ST4234: Bayesian Statistics

Take-home Midterm Exam Solutions, AY 19/20

1.

(a) **(3 marks)** The likelihood function is

$$p(\boldsymbol{y}|\theta) = \prod_{i=1}^{n} \theta(1-\theta)^{y_i} = \theta^n (1-\theta)^{\sum_{i=1}^n y_i}.$$

A class of conjugate prior is of the form $p(\theta) \propto \theta^{c_1}(1-\theta)^{c_2}$ for $\theta \in (0,1)$. We can recognize that this is the family of beta distributions. Therefore, a class of conjugate priors is the family of beta distributions $\mathrm{Beta}(a,b)$ with the pdf $p(\theta) = \frac{1}{\mathrm{B}(a,b)}\theta^{a-1}(1-\theta)^{b-1}$ for $\theta \in (0,1)$. The posterior of θ is

$$\begin{aligned}
p(\theta|\boldsymbol{y}) &\propto p(\boldsymbol{y}|\theta)p(\theta) \\
&\propto \theta^n(1-\theta)^{\sum_{i=1}^n y_i} \cdot \theta^{a-1}(1-\theta)^{b-1} \\
&\propto \theta^{n+a-1}(1-\theta)^{\sum_{i=1}^n y_i + b - 1}.
\end{aligned}$$

Hence, the posterior distribution of θ is $\mathrm{Beta}\left(n+a, \sum_{i=1}^n y_i + b\right)$.

(b) **(3 marks)** The prior mean of $\mathrm{Beta}(a,b)$ is $\frac{a}{a+b}$. The posterior mean is $\frac{n+a}{a+b+\sum_{i=1}^n y_i + n}$.

(c) **(4 marks)** For $y_{n+1} = 0, 1, 2, \ldots$, the posterior predictive density is

$$\begin{aligned}
&p(y_{n+1}|\boldsymbol{y}) \\
&= \int_0^1 p(y_{n+1}|\theta)p(\theta|\boldsymbol{y})\mathrm{d}\theta \\
&= \int_0^1 \theta(1-\theta)^{y_{n+1}} \cdot \frac{1}{\mathrm{B}\left(n+1, \sum_{i=1}^n y_i + 1\right)}\theta^n(1-\theta)^{\sum_{i=1}^n y_i}\mathrm{d}\theta \\
&= \frac{1}{\mathrm{B}\left(n+1, \sum_{i=1}^n y_i + 1\right)} \int_0^1 \theta^{n+1}(1-\theta)^{\sum_{i=1}^{n+1} y_i}\mathrm{d}\theta \\
&= \frac{\mathrm{B}\left(n+2, \sum_{i=1}^{n+1} y_i + 1\right)}{\mathrm{B}\left(n+1, \sum_{i=1}^n y_i + 1\right)} \qquad \text{(Full marks up to here)}
\end{aligned}$$

$$= \frac{\frac{\Gamma(n+2)\Gamma\left(\sum_{i=1}^{n+1} y_i+1\right)}{\Gamma\left(\sum_{i=1}^{n+1} y_i+n+3\right)}}{\frac{\Gamma(n+1)\Gamma\left(\sum_{i=1}^{n} y_i+1\right)}{\Gamma\left(\sum_{i=1}^{n} y_i+n+2\right)}}$$

$$= \frac{(n+1)\prod_{k=0}^{y_{n+1}-1}\left(\sum_{i=1}^{n} y_i + 1 + k\right)}{\prod_{k=0}^{y_{n+1}}\left(\sum_{i=1}^{n} y_i + n + 2 + k\right)},$$

for $y_{n+1} = 0, 1, 2, \ldots$. This distribution is called beta negative binomial distribution.

(d) **(4 marks)** First, we calculate the expected Fisher information. The log-likelihood function and its derivatives are:

$$\ell = \log p(\boldsymbol{y}|\theta) = n\log\theta + \sum_{i=1}^{n} y_i \log(1-\theta)$$

$$\frac{\partial \ell}{\partial \theta} = \frac{n}{\theta} - \frac{\sum_{i=1}^{n} y_i}{1-\theta}$$

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

$$I(\theta) = -\mathrm{E}_{\boldsymbol{y}|\theta}\left[\frac{\partial^2 \ell}{\partial \theta^2}\right] = \frac{n}{\theta^2} + \frac{\sum_{i=1}^{n} \mathrm{E}(y_i|\theta)}{(1-\theta)^2} = \frac{n}{\theta^2} + \frac{n\frac{1-\theta}{\theta}}{(1-\theta)^2}$$

$$= n\left(\frac{1}{\theta^2} + \frac{1}{\theta(1-\theta)}\right) = \frac{n}{\theta^2(1-\theta)}.$$

Hence, the Jeffreys prior for θ is given by

$$p(\theta) \propto \sqrt{I(\theta)} = \sqrt{\frac{n}{\theta^2(1-\theta)}} \propto \theta^{-1}(1-\theta)^{-1/2}.$$

Since $\theta \in (0,1)$, this becomes the density of Beta$(0, 1/2)$, which is not a proper density. We can also see this from the fact that $\int_0^1 \frac{1}{\theta\sqrt{1-\theta}} \mathrm{d}\theta = +\infty$. So the Jeffreys prior for θ is an improper prior.

2.

(a) **(5 marks)** The likelihood function is

$$p(\boldsymbol{x}, \boldsymbol{y}|\theta_1, \theta_2, \sigma^2) = p(\boldsymbol{x}|\theta_1, \theta_2, \sigma^2)p(\boldsymbol{y}|\theta_1, \theta_2, \sigma^2)$$

$$
\begin{aligned}
&= \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x_i-\theta_1)^2}{2\sigma^2}\right\} \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_j-\theta_2)^2}{2\sigma^2}\right\} \\
&\propto (\sigma^2)^{-\frac{m+n}{2}} \exp\left\{-\frac{m(\theta_1-\bar{x})^2}{2\sigma^2}\right\} \exp\left\{-\frac{n(\theta_2-\bar{y})^2}{2\sigma^2}\right\} \\
&\quad \times \exp\left\{-\frac{(m-1)s_x^2+(n-1)s_y^2}{2\sigma^2}\right\},
\end{aligned}
$$

where $\bar{x}, \bar{y}, s_x^2, s_y^2$ are as defined in Part (c).

The joint posterior density of $(\theta_1, \theta_2, \sigma^2)$ is

$$
\begin{aligned}
&p(\theta_1, \theta_2, \sigma^2|\boldsymbol{x}, \boldsymbol{y}) \propto p(\boldsymbol{x}, \boldsymbol{y}|\theta_1, \theta_2, \sigma^2) p(\theta_1, \theta_2, \sigma^2) \\
&\propto (\sigma^2)^{-\frac{m+n}{2}} \exp\left\{-\frac{m(\theta_1-\bar{x})^2}{2\sigma^2}\right\} \exp\left\{-\frac{n(\theta_2-\bar{y})^2}{2\sigma^2}\right\} \\
&\quad \times \exp\left\{-\frac{(m-1)s_x^2+(n-1)s_y^2}{2\sigma^2}\right\} \times (\sigma^2)^{-2} \\
&\propto \left[(\sigma^2)^{-1/2} \exp\left\{-\frac{m(\theta_1-\bar{x})^2}{2\sigma^2}\right\}\right] \\
&\quad \times \left[(\sigma^2)^{-1/2} \exp\left\{-\frac{n(\theta_2-\bar{y})^2}{2\sigma^2}\right\}\right] \\
&\quad \times \left[(\sigma^2)^{-\frac{m+n}{2}-1} \exp\left\{-\frac{(m-1)s_x^2+(n-1)s_y^2}{2\sigma^2}\right\}\right].
\end{aligned}
$$

In the last expression, the first bracket is $p(\theta_1|\sigma^2, \boldsymbol{x}, \boldsymbol{y})$, the second bracket is $p(\theta_2|\sigma^2, \boldsymbol{x}, \boldsymbol{y})$, and hence the third bracket is $p(\sigma^2|\boldsymbol{x}, \boldsymbol{y})$. They can be recognized as the following distributions:

$$
\begin{aligned}
\theta_1|\sigma^2, \boldsymbol{x}, \boldsymbol{y} &\sim \mathrm{N}\left(\bar{x}, \frac{\sigma^2}{m}\right), \\
\theta_2|\sigma^2, \boldsymbol{x}, \boldsymbol{y} &\sim \mathrm{N}\left(\bar{y}, \frac{\sigma^2}{n}\right), \\
\sigma^2|\boldsymbol{x}, \boldsymbol{y} &\sim \text{Inv-Gamma}\left(\frac{m+n}{2}, \frac{(m-1)s_x^2+(n-1)s_y^2}{2}\right).
\end{aligned}
$$

Meanwhile, the derivation above has shown that

$$
p(\theta_1, \theta_2, \sigma^2|\boldsymbol{x}, \boldsymbol{y}) \propto p(\theta_1|\sigma^2, \boldsymbol{x}, \boldsymbol{y}) \times p(\theta_2|\sigma^2, \boldsymbol{x}, \boldsymbol{y}) \times p(\sigma^2|\boldsymbol{x}, \boldsymbol{y}).
$$

(b) **(4 marks)** Note that conditional on σ^2, the posterior distributions of θ_1 and θ_2 are independent normals. Therefore, for $\delta = \theta_1 - \theta_2$, we have that

$$\begin{aligned}
p(\delta|\boldsymbol{x},\boldsymbol{y}) &= \int_0^\infty p(\delta,\sigma^2|\boldsymbol{x},\boldsymbol{y})\mathrm{d}\sigma^2 \\
&= \int_0^\infty p(\delta|\sigma^2,\boldsymbol{x},\boldsymbol{y})p(\sigma^2|\boldsymbol{x},\boldsymbol{y})\mathrm{d}\sigma^2 \\
&\propto \int_0^\infty (\sigma^2)^{-1/2}\exp\left\{-\frac{[\delta-(\bar{x}-\bar{y})]^2}{2\sigma^2(m^{-1}+n^{-1})}\right\} \\
&\quad \times (\sigma^2)^{-\frac{m+n}{2}-1}\exp\left\{-\frac{S}{2\sigma^2}\right\}\mathrm{d}\sigma^2 \\
&\propto \Gamma\left(\frac{m+n+1}{2}\right)\left[\frac{[\delta-(\bar{x}-\bar{y})]^2/(m^{-1}+n^{-1})+S}{2}\right]^{-\frac{m+n+1}{2}} \\
&\propto \left[[\delta-(\bar{x}-\bar{y})]^2/(m^{-1}+n^{-1})+S\right]^{-\frac{m+n+1}{2}} \\
&\propto \left[1+\frac{1}{m+n}\left(\frac{\delta-(\bar{x}-\bar{y})}{\sqrt{S/(mn)}}\right)^2\right]^{-\frac{m+n+1}{2}}.
\end{aligned}$$

Therefore, the transformed variable $t = \frac{[\delta-(\bar{x}-\bar{y})]}{\sqrt{S/(mn)}}$ satisfies $t|\boldsymbol{x},\boldsymbol{y} \sim t_{m+n}$, where $S = (m-1)s_x^2 + (n-1)s_y^2$.

(c) **(4 marks)** We can plug in the numbers to obtain that

$$\begin{aligned}
&\sigma^2|\boldsymbol{x},\boldsymbol{y} \sim \text{Inv-Gamma}(15, 11.8545), \\
&t = \frac{\delta-(\bar{x}-\bar{y})}{\sqrt{S/(mn)}} = \frac{\delta-(-4.11)}{0.3443}, \quad t|\boldsymbol{x},\boldsymbol{y} \sim t_{30}.
\end{aligned}$$

For the 90% HPD interval of σ^2, we can use the `hpd` function to get $[0.4918,\ 1.1908]$.

For the 90% HPD interval of δ, since the t_{30} distribution is symmetric, we can simply use

$$\left[(\bar{x}-\bar{y}) - q_{0.95,30}\sqrt{\frac{S}{mn}},\ \ (\bar{x}-\bar{y}) + q_{0.95,30}\sqrt{\frac{S}{mn}}\right],$$

which is $[-4.11 - 0.3443q_{0.95,30}, -4.11 + 0.3443q_{0.95,30}]$, where $q_{0.95,30} = 1.6973$ is the 0.95 quantile of t_{30}. The 90% HPD interval of δ is therefore $[-4.6944,\ -3.5256]$.

```
m <- 10
bar.x <- 1.656
s2.x <- 0.6750
n <- 20
bar.y <- 5.766
s2.y <- 0.9281

# posterior parameters for sigma2 ~ InvGamma(a1,b1)
(a1 <- (m+n)/2)

## [1] 15

(b1 <- ((m-1)*s2.x+(n-1)*s2.y)/2)

## [1] 11.85445

# HPD interval for sigma2
require(TeachingDemos)

## Loading required package:  TeachingDemos

## Warning:  package 'TeachingDemos' was built under R version 3.6.2

require(invgamma)

## Loading required package:  invgamma

hpd(qinvgamma, shape=a1, rate=b1, conf=0.90)

## [1] 0.4917925 1.1908722

# posterior center and scale for delta
(center.delta <- bar.x - bar.y)
```

```
## [1] -4.11

(scale.delta <- sqrt(((m-1)*s2.x+(n-1)*s2.y)/(m*n)))

## [1] 0.3443029

radius <- qt(0.95, df=m+n)
c(center.delta - scale.delta*radius, center.delta + scale.delta*radius)

## [1] -4.694372 -3.525628
```

3.

(a) **(3 marks)** Let Y_A and Y_B denote the number of tumour counts in type A and type B mice respectively. Then

$$Y_A|\theta_A \sim \text{Poisson}(\theta_A) \quad \text{and} \quad Y_B|\theta_B \sim \text{Poisson}(\theta_B).$$

Given $p(\theta_A, \theta_B) = p(\theta_A, \theta_B)$,

$$\begin{aligned} p(\theta_A, \theta_B|\boldsymbol{y}_A, \boldsymbol{y}_B) &\propto p(\boldsymbol{y}_A, \boldsymbol{y}_B|\theta_A, \theta_B)p(\theta_A, \theta_B|\boldsymbol{y}_A, \boldsymbol{y}_B) \\ &\propto p(\boldsymbol{y}_A, |\theta_A)p(\theta_A)p(\boldsymbol{y}_B|\theta_B)p(\theta_B) \\ & p(\theta_A|\boldsymbol{y}_A)p(\theta_B|\boldsymbol{y}_B) \end{aligned}$$

Therefore $p(\theta_A, \theta_B|\boldsymbol{y}_A, \boldsymbol{y}_B) = p(\theta_A|\boldsymbol{y}_A)p(\theta_B|\boldsymbol{y}_B)$. From lecture,

$$\boxed{\begin{array}{r} \theta \sim \text{Gamma}(a_0, b_0) \\ Y_1, \ldots, Y_n|\theta \overset{\text{i.i.d.}}{\sim} \text{Poisson}(\theta) \end{array} \quad \Rightarrow \quad \theta|\boldsymbol{y} \sim \text{Gamma}(a_0 + \sum_{i=1}^{n} y_i, b_0 + n)}$$

Therefore,

$$\begin{aligned} \theta_A|\boldsymbol{y}_A &\sim \text{Gamma}(120 + 117, 10 + 10) = \text{Gamma}(237, 20), \\ \theta_B|\boldsymbol{y}_B &\sim \text{Gamma}(12n_0 + 113, n_0 + 13). \end{aligned}$$

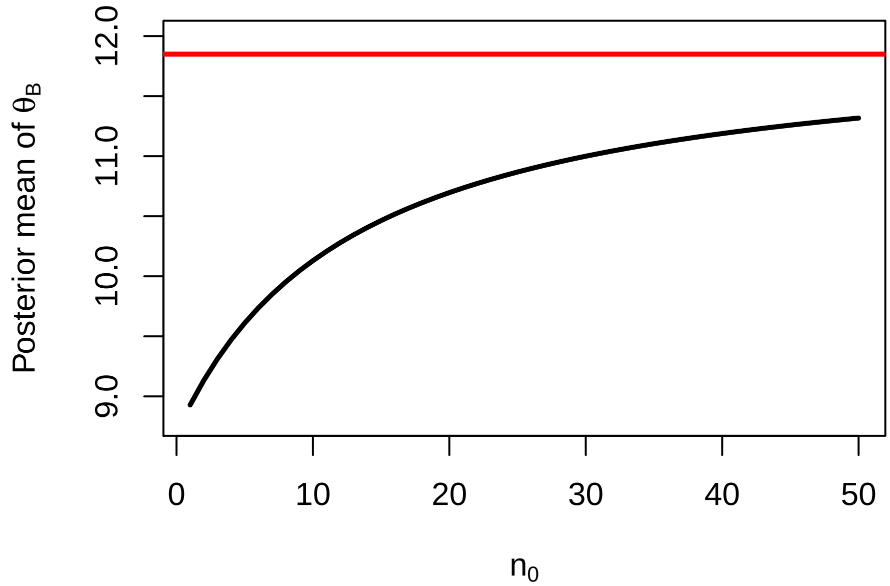
```r
yA <- c(12, 9, 12, 14, 13, 13, 15, 8, 15, 6)
yB <- c(11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7)
nA <- length(yA); nB <- length(yB)
a0A <- 120; b0A <- 10             # prior parameters for theta.A
aA <- a0A + sum(yA); bA <- b0A +nA      # posterior parameters for theta.A
```

(b) **(4 marks)** The posterior mean of θ_A is $237/20 = 11.85$. The posterior mean of θ_B is

$$\frac{12n_0 + 113}{n_0 + 13}.$$

The figure below plots the posterior means of θ_B against n_0. The posterior mean of θ_A is marked in the red horizontal line. The plot shows that the posterior mean of θ_B increases with n_0 and approaches the posterior mean of θ_A which is 237/20=11.85.

```r
posteriormeanA <- aA/bA
n0 <- 1:50
M <- length(n0)
posteriormeanB <- (12*n0+sum(yB))/(n0+nB)
plot(posteriormeanB, type="l", xlab=expression(n[0]), lwd=2.5,
     ylab=expression(paste("Posterior mean of ", theta[B])),
     ylim=c(8.8,12))
abline(h=posteriormeanA, col="red", lwd=2.5)
```



(c) **(3 marks)** Using Monte Carlo sampling, let $S = 10000$ and sample $\{\theta_A^{(1)}, \ldots, \theta_A^{(S)}\} \overset{\text{i.i.d.}}{\sim} p(\theta_A|\boldsymbol{y}_A)$ and $\{\theta_B^{(1)}, \ldots, \theta_B^{(S)}\} \overset{\text{i.i.d.}}{\sim} p(\theta_B|\boldsymbol{y}_B)$.

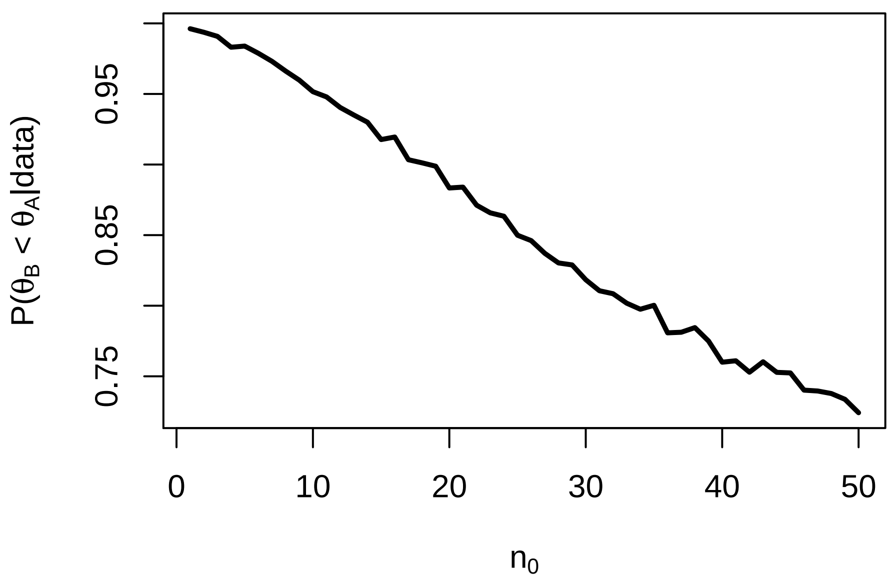
Then

$$\mathrm{P}(\theta_B < \theta_A|\boldsymbol{y}_A, \boldsymbol{y}_B) \approx \frac{1}{S}\sum_{s=1}^{S} \mathbb{1}\{\theta_B^{(s)} < \theta_A^{(s)}\}.$$

The figure below plots the estimated value of $\mathrm{P}(\theta_B < \theta_A|\boldsymbol{y}_A, \boldsymbol{y}_B)$. $\mathrm{P}(\theta_B < \theta_A|\boldsymbol{y}_A, \boldsymbol{y}_B)$ decreases steadily as n_0 increases, and appears to be quite sensitive to the prior on θ_B, ranging from 1 to about 0.75.

```
set.seed(4234)
S <- 10000
prob <- rep(0, M)
for (m in 1:M){
```

```r
        a0B <- 12*n0[m]
        b0B <- n0[m]
        aB <- a0B + sum(yB)
        bB <- b0B +nB
        thetaA.samples <- rgamma(S,aA,bA)
        thetaB.samples <- rgamma(S,aB,bB)
        prob[m] <- mean(thetaB.samples < thetaA.samples)
}
plot(prob, type="l", xlab=expression(n[0]), lwd=2.5,
     ylab=expression(paste("P(",theta[B]," < ",theta[A],"|data)")))
```



(d) **(3 marks)** Since the posterior predictive distribution is given by

$$p(\tilde{y}_A, \tilde{y}_B|\boldsymbol{y}_A, \boldsymbol{y}_B) = \int \underbrace{p(\tilde{y}_A, \tilde{y}_B|\theta_A, \theta_B)}_{p(\tilde{y}_A|\theta_A)p(\tilde{y}_B|\theta_B)} \underbrace{p(\theta_A, \theta_B|\boldsymbol{y}_A, \boldsymbol{y}_B)}_{p(\theta_A|\boldsymbol{y}_A)p(\theta_B|\boldsymbol{y}_B)} \, \mathrm{d}\theta_A \mathrm{d}\theta_B$$

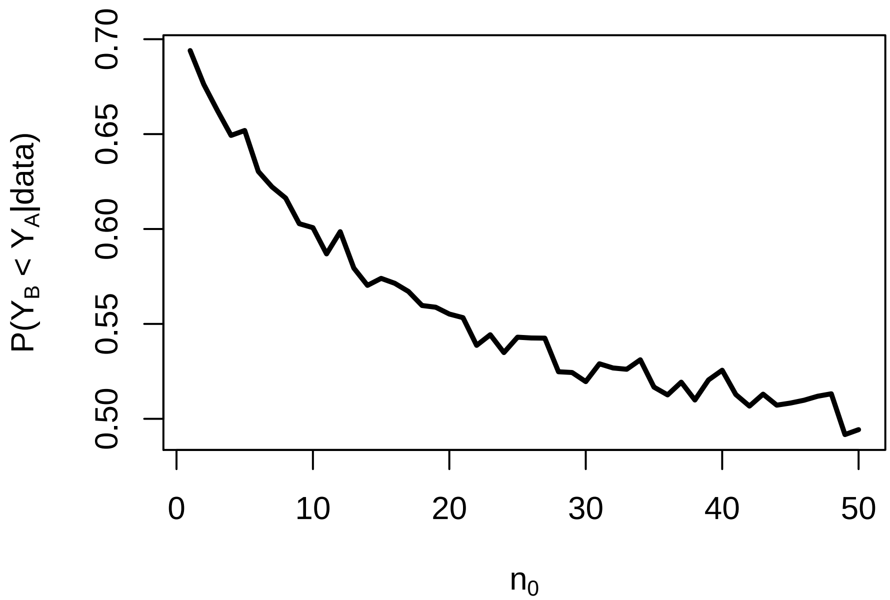
We can proceed in two steps. For $s = 1, \ldots, S = 10000$, sample $\{\theta_A^{(1)}, \ldots, \theta_A^{(S)}\} \overset{\text{i.i.d.}}{\sim} p(\theta_A|\boldsymbol{y}_A)$ and $\{\theta_B^{(1)}, \ldots, \theta_B^{(S)}\} \overset{\text{i.i.d.}}{\sim} p(\theta_B|\boldsymbol{y}_B)$.

Then for $s = 1, \ldots, S = 10000$, sample $\tilde{Y}_A^{(s)} \sim \text{Poisson}(\theta_A^{(s)})$ and $\tilde{Y}_B^{(s)} \sim \text{Poisson}(\theta_B^{(s)})$. Finally,

$$P(\tilde{Y}_B < \tilde{Y}_A|\boldsymbol{y}_A, \boldsymbol{y}_B) \approx \frac{1}{S}\sum_{s=1}^{S} \mathbb{1}\{\tilde{Y}_B^{(s)} < \tilde{Y}_A^{(s)}\}.$$

The figure below plots the estimated value of $\text{P}(\tilde{Y}_B < \tilde{Y}_A|\boldsymbol{y}_A, \boldsymbol{y}_B)$. $\text{P}(\tilde{Y}_B < \tilde{Y}_A|\boldsymbol{y}_A, \boldsymbol{y}_B)$ also decreases steadily as n_0 increases, and appears to be quite sensitive to the prior on θ_B, ranging from 0.7 to about 0.5. The similar decreasing pattern is observed as in Part (c).

```r
set.seed(4234)
predprob <- rep(0, M)
for (m in 1:M){
        a0B <- 12*n0[m]
        b0B <- n0[m]
        aB <- a0B + sum(yB)
        bB <- b0B +nB
        thetaA.samples <- rgamma(S,aA,bA)
        ytildeA.samples <- rpois(S,thetaA.samples)
        thetaB.samples <- rgamma(S,aB,bB)
        ytildeB.samples <- rpois(S,thetaB.samples)
        predprob[m] <- mean(ytildeB.samples < ytildeA.samples)
}
plot(predprob, type="l",xlab=expression(n[0]), lwd=2.5,
     ylab=expression(paste("P(",tilde(Y)[B]," < ",tilde(Y)[A],"|data)")))
```



(e) **(5 marks)** The figure below shows the histograms of $t^{(s)}$ for type A mice (left) and type B mice (right). The red line marks the observed values of t for the two types of mice.

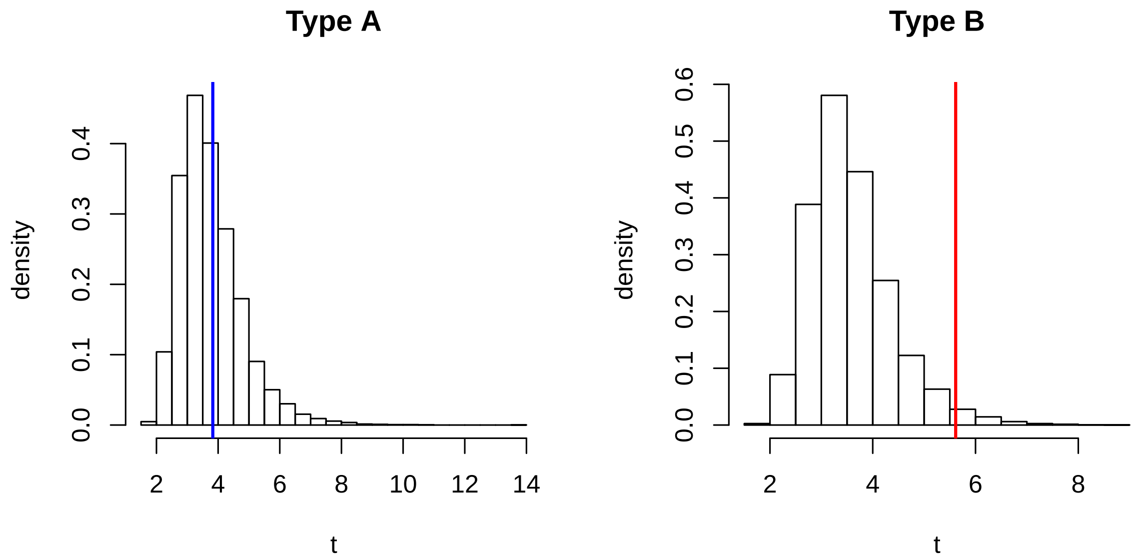
For type A, the Poisson model seems adequate since the observed value of t has a high posterior predictive probability, i.e. it is quite likely that a dataset predicted by our model is similar to the observed one in terms of t.

For type B, the Poisson model seems inadequate, however, as the observed value of t has a very low posterior predictive probability. Only around 2% of the Monte Carlo datasets had values of t equaling or exceeding the observed value. The datasets predicted by our Poisson model tend to have greater variation about the mean than the observed dataset. Hence the assumption of the Poisson model that the mean and variance are similar may not be appropriate for type B mice.

```
set.seed(4234)
tA.obs <- mean(yA)/sd(yA)
```

```
tB.obs <- mean(yB)/sd(yB)
S <- 10000
a0A <- 120; b0A <- 10           # prior parameters for theta.A
aA <- a0A + sum(yA); bA <- b0A + nA     # posterior parameters for theta.A
a0B <- 12*50; b0B <- 50          # prior parameters for theta.B
aB <- a0B + sum(yB); bB <- b0B + nB     # posterior parameters for theta.B
thetaA.samples <- rgamma(S,aA,bA)
thetaB.samples <- rgamma(S,aB,bB)
tA.samples <- rep(0,S)
tB.samples <- rep(0,S)

for (s in 1:S){
        YA.samples <- rpois(nA, thetaA.samples[s])
        YB.samples <- rpois(nB, thetaB.samples[s])
        tA.samples[s] <- mean(YA.samples)/sd(YA.samples)
        tB.samples[s] <- mean(YB.samples)/sd(YB.samples)
}
par(mfrow=c(1,2))
hist(tA.samples, freq=FALSE, breaks=20, main="Type A",ylab="density",xlab="t")
abline(v=tA.obs, col="blue", lwd=2)
hist(tB.samples, freq=FALSE, breaks=20, main="Type B",ylab="density",xlab="t")
abline(v=tB.obs, col="red", lwd=2)
```




```
mean(tA.samples >= tA.obs)

## [1] 0.3964

mean(tB.samples >= tB.obs)

## [1] 0.0221
```