PHSX 315 Assignment 5

Due by Fri. March 31st at 1 PM

Learning Goals: Fitting models to data using function minimization software with consequent hypothesis testing, and parameter estimation. Adoption of appropriate input argument handling infrastructure. Use of χ^2 and likelihood. It is expected that you use iminuit to access the Minuit function minimization software.

Problem 1 Adapt the Fitting/Chisqfit.py example in ClassExamples so that at least the four main parameters defining the toy data and underlying true model are configurable at run time with the default values being picked up if none are specified. Highly recommend using argparse to achieve this. The four main parameters are

- the random number seed
- the number of data points
- the true value of the model intercept parameter (c)
- the true value of the model slope parameter (s)

Figure out how to customize the plot to also include a plot title, and x-axis, and y-axis labels, and include the plot in your writeup. What is the result of your fit for a random number seed of 102. Run the program 100 separate times with random number seeds ranging from 1 to 100. How many times do you find a p-value of less than 0.05?, and what are the corresponding random number seeds? What is the smallest p-value that you observe? Is there any good reason to think that such results mean that the straight-line model is inappropriate?

Problem 2
The period of a simple pendulum does depend a little on the oscillation amplitude. The effect is relatively small and especially so for small amplitudes.

In the files, LowPrecisionPeriodData.dat, and HighPrecisionPeriodData.dat, you will find two data-sets. Each has measurements of the period (in seconds) of a simple pendulum for different launch angles. The low precision data set has measurement uncertainties on the period ranging from 0.01s to 0.02s while the uncertainties on the high precision data set are one micro-second. Both data sets use the same underlying exact model.

You should fit the data with appropriate models using the fit chi-squared to judge what is acceptable. For each model you fit you should report the fit chi-squared, number of degrees of freedom, and p-value. You will likely find that the low precision data set can be fitted satisfactorily relatively easily and you should refrain from "over-fitting".

On the other hand for the high precision data, it is expected that all but the correct model will need several parameters. The correct model only needs one parameter!

Problem 3

Poisson probability distributions are often mandatory in experiments where small numbers of events are expected. The following table (Table 1) summarizes the "event-counting" experimental results from Table 5 of the paper entitled "Measurement of Single Photon Production in Electron-Positron Collisions near the Z^0 Resonance" by the OPAL Collaboration, published in Z. Phys. C 65 (1995) 47-66. I was the primary author of this paper. It took four years and is probably my most useful scientific contribution. The measurement of the number of events produced consistent with a single photon and nothing else (assumed to be invisible particles) was used in figuring out how many neutrinos there are in the Universe. You can too!

Year	E_{CM} (GeV)	n_{obs}	μ_{bkg}	$\mu_{\mathrm{SM}}^{N_\nu=1}$	μ_{W}
1990	88.252	2	0.3	0.27	0.13
	89.256	1	0.2	0.30	0.16
	90.254	2	0.2	0.27	0.11
	91.244	21	2.5	6.22	1.70
	92.234	7	0.4	2.01	0.30
	93.238	11	0.5	4.24	0.33
	94.234	22	0.5	5.99	0.19
1991	88.480	3	0.5	0.54	0.25
	89.470	6	0.6	0.74	0.35
	90.226	11	0.8	1.15	0.50
	91.243	80	7.9	19.35	5.28
	91.970	16	0.8	3.31	0.61
	92.968	22	0.8	6.46	0.58
	93.716	35	1.0	10.22	0.35
1992	91.299	208	20.1	56.29	14.79

Table 1: Single photon event counts (n_{obs}) from the OPAL experiment at 15 different center-of-mass energies. For each data-taking point, the expected number of events from background processes, μ_{bkg}, and the number of expected events per neutrino generation, $\mu_{\mathrm{SM}}^{N_\nu=1}$, from Z decays alone, and expected contributions from W related processes, μ_{W}, are listed.

The expected total number of events expected at each of the 15 data-taking points, i, is given by,

$$\mu_i(N_\nu) = \mu_{\mathrm{bkg},i} + \mu_{\mathrm{W},i} + N_\nu\ \mu_{\mathrm{SM},i}^{N_\nu=1}$$

These expected event counts, depend on one single parameter, the number of light neutrino generations, N_ν, that is expected to be exactly equal to 3 within the standard model of particle physics. This can then be used in a Poisson likelihood consisting of terms where one evaluates the Poisson probability of observing n_{obs} events at each point i given the expected event count, μ_i, which depends on N_ν. By maximizing the likelihood one can determine the best estimate for the real number N_ν and its uncertainty. Remember when minimizing the negative log-likelihood, $-\ln L$, one should set the error definition to a change in the objective function of 0.5 (or more conveniently choose to minimize, $-2\ln L$, where the error definition is identical to the chi-squared case and corresponds to a change in the objective function of 1.0.) You may want to start from ClassExamples/Fitting/NLLfit.py.

Problem 4

Use problem 1 as the basis for this. Define a new toy model using a quadratic, namely,

$$y(x) = ax^2 + bx + c$$

with model parameters, a, b, c. Set $a = 1$, $b = 2$, $c = 3$, and see how well one would measure the three parameters when fitting data that are generated from this model.

Sometimes it can be advantageous to use orthogonal polynomials when fitting with polynomial function models. A common choice appropriate to [-1, 1] is to use Legendre polynomials. Fit the same data with the sum of the first three Legendre polynomials. Here the fitted model parameters will be the coefficients of each Legendre polynomial. Do you get very different results for the parameter correlations?

You should upload to your repository a summary of your findings, a copy of your code, example results from running the code, and appropriate figures. In these cases plots of the data with error bars and super-imposed curves and fit parameters corresponding to the models would be appropriate.