Inference from Scientific Data, 2020 - Worksheet 2

Canvas submission deadline: Wednesday, 18 November
Marks out of a total of 20 are shown in brackets

Question 1: Consider the Poisson problem of estimating the density of stars S (brighter than a certain threshold magnitude) per square degree on the sky. You survey $10\,\mathrm{deg}^2$ of the sky and count 50 stars.

a) Write down the posterior PDF for S, assuming a uniform prior on S between 0 and 100. **[2]**

b) Plot the posterior for S and compare the mean and mode of the distribution. **[2]**

c) 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. Incorporate this new information into your Bayesian analysis and plot the resulting posterior distribution. What is the most probable value of S now? **[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

$$d[t_j] = s[t_j] + n_j\,, \tag{1}$$
$$\text{where} \quad n_j \sim \mathcal{N}(0,1)\,;$$
$$\text{i.e.} \quad d[t_j] \sim \mathcal{N}(s[t_j], 1)\,. \tag{2}$$

The model $s[t]$ is one of the three possibilities described below.

Your job is to analyse the optical data as they come in and to compare the three different hypotheses:

1. **Model 0:** There is no optical source at the position reported by the team, just the usual galaxy:
$$s[t_j] = B\,.$$
2. **Model 1:** There is a source characterised by a constant excess luminosity L_0 at this position:
$$s[t_j] = B + L_0\,.$$
3. **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 analysis 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`.

Conduct the necessary Bayesian analyses at each of these time (10, 24 and 100 hours) to compare the three models. You may assume that $0 < L_0 < 10$, and $0 < A_0 < 10$, and $1 < \tau < 100$ with uniform priors on each. You may assume the various models have equal prior probability; i.e. $P(\text{Model}\,0) = P(\text{Model}\,1) = P(\text{Model}\,2) = 1/3$. Your answer should include the following.

a) The equations required to compute the posterior odds ratios $\mathcal{O}_{1,0}$, $\mathcal{O}_{2,0}$, $\mathcal{O}_{2,1}$, where $\mathcal{O}_{i,j}$ is the odds between model i and j. **[2]**

b) Numerical values for each of these three odds ratios at each of the three times considered, and a discussion of the conclusions you can draw from these values. **[6]**

c) For each of the three times considered, if either of the models that includes a source is preferred over Model 0, then include suitable plot(s) of the posterior probability distribution for the model parameters. **[3]**

d) For the longest dataset (100 hours), a plot of the data as a function of time on the same axes as plots of the predictions from a selection of models. (Exactly which models, and with which parameters, you choose to use for these plot predictions is up to you, but your choices should be explained clearly.) **[2]**

e) A summary of your conclusions about any possible optical source. **[1]**