MAST30027: Modern Applied Statistics

Assignment 5

Due: 1:00 pm Fri 16 October (week 11)

This assignment is worth 3 1/3% of your total mark.

1. Recall the multinomial logistic regression model, with a single covariate x,

$$\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{id}) \sim \text{multinomial}(m_i, \mathbf{p}_i) \text{ for } i = 1, \ldots, n,$$

where $\mathbf{p}_i = (p_{i1}, \ldots, p_{id})$ can be written as

$$p_{ij} = \frac{\phi_{ij}}{\sum_{k=1}^{d} \phi_{ik}} = \frac{\exp(\eta_{ij})}{\sum_{k=1}^{d} \exp(\eta_{ik})} = \frac{\exp(\alpha_j + \beta_j x_i)}{\sum_{k=1}^{d} \exp(\alpha_k + \beta_k x_i)}$$

So that the model is not over parameterised, we put $\alpha_1 = \beta_1 = 0$ (that is $\phi_{i1} = 1$ or $\eta_{i1} = 0$).

A Bayesian formulation of this model typically puts vague priors on the α_j and β_j, $j = 2, \ldots, d$.

In 2003 Briggs, Ades and Price reported on a trial for the treatment of asthma. Patients received one of two treatments (seretide or fluticasone), and their status was monitored from week to week. Possible states were

STW Successfully treated week

UTW Unsuccessfully treated week

HEX Hospital managed exacerbation

PEX Primary-care managed exacerbation

TF Treatment failure (treatment ceased and patient removed from the trial)

There were 372 transitions from the state STW. 272 of these patients were being treated with seretide and 100 with fluticasone. Here are the numbers

	To					
	STW	UTW	HEX	PEX	TF	Total
Seretide	210	60	0	1	1	272
Fluticasone	66	32	0	0	2	100

Here is a WinBUGS coding of a multinomial logistic regression model for this data. Missing values are indicated by ?

```
model
{
  for (i in 1:2) {
    count[i, 1:5] ~ dmulti(q[i, 1:5], M[i])
    for (r in 1:5) {
      q[i, r] <- phi[i, r]/sum(phi[i,])
      log(phi[i, r]) <- a[r] + b.treat[r]*treat[i]
    }
  }
  a[1] <- 0
```

```
  for (r in 2:5) {
    a[r] ~ dnorm(0, 0.000001)
  }
  b.treat[1] <- 0
  b.treat[2] ~ dnorm(0, 0.000001)
  for (r in 3:5) {
    b.treat[r] <- 0
  }
}

# data
list(
  M = c(?, ?),
  count = structure(.Data = c(?, ?, ?, ?, ?, ?, ?, ?, ?, ?), .Dim = c(2, 5)),
  treat = c(0, 1)
)

# initial values
list(a = c(NA, ?, ?, ?, ?), b.treat = c(NA, ?, NA, NA, NA))
```

(a) What are d and the priors on $\alpha_1, \ldots, \alpha_d$ and $\beta_1, \ldots, \beta_d$?
What is n, the total number of multinomial observations, and $m_1, \ldots, m_n$, the sizes of each multinomial? Also, what are $x_1, \ldots, x_n$?
Solution: We have $d = 5$, $\alpha_1 = 0$, $\alpha_2, \ldots, \alpha_5 \sim N(0, 1000^2)$, $\beta_1 = \beta_3 = \beta_4 = \beta_5 = 0$, $\beta_2 \sim N(0, 1000^2)$, $n = 2$, $m_1 = 272$, $m_2 = 100$, $x_1 = 0$, and $x_2 = 1$.
The multinomial logistic regression is being used in a non-standard way here. The x variable is being used to "turn on" or "turn off" the β coefficients, so that we can have two different types of multinomial response. The priors on the β's then let us control how much of a difference we expect in the rates for a particular element of the response. In our case there isn't enough data to sensibly get different rates for HEX, PEX and TF, so we just set the corresponding β's to zero.

(b) Fit the model with WinBUGS. You will need to write the data file and choose appropriate initial values for the unobserved variables.
Give posterior means and 95% credible intervals for the α_j and β_j.
Solution: The data and a suitable starting value are

```
# data
list(
  M = c(272, 100),
  count = structure(.Data = c(210, 60, 0, 1, 1,66, 32, 0, 0,  2), .Dim = c(2, 5)),
  treat = c(0, 1)
)

# initial values
list(a = c(NA, 0, 0, 0, 0), b.treat = c(NA, 0, NA, NA, NA))
```

Note that WinBUGS reads arrays row by row, unlike R which reads them col by col (by default).
Using a burn-in of 4000 and a sample of size 10,000 I got the following results (no burn-in period was asked for, but convergence of the chain should always be a consideration).

node	mean	2.5%	97.5%
a[2]	-1.26	-1.54	-0.98
a[3]	-797	-2210	-32
a[4]	-6.22	-9.35	-4.33
a[5]	-4.70	-6.08	-3.65
b.treat[2]	0.507	-0.011	1.024

Looking at `b.treat` it looks like the difference between the two treatments is borderline significant. We test this again in the next part.

(c) The odds of going from state STW to state UTW are a measure of the treatment effectiveness.

Plot a density estimate of the odds ratio for fluticasone over seretide. How strong is the evidence that seretide performs better?

Hint: you will need to add a logical node to the model, equal to the desired odds ratio, and plot its posterior distribution.

Solution: We are interested in the posterior of

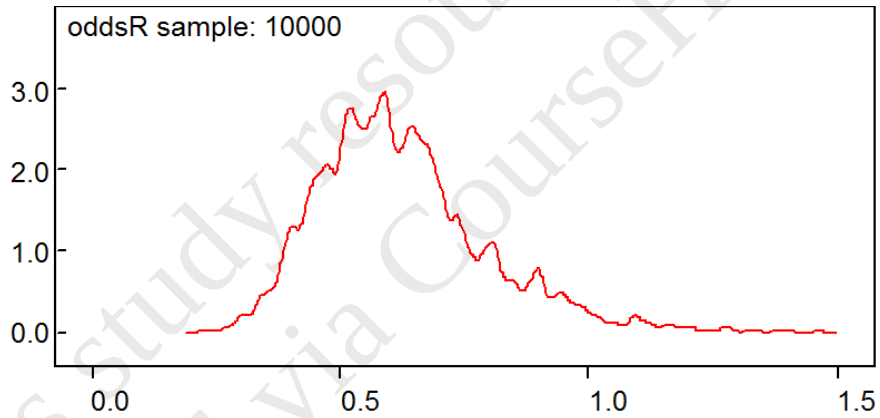
$$\frac{p_{12}}{1-p_{12}} \Big/ \frac{p_{22}}{1-p_{22}} = \frac{\phi_{12}}{\sum_{j\neq 2}\phi_{1j}} \Big/ \frac{\phi_{22}}{\sum_{j\neq 2}\phi_{2j}} = \frac{1}{e^{\beta_2}}.$$

(The last equality follows only because $\beta_1 = \beta_3 = \beta_4 = \beta_5 = 0$.) Smaller odds indicate better performance, so if seretide performs better then we would expect to see the odds ratio less than one

We add the following code to the model. The posterior mean of `oddR1` will give us the posterior probability that the odds ratio is less than 1.

```
oddsR <- (q[1, 2] * (1 - q[2, 2])) / ((1 - q[1, 2]) * q[2, 2])
oddsR1 <- step(1 - oddsR)
```

We get an estimated posterior probability of 0.9725, so we have reasonable evidence that seretide performs better than fluticasone. A plot of the posterior distribution also shows that the bulk of the posterior of the odds ratio is less than 1.



Finally, as a check, we compare the posterior mean and median of the odds ratio, namely 0.6236 and 0.5987, to the empirical odds ratio

$$\frac{60/212}{32/68} = 0.601.$$

All three are reassuringly close.

2. Consider the following algorithm for sampling from a beta$(a, 1)$ distribution. Starting with any $X(0) \in (0, 1)$, for $t = 1, 2, \ldots$

   1° With probability $X(t-1)$ set $X(t) \sim \text{beta}(a+1, 1)$.

   2° Otherwise set $X(t) = X(t-1)$.

   (This is similar to, but not quite, a Metropolis-Hastings algorithm.)

(a) Show that the algorithm above yields a Markov process with transition density

$$p(x,y) = xf_{(a+1,1)}(y) + (1-x)\delta_x(y)$$

where δ_x is the Dirac delta at x, and $f_{(a+1,1)}$ is the density of a beta$(a+1,1)$ r.v.
Solution: As the distribution of $X(t)$ depends only on $X(t-1)$, the process is clearly Markov and time-homogeneous. We have

$$\begin{aligned} p(x,y)dy &= \mathbb{P}(X(t) \in [y, y+dy) | X(t-1) = x) \\ &= \mathbb{P}(X(t) \in [y, y+dy) | X(t-1) = x, \text{ we choose the beta}) x \\ &\quad + \mathbb{P}(X(t) \in [y, y+dy) | X(t-1) = x, \text{ we don't choose the beta})(1-x) \\ &= xf_{(a+1,1)}(y)dy + (1-x)1_{\{x=y\}} \end{aligned}$$

Dividing both sides by dy we get $p(x,y) = xf_{(a+1,1)}(y) + (1-x)1_{\{x=y\}}/dy$, where $1_{\{x=y\}}/dy = \delta_x(y)$, as required.

(b) Show that the Markov process defined above is reversible w.r.t. a beta$(a,1)$ distribution. Hence explain why the distribution of $X(n)$ converges to a beta$(a,1)$ distribution.
Solution: We need to show that $f_{(a,1)}(x)p(x,y) = f_{(a,1)}(y)p(y,x)$. It is clearly true for $x = y$, while for $x \neq y$ we have

$$\begin{aligned} f_{(a,1)}(x)p(x,y) &= f_{(a,1)}(x)xf_{(a+1,1)}(y) \\ &= \beta(a,1)^{-1}x^{a-1}x\beta(a+1,1)^{-1}y^a \\ &= \beta(a,1)^{-1}y^{a-1}y\beta(a+1,1)^{-1}x^a \\ &= f_{(a,1)}(y)yf_{(a+1,1)}(x) \\ &= f_{(a,1)}(y)p(y,x) \end{aligned}$$

Clearly we have $\mathbb{P}(X(t) = 0 | X(t-1) \neq 0) = 0$ and thus $\mathbb{P}(X(t) = 0 \text{ some } t | X(0) \neq 0) = 0$. We will assume $\mathbb{P}(X(0) = 0) = 0$ from here on.
Now the process is aperiodic because there is always a chance that it stays where it is. Also, it is irreducible, since for any interval $(a,b) \subset [0,1]$, $a < b$, we have

$$\mathbb{P}(X(t) \in (a,b) | X(t-1) = x) = x\int_a^b f_{(a+1,1)}(y)dy + (1-x)1_{(a,b)}(x) > 0.$$

(This follows since $f_{(a+1,1)}(y) > 0$ on $(0,1)$.) Thus any stationary distribution is unique and limiting. But because the process is reversible w.r.t. $f_{(a,1)}$, it is a stationary distribution and thus limiting, as required.

(c) Show that the probability that the process stays where it is from time t to $t+1$, converges to $1/(1+a)$ as $t \to \infty$.
Solution: Let $f_{X(t)}$ be the density of $X(t)$, then

$$\begin{aligned} \mathbb{P}(X(t) = X(t-1)) &= \int_0^1 \mathbb{P}(X(t) = X(t-1) | X(t-1) = x) f_{X(t-1)}(x)dx \\ &\to \int_0^1 \mathbb{P}(X(t) = X(t-1) | X(t-1) = x) f_{(a,1)}(x)dx \text{ as } n \to \infty \\ &= \int_0^1 (1-x) f_{(a,1)}(x)dx \\ &= 1 - \mathbb{E}Y \text{ where } Y \sim \text{beta}(a,1) \\ &= 1 - a/(a+1) \quad = \quad 1/(1+a) \end{aligned}$$

as required.