

# EAX\_spring2022

January 17, 2022

*Updated by LLP on 1/17/22*

*Updated by WLH on 8/24/21*

## 0.1 EAX - Error Analysis Exercise

### 0.1.1 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 (you need to use your Berkeley email to access this). The Error Analysis Exercise due date is listed under Advanced Lab Report Due Dates.

### 0.1.2 References

[Books available online with UC Berkeley authentication]

P. Bevington, "Data Reduction and Error Analysis for the Physical Sciences", McGraw-Hill. [An old standard that is pretty dry but straightforward. Chapter 5 is particularly important.]

A. C. Melissinos and J. Napolitano, ["Experiments in Modern Physics, 2nd Edition"], Academic Press (2003).

W. H. Press, et al., ["Numerical Recipes in C:] The Art of Scientific Computing, 2nd Edition", Cambridge University Press (1992); refer to Ch. 14—"Modeling of Data". [The Numerical Recipes in Pascal or FORTRAN books contain identical information. This book is the standard reference for doing scientific work on computers. Chapter 14 has a good introduction to the method of maximum likelihood, chi-square fitting, modeling data in general, error estimates of fit parameters, and, important for later experiments, the Monte Carlo method of simulation.]

I. G. Hughes and T. P. A. Hase, Measurements and their Uncertainties, Oxford University Press (2010). [This is a well-written thin book that covers many basic concepts of statistics, extremely useful for this course.]

Louis Lyons, "A Practical Guide to Data Analysis for Physical Science Students" (1991) Cambridge Press; QC33.L9 1991

Yardley Beers, "Introduction to the Theory of Error"; ADDISON-WESLEY PUBLISHING (1957) QA275 B4 1957;

Reprints and other information can be found on the Physics 111 Library Site.

### 0.1.3 Introduction

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 Python computing environment, which you will be using throughout the semester.

Before starting on EAX, please look over either the Python Tutorials [https://github.com/avirukt/intro\\_python](https://github.com/avirukt/intro_python) or the Intro to Matlab section. Additional resources for using Python are <http://pythontutor.com/> and <https://datahub.berkeley.edu/>

## 0.2 Problem Set:

**Problem 1: Poisson statistics** We want to measure the activity (number of decays per second, a unit known as Becquerel) 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 obtain 1000 decays in 10 minutes. How long does it take (in seconds) in order to determine the activity with a statistical uncertainty of 2%? Explain.

Enter the calculation in the cell below. Add a "Markdown" cell to explain your calculation

### Problem 2: Error propagation

- a) You are given the measurements of two sides of a rectangle  $A$  and  $B$  with the associate errors  $\sigma_A$  and  $\sigma_B$ , respectively. Assuming  $A$  and  $B$  are uncorrelated, calculate the error in the: > (i) sum  $A + B$  (ii) difference  $A - B$  (iii) the perimeter  $2A + 2B$  (iv) the area  $A \times B$  (v) the aspect ratio  $A/B$
- b) 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  $(x_i, \log y_i)$  is plotted? Explain and illustrate. what happens, quantitatively? Assume  $y_i$  is reasonably large.

- c) In a separate experiment, you find that  $\log_{10} E_0 = 1.7 \pm 0.7$  (at 68% confidence level, CL). What is the value of  $E_0$  and the experimental bounds at 68% CL? (Note that 0.7 is not small compared to 1.7).

Write down the calculation in the "Markdown" cell below, or enclose a separate file. Show detail of your calculation.

**Problem 3: Central Limit Theorem** Here we will verify the Central Limit Theorem and reproduced a wikipedia plot ([https://en.wikipedia.org/wiki/Central\\_limit\\_theorem#/media/File:Dice\\_sum\\_central\\_limit\\_theorem.svg](https://en.wikipedia.org/wiki/Central_limit_theorem#/media/File:Dice_sum_central_limit_theorem.svg))

- Write a function that returns  $n$  integer random numbers, uniformly distributed between 1 and 6, inclusively. This represents  $n$  throws of a fair 6-sided die. The value that comes up at each throw will be called the "score".
- Generate a distribution of 1000 dice throws and plot it as a histogram normalized to unit area. Compute the mean  $\mu_1$  and standard deviation  $\sigma_1$  of this distribution. Compare your numerical result to the analytical calculation.
- Generate 1000 sets of throws of  $N = 2, 3, 4, 5, 10, 20, 30$  dice, computing the total sum of dice scores for each set. For each value of  $N$ , plot the distribution of total scores, and compute the mean  $\mu_N$  and standard deviation  $\sigma_N$  of each distribution. This should be similar to the plot at the link above.
- Plot the standard deviation  $\sigma_N$  as a function of  $N$ . Does it follow the Central Limit Theorem?

**Problem 4: gamma peak** You are given a dataset (peak.dat, in the current directory) from a gamma-ray experiment consisting of  $\sim 1000$  hits. Each line in the file corresponds to one recorded gamma-ray event, and stores the the measured energy of the gamma-ray. We will assume that the energies are randomly distributed about a common mean, and that each event is uncorrelated to others. Read the dataset from the enclosed file and:

- Compute the mean and standard deviation of the given distribution of measured energies. These are estimates of the mean and standard deviation of the underlying parent distribution. Determine the uncertainties for these estimates of the mean and variance of the parent distribution. (Hint: The fractional uncertainty in the estimate of the standard deviation for Gaussian distributed data is given by Equation 2.8 in Reference 4, I. G. Hughes and T. P. A. Hase, "Measurements and their Uncertainties". For the purposes of this calculation, you can assume the sample data is drawn from a Gaussian distribution).
- 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.
- Fit the distribution to a Gaussian function using an unbinned fit (Hint: use `scipy.stats.norm.fit()` function), and compare the parameters of the fitted Gaussian with the mean and standard deviation computed in (b)
- Fit the distribution to a Gaussian function using a binned least-squares fit (Hint: use `scipy.optimize.curve_fit()` function), and compare the parameters of the fitted Gaussian and

their uncertainties to the parameters obtained in the unbinned fit above.

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

**Problem 5: Optical Pumping** In the optical pumping experiment (OPT) we measure the resonant frequency of a Zeeman transition as a function of the applied current (proportional to the applied magnetic field). Consider a mock data set:

Current I (Amps)

0.0

0.2

0.4

0.6

0.8

1.0

1.2

1.4

1.6

1.8

2.0

2.2

Frequency f (MHz)

0.13

0.60

1.21

1.94

2.47

3.07

3.83

4.16

4.68

5.60

6.31

6.55

- a) Plot a graph of the pairs of values. Assuming a linear relationship between  $I$  and  $f$ , 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.
- b) From what s/he knows about the equipment used to measure the resonant frequency, your lab partner hastily estimates the uncertainty in the measurement of  $f$  to be  $\sigma(f) = 0.07$  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. (Hint: use `scipy.stats.chi2` class to compute the quantile of the chi2 distribution). What can you conclude from these results?
- c) Assume that the best-fit line found in Part 1 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.
- d) Now assume that the uncertainty in each value of  $f$  grows with  $f$ :  $\sigma(f) = 0.07 + 0.03 * f$  (MHz). Determine the slope and the intercept of the best-fit line using the least-squares method with unequal weights (weighted least-squares fit)