NumRep Report: Generating Data and fitting using a 2-dimensional PDF

Semester 1 2018

Due in by: 30th November 2018 by 23.59

1 Introduction

The theme of this report is to show you the difference between ”well behaved fits” and ”badly behaved fits.

This checkpoint is based around a data set corresponding to the decay of a radioactive particle X to a daughter particle D. This decay $X \to D$ has two components contributing, each with a slightly different lifetime.

These two components not only have a different lifetime, but also the angular distribution of D with respect to X differs for each component.

Therefore there are 2 quantities observed for each event:

- t: the decay time of X
- θ: the decay angle of D with respect to X

The first component has a PDF given by

$$P_1(t, \theta; \tau_1) = \frac{1}{N_1(\tau_1)}(1 + \cos^2\theta) \times \exp(-t/\tau_1)$$

The second component has a PDF given by

$$P_2(t, \theta; \tau_2) = \frac{1}{N_2(\tau_2)} 3\sin^2\theta \times \exp(-t/\tau_2)$$

where N_i are normalisation functions to be found.

The PDF of the decay is therefore given by a fraction F of P_1 and the remainder of P_2.

It is unimportant exactly what θ is geometrically - all that is important is that we observe 2 observable quantities per event and that there is a PDF which describes the distributions we observe.

2 Part 1: Generate and plot simulated data

Determine the normalisation functions N_1 and N_2

Write code to generate a set of 10000 random events with a distribution of t, θ given by the PDF described above.

The values of the physics parameters you should use are $F = 0.5$, $\tau_1 = 1.0$ and $\tau_2 = 2.0$. The limits of each observable are $t : [0, 10]$ and $\theta : [0, 2\pi]$.

Plot the t and θ distributions of the data you have generated.

Generate data for each component separately, i.e. with F set to 0 and then F set to 1. In each case, plot the t and θ distributions of the data you have generated. Comment on how easily (or not) a fit would be able to distinguish the components from each distribution alone.

3 Part 2: Perform a maximum likelihood fit to only the t data

For this part you should use the data set supplied to you in the file called **datafile-Xdecay.txt.**

The true values of (F, τ_1, τ_2) are of course unknown to you (mystery data).

Using only the t information, perform a maximum likelihood fit to determine the best fit values of the physics parameters (F, τ_1, τ_2) and their errors. To determine the errors in this part you should use the simplified method described in the lectures (varying one parameter whilst keeping all others fixed). Present the results in a table.

Note: The fit may try to make the fraction F go outside the physical range [0,1] This can lead to negative probabilities and hence the fit may crash if you take the log of a negative number. You may need to project against this, and if so you should describe how you do so in the report.

Note: The fit to the t data only may exhibit unstable behaviour. This is because the best fit values of the two lifetime parameters are very strongly (anti) correlated (i.e. if one goes down the other will tend to go up to compensate to achive the same NLL). Minimisers have a harder time with correlated parameters. Be prepared for this and if you have problems think of moving on to part 4 which is much better behaved, and then coming back to part 3 later.

4 Part 3: Perform a maximum likelihood fit to the full t, θ data

For this part you should use the data set supplied to you in the file called **datafile-Xdecay.txt.**

Using all of the t and θ information, perform a 2-dimensional maximum likelihood fit to determine the best fit values of the physics parameters (F, τ_1, τ_2) and their errors.

Present the results in a table in comparison to the results from part 3. Compare and comment on the results obtained here to those obtained in part 3

5 Part 4: Calculate the proper errors

The method of calculating errors you have used so far is simplistic. It works ok only if the physics parameters are totally uncorrelated. In this case however it is clear that the best fit values of the two lifetimes are correlated.

The correct way (always) to determine the error on a parameter is to vary that parameter about its best fit value as before, but then at every point re-perform the minimisation of all the other parameters so that the NLL can get as low as possible. You continually do this re-minimisation as you seek the point at which the NLL rises by 0.5.

If it helps you can think of it like this: the fit does its best to give you the best fit parameters. If now you tell it that you want one of the parameters to be something else, then the fit has every right to say "if you had told me that at the outset then I would have done a better job on the other parameters".

Determine the correct errors on the parameters determined in parts 3 and 4 by re-minimising as described above. Compare these errors to the simplistic errors you obtained earlier and comment on any differences you see.

You should write your own code to do this. If you are using Minuit it will give you these proper errors (it doesnt do the simplistic ones) along with the correlation matrix. You are free to check your results against this, and include this information in the report, but you cannot just submit the Minuit results alone.

6 Notes

1. You can use any minimiser, including scipy.optimize or Minuit or your own if you wish to.

7 Details of the report

The basic style should be that of

- Abstract
- Introduction
- Appropriate sections describing context, problem, algorithms and methods employed.
- Results
- Summary
- Appendix

The report is to be written so that it would be understandable by a scientist who has no knowledge of the course. Therefore it must describe the context, problem, methods applied and results in a self-standing way, which does not rely upon any other knowledge of the problem **or of the course**.

The report must have an appendix containing all of code written by you (there is no need to include any supplied software). This must be well documented, and there should be an explanation at the top of each main function and class. Other functions and methods should be documented in the code.

The report should be submitted on LEARN as a single pdf file including the Appendix.
The main body of the report (i.e. excluding the code appendix) should not exceed 8 pages.

Report marking criteria include:

- Structure of report
- Quality of communication to reader with no prior knowledge.
- Depth to which the problem has been addressed.
- Correctness of the results
- Suitable plots, tables and histograms
- Quality and completeness of presentation of results.
- Quality and documentation of the code