HW5, Bayesian Modelling B 2016/17

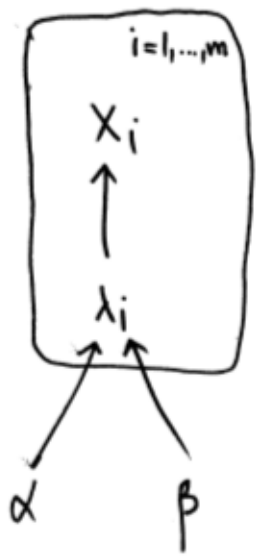
Jonathan Rougier
School of Mathematics
University of Bristol UK

In the homeworks, questions with marks are officially 'exam-style', although you can expect any homework question to appear as an exam question, unless it is explicitly 'not examinable'.

Hand in Q1–Q4.

1. Let X_i be the number of eruptions of volcano i during time interval $(t_i, \text{today}]$, measured in years, for $i = 1, \dots, m$. Let each volcano have its own eruption rate λ_i (units of /yr) and model the eruption rates as exchangeable. Write down the DAG for this model, using Θ for the hyperparameters (hint: use a plate). Write the extensive form of the model, making sensible choices for Θ, and the marginal and conditional distributions. [10 marks]

   Answer. Here is the DAG (I did not use Θ but wrote it for the actual model below):



   I will let $\Theta = (\alpha, \beta)$.

$$
\begin{aligned}
X_i \mid \lambda &\sim \text{Pois}(\lambda_i \cdot (\text{today} - t_i)) && i = 1, \dots, m \\
\lambda_i \mid \alpha, \beta &\sim \text{Gam}(\alpha, \beta) && i = 1, \dots, m \\
\alpha &\sim \text{Gam}(0.001, 0.001) \\
\beta &\sim \text{Gam}(0.001, 0.001)
\end{aligned}
$$

2. Write down the contents of the Volcanoes.bug file, and the command in rjags used to initialise the model in R, based on observations in the vectors `xobs` and `times`, and the value `today = 2016`. Specify four chains. [10 marks]

Answer.

```
#### write out to Volcanoes.bug

cat("
#### Volcanoes model

model {
for (i in 1:m) {
  X[i] ~ dpois(lambda[i] * (today - times[i]))
  lambda[i] ~ dgamm(alpha, beta)
}
alpha ~ dgamm(0.001, 0.001)
beta ~ dgamm(0.001, 0.001)
}
", file = "Volcanoes.bug")

#### here is the jags model

mydata <- list(X = xobs, m = length(xobs), times = times, today = 2016)
myvolc <- jags.model("Volcanoes.bug", data = mydata, n.chains = 4)
```

3. Explain how Gibbs sampling is used in this model, to estimate expectations conditional on $X = x^{\text{obs}}$. [10 marks]

Answer. We are given x^{obs}, and so our target distribution is

$$\begin{aligned} \mathrm{p}(\alpha, \beta, \lambda_{1:m} \mid x^{\text{obs}}) &\propto \mathrm{p}(\alpha, \beta, \lambda_{1:m}, x^{\text{obs}}) \\ &= \mathrm{p}(\alpha)\,\mathrm{p}(\beta) \prod_{i=1}^{m} \mathrm{p}(\lambda_i \mid \alpha, \beta) \cdot \prod_{i=1}^{m} \mathrm{p}(x_i^{\text{obs}} \mid \lambda_i), \end{aligned}$$

where we have explicit choices for each of the marginal and conditional distributions. Gibbs sampling cycles through the full conditionals of the $m+2$ random quantities. Hence, for current value $(\alpha_t, \beta_t, \lambda_{1:m,t})$ one full iteration of the Gibbs sampler is

(a) Sample a new value for α:

$$\begin{aligned} \alpha_{t+1} &\sim \mathrm{p}(\alpha \mid \beta_t, \lambda_{1:m,t}, x^{\text{obs}}) \\ &\propto \mathrm{p}(\alpha, \beta_t, \lambda_{1:m,t}, x^{\text{obs}}) \\ &\propto \mathrm{p}(\alpha) \prod_{i=1}^{m} \mathrm{p}(\lambda_{it} \mid \alpha, \beta_t), \end{aligned}$$

after dropping multiplicative constants that do not depend on α.

(b) Sample a new value for β:

$$\begin{aligned}\beta_{t+1} &\sim \mathrm{p}(\beta \mid \alpha_{t+1}, \lambda_{1:m,t}, x^{\mathrm{obs}}) \\ &\propto \mathrm{p}(\alpha_{t+1}, \beta, \lambda_{1:m,t}, x^{\mathrm{obs}}) \\ &\propto \mathrm{p}(\beta) \prod_{i=1}^{m} \mathrm{p}(\lambda_{it} \mid \alpha_{t+1}, \beta),\end{aligned}$$

after dropping multiplicative constants that do not depend on β.

(c) Sample a new value for λ_1:

$$\begin{aligned}\lambda_{1,t+1} &\sim \mathrm{p}(\lambda_1 \mid \alpha_{t+1}, \beta_{t+1}, \lambda_{2:m,t}, x^{\mathrm{obs}}) \\ &\propto \mathrm{p}(\alpha_{t+1}, \beta_{t+1}, \lambda_1, \lambda_{2:m,t}, x^{\mathrm{obs}}) \\ &\propto \mathrm{p}(\lambda_1 \mid \alpha_{t+1}, \beta_{t+1})\, \mathrm{p}(x_1^{\mathrm{obs}} \mid \lambda_1),\end{aligned}$$

after dropping multiplicative constants that do not depend on λ_1.

(d) Same for $\lambda_{2:m}$:

$$\lambda_{i,t+1} \sim \mathrm{p}(\lambda_i \mid \alpha_{t+1}, \beta_{t+1})\, \mathrm{p}(x_i^{\mathrm{obs}} \mid \lambda_i), \qquad i = 2, \dots, m.$$

This gives us a new value $(\alpha_{t+1}, \beta_{t+1}, \lambda_{1:m,t+1})$.

After going through these steps for thousands of iterations, and checking convergence, the resulting sample is used to estimate expectations using Cesàro averages, which converge in mean square to the true expectations under the target distribtution $\mathrm{p}(\cdot \mid x^{\mathrm{obs}})$, according to the Ergodic Theorem.

4. For each unobserved random quantity in your model, write down the kernel of the full conditional distribution, conditional on $X = x^{\mathrm{obs}}$, and identify those full conditionals which have recognisable distributions. [10 marks]

(In an exam, you would be given an explicit extensive form for the model, but in this case there is only one sensible choice.)

Answer. Using the same steps as above:

(a)

$$\begin{aligned}\mathrm{p}(\alpha \mid \beta, \lambda_{1:m}, x^{\mathrm{obs}}) &\propto \mathrm{p}(\alpha) \prod_{i=1}^{m} \mathrm{p}(\lambda_i \mid \alpha, \beta) \\ &\propto \alpha^{0.001-1} e^{-0.001\alpha} \cdot \prod_{i=1}^{m} \lambda_i^{\alpha-1} e^{-\beta\lambda_i} \\ &\propto \alpha^{0.001-1} e^{-0.001\alpha} (\lambda_1 \times \cdots \times \lambda_m)^{\alpha} \\ &= \alpha^{0.001-1} (e^{-0.001} \times \lambda_1 \times \cdots \times \lambda_m)^{\alpha}\end{aligned}$$

which is not a recognisable distribution.

(b)

$$
\begin{aligned}
\mathrm{p}(\beta \mid \alpha, \lambda_{1:m}, x^{\text{obs}}) &\propto \mathrm{p}(\beta) \cdot \prod_{i=1}^{m} \mathrm{p}(\lambda_i \mid \alpha, \beta) \\
&\propto \beta^{0.001-1} e^{-0.001\beta} \cdot \prod_{i=1}^{m} \lambda_i^{\alpha-1} e^{-\beta\lambda_i} \\
&\propto \beta^{0.001-1} e^{-0.001\beta} \cdot \prod_{i=1}^{m} e^{-\beta\lambda_i} \\
&= \beta^{0.001-1} e^{-(0.001+\lambda_1+\cdots+\lambda_m)\beta},
\end{aligned}
$$

which is the kernel of a $\text{Ga}(0.001, 0.001 + \lambda_1 + \cdots + \lambda_m)$ distribution.

(c)

$$
\begin{aligned}
\mathrm{p}(\lambda_1 \mid \alpha, \beta, \lambda_{2:m}, x^{\text{obs}}) &\propto \mathrm{p}(\lambda_1 \mid \alpha, \beta)\, \mathrm{p}(x_1^{\text{obs}} \mid \lambda_1) \\
&\propto (\lambda_1)^{\alpha-1} e^{-\beta\lambda_1} \cdot e^{-d_1\lambda_1} \frac{(d_1\lambda_1)^{x_1^{\text{obs}}}}{(x_1^{\text{obs}})!} \\
&\propto (\lambda_1)^{\alpha+x_1^{\text{obs}}-1} e^{-(\beta+d_1)\lambda_1},
\end{aligned}
$$

where $d_1 = \text{today} - t_1$. This is the kernel of a $\text{Ga}(\alpha + x_1^{\text{obs}}, \beta + d_1)$ distribution.

(d) Same for $\lambda_{2:m}$.

So $m+1$ of the full conditionals have Gamma distributions, but that of α has an unrecognisable distribution.

5. (Not examinable) Investigate how JAGS deals with full conditionals which do not have recognisable distributions.

Answer. If you dig around in the manual you will find that JAGS uses *slice sampling*, see `https://en.wikipedia.org/wiki/Slice_sampling`.