A comparison of well and badly behaved minima in parameter estimation

Sean O'Brien

Abstract

This is a meta-analysis of parameter finding techniques. Specifically maximum likelyhood fitting, which finds parameters of a PDF by minimising the NLL function. This technique has very broad application in scientific data analysis and its reliability and pitfalls should be understood.

In particular, the situation of mismatched PDF complexity to data content was examined. The behaviour and stability of fit minima was tested when full and partial data was used to estimate the parameters of a relatively complex PDF. The findings showed that when only the partial data was used the parameter estimation values differed significantly from when the full data was used and that the minima from the partial data was unstable i.e. the results depended on the minimisation initial guess. When the full data was used and the PDF complexity was matched by the data, the fit was stable and gave repeatable results.

Introduction

The data consisted of 100,000 points, each consisting of a t and θ measurement, the significance of which is irrelevant. The 3 parameter PDF (PROBABILITY DENSITY FUNCTION) $P(t, \theta; F, \tau_1, \tau_2)$ used to model the data was a linear mixing of 2 PDFs $P_1(t, \theta; \tau_1)$ and $P_2(t, \theta; \tau_2)$

$$P_1(t, \theta; \tau_1) = \frac{1}{3\pi\tau_1}(1 + \cos^2\theta)e^{-\frac{t}{\tau_1}} \qquad P_2(t, \theta; \tau_2) = \frac{1}{\pi\tau_2}\sin^2 e^{-\frac{t}{\tau_2}}$$

$$P(t, \theta; F, \tau_1, \tau_2) = FP_1(t, \theta; \tau_1) + (1 - F)P_2(t, \theta; \tau_2)$$

The paramters estimated were F, τ_1, τ_2. This was done by performing a maximum likelyhood fit. Which involves minimising the NLL (NEGATIVE LOG LIKELYHOOD) related to the parameters in question.

$$NLL(F, \tau_1, \tau_2) = -\sum_i \log P(t_i, \theta_i; F, \tau_1, \tau_2)$$

- Where each t_i, θ_i are points in our dataset.

Notes

- When appropriate, a method or expression used will be explained in the glossary and in first reference will be in THIS FONT
- Any function minimisation was performed my Minuit, packaged for python as iminuit `https://iminuit.readthedocs.io/` All other calculations were performed by my own python code, which is included in the appendix of this report.
- The code, data and plots are also available at `https://github.com/mrseanman/NumRecProject`

1 Investigation of $P(t, \theta)$

10,000 random (t, θ) data points were generated simulating the $P(t, \theta)$ PDF with various values for F, τ_1, τ_1 The random generator employed the BOX METHOD with the ranges set as $t \in [0, 10],\ \theta \in [0, 2\pi]$. The finite range of t meant that P_1 and P_2 had to be normalised differently such that. $P_1(t, \theta; \tau_1) = [3\pi\tau_1(1 - e^{\frac{-10}{\tau_1}})]^{-1}(1 + \cos^2\theta)e^{-\frac{t}{\tau_1}}$ and $P_2(t, \theta; \tau_1) = [\pi\tau_2(1 - e^{\frac{-10}{\tau_2}})]^{-1}\sin^2\theta e^{-\frac{t}{\tau_2}}$ after which $P(t, \theta)$ was defined as normal.

Below are plots of the simulated data.

comments on how easily a fit would distinguish components

text

text

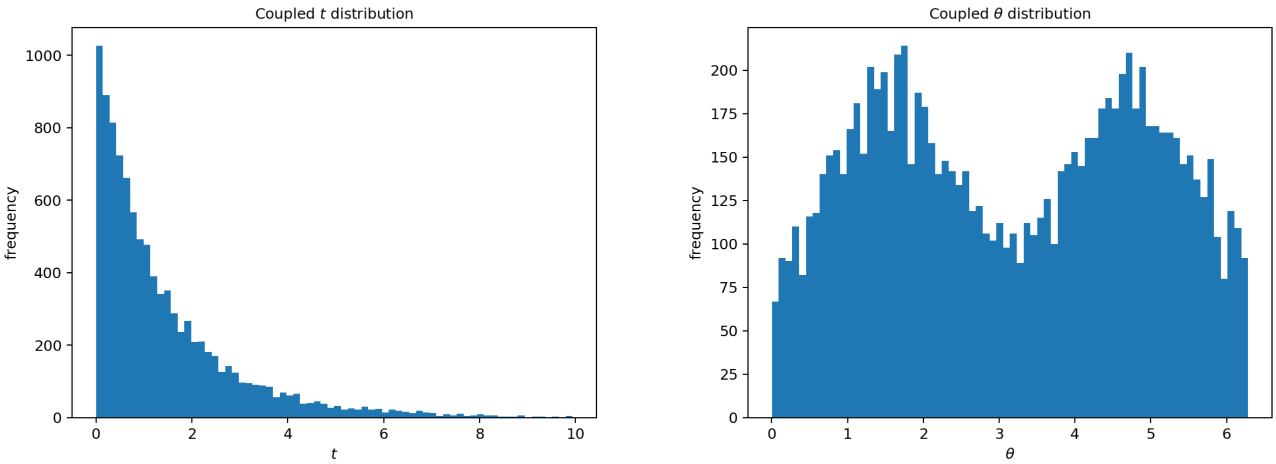

Figure 1: Coupled distributions $F = 0.5, \tau_1 = 1.0, \tau_2 = 2.0$

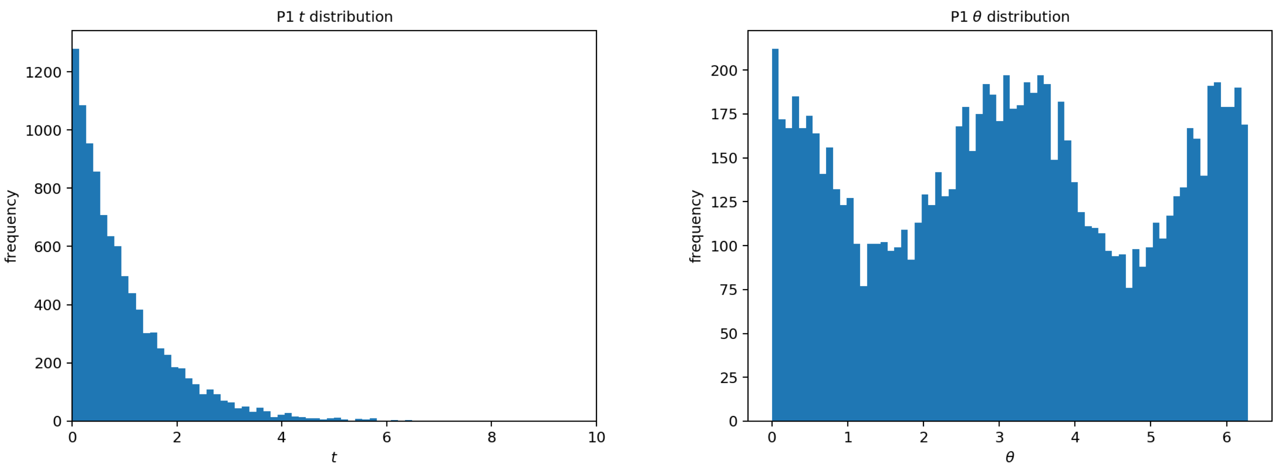

Figure 2: P_1 only distributions $F = 1.0, \tau_1 = 1.0, \tau_2 = 2.0$

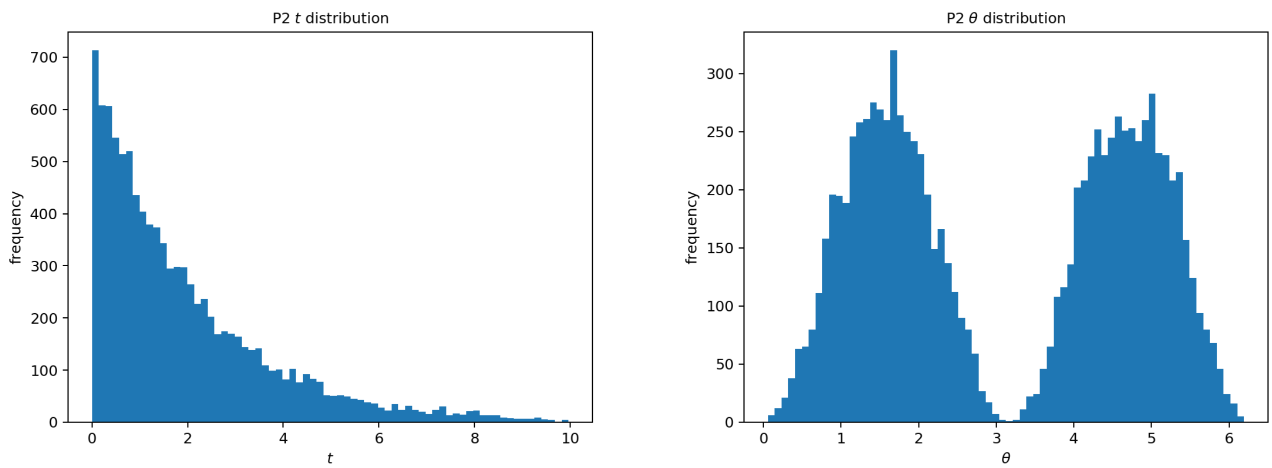

Figure 3: P_2 only distributions $F = 0.0, \tau_1 = 1.0, \tau_2 = 2.0$

Table 1: t data fits

	Initial			Fit			Errors		
run	F	$\tau_1(s)$	$\tau_2(s)$	F	$\tau_1(s)$	$\tau_2(s)$	F	$\tau_1(s)$	$\tau_2(s)$
(1)	0.2	3.5	2.3	0.038	0.512	1.92	0.0029	0.036	0.0064
(2)	0.9	1.9	0.5	0.962	1.925	0.513	0.0029	0.0064	0.036
*(3)	0.3	2.0	0.3	0.809	1.87	1.87	-	-	-

2 Maximum Likelyhood Fit on t Data Only

A maximum likelyhood fit was performed on the (t) data only. To do this the PDF P had to be altered to 'simulate' no θ dependence. This was accomplished by integrating $P(t,\theta)$ over all θ $[0,2\pi]$ to give us $P_t(t)$. The resulting PDFs were

$$P_{1_t}(t) = \frac{1}{\tau_1}e^{-\frac{t}{\tau_1}} \qquad P_{2_t}(t) = \frac{1}{\tau_2}e^{-\frac{t}{\tau_2}}$$

and similar to before

$$P_t(F,t) = FP_{1_t}(t) + (1-F)P_{2_t}(t)$$

For infinite data points this approach is equivalent to leaving $P(t,\theta)$ as it is defined in the introduction and choosing random θ from a uniform distribution on $[0,2\pi]$. Although in this case there are many data points, using P_t was chosen as it was more deterministic and afforded more confidence when comparing results of different runs. Having the integrated form also showed analytically the behaviour of the new PDF.

Several of these fits were run with different starting values, which resulted in some different, physically reasonable parameter estimations. The Initial values and resulting parameter estimations along with their errors for some of these runs are shown in Table 1.

The errors shown in Table 1 are the mean of the modulo of the positive and negative errors calculated using the SIMPLISTIC ERROR METHOD.

Plots of ΔNLL around the parameter minimas in run(2) are given in Figure 4

2.1 Comments on the results

- A large range of starting parameters were tested and all the resulting minima were one of the runs in Table 1. Although by no means was an exhaustive search for all (physically reasonable) local minima carried out.
- The result of run(2) could have been predicted from run(1). Going from run(1) to run(2) we see they relate by swapping τ_1 and τ_2, and mapping $F \to (1-F)$. Inspection of the form of P_t shows that P_t is indeed symmetric under such a change.
- What has occured in run(3) is that a local minimum in which $\tau_1 = \tau_2$ has been found. In this case P_t is independent of F. Thus the error of F for this run in infinite. This is a particularily bad case and in fact inspecing `minuit.is_valid` would have shown that this wasn't even a valid minimum and should be discarded. Because of the nature of this report it is left in.

3 Maximum Likelyhood Fit on Full t,θ Data

A maximum likelyhood fit was performed finding the parameters of $P(t,\theta)$ using the full (t,θ) data. A comparison of the fit with full (t,θ) data and (t) only is shown in Table 2. The (t) only fit is run(2) in Table 1. As in section 2, the errors for the full (t,θ) fit are calculated using the simplistic errors method.

All initial parameters for the (t,θ) fit resulted in the same fit values so a comparison of initial parameters is ommited from Table 2

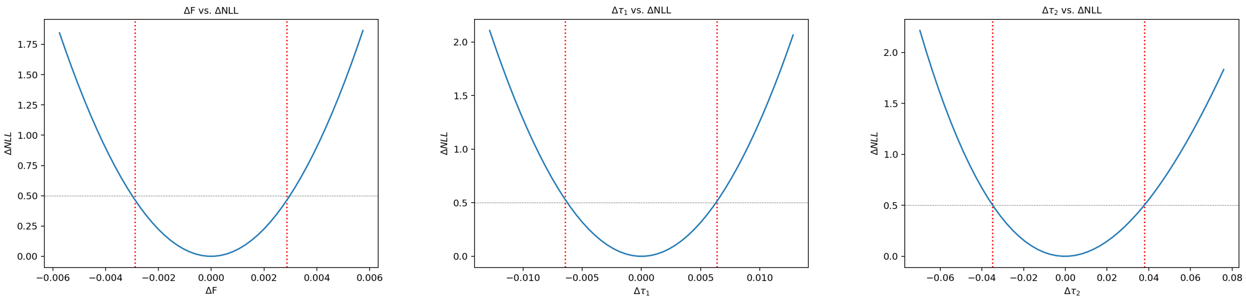

Figure 4: ΔNLL around each parameter minimum

Table 2: Full (t, θ) and (t) only fits

	Fit			Errors		
Data	F	$\tau_1(s)$	$\tau_2(s)$	F	$\tau_1(s)$	$\tau_2(s)$
(t, θ)	0.540	1.466	2.35	0.0030	0.0087	0.013
(t) run(2)	0.962	1.925	0.513	0.0029	0.0064	0.036

3.1 Comments on the results

- The fit values for both cases in Table 2 are very different. The t data alone is not sufficient to gather an estimate of our fit parameters.
- It is a reasonable assumption that given aribitrarily many points of t data, the fit values would not approach the values for the full (t, θ) fit. It is the fact that the θ values break the symmetry between the P_1 and P_2 PDFs that allows for a proper estimation of the parameters.
- The errors on the two fits are quite similar

4 Full Errors on Fits

The simplistic error method used in Section 2 and Section 3 is only accurate if the the parameters are uncorrelated. In this case the parameters are correlated, especially for the case in Section 2 and the t data only fit.

What must be considered when calculating errors is if one parameter is fixed away from the minimum, how would the other parameters change to re-minimise the NLL. For example, let F, τ_1, τ_2 be set at the fit values for the fit in Section 2. Now increase τ_1 by a small ammount. By observing the form of P_t one can see that if τ_1 is now fixed, F and τ_2 should decrease to re-minimise the NLL. So in this case τ_1 is negatively correlated to F and τ_2.

When the parameters are correlated, errors should be calculated using the full errors method. These full error values were calculated for the fits in Section 2 and Section 3 and are shown in Table 3.

Error Contour

- For the errors in this fit, the `minuit.hesse()` error calculator was not adequate as is warned about in the iminuit documentation. The values in Table 1 were calculated using my own implementation of the full error method and were checked against the more thorough `minuit.minos()` method.

Glossary

- **Fit:**
- **PDF (Pobability Density Function):**
- **NLL (Negative Log Likelyhood)**:
- **Minima**
- **Parameter Estimations**

Methods

- **Box Method:**
- **Simplistic Error Method**
- **Full Error Method**
- Minima
- Parameter Estimations
- Box Method
- Throughout when 'minima' are mentioned it refers to (local unless mentioned) minima of $NLL(F, \tau_1, \tau_2)$ and 'parameter estimations' refer to values of F, τ_1, τ_2 corresponding to these NLL minima. Any 'fit' involves finding parameter estimations of P (or similar) by means of minimising $NLL(F, \tau_1, \tau_2)$ as described above.