# Bayesian Modelling and Data Analysis

## CAT II

124384 - Luycer Bosire

3/14/2021

a).

```r
library('heemod')

## Warning: package 'heemod' was built under R version 4.0.4

## Registered S3 method overwritten by 'pryr':
##   method      from
##   print.bytes Rcpp

library('diagram')

## Loading required package: shape

mat_dim <- define_transition(
state_names = c('2', '1', '3'),x, 0, 0, "6x", "3x", x,x, "3x", x);
curves <- matrix(nrow = 3, ncol = 3, 0.05)
plot(mat_dim, curve=curves, self.shiftx = c(0.1,-0.1,-0.15), self.shifty =
c(-0.1,-0.1,0.06), self.arrpos = c(1.4,1.7,1.5))
```
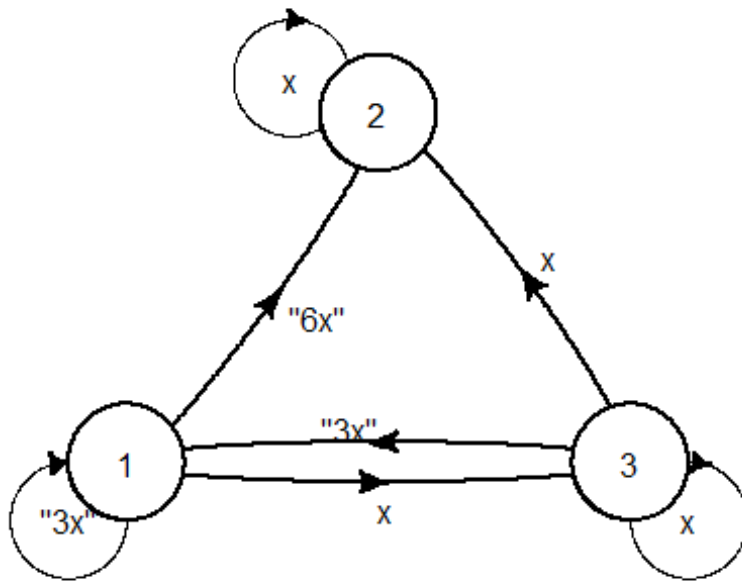
From the MCMC methods, the diagram has 3 transition states;

$$\begin{bmatrix} 3x & 6x & x \\ 0 & x & 0 \\ 3x & x & x \end{bmatrix}$$

We solve for each row and substitute the value of x in each respective row.

Solving for row 1;

$$3x + 6x + x = 110x = 1x = 0.1$$

Solving for row 2;

$$0 + x + 0 = 1x = 1$$

Solving for row 3;

$$3x + x + x = 15x = 1x = 0.2$$

Therefore;

$$\begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0 & 1 & 0 \\ 0.6 & 0.2 & 0.2 \end{bmatrix}$$

b). The number of the lions $y = 1,2,3, \dots$ breaking out of a Nairobi national park within the last one month follow the distribution $f(\theta) = \theta(1 - \theta)^{(y-1)}; y = 1,2,3\dots; 0 < \theta < 1$. Find the Jeffrey's prior distribution of $\theta$ and hence or otherwise its posterior distribution.

## solution

Using the distribution function of a binomial distribution,

$$f(y) = \theta^y(1 - \theta)^{n-y}$$

The likelihood is given by

$$L(\theta) = \prod_{i=1}^{n} f(y) = \prod_{i=1}^{n} \theta^y (1 - \theta)^{n-y} = \theta^{\Sigma y_i}(1 - \theta)^{n-\Sigma y_i}$$

The Log-likelihood is given by

$$\ell(\theta) = \log L(\theta) = \log\{\theta^{\Sigma y_i}(1 - \theta)^{n-\Sigma y_i}\} = \Sigma y_i \log\theta + (n - \Sigma y_i)\log(1 - \theta)$$

We obtain the first derivative of the log-likelihood w.r.t $\theta$

$$\frac{\partial}{\partial\theta}\ell(\theta) = \frac{\Sigma y_i}{\theta} - \frac{(n - \Sigma y_i)}{(1 - \theta)}$$

Further we obtain the second derivative w.r.t $\theta$ as

$$\frac{\partial^2}{\partial^2\theta}\ell(\theta) = -\frac{\Sigma y_i}{\theta^2} - \frac{(n - \Sigma y_i)}{(1 - \theta)^2}$$

$$I(\theta) = -E\{\frac{\partial^2}{\partial^2\theta}\ell(\theta)\} = \frac{n\theta}{\theta^2} + \frac{(n - n\theta)}{(1 - \theta)^2} = \frac{n}{\theta} + \frac{n}{(1 - \theta)} = \frac{n}{\theta(1 - \theta)}$$

Therefore the prior,

$$P(\theta) = I(\theta)^{\frac{1}{2}} = n^{\frac{1}{2}}\theta^{-\frac{1}{2}}(1 - \theta)^{-\frac{1}{2}}$$

$\propto \theta^{-\frac{1}{2}}(1 - \theta)^{-\frac{1}{2}}$ which is in the form of a $Beta(0.5,0.5)$ density

.

The posterior is given by

$$Posterior \propto Prior * Likelihood$$

Hence

$$Posterior = \theta^{-\frac{1}{2}}(1 - \theta)^{-\frac{1}{2}} * \theta^{\Sigma y_i}(1 - \theta)^{n-\Sigma y_i} = \theta^{\Sigma y_i - \frac{1}{2}}(1 - \theta)^{n-\Sigma y_i - \frac{1}{2}}$$

Therefore

$$Posterior = Beta(\sum y_i + \frac{1}{2}, n - \sum y_i + \frac{1}{2})$$

c).Find the long term trend of the transition matrix

## solution

$$\begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

The long term trend refers to the likelihood that our path will be absorbed at a certain state.

We seek to find the Fundamental Matrix given by:

$$\begin{bmatrix} I & O \\ R & Q \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

$$FR = (I_1 - Q)^{-1}$$

$$R = [0.2, 0.5]$$

$$Q = [0.3]$$

$$F = [1 - 0.3]^{-1} = [0.7]^{-1} = 1/0.7 = 10/7$$

$$FR = 10/7[0.2, 0.5] = [2/7, 5/7]$$

If the system starts in nonabsorbing state 1, there is 2/7 chance of ending up in the absoption state 2. And 5/7 chance of ending in the absorption state 3.

d). Suppose that 1 in 200 people in a population carry a certain gene. A test for the gene has been developed such that 90% of those who carry the gene test positively and 99% of those who do not carry the gene test negatively. If you test negative, what is the probability that you carry the gene? If you test positive, what is the probability that you do not carry the gene

## solution

```
## Warning: package 'igraph' was built under R version 4.0.4
```
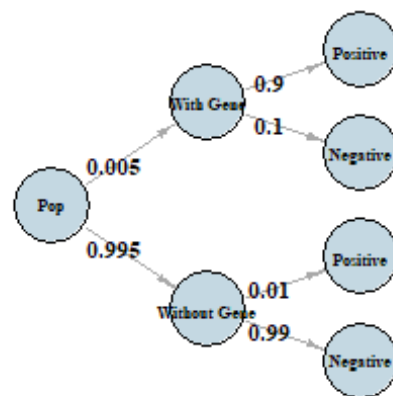


It follows that

$$P(Negative|WithGene) = 0.005 * 0.1 = 0.0005$$

$$P(Positive|WithoutGene) = 0.995 * 0.01 = 0.00995$$

e). Briefly descirbe Gibbs sampler for parameters $\theta_1, \theta_2, \theta_3$ with joint distribution $p(y|\theta_1, \theta_2, \theta_3)$.

## solution

We are interested in sampling from the posterior $p(\theta|y)$ where $\theta$ is a vector of three parameters $\theta_1, \theta_2, \theta_3$.

Step 1: Pick a vector of starting values $\theta^{(0)}$ (Defining a starting distribution $\pi^{(0)}$ and drawing $\theta^{(0)}$ from it).

Step 2: Start with any $\theta$. The order does not matter, but we will start with $\theta$ for convenience. We draw a value $\theta_1^{(1)}$ from the full conditional $p(\theta_1|\theta_1^{(1)}, \theta_2^{(1)}, y)$.

Step 3: We draw a value $\theta_3^{(1)}$ from the full conditional $p(\theta_2|\theta_1^{(1)}, \theta_3^{(1)}, y)$.Again, the order does not matter. We must use the updated value of $theta_1^{(1)}$.

Step 4: We draw a value $\theta_3^{(1)}$ from the full conditional $p(\theta_3|\theta_1^{(1)}, \theta_2^{(1)}, y)$ using updated values. Steps 2-4 are analogous to multiplying $\pi^{(0)}$ and P to get $\pi^{(0)}$ and then drawing $\theta^{(1)}$ from $\pi^{(0)}$.

Step 5: Draw $\theta^{(2)}$ using $\theta^{(1)}$ and continuously using the most updated values.

Step 6: Repeat until we get m draws with each draw being a vector of $\theta^{(t)}$.

Step 7: Optional burn-in and/or thinning.

Our result is a Markov chain with a bunch of draws of $\theta$ that are approximately from our posterior. We can do Monte Carlo Integration on the draws to get quantities of interest.