Inference from Scientific Data, 2021 - Worksheet 2

Canvas submission deadline: Wednesday, 19 November
Marks out of a total of 20 are shown in brackets

Question 1: Consider the problem of estimating the density of stars S (brighter than a certain threshold magnitude) per square degree on the sky (i.e. units [deg^{-2}]). After surveying $10\,\text{deg}^2$ of the sky, you count 50 such stars.

a) Write down the posterior PDF for S, assuming a Jerreys prior on S between $1\,\text{deg}^{-2}$ and $100\,\text{deg}^{-2}$. **[2]**

b) Sketch the posterior distribution for S. Find the maximum *a posteriori* value of the density parameter, $\hat{S}$. **[2]**

For the remainder of the question you are encouraged to work numerically (e.g. using Python).

c) Find the mean of the posterior distribution for S. Compare this to the maximum value found above. **[1]**

d) Estimate the uncertainty on your measurement of the parameter S. **[1]**

e) Later you read of a much larger survey, but its error is dominated by an uncertainty in brightness calibration. The published estimate of S is 7.7 ± 0.3 where the error is believed to be Gaussian. Perform a new Bayesian analysis incorporate this new information into your prior. What is the most probable value of S now? Plot the the two posterior distributions you have found (from parts b and e) on the same set of axes and comment on how they compare. Do you think this new prior has lead to an improved measurement of S? **[2]**

Question 2: A team of gamma-ray astronomers claim to have detected a gamma-ray burst in a distant galaxy. You are part of a team of optical astronomers who have followed up this source, measuring its luminosity once per hour. The galaxy itself is known to have a background luminosity of $B = 5$, in arbitrary units. The data are assumed to be subject to uncorrelated Gaussian measurement errors with standard deviation $\sigma = 1$. Thus, the j-th data point $d[t_j]$, measured at time $t_j = 0, 1, 2, \ldots$ hours after the alert is assumed to be

$$\begin{aligned} &d[t_j] = s[t_j] + n_j, \quad \text{where} \quad n_j \sim \mathcal{N}(0,1)\,; \\ &\text{i.e.} \quad d[t_j] \sim \mathcal{N}(s[t_j], 1)\,. \end{aligned}$$

The model $s[t]$ is one of the three possibilities described below. Your job is to analyse the optical data as they come in using two different models.

1. **Model 1:** There is a source characterised by a constant excess luminosity L_0 at this position:

$$s[t_j] = B + L_0\,.$$

2. **Model 2:** There is a source characterised by an exponentially decaying luminosity, with initial amplitude A_0 and decay constant τ:

$$s[t_j] = B + A_0 \exp(-t_j/\tau)\,.$$

You are to perform the analysis several times, as more data gradually becomes available. Perform the analyses after:

- 10 observations (or 10 hours),
- 24 observations (or 24 hours),
- 100 observations (or 100 hours).

The data you need to perform these analyses are provided on the course Canvas page in the files named `data_10hr.txt`, `data_24hr.txt`, and `data_100hr.txt`.

a) For model 1, identify the parameter(s) of the model and choose suitable priors for these, making your choice clear. Write down an expression for the likelihood of the data for this model. Perform a Bayesian analysis of the data to obtain the posterior distribution on the model parameter(s). Repeat the analysis using the 10 hour, 24 hour and 100 hour data sets; compare how your measurements of the model parameter(s) change as you analyse more data. **[4]**

b) Repeat for model 2. **[8]**