Computational Physics: χ^2 Fitting

Ujwal Kumar

March 24, 2021

1 Introduction

In this report, we discuss a goodness of fit measure called the χ^2 test and its application to modelling data. When fitting models to data, we aim to also understand how good or bad that fit is. The χ^2 test measures the deviation of each datapoint (including the y-uncertainty on the datapoint) from the model's prediction at that point.

We define the χ^2 value to thus be:

$$\chi^2 = \Sigma_i \frac{(P(x_i) - y_i)^2}{\sigma_i^2} \tag{1}$$

In an ideal scenario, the χ^2 test equals 1 per degree of freedom (number of data points - number of variables). In other words, the "reduced χ^2 " is 1. Values much larger than 1 may indicate a bad fit whereas values smaller than 1 may either indicate over-fitting or overly-conservative uncertainties. We thus design a program that implements a method to minimize the χ^2 value for a random set of datapoints with coefficients and uncertainties defined in the following polynomial form:

$$P(x) = \Sigma_k^M \alpha_k x^k \tag{2}$$

We will study this problem in the context of a cannonball dropped out of a plane.

2 Problems

2.1 Question 1

In this problem, we fit the first six seconds of data to the form:

$$y = y_0 + v_0 t - \frac{1}{2} g t^2$$

Since the datafile, cannonball.dat, starts at t=0 and is incremented by 0.5s, the first 6 seconds is represented by the first 13 lines in the code. This yields a best-fit polynomial:

$$y = 16380.4 + 4.57t - 5.52641t^2$$

With a χ^2 value of 6.91. We can extract the gravitational constant using the coefficient of t^2 to get $g \approx 11.1 m/s^2$.

This dataset contains 13 datapoints and there are 3 parameters to fit in the model yielding 10 degrees of freedom. Therefore, the reduced χ^2 is 0.691 which indicates that the code has overfit the data.

We estimate confidence using the distribution of ^{2} for varying samples. This is given by the formula:

$$\int_{\chi^2} f(\chi^2) d\chi^2 \tag{3}$$

This can be evaluated using Mathematica's `CDF[ChiSquareDistribution[i], \Chi^2]` yielding a value of 0.670 or a 67% confidence in the data.

2.2 Question 2

Next, we utilize the entire dataset and repeat the procedure.

We then get a best fit polynomial:

$$y = 16395.6 - 3.4245t - 4.692395t^2$$

With a χ^2 value of 140.23 and a reduced χ^2 of 1.219 providing a better fit than before, but still overfitting. We can extract the gravitational constant using the coefficient of t^2 to get $g \approx 9.38479 m/s^2$, which is closer to the expected value of $9.81 m/s^2$, but is now underestimating the value. The confidence yielded through the same procedure in Mathematica was only 5.5% which means that certain variables that model this data were not considered perhaps such as air resistance which can factor in more apparently over a larger timescale.

3 Conclusion

This report helped me build on previous studies I had conducted looking at ^{2} fits to various models such as the broadband model, thermal black body models, or other models to characterize gamma ray bursts in astrophysics. We used xspec as a framework to do that but I gained an understanding of how the fitting mechanism works from a backend perspective. I would love to delve into such problems deeper and I enjoy how previous use of linear algebra routines are still usable in the same way in these broader problems.