded. We will assume that the energies are randomly distributed about a common mean, and that each hit is uncorrelated to others. Read the dataset from the enclosed file and:

1. Produce a histogram of the distribution of energies. Choose the number of bins wisely, i.e. so that the width of each bin is smaller than the width of the peak, and at the same time so that the number of entries in the most populated bin is relatively large. Since this plot represents randomly-collected data, plotting error bars would be appropriate (hint: use errorbar function in Matlab or Matplotlib)
2. Compute the mean and standard deviation of the distribution of energies and their statistical uncertainties.

3. Fit the distribution to a Gaussian function, and compare the parameters of the fitted Gaussian with the mean and standard deviation computed above.

4. How consistent is the distribution with a Gaussian? In other words, compare the histogram from (1) to the fitted curve, and compute a goodness-of-fit value, such as $\chi^2 / df$.

**Problem 6**

In the optical pumping experiment we measure the resonant frequency as a function of the applied current (local magnetic field). Consider a mock data set:

<table>
<thead>
<tr>
<th>Current $I$ (Amps)</th>
<th>0.0</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
<th>1.2</th>
<th>1.4</th>
<th>1.6</th>
<th>1.8</th>
<th>2.0</th>
<th>2.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency $f$ (MHz)</td>
<td>0.12</td>
<td>0.63</td>
<td>1.18</td>
<td>1.97</td>
<td>2.37</td>
<td>3.17</td>
<td>3.63</td>
<td>4.36</td>
<td>4.48</td>
<td>5.40</td>
<td>6.21</td>
<td>6.98</td>
</tr>
</tbody>
</table>

1. Plot a graph of the pairs of values. Assuming a linear relationship between $f$ and $I$, determine the slope and the intercept of the best-fit line using the least-squares method with equal weights, and draw the best-fit line through the data points in the graph.

2. From what your lab partner knows about the equipment used to measure the resonant frequency, they hastily estimate the uncertainty in the measurement of $f$ to be $\sigma_f = 0.024$ MHz. Estimate the probability that the straight line you found is an adequate description of the observed data if it is distributed with the uncertainty guessed by your lab partner. (Use Matlab to calculate it or look it up in a table. For example, see table C-4 in Bevington). What can you conclude from these results? Repeat the analysis assuming your partner estimated the uncertainty to be $\sigma_f = 0.16$ MHz. What can you conclude from these results?

3. Assume that the best-fit line found in the previous exercise is a good fit to the data. Estimate the uncertainty in measurement of $y$ from the scatter of the observed data about this line. Again, assume that all the data points have equal weight. Use this to estimate the uncertainty in both the slope and the intercept of the best-fit line. This is the technique you will use in the Optical Pumping lab to determine the uncertainties in the fit parameters.

4. Now assume that the uncertainty in each value of $f$ grows with $f$: $\sigma_f = 0.06 + 0.05 \times f$ (MHz). Determine the slope and the intercept of the best-fit line using the least-squares method with unequal weights (weighted least-squares fit).

**Problem 7**

1. In the muon lifetime experiment we obtain a histogram for the decay rate as a function of the time after the muon enters the detector and announces its presence. We expect the distribution (the histogram) to be described by an exponential function. Rather than fitting with an exponential function, it is more convenient to plot the logarithm of the decay rate as a function of time and then fit a straight line to it. Since each data point $(x_i, y_i)$ has a statistical error, $\sigma_i$, associated with it, qualitatively, what happens to these errors when the semi-log histogram is plotted? Explain and illustrate. what happens, quantitatively? Assume $\sigma_i$ is reasonably large.

2. In a separate experiment, you find that $\log E_0 = 1.9 \pm 0.5$. What is the value of
\( E_0 \) and the experimental bounds? (Note that 0.5 is not small compared to 1.9).

Source URL: http://experimentationlab.berkeley.edu/EAX

Links
[1] http://experimentationlab.berkeley.edu/ErrorAnalysisNotes#Data_Analysis