# 1 Poisson population comparisons:

6.1 (pg 237) Let's reconsider the number of children data of Exercise 4.8. We'll assume Poisson sampling models for the two groups as before, but now we'll parameterize $\theta_A$ and $\theta_B$ as $\theta_A = \theta$, $\theta_B = \theta \times \gamma$. In this parameterization, $\gamma$ represents the relative rate $\theta_B/\theta_A$. Let $\theta \sim gamma(a_\theta, b_\theta)$ and let $\gamma \sim gamma(a_\gamma, b_\gamma)$.

a) Are $\theta_A$ and $\theta_B$ independent or dependent under this prior distribution? In what situations is such a joint prior distribution justified?

$\theta_A$ and $\theta_B$ be the average number of children of men in their 30s with and without bachelor's degrees, respectively. In order to prove $\theta_A$ and $\theta_B$ are independent or not, we can check

$\theta \sim gamma(a_\theta, b_\theta)$, $E[\theta] = \frac{a_\theta}{b_\theta}$, $V[\theta] = \frac{a_\theta}{b_\theta^2}$,

$E[\theta^2] = E[\theta]^2 + V[\theta] = (\frac{a_\theta}{b_\theta})^2 + \frac{a_\theta}{b_\theta^2} = \frac{a_\theta(a_\theta+1)}{b_\theta^2}$

$\gamma \sim gamma(a_\gamma, b_\gamma)$, $E[\gamma] = \frac{a_\gamma}{b_\gamma}$

$$E[\theta_A]E[\theta_B] = E[\theta]E[\theta \cdot \gamma] \underset{\theta \perp \gamma}{=} E[\theta]^2 \cdot E[\gamma] = \frac{a_\theta^2}{b_\theta^2}\frac{a_\gamma}{b_\gamma}$$

$$E[\theta_A\theta_B] = E[\theta \cdot \theta \cdot \gamma] \underset{\theta \perp \gamma}{=} E[\theta^2] \cdot E[\gamma] = \frac{a_\theta(a_\theta+1)}{b_\theta^2}\frac{a_\gamma}{b_\gamma}$$

For $a_\theta \neq a_\theta + 1$, $E[\theta_A\theta_B] \neq E[\theta_A]E[\theta_B]$

Thus, $\theta_A$ and $\theta_B$ are dependent under the prior distribution $\theta$ and $\gamma$.

We assume the model includes two parameters $\theta_A$ and $\theta_B$. $\theta_A$ follows a Gamma distribution. $\theta_B$ equal $\theta_A$ times another Gamma distributed parameter $\gamma$ which represents the relative rate.

b) Obtain the form of the full conditional distribution of $\theta$ given $y_A$, $y_B$ and $\gamma$.

Denote # of $y_A$ is $n_A$, # of $y_B$ is $n_B$

$$\begin{aligned}
p(\theta|y_A, y_B, \gamma) &\propto p(y_A, y_B|\theta, \gamma)p(\theta) \\
&\propto p(y_A|\theta)p(y_B|\theta, \gamma)p(\theta) \\
&\propto \prod_{i=1}^{n_A}(\theta^{y_A}e^{-\theta}) \cdot \prod_{i=1}^{n_B}((\theta\gamma)^{y_B}e^{-\theta\gamma}) \cdot \theta^{a_\theta-1}e^{-b_\theta\theta} \\
&\propto \theta^{\sum_{i=1}^{n_A} y_A}e^{-n_A\theta} \cdot (\theta\gamma)^{\sum_{i=1}^{n_B} y_B}e^{-n_B\theta\gamma} \cdot \theta^{a_\theta-1}e^{-b_\theta\theta} \\
&\propto \theta^{n_A\bar{y}_A}e^{-n_A\theta} \cdot \theta^{n_B\bar{y}_B}e^{-n_B\theta\gamma} \cdot \theta^{a_\theta-1}e^{-b_\theta\theta} \\
&\propto \theta^{n_A\bar{y}_A+n_B\bar{y}_B+a_\theta-1}e^{-\theta(n_A+n_B\gamma+b_\theta)} \\
\theta|y_A, y_B, \gamma &\sim Gamma(n_A\bar{y}_A + n_B\bar{y}_B + a_\theta, n_A + n_B\gamma + b_\theta)
\end{aligned}$$

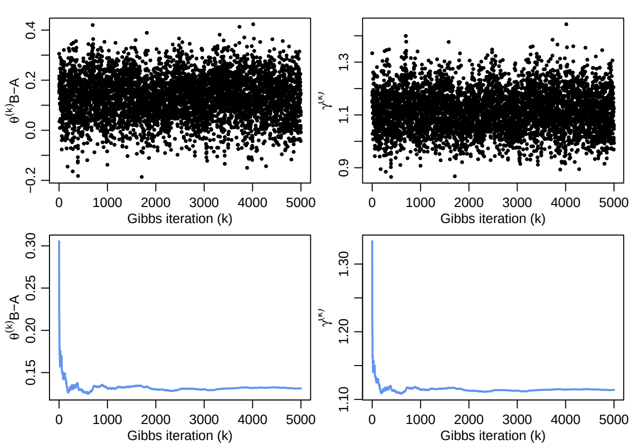
c) Obtain the form of the full conditional distribution of $\gamma$ given $y_A$, $y_B$ and $\theta$.

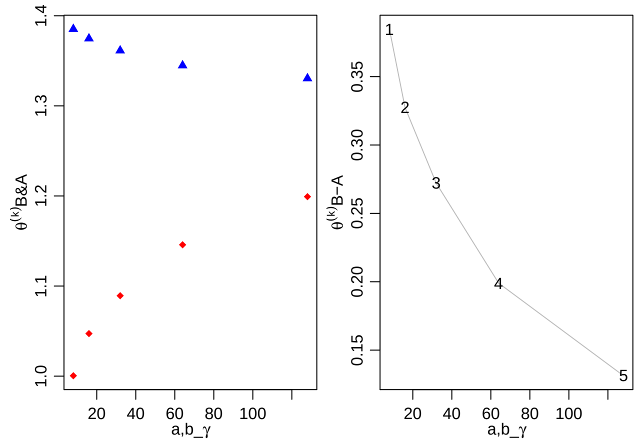
$$
\begin{aligned}
p(\gamma|y_A, y_B, \theta) &\propto p(y_B|\theta,\gamma)p(\gamma) \\
&\propto ((\theta\gamma)^{n_B\bar{y}_B} e^{-n_B\theta\gamma}) \cdot \gamma^{a_\gamma - 1} e^{-b_\gamma \gamma} \\
&\propto \gamma^{n_B\bar{y}_B} e^{-n_B\theta\gamma} \cdot \gamma^{a_\gamma - 1} e^{-b_\gamma \gamma} \\
&\propto \gamma^{n_B\bar{y}_B + a_\gamma - 1} e^{-\gamma(n_B\theta + b_\gamma)} \\
\gamma|y_A, y_B, \theta &\sim Gamma(n_B\bar{y}_B + a_\gamma, n_B\theta + b_\gamma)
\end{aligned}
$$

d) Set $a_\theta = 2$ and $b_\theta = 1$. Let $a_\gamma = b_\gamma \in \{8, 16, 32, 64, 128\}$. For each of these five values, run a Gibbs sampler of at least 5,000 iterations and obtain $E[\theta_B - \theta_A|y_A, y_B]$. Describe the effects of the prior distribution for $\gamma$ on the results

```r
rm(list = ls())
y.A <- scan("menchild30bach.dat"); y.B <- scan("menchild30nobach.dat")
nA <- length(y.A); nB <- length(y.B)
sumA <- sum(y.A); sumB <- sum(y.B)
a_theta<- 2; b_theta<- 1
S = 5000
ybar_a = mean(y.A); ybar_b = mean(y.B)
# Starting values
a_gamma <- b_gamma <-  c(8, 16, 32, 64, 128)
theta_diff <-A_theta <-B_theta <- numeric(5)
PHI<-matrix(nrow=S,ncol=3)
PHI[1,]<-phi<-c(ybar_a,ybar_b, ybar_b/ybar_a)

set.seed(1)
for (i in 1:5) {
  for (s in 1:S) {
# Sample theta_A from Gamma1
phi[1]= rgamma(1, a_theta + nA * ybar_a + nB * ybar_b,b_theta + nA + nB *phi[3])
# Sample gamma from Gamma2
phi[3]= rgamma(1,a_gamma[i] + nB * ybar_b,b_gamma[i] + nB * phi[1])
# Calculate theta_B
phi[2]=phi[1]*phi[3]
PHI[s,]<-phi
  }
A_theta[i]<- mean(PHI[,1]);B_theta[i]<- mean(PHI[,2])
theta_diff[i] <-   mean(PHI[,2]-PHI[,1])
}
```






For $a_\gamma = b_\gamma$, the 5 sets of $a_\gamma$ and $b_\gamma$ give same expected $\gamma$ value. However, the larger $a_\gamma$ and $b_\gamma$ will give smaller mean posterior difference between $\theta_B$ and $\theta_A$.

## 2. Gibbs sampling

Gibbs sampling is incredibly flexible and can easily handle hierarchical models, particularly when each piece of the model involves a conjugate (or at least semi-conjugate) prior. Hierarchical models are particularly useful when we don't have suffient information to specify meaningfully some parameters in the priors, or when there is an inherent hierarchical structure to the relationships between the data, latent variables, and parameters. So in these cases, we may want to put a prior not only on the parameters, but also on the *hyperparameters*. Priors on the prior parameters are often called *hyperpriors*.
Consider the Normal model with the semi-conjugate prior we defined in class, and assume that $\boldsymbol{\lambda}_0 \sim \text{Gamma}(\nu/2, \nu/1000)$, so that the model is defined by:

$$
\begin{aligned}
\boldsymbol{\lambda}_0 &\sim \text{Gamma}(\nu/2, \nu/1000) \\
\boldsymbol{\theta}|\lambda_0 &\sim \text{N}(\mu_0, \lambda_0^{-1}) \\
\boldsymbol{\lambda} &\sim \text{Gamma}(1/2, 1/2) \\
Y_1, \ldots, Y_n &\overset{iid}{\sim} \text{N}(\theta, \lambda^{-1})
\end{aligned}
$$

- Derive the Gibbs sampling algorithm to sample from the posterior $p(\theta, \lambda, \lambda_0 | y_{1:n})$.

$$
\begin{aligned}
p(\lambda_0|\theta,\lambda) &\propto p(\theta|\lambda_0)p(\lambda_0) \\
&\propto \lambda_0^{\frac{1}{2}} \exp[-\frac{1}{2}\lambda_0(\theta-\mu_0)^2] \cdot \lambda_0^{\frac{\nu}{2}-1} \exp[-\frac{\nu}{1000}\lambda_0] \\
&\propto \lambda_0^{\frac{1}{2}+\frac{\nu}{2}-1} \exp[-\lambda_0(\frac{1}{2}(\theta-\mu_0)^2 + \frac{\nu}{1000})] \\
\lambda_0|\theta,\lambda &\sim Gamma\left(\frac{1}{2}+\frac{\nu}{2}, \frac{1}{2}(\theta-\mu_0)^2 + \frac{\nu}{1000}\right)
\end{aligned}
$$

$$
\begin{aligned}
p(\theta|\lambda_0,\lambda,y_{1:n}) &\propto p(y_{1:n}|\theta,\lambda_0,\lambda)p(\theta|\lambda_0)\\
&\propto \prod_{i=1}^{n}\lambda^{\frac{1}{2}}\exp[-\frac{1}{2}\lambda(y_i-\theta)^2]\cdot\lambda_0^{\frac{1}{2}}\exp[-\frac{1}{2}\lambda_0(\theta-\mu_0)^2]\\
&\propto \lambda^{\frac{n}{2}}\exp[-\frac{1}{2}\lambda\sum_{i=1}^{n}(y_i-\theta)^2]\cdot\lambda_0^{\frac{1}{2}}\exp[-\frac{1}{2}\lambda_0(\theta^2-2\mu_0\theta)]\\
&\propto \lambda^{\frac{n}{2}}\lambda_0^{\frac{1}{2}}\exp[-\frac{1}{2}(n\lambda+\lambda_0)\theta^2+(n\bar{y}\lambda+\mu_0\lambda_0)\theta]\\
&\propto (n\lambda+\lambda_0)^{\frac{1}{2}}\exp[-\frac{n\lambda+\lambda_0}{2}(\theta-\frac{n\bar{y}\lambda+\mu_0\lambda_0}{n\lambda+\lambda_0})^2]\\
\theta|\lambda_0,\lambda,y_{1:n} &\sim Normal\left(\frac{n\bar{y}\lambda+\mu_0\lambda_0}{n\lambda+\lambda_0},(n\lambda+\lambda_0)^{-1}\right)
\end{aligned}
$$

$$
\begin{aligned}
p(\lambda|\lambda_0,\theta,y_{1:n}) &\propto p(y_{1:n}|\theta,\lambda_0,\lambda)p(\lambda)\\
&\propto \prod_{i=1}^{n}\lambda^{\frac{1}{2}}\exp[-\frac{1}{2}\lambda(y_i-\theta)^2]\cdot\lambda^{\frac{1}{2}-1}\exp[-\frac{1}{2}\lambda]\\
&\propto \lambda^{\frac{n}{2}}\exp[-\frac{1}{2}\lambda\sum_{i=1}^{n}(y_i-\theta)^2]\cdot\lambda^{\frac{1}{2}-1}\exp[-\frac{1}{2}\lambda]\\
&\propto \lambda^{\frac{n}{2}+\frac{1}{2}-1}\exp[-\lambda(\frac{1}{2}+\frac{1}{2}\sum_{i=1}^{n}(y_i-\theta)^2)]\\
&\propto \lambda^{\frac{n}{2}+\frac{1}{2}-1}\exp[-\lambda(\frac{1}{2}+\frac{1}{2}((n-1)S_n^2+\frac{n}{2}(\bar{y}-\theta)^2))]
\end{aligned}
$$

where

$$
\begin{aligned}
\sum_{i=1}^{n}(y_i-\theta)^2 =& \sum_{i=1}^{n}(y_i-\bar{y}+\bar{y}-\theta)^2\\
=& \sum_{i=1}^{n}(y_i-\bar{y})^2+2\sum_{i=1}^{n}(y_i-\bar{y})(\bar{y}-\theta)+\sum_{i=1}^{n}(\bar{y}-\theta)^2\\
=& \sum_{i=1}^{n}(y_i-\bar{y})^2+2(\bar{y}-\theta)\underbrace{\sum_{i=1}^{n}(y_i-\bar{y})}_{n\bar{y}-n\bar{y}=0}+n(\bar{y}-\theta)^2\\
=& \sum_{i=1}^{n}(y_i-\bar{y})^2+n(\bar{y}-\theta)^2\\
=& (n-1)S_n^2+n(\bar{y}-\theta)^2
\end{aligned}
$$

$$
\lambda|\lambda_0,\theta,y_{1:n}\sim Gamma\left(\frac{n+1}{2},\frac{1+(n-1)S_n^2+n(\bar{y}-\theta)^2}{2}\right)
$$

- Assume that:

$$
n=50,\quad \bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i=150\ \text{ and }\ S^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})=81.
$$

Implement and run the Gibbs sampler for $S = 10^4$ iterations, with each of the following: $(\mu_0 = 110, \nu = 1)$, $(\mu_0 = 0, \nu = 1)$, and $(\mu_0 = 110, \nu = 10)$. Comment on your results.

```
rm(list = ls()); set.seed(1)
# data
n<-50 ; mean.y<-150 ; var.y<- 81
# priors
mu0<-c(110,0,110)  ; nu0<-c(1,1,10)
S<-10^4
alpha<-1/2; beta<-1/2
lambda0 <- 500
PHI<-matrix(nrow=S,ncol=3); Theta<-Lambda <- numeric(3)
PHI[1,]<-phi<-c(lambda0, mean.y, 1/var.y)
```

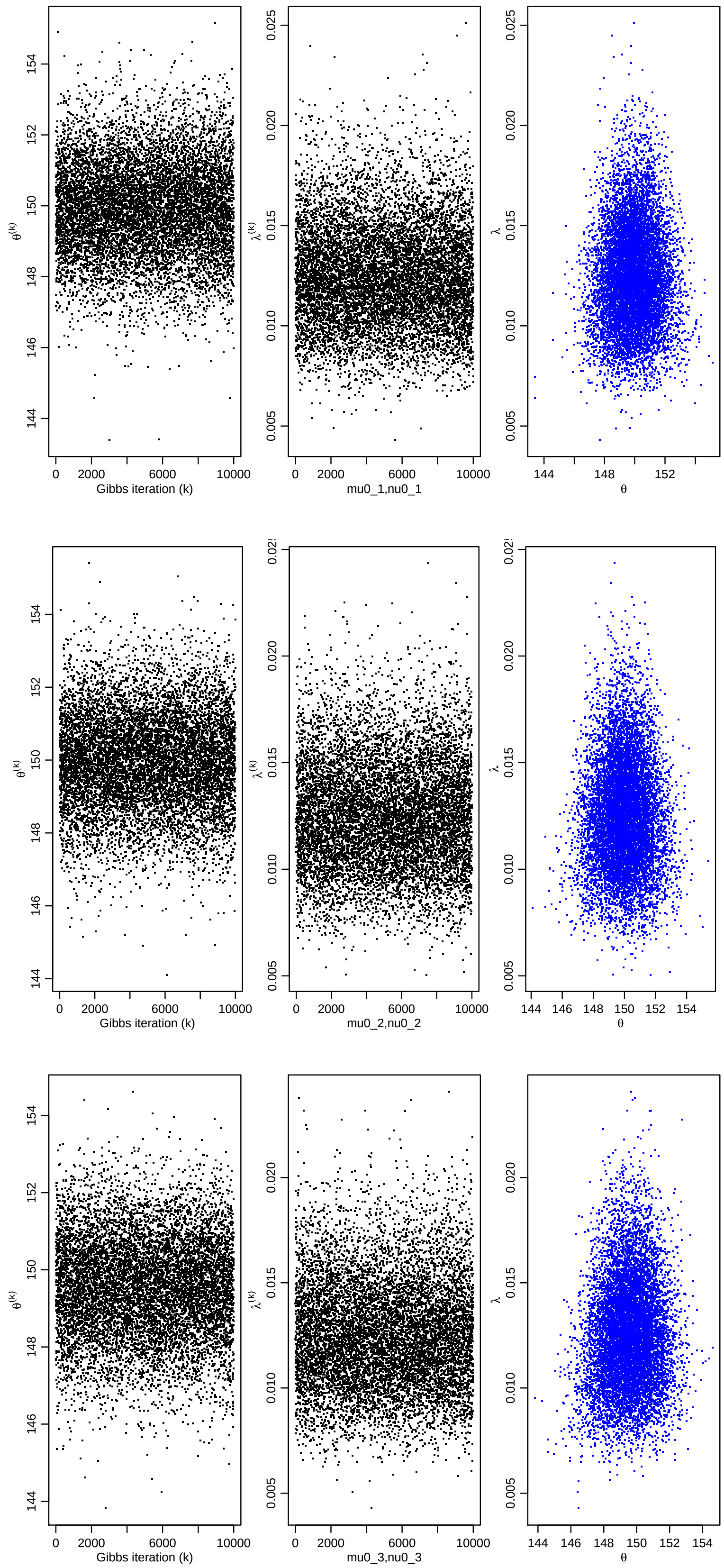
```
for(i in 1:3) {
alpha0<-nu0[i]/2; beta0<-nu0[i]/1000
# Gibbs sampling
 for(s in 2:S) {
# generate a new lambda0 value from its full conditional
  phi[1] <- rgamma(1, alpha0+alpha, beta0+beta*(phi[2]-mu0[i])^2 )
# generate a new theta value from its full conditional
  mu.n<- (mu0[i]*phi[1]+ n*mean.y*phi[3])/(phi[1]+ n*phi[3])
  lambda.n<- (phi[1]+ n*phi[3])
  phi[2]<- rnorm(1, mu.n, sqrt(1/lambda.n))
# generate a new lambda value from its full conditional
  alpha.n<- alpha+n/2
  beta.n<- beta + 1/2*((n-1)*var.y + n*(mean.y-phi[2])^2 )
  phi[3]<- rgamma(1, alpha.n, beta.n)
  PHI[s,]<-phi
 }
Theta[i]<- mean(PHI[,2]); Lambda[i]<- mean(PHI[,3])
}
Theta;Lambda
```

```
## [1] 149.9137 149.9679 149.5470
```

```
## [1] 0.01256768 0.01256851 0.01254272
```

It shows that the starting value of $\mu_0$ and $\nu = 10$ will not affects the results, if runing the Gibbs sampler for a large-enough number of iterations.

- For each of these values of $(\mu_0, \nu)$, make traceplots for $\theta$ and $\lambda$, and scatterplots (i.e., xy-plots) for $(\theta, \lambda)$.




- Calculate the means and 95% credible intervals for $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$.

```
### Gibbs based credible intervals
CI.theta <- round(quantile(PHI[1:S,2],probs=c(0.025,0.975)),4)
CI.lambda <- round(quantile(PHI[1:S,3],probs=c(0.025,0.975)),4)
```

```
CI.theta1;CI.lambda1;CI.theta2;CI.lambda2;CI.theta3;CI.lambda3
```

```
##     2.5%    97.5%
## 147.3491 152.5017

##   2.5%  97.5%
## 0.0082 0.0179

##     2.5%    97.5%
## 147.4494 152.5265

##   2.5%  97.5%
## 0.0081 0.0179

##     2.5%    97.5%
## 146.9954 152.0275

##   2.5%  97.5%
## 0.0082 0.0181
```