

ST4234 Bayesian Statistics
AY 19/20 Semester 2 Final Assignment

Tan Jun Yu (A0183140R)

Question 1

(a) We have $\mathbf{y} = \{y_1, \dots, y_n\}$, $p(y|\alpha, \beta) = \frac{\alpha\beta^\alpha}{(y+\beta)^{\alpha+1}}$, $y > 0, \alpha > 0, \beta > 0$.

$$p(\mathbf{y}|\alpha, \beta) = \prod_{i=1}^n \frac{\alpha\beta^\alpha}{(y_i + \beta)^{\alpha+1}} = (\alpha\beta^\alpha)^n \prod_{i=1}^n \frac{1}{(y_i + \beta)^{\alpha+1}}$$

(b) (i) $\beta = 1$.

$$\begin{aligned} p(\mathbf{y}|\alpha) &= \alpha^n \prod_{i=1}^n \frac{1}{(y_i + 1)^{\alpha+1}} = \alpha^n c^{\alpha+1} \\ &\propto \alpha^n \exp\{\log(c^{\alpha+1})\} \\ &\propto \alpha^n \exp\{\alpha \log c\} \end{aligned}$$

Propose $\alpha \sim \text{Gamma}(a, b)$ as the prior, so $p(\alpha) \propto \alpha^{a-1} \exp\{-b\alpha\}$.

$$\begin{aligned} p(\alpha|\mathbf{y}) &\propto p(\mathbf{y}|\alpha)p(\alpha) \\ &\propto \alpha^n \exp\{\alpha \log c\} \alpha^{a-1} \exp\{-b\alpha\} \\ &\propto \alpha^{n+a-1} \exp\{-\alpha(b - \log c)\} \end{aligned}$$

Therefore

$$\theta|\mathbf{y} \sim \text{Gamma}(n + a, b - \log c), \text{ where } c = \prod_{i=1}^n \frac{1}{(y_i + 1)}$$

We can also do this via the one-parameter exponential family, any model whose density can be expressed as $p(y|\theta) = h(y)g(\theta) \exp\{\eta(\theta)t(y)\}$.

$$\begin{aligned} p(\mathbf{y}|\alpha) &= \alpha^n \prod_{i=1}^n \frac{1}{(y_i + 1)^{\alpha+1}} = \alpha^n c^{\alpha+1} \\ &= \alpha^n \exp\left\{\log\left(\prod_{i=1}^n \frac{1}{(y_i + 1)^{\alpha+1}}\right)\right\} \\ &= \alpha^n \exp\left\{-(\alpha + 1) \sum_{i=1}^n \log(y_i + 1)\right\} \end{aligned}$$

where $h(\mathbf{y}) = 1, g(\alpha) = \alpha^n, \eta(\alpha) = -(\alpha + 1), t(\mathbf{y}) = \sum_{i=1}^n \log(y_i + 1)$. The conjugate prior is given by

$$\begin{aligned} p(\alpha) &\propto g(\alpha)^\nu \exp\{\eta(\alpha)\tau\} \\ &\propto (\alpha^n)^\nu \exp\{-(\alpha + 1)\tau\} \\ &\propto \alpha^{n\nu} \exp\{-\alpha\tau\} \\ \alpha &\sim \text{Gamma}(a, b), \text{ where } a = n\nu + 1, b = \tau \end{aligned}$$

This conjugate prior is part of the $\text{Gamma}(a, b)$ family, where $a = n\nu + 1, b = \tau$. The posterior density is

$$\begin{aligned} p(\alpha|\mathbf{y}) &\propto p(\mathbf{y}|\alpha)p(\alpha) \\ &= \alpha^n \exp\left\{-\alpha \sum_{i=1}^n \log(y_i + 1)\right\} \alpha^{n\nu} \exp\{-\alpha\tau\} \\ &\propto \alpha^{n+n\nu} \exp\left\{-\alpha \left(\tau + \sum_{i=1}^n \log(y_i + 1)\right)\right\} \\ \alpha|\mathbf{y} &\sim \text{Gamma}\left(n + n\nu + 1, \tau + \sum_{i=1}^n \log(y_i + 1)\right) \\ &\sim \text{Gamma}(n + a, \tau + t(\mathbf{y})) \end{aligned}$$

(ii)

$$\begin{aligned} \ell(\alpha) &= \log p(\mathbf{y}|\alpha) = \log \left[\alpha^n \prod_{i=1}^n (y_i + 1)^{-(\alpha+1)} \right] \\ &= n \log \alpha - (\alpha + 1) \sum_{i=1}^n \log(y_i + 1) \\ \ell'(\alpha) &= \frac{n}{\alpha} - \sum_{i=1}^n \log(y_i + 1), \quad \ell''(\alpha) = -\frac{n}{\alpha^2} \\ p(\alpha) &\propto \sqrt{I(\alpha)} = \sqrt{-\mathbf{E}_{y|\alpha} \left(-\frac{n}{\alpha^2} \right)} = \frac{n}{\alpha^2} \propto \frac{1}{\alpha} \end{aligned}$$

This is an improper prior as it does not integrate to 1.

$$\begin{aligned} p(\alpha|\mathbf{y}) &\propto p(\mathbf{y}|\alpha)p(\alpha) \\ &= \alpha^n \prod_{i=1}^n \frac{1}{(y_i + 1)^{\alpha+1}} \times \frac{1}{\alpha} \\ &\propto \alpha^{n-1} c^\alpha \end{aligned}$$

Since $\int_0^\infty \alpha^a b^\alpha d\alpha < \infty$, with $a = n - 1 \geq 0$, and $b = \prod_{i=1}^n \frac{1}{(y_i + 1)^\alpha}$, the posterior is a proper probability distribution. Also, $p(\alpha|\mathbf{y}) \propto \alpha^{n-1} \exp\{\alpha \log c\}$, so $\alpha|\mathbf{y} \sim \text{Gamma}(n, -\log c)$ which is a proper density.

(c) $\alpha = 1$. We have

$$p(\mathbf{y}|\beta) = \prod_{i=1}^n \frac{\beta}{(y_i + \beta)^2}, \quad p(\beta) = \frac{1}{(1 + \beta)^2}, \beta > 0$$

$$p(\beta|\mathbf{y}) \propto p(\mathbf{y}|\beta)p(\beta) = \left(\prod_{i=1}^n \frac{\beta}{(y_i + \beta)^2} \right) \times \frac{1}{(1 + \beta)^2}$$

Candidate proposal 1: $g(\beta) = \lambda \exp(-\lambda\beta)$

We want to find a finite positive bounding constant M such that

$$\frac{p(\beta|\mathbf{y})}{g(\beta)} = \frac{\left(\prod_{i=1}^n \frac{\beta}{(y_i + \beta)^2} \right) \frac{1}{(1 + \beta)^2}}{\lambda \exp(-\lambda\beta)} \leq M$$

Both $p(\beta|\mathbf{y})$ and $g(\beta)$ are decreasing functions, when $\beta \rightarrow \infty$, $g(\beta)$ decreases faster as it exponential, compared to $p(\beta|\mathbf{y})$ which is polynomial. This means that for large values of β , we are not able to obtain a finite M , and the posterior density cannot be enveloped by the proposal density. Therefore this candidate is inappropriate.

Candidate proposal 2: $p(\beta) = \frac{1}{(1 + \beta)^2}$

Again, we want to find a finite positive bounding constant M such that

$$\frac{p(\beta|\mathbf{y})}{p(\beta)} = \frac{\left(\prod_{i=1}^n \frac{\beta}{(y_i + \beta)^2} \right) \frac{1}{(1 + \beta)^2}}{\frac{1}{(1 + \beta)^2}} = \prod_{i=1}^n \frac{\beta}{(y_i + \beta)^2} \leq M$$

This likelihood function is positive. Since $y_i > 0$ for all i, \dots, n , $(y_i + \beta)^2 > \beta$ for all $\beta > 0$, this likelihood approaches 0^+ when $\beta \rightarrow \infty$, indicating that the proposal has larger tails than the posterior. The proposal is appropriate and we will define M as:

$$M = \frac{p(\beta|\mathbf{y})}{p(\beta)} = \sup_{\beta} \prod_{i=1}^n \frac{\beta}{(y_i + \beta)^2}$$

We can obtain draws from $p(\beta|by)$ using the rejection sampling algorithm:

- (1) Generate $\beta^{(s)} \sim p(\beta) = \frac{1}{(1 + \beta)^2}$
 - (2) Generate $U \sim \text{Uniform}[0, 1]$
 - (3) If $U < \frac{f(\beta^{(s)})}{Mg(\beta^{(s)})}$, accept $\beta^{(s)}$, otherwise reject $\beta^{(s)}$
 - (4) Repeat steps 1-3 until the desired sample $\{\beta^{(s)} : s = 1, \dots, S\}$ is obtained.
The members of this sample will be random variables from $p(\beta|\mathbf{y})$.
- (d) Equation (2) gives us $p(y, \lambda|\alpha, \beta) = p(y|\lambda)p(\lambda|\alpha, \beta)$. To obtain Equation (1), which

is $p(y|\alpha, \beta)$, we have to integrate λ out of $p(y, \lambda|\alpha, \beta)$.

$$\begin{aligned}
p(y|\alpha, \beta) &= \int p(y, \lambda|\alpha, \beta) d\lambda \\
&= \int p(y|\lambda) p(\lambda|\alpha, \beta) d\lambda \\
&= \int \lambda \exp(-\lambda y) \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta \lambda) d\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int \lambda^\alpha \exp\{-\lambda(\beta + y)\} d\lambda \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha + 1)}{(\beta + y)^{\alpha+1}} \\
&= \frac{\alpha \beta^\alpha}{(y + \beta)^{\alpha+1}} \quad \text{which is Equation (1)}
\end{aligned}$$

Question 2

(a) We have $p(y|\theta) = \theta(1-\theta)^y$ and $p(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}$, so the posterior density is

$$\begin{aligned} p(\theta|\mathbf{y}) &\propto p(\mathbf{y}|\theta)p(\theta) \\ &\propto \left(\prod_{i=1}^n \theta(1-\theta)^{y_i} \right) \theta^{a-1}(1-\theta)^{b-1} \\ &\propto \theta^n(1-\theta)^{\sum_{i=1}^n y_i} \theta^{a-1}(1-\theta)^{b-1} \end{aligned}$$

$$\begin{aligned} \ell(\theta) &= \log p(\theta|\mathbf{y}) = (n+a-1)\log \theta + \left(b-1 + \sum_{i=1}^n y_i \right) \log(1-\theta) \\ \ell'(\theta) &= \frac{n+a-1}{\theta} - \frac{b-1 + \sum_{i=1}^n y_i}{1-\theta} \\ \ell'(\theta) = 0 &\implies \frac{n+a-1}{\theta} = \frac{b-1 + \sum_{i=1}^n y_i}{1-\theta} \\ \hat{\theta} &= \frac{n+a-1}{n+a+b-2 + \sum_{i=1}^n y_i} \\ \ell''(\theta) &= -\frac{n+a-1}{\theta^2} - \frac{b-1 + \sum_{i=1}^n y_i}{(1-\theta)^2} \\ -\ell''(\theta) &= \frac{n+a-1}{\left(\frac{n+a-1}{n+a+b-2 + \sum_{i=1}^n y_i} \right)^2} + \frac{b-1 + \sum_{i=1}^n y_i}{\left(\frac{b-1 + \sum_{i=1}^n y_i}{n+a+b-2 + \sum_{i=1}^n y_i} \right)^2} \\ &= \frac{(n+a+b-2 + \sum_{i=1}^n y_i)^2}{n+a-1} + \frac{(n+a+b-2 + \sum_{i=1}^n y_i)^2}{b-1 + \sum_{i=1}^n y_i} \\ &= \frac{(n+a+b-2 + \sum_{i=1}^n y_i)^3}{(n+a-1)(b-1 + \sum_{i=1}^n y_i)} \end{aligned}$$

The normal approximation to the posterior of θ is

$$\theta|\mathbf{y} \stackrel{\text{approx}}{\sim} \text{N} \left(\frac{n+a-1}{n+a+b-2 + \sum_{i=1}^n y_i}, \frac{(n+a-1)(b-1 + \sum_{i=1}^n y_i)}{(n+a+b-2 + \sum_{i=1}^n y_i)^3} \right)$$

(b) Take $g(\theta) = \theta$.

$$\begin{aligned} h^*(\theta) &= -n^{-1}[\log g(\theta) + \ell(\theta)] \\ &= -n^{-1} \left[\log \theta + (n+a-1)\log \theta + \left(b-1 + \sum_{i=1}^n y_i \right) \log(1-\theta) \right] \\ &= -n^{-1} \left[(n+a)\log \theta + \left(b-1 + \sum_{i=1}^n y_i \right) \log(1-\theta) \right] \end{aligned}$$

$$\begin{aligned}
\frac{dh^*(\theta)}{d\theta} &= -n^{-1} \left[\frac{n+a}{\theta} - \frac{b-1 + \sum_{i=1}^n y_i}{1-\theta} \right] \\
\frac{dh^*(\theta)}{d\theta} = 0 &\implies \frac{n+a}{\theta^*} = \frac{b-1 + \sum_{i=1}^n y_i}{1-\theta^*} \\
\theta^* &= \frac{n+a}{a+b+n-1 + \sum_{i=1}^n y_i} \\
\frac{d^2h^*(\theta)}{d\theta^2} &= -n^{-1} \left[\frac{n+a}{\theta^2} + \frac{b-1 + \sum_{i=1}^n y_i}{(1-\theta)^2} \right]
\end{aligned}$$

$$\begin{aligned}
\hat{\sigma}^{*2} &= \left[\frac{d^2h^*(\theta)}{d\theta^2} \right]^{-1} \Big|_{\theta=\theta^*} = n \left[\frac{n+a}{\theta^{*2}} + \frac{b-1 + \sum_{i=1}^n y_i}{(1-\theta^*)^2} \right]^{-1} \\
&= n \left[\frac{(a+b+n-1 + \sum_{i=1}^n y_i)^2}{n+a} + \frac{(a+b+n-1 + \sum_{i=1}^n y_i)^2}{b-1 + \sum_{i=1}^n y_i} \right]^{-1} \\
&= n \frac{(n+a)(b-1 + \sum_{i=1}^n y_i)}{(a+b+n-1 + \sum_{i=1}^n y_i)^3}
\end{aligned}$$

$$\begin{aligned}
h(\theta) &= -n^{-1} \ell(\theta) \\
&= -n^{-1} \left[(n+a-1) \log \theta + \left(b-1 + \sum_{i=1}^n y_i \right) \log(1-\theta) \right]
\end{aligned}$$

$$\begin{aligned}
\frac{dh(\theta)}{d\theta} &= -n^{-1} \left[\frac{n+a-1}{\theta} - \frac{b-1 + \sum_{i=1}^n y_i}{1-\theta} \right] \\
\frac{dh(\theta)}{d\theta} = 0 &\implies \frac{n+a-1}{\hat{\theta}} = \frac{b-1 + \sum_{i=1}^n y_i}{1-\hat{\theta}} \\
\hat{\theta} &= \frac{n+a-1}{a+b+n-2 + \sum_{i=1}^n y_i} \\
\frac{d^2h(\theta)}{d\theta^2} &= -n^{-1} \left[\frac{n+a-1}{\theta^2} + \frac{b-1 + \sum_{i=1}^n y_i}{(1-\theta)^2} \right]
\end{aligned}$$

$$\begin{aligned}
\hat{\sigma}^2 &= \left[\frac{d^2h(\theta)}{d\theta^2} \right]^{-1} \Big|_{\theta=\hat{\theta}} = n \left[\frac{n+a-1}{\hat{\theta}^2} + \frac{b-1 + \sum_{i=1}^n y_i}{(1-\hat{\theta})^2} \right]^{-1} \\
&= n \left[\frac{(a+b+n-2 + \sum_{i=1}^n y_i)^2}{n+a-1} + \frac{(a+b+n-2 + \sum_{i=1}^n y_i)^2}{b-1 + \sum_{i=1}^n y_i} \right]^{-1} \\
&= n \frac{(n+a-1)(b-1 + \sum_{i=1}^n y_i)}{(a+b+n-2 + \sum_{i=1}^n y_i)^3}
\end{aligned}$$

$$\begin{aligned}
E(\theta|y) &\approx \frac{\hat{\sigma}^* \theta^* p(y|\theta^*) p(\theta^*)}{\hat{\sigma} p(y|\hat{\theta}) p(\hat{\theta})} \\
&= \frac{\sqrt{n \frac{(n+a)(b-1+\sum_{i=1}^n y_i)}{(a+b+n-1+\sum_{i=1}^n y_i)^3}} \cdot \theta^* \cdot (\theta^*)^{n+a-1} (1-\theta^*)^{b-1+\sum_{i=1}^n y_i}}{\sqrt{n \frac{(n+a-1)(b-1+\sum_{i=1}^n y_i)}{(a+b+n-2+\sum_{i=1}^n y_i)^3}} \cdot \hat{\theta}^{n+a-1} (1-\hat{\theta})^{b-1+\sum_{i=1}^n y_i}} \\
&= \frac{(n+a)^{n+a+\frac{1}{2}} (a+b+n-2+\sum_{i=1}^n y_i)^{a+b+n-\frac{1}{2}+\sum_{i=1}^n y_i}}{(n+a-1)^{n+a-\frac{1}{2}} (a+b+n-1+\sum_{i=1}^n y_i)^{a+b+n+\frac{1}{2}+\sum_{i=1}^n y_i}}
\end{aligned}$$

```

(c) set.seed(0183140)

n <- 5
sy <- 18
a <- 2
b <- 2
S <- 100

theta.samples <- runif(S)

logf <- function(theta, sy, n, a, b) {
  (n+a-1)*log(theta) + (b-1+sy)*log(1-theta)
}

logw <- logf(theta.samples, sy, n, a, b) -
  dunif(theta.samples, log = TRUE)
w <- exp(logw - max(logw))
W <- w/sum(w)
(est <- sum(W*theta.samples))

## [1] 0.2675

(se <- sqrt(sum((W*(theta.samples - est))^2)))

## [1] 0.0099

```

From the output above, the important sampling estimate of $E(\theta|by) = 0.2675$ and with standard error of 0.0099.

Question 3

- (a) The unnormalized posterior density of (θ_1, θ_2) is

$$\begin{aligned}
 p(\theta_1, \theta_2 | \mathbf{y}) &\propto p(\mathbf{y} | \theta_1, \theta_2) p(\theta_1, \theta_2) \\
 &\propto \left(\frac{\theta_1}{4} + \frac{1}{8}\right)^{y_1} \left(\frac{\theta_1}{4}\right)^{y_2} \left(\frac{\theta_2}{4}\right)^{y_3} \left(\frac{\theta_2}{4} + \frac{3}{8}\right)^{y_4} \left(\frac{1 - \theta_1 - \theta_2}{2}\right)^{y_5} \\
 &\propto \left(\frac{2\theta_1 + 1}{8}\right)^{y_1} \left(\frac{\theta_1}{4}\right)^{y_2} \left(\frac{\theta_2}{4}\right)^{y_3} \left(\frac{2\theta_2 + 3}{8}\right)^{y_4} \left(\frac{1 - \theta_1 - \theta_2}{2}\right)^{y_5} \\
 &\propto (2\theta_1 + 1)^{y_1} \theta_1^{y_2} \theta_2^{y_3} (2\theta_2 + 3)^{y_4} (1 - \theta_1 - \theta_2)^{y_5}
 \end{aligned}$$

- (b) We split the first and fourth category into two new categories each, augmenting the data into seven categories in total, with probabilities

$$\left(\frac{\theta_1}{4}, \frac{1}{8}, \frac{\theta_1}{4}, \frac{\theta_2}{4}, \frac{\theta_2}{4}, \frac{3}{8}, \frac{1 - \theta_1 - \theta_2}{2}\right)$$

Let the counts of the first two categories be $h, y_1 - h$ and the counts of the fifth and sixth categories be $k, y_4 - k$, so we have

$$\mathbf{y} = (h, y_1 - h, y_2, y_3, k, y_4 - k, y_5)$$

The full posterior is then

$$\begin{aligned}
 p(\theta_1, \theta_2, h, k | \mathbf{y}) &= \frac{p(\theta_1, \theta_2, h, k, \mathbf{y})}{p(\mathbf{y})} \propto p(\theta_1, \theta_2, h, k, \mathbf{y}) \\
 &= p(y, h, k | \theta_1, \theta_2) p(\theta_1, \theta_2) \\
 &= p(h, y_1 - h, y_2, y_3, k, y_4 - k, y_5 | \theta_1, \theta_2) \times 1 \\
 &= \frac{n!}{h!(y_1 - h)!y_2!y_3!k!(y_4 - k)!y_5!} \times \\
 &\quad \left(\frac{\theta_1}{4}\right)^h \left(\frac{1}{8}\right)^{y_1 - h} \left(\frac{\theta_1}{4}\right)^{y_2} \left(\frac{\theta_2}{4}\right)^{y_3} \left(\frac{\theta_2}{4}\right)^k \left(\frac{3}{8}\right)^{y_4 - k} \left(\frac{1 - \theta_1 - \theta_2}{2}\right)^{y_5} \\
 &\propto \frac{1}{h!(y_1 - h)!k!(y_4 - k)!} \times \\
 &\quad \left(\frac{1}{8}\right)^{y_1 - h} \left(\frac{3}{8}\right)^{y_4 - k} \left(\frac{\theta_1}{4}\right)^{h + y_2} \left(\frac{\theta_2}{4}\right)^{y_3 + k} \left(\frac{1 - \theta_1 - \theta_2}{2}\right)^{y_5}
 \end{aligned}$$

We find the four conditional posteriors:

$$\begin{aligned}
 p(\theta_1 | h, k, \theta_2, y) &\propto \left(\frac{\theta_1}{4}\right)^{h + y_2} \left(\frac{1 - \theta_1 - \theta_2}{2}\right)^{y_5} \\
 &\propto \theta_1^{h + y_2} (1 - \theta_2 - \theta_1)^{y_5} \\
 \implies \theta_1 | h, k, \theta_2, \mathbf{y} &\sim (1 - \theta_2) S_1, \quad S_1 \sim \text{Beta}(h + y_2 + 1, y_5 + 1)
 \end{aligned}$$

$$\begin{aligned}
 p(\theta_2 | h, k, \theta_1, y) &\propto \left(\frac{\theta_2}{4}\right)^{y_3 + k} \left(\frac{1 - \theta_1 - \theta_2}{2}\right)^{y_5} \\
 &\propto \theta - 2^{y_3 + k} (1 - \theta_2 - \theta_1)^{y_5} \\
 \implies \theta_2 | h, k, \theta_1, \mathbf{y} &\sim (1 - \theta_1) S_2, \quad S_2 \sim \text{Beta}(y_3 + k + 1, y_5 + 1)
 \end{aligned}$$

$$\begin{aligned}
p(h|k, \theta_1, \theta_2, \mathbf{y}) &\propto \frac{1}{h!(y_1 - h)!} \left(\frac{1}{8}\right)^{y_1 - h} \left(\frac{\theta_1}{4}\right)^{h + y_2} \\
&\propto \binom{y_1}{h} \left(\frac{1}{8}\right)^{y_1 - h} \left(\frac{\theta_1}{4}\right)^h \\
&\propto \binom{y_1}{h} \left(\frac{\frac{1}{8}}{\frac{1}{8} + \frac{\theta_1}{4}}\right)^{y_1 - h} \left(\frac{\frac{\theta_1}{4}}{\frac{1}{8} + \frac{\theta_1}{4}}\right)^h \\
&\propto \binom{y_1}{h} \left(\frac{1}{1 + 2\theta_1}\right)^{y_1 - h} \left(\frac{2\theta_1}{1 + 2\theta_1}\right)^h \\
\Rightarrow h|k, \theta_1, \theta_2, \mathbf{y} &\sim \text{Binomial}\left(y_1, \frac{2\theta_1}{1 + 2\theta_1}\right)
\end{aligned}$$

$$\begin{aligned}
p(k|h, \theta_1, \theta_2, \mathbf{y}) &\propto \frac{1}{k!(y_4 - k)!} \left(\frac{3}{8}\right)^{y_4 - k} \left(\frac{\theta_2}{4}\right)^{y_3 + k} \\
&\propto \binom{y_4}{k} \left(\frac{3}{8}\right)^{y_4 - k} \left(\frac{\theta_2}{4}\right)^k \\
&\propto \binom{y_4}{k} \left(\frac{\frac{3}{8}}{\frac{3}{8} + \frac{\theta_2}{4}}\right)^{y_4 - k} \left(\frac{\frac{\theta_2}{4}}{\frac{3}{8} + \frac{\theta_2}{4}}\right)^k \\
&\propto \binom{y_4}{k} \left(\frac{3}{3 + 2\theta_2}\right)^{y_4 - k} \left(\frac{2\theta_2}{3 + 2\theta_2}\right)^k \\
\Rightarrow k|h, \theta_1, \theta_2, \mathbf{y} &\sim \text{Binomial}\left(y_4, \frac{2\theta_2}{3 + 2\theta_2}\right)
\end{aligned}$$

Now we can use Gibbs sampler to draw T samples from the posterior $p(\theta_1, \theta_2|y)$.

- (i) Initialize $(\theta_1^{(0)}, \theta_2^{(0)})$ such that $0 \leq \theta_1^{(0)} \leq 1$, $0 \leq \theta_2^{(0)} \leq 1$, $0 \leq \theta_1^{(0)} + \theta_2^{(0)} \leq 1$
- (ii) For $t = 1, \dots, T$,

- (1) Draw $h^{(t)}$ from $\text{Binomial}\left(y_1, \frac{2\theta_1^{(t-1)}}{1 + 2\theta_1^{(t-1)}}\right)$
- (2) Draw $k^{(t)}$ from $\text{Binomial}\left(y_4, \frac{2\theta_2^{(t-1)}}{3 + 2\theta_2^{(t-1)}}\right)$
- (3) Draw $\theta_1^{(t)}$ from $(1 - \theta_2^{(t-1)})\text{Beta}(h^{(t)} + y_2 + 1, y_5 + 1)$
- (4) Draw $\theta_2^{(t)}$ from $(1 - \theta_1^{(t-1)})\text{Beta}(y_3 + k^{(t)} + 1, y_5 + 1)$

Question 4

- (a) We have $y_i \stackrel{\text{iid}}{\sim} \text{Binomial}(n_i, p_i)$ and $p(\beta_0, \beta_1) \propto 1$. Also,

$$\log \frac{p_i}{1 - p_i} = \beta_0 + \beta_1 x_i \implies p_i = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$$

The posterior density is

$$\begin{aligned} p(\beta_0, \beta_1 | \mathbf{y}) &\propto p(\mathbf{y} | \beta_0, \beta_1) p(\beta_0, \beta_1) \\ &\propto \prod_{i=1}^n \left(\frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)} \right)^{y_i} \left(\frac{1}{1 + \exp(\beta_0 + \beta_1 x_i)} \right)^{n_i - y_i} \end{aligned}$$

The log posterior density is

$$\begin{aligned} \log p(\beta_0, \beta_1 | \mathbf{y}) &\propto \sum_{i=1}^n y_i [(\beta_0 + \beta_1 x_i) - \log(1 + \exp(\beta_0 + \beta_1 x_i))] \\ &\quad + (n_i - y_i) [-\log(1 + \exp(\beta_0 + \beta_1 x_i))] \\ &\propto \sum_{i=1}^n \{y_i(\beta_0 + \beta_1 x_i) - n_i \log(1 + \exp(\beta_0 + \beta_1 x_i))\} \end{aligned}$$

- (b) The covariance for the normal approximation is the negative inverse of the Hessian from the output.

$$\begin{aligned} \det(H) &= -17.4066 \times -9017.5273 - (-393.5803 \times -393.5803) \\ &= 2059.038152 \end{aligned}$$

$$\begin{aligned} - \begin{bmatrix} -17.4066 & -393.5803 \\ -393.5803 & -9017.5273 \end{bmatrix}^{-1} &= \frac{1}{2059.038152} \begin{bmatrix} 9017.5273 & -393.5803 \\ -393.5803 & 17.4066 \end{bmatrix} \\ &= \begin{bmatrix} 4.3794963 & -0.191148133 \\ -0.1911481 & 0.008453774 \end{bmatrix} \end{aligned}$$

The normal approximation to the posterior $p(\beta_0, \beta_1 | \mathbf{y})$ is

$$\beta_0, \beta_1 | \mathbf{y} \sim N \left(\begin{bmatrix} -5.366289 \\ 0.238423 \end{bmatrix}, \begin{bmatrix} 4.3794963 & -0.191148133 \\ -0.1911481 & 0.008453774 \end{bmatrix} \right)$$

- (c) (i) Algorithm 2 uses a scale of 4 but Algorithm 1 has no scaling factor. This allows the random walk in Algorithm 1 to take larger steps, converging towards the posterior mode using fewer iterations.
- (ii) Algorithm 2 uses the covariance matrix in the normal approximation, so it models the posterior more closely and hence its acceptance rate will be closer to the optimal acceptance rate. We learnt that the optimal acceptance rate of a random walk Metropolis can be achieved by setting $g(\beta^* | \beta^{(t-1)})$ as the density $N(\theta^{(t-1)}, c^2(-H)^{-1})$, where H is the Hessian and $c = 2.4/\sqrt{d}$, for dimensionality of β, d , where $d \geq 3$.

Algorithm 1 assumes independence in (β_2, β_1) , but as we can observe from the covariance matrix in the normal approximation, this is not actually the case. Each sample for β_0 will affect β_1 and vice versa. If we use Algorithm 1, we will converge slowly to the posterior mode, because at each iteration β_0 and β_1 would not move in an optimal direction towards the posterior mode. This is also exacerbated by the fact that Algorithm 1 takes smaller steps for each iteration. Another downside of this is that Algorithm 1 will not explore the parameter space as much as Algorithm 2, which limits its ability to identify potential neighbouring peaks.

- (d) We can observe drastic changes in values for β_0 and β_1 for the first 200 or so iterations, these values are not distributed from the posterior distribution, but initial steps to approach the posterior mode given the starting values. We can discard the first 500 values as burn in. Alternatively, since starting values of around $(0, -1.25)$ were used according to the plots, we can use choose more appropriate starting values: the estimated posterior mode itself using `out$par` from the given output.

The traceplots do not show wild or rapid fluctuation; there are observable upward and downward trends. Thus, there are considerable serial dependence in both β_0 and β_1 , indicating that the Markov chain is not mixing well. We can use a larger scale so that wider steps are taken each iteration to induce greater fluctuation. Alternatively, we can also use the method of “thinning” only keep the sample at every n th iteration to reduce autocorrelation. However, this may lead to inefficiency without solving the root problem of insufficient step size.