**CSC 2541, Assignment #1, due in class on October 20. Worth 12% of the mark.**

The Department of Uncertainty and Machine Forgetting (DUMF) offers five courses each year. Some of the students in these courses are in DUMF's one-year Professional Masters Programme, who are required to take all five courses. The remaining students are from other departments, which allow students to take one of these courses as an option. The five courses are all equally attractive to such students.

We have data on the enrollments in all of these courses for a six year period, as follows:

| | | | | |
|---|---|---|---|---|
| 9 | 10 | 8 | 8 | 10 |
| 16 | 16 | 14 | 16 | 19 |
| 7 | 7 | 10 | 10 | 10 |
| 11 | 7 | 9 | 9 | 10 |
| 15 | 15 | 17 | 17 | 18 |
| 18 | 17 | 18 | 18 | 18 |

We don't have data on the number of students in the Professional Masters Programme each year.

We decide to model this data assuming that a large pool of potential students decide whether or not to enter the Professional Masters Programme or take an optional course independently, and that any changes from year to year are due only to chance, not to any trends.

Writing $X_{ti}$ for the enrollment in the $i$'th course in year $t$, these assumptions give rise to the following model:

$$\begin{aligned} X_{ti} \mid \lambda_2, Y_t &\sim Y_t + \text{Poisson}(\lambda_2) \\ Y_t \mid \lambda_1 &\sim \text{Poisson}(\lambda_1) \\ \lambda_1 &\sim \text{Exp}(8) \\ \lambda_2 &\sim \text{Exp}(4) \end{aligned}$$

Here, $\text{Exp}(\mu)$ represents the exponential distribution with mean $\mu$, whose density function (over the positive reals) is $(1/\mu)\exp(-\lambda/\mu)$. The unobserved variables $Y_t$ represent the numbers of students enrolled in the Professional Masters Programme each year.

First, write a Gibbs sampling program to find a sample of values from the posterior distribution for this model given the data above, with the state of the Markov chain being $\lambda_1$, $\lambda_2$, and $Y_1, \ldots, Y_6$. Produce a picture of the joint posterior distribution of $\lambda_1$ and $\lambda_2$ by making a scatterplot of their values in your Gibbs sampling run (after discarding a suitable burn-in period).

Next, write a Gibbs sampling program that solves this problem with $\lambda_1$ and $\lambda_2$ integrated away analytically — ie, with the Markov chain state consisting of $Y_1, \ldots, Y_6$ only. Figure out how to produce the same sort of plot as you did before, and compare (roughly) the efficiency of the two methods, accounting for any possible difference in how well the two Markov chains move around the state space.