DASC 5410 Assignment1

Hao Xu T00732492

2024-02-28

Problem 1: Exercise 2.4

Symbolic manipulation: Prove the following form of Bayes' rule:

$$\Pr(H_j|E) = \frac{\Pr(E|H_j)\Pr(H_j)}{\sum_{k=1}^{K} \Pr(E|H_k)\Pr((H_k)}$$

where E is any event and $\{H_1, \ldots, H_K\}$ form a partition. Prove this using only axioms **P1-P3** from this chapter, by following steps a)-d) below:

P1: $0 = \Pr(\text{not}H|H) \leq \Pr(F|H) \leq \Pr(H|H) = 1$

P2: $\Pr(F \cap G|H) = \Pr(F|H) + \Pr(G|H)$ if $F \cap G = \emptyset$

P3: $\Pr(F \cap G|H) = \Pr(G|H)\Pr(F|G \cap H)$

a) Show that $\mathbf{Pr}(H_j|E)\mathbf{Pr}(E) = \mathbf{Pr}(E|H_j)\mathbf{Pr}(H_j)$

$$LHS = \Pr(H_j|E \cap \Omega) \cdot \Pr(E|\Omega) = \Pr(H_j \cap E|\Omega)$$
$$RHS = \Pr(E|H_j \cap \Omega) \cdot \Pr(H_j|\Omega) = \Pr(E \cap H_j|\Omega)$$
$$\therefore LHS = RHS \qquad (\text{Using } \mathbf{P3})$$

b) Show that $\mathbf{Pr}(E) = \mathbf{Pr}(E \cap H_1) + \mathbf{Pr}(E \cap \{\bigcup_{k=2}^{K} H_k\})$

$$\because \{H_1, \ldots, H_K\} \text{ form a partition}$$

$$\therefore H_1 \cap \{\bigcup_{k=2}^{K} H_k\} = \emptyset \qquad \text{and} \qquad H_1 \cup \{\bigcup_{k=2}^{K} H_k\} = \Omega$$

$$\therefore E \cap H_1 \cap \{\bigcup_{k=2}^{K} H_k\} = \emptyset.$$

$$\therefore \Pr(E) = \Pr(E|\Omega) = \Pr(E \cap H_1|\Omega) + \Pr(E \cap \{\bigcup_{k=2}^{K} H_k\}|\Omega) \qquad (\text{Using } \mathbf{P2})$$

$$= \Pr(E \cap H_1) + \Pr(E \cap \{\bigcup_{k=2}^{K} H_k\})$$

c) Show that $\mathbf{Pr}(E) = \sum_{k=1}^{K} \mathbf{Pr}(E \cap \{\bigcup_{k=1}^{K} H_k\})$

Prove by induction:

1. When $K = 1$, $H_1 = \Omega$, $\sum_{k=1}^{1} \Pr(E \cap \{\bigcup_{k=1}^{1} H_k\} = \Pr(E \cap H_1) = \Pr(E \cap \Omega) = \Pr(E)$

2. Assume the statement is true for $K = i$, which yields:

$$\Pr(E) = \sum_{k=1}^{i} \Pr(E \cap \{\bigcup_{k=1}^{i} H_k\})$$

3. For $K = i + 1$

$$\sum_{k=1}^{i+1} \Pr(E \cap \{\bigcup_{k=1}^{i+1} H_k\}) = \sum_{k=1}^{i} \Pr(E \cap H_i) + \Pr((E \cap H_{i+1})$$

$$= \Pr(E \cap \{\bigcup_{k=1}^{i} H_k\}) + \Pr(E \cap H_{i+1})$$

$$= \Pr(E \cap \{\bigcup_{k=1}^{i} H_k\}|\Omega) + \Pr(E \cap H_{i+1}|\Omega)$$

$$= \Pr\left((E \cap \{\bigcup_{k=1}^{i} H_k\}) \cap (E \cap H_{i+1})\right) \qquad \text{(Using \textbf{P2})}$$

$$= \Pr\left(E \cap \left(\{\bigcup_{k=1}^{i} H_k\} \cup H_{i+1}\right)\right)$$

$$= \Pr\left(E \cap \{\bigcup_{k=1}^{i+1} H_k\}\right)$$

d) Put it all together to show Bayes' rule, as described above.

$$\Pr(H_j|E) = \frac{\Pr(E|H_j)\Pr(H_j)}{\Pr(E)}$$

$$= \frac{\Pr(E|H_j)\Pr(H_j)}{\sum_{k=1}^{K} \Pr(E \cap H_k)}$$

$$= \frac{\Pr(E|H_j)\Pr(H_j)}{\sum_{k=1}^{K} \Pr(E|H_k)\Pr(H_k)}$$

Problem 2: Exercise 3.7

Posterior prediction: Consider a pilot study in which $n_1 = 15$ children enrolled in special education classes were randomly selected and tested for a certain type of learning disability. In the pilot study, $y_1 = 2$ children tested positive for the disability.

Part a)

Using a uniform prior distribution, find the posterior distribution of θ, the fraction of students in special education classes who have the disability. Find the posterior mean, mode and standard deviation of θ, and plot the posterior density.

"Uniform prior" $\Rightarrow \theta \sim \text{beta}(1,1)$

$$
\begin{aligned}
P(\theta|y_1) &= \frac{P(\theta)P(y_1|\theta)}{P(y_1)} \\
&= \frac{1}{P(y_1)} \cdot \frac{\Gamma(1+1)}{\Gamma(1)\Gamma(1)} \cdot \theta^{1-1}(1-\theta)^{1-1} \binom{n_1}{y_1} \theta^{y_1}(1-\theta)^{n-y_1} \\
&= c(n_1, y_1, 1, 1) \cdot \theta^{1+y_1-1} \cdot (1-\theta)^{1+n-y_1-1} \\
\therefore P(\theta|y_1) &\sim \text{beta}(1+y_1, 1+n_1-y_1) \Leftrightarrow P(\theta|y_1) \sim \text{beta}(3,14)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}[\theta|y_1] &= \frac{1+y_1}{1+1+n_1} = \frac{3}{17} \approx .1764706 \\
\text{mode}[\theta|y_1] &= \frac{1+y_1-1}{1+1+n_1-2} = \frac{2}{15} \approx .1333333 \\
\mathbb{E}[1-\theta|y_1] &= \frac{1}{B(1+y_1, 1+n_1-y_1)} \int_0^{+\infty} (1-\theta)\theta^{y_1}(1-\theta)^{n_1-y_1} d\theta \\
&= \frac{\Gamma(1+y_1, 1+n_1-y_1)}{\Gamma(1+y_1)\Gamma(1+n_1-y_1)} \int_0^{+\infty} \theta^{y_1}(1-\theta)^{n_1-y_1+1} d\theta \\
&= \frac{\Gamma(3+14)}{\Gamma(3)\Gamma(14)} \cdot \frac{\Gamma(3)\Gamma(15)}{\Gamma(3+15)} \\
&= \frac{\Gamma(17\sqrt{21})\Gamma(15)}{\Gamma(14)\Gamma(18)} \\
&= \frac{14}{17} \approx .8235294 \\
\text{sd}[\theta|y_2] &= \sqrt{\frac{\mathbb{E}[\theta|y_1]\mathbb{E}[1-\theta|y_1]}{1+1+n_1+1}} = \frac{\sqrt{21}}{51} \approx .0898544
\end{aligned}
$$

```r
# Posterior density plot for Exercise 3.7 part a
#pdf("Exercise_3.7.pdf",family="Times",height=4.4,width=7)
#par(mar=c(3,3,1,1),mgp=c(1.75,.75,0),oma=c(0,0,.5,0))
theta<-seq(0,1,length=1000)

a<-1; b<-1
n<-15 ; y<-2
plot(theta,dbeta(theta,a+y,b+n-y),type="l",ylab=
     expression(paste(italic("p("),theta,"|y)",sep="")),
     xlab=expression(theta), lwd=2)
mtext(expression(paste("beta(1,1) prior,  ", italic("n"),"=15  ",
     italic(sum(y[i])),"=2",sep="")), side=3,line=.1)
lines(theta,dbeta(theta,a,b),type="l",col="gray",lwd=2)
legend(.45,2.4,legend=c("prior","posterior"),lwd=c(2,2),
       col=c("gray","black"), bty="n")
```

beta(1,1) prior, $n=15$ $\sum y_i=2$
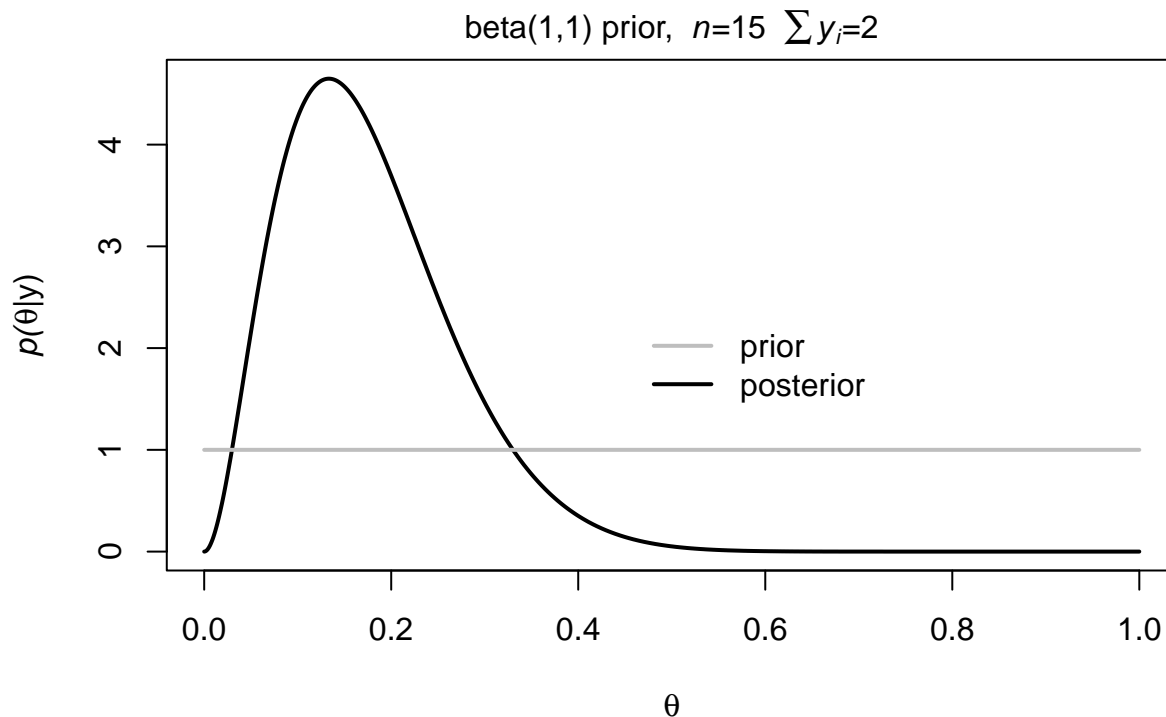
Researchers would like to recruit students with the disability to participate in a long-term study, but first they need to make sure they can recruit enough students. Let $n_2 = 278$ be the number of children in special education classes in this particular school district, and let Y_2 be the number of students with the disability.

Part b)

Find $\Pr(Y_2 = y_2|Y_1 = 2)$, the posterior predictive distribution of Y_2, as follows:

1. Discuss what assumptions are needed about the joint distribution of (Y_1, Y_2) such that the following is true:

$$\Pr(Y_2 = y_2|Y_1 = 2) = \int_0^1 \Pr(Y_2 = y_2|\theta)p(\theta|Y_1 = 2)d\theta$$

Y_1 and Y_2 are NOT independent.

2. Now plug in the forms for $\Pr(Y_2 = y_2|\theta)$ and $p(\theta|Y_1 = 2)$ in the above integral.

$$\Pr(Y_2 = y_2|\theta) = \binom{n_2}{y_2}\theta^{y_2}(1-\theta)^{n_2-y_2} = \binom{278}{y_2}\theta^{y_2}(1-\theta)^{278-y_2}$$

$$p(\theta|Y_1 = 2) = \frac{1}{B(3,14)}\theta^2(1-\theta)^{13}$$

$$\Pr(Y_2 = y_2|Y_1 = 2) = \int_0^1 \Pr(Y_2 = y_2|\theta)p(\theta|Y_1 = 2)d\theta$$

$$= \int_0^1 \binom{278}{y_2}\theta^{y_2}(1-\theta)^{278-y_2} \cdot \frac{1}{B(3,14)}\theta^2(1-\theta)^{13}$$

$$= \binom{278}{y_2}\frac{1}{B(3,14)}\int_0^1 \theta^{y_2+2}(1-\theta)^{291-y_2}d\theta$$

$$= \binom{278}{y_2}\frac{B((y_2+3),(292-y_2))}{B(3,14)}$$

4

3. Figure out what the above integral must be by using the calculus result discussed in Section 3.1.

$$(Y_2|Y_1 = 2) \sim \text{BetaBin}\left(n = 278, \alpha = 3, \beta = 14\right)$$

Beta-binomial distribution.

Part c)

Plot the function $\Pr(Y_2 = y_2|Y_1 = 2)$ as a function of y_2. Obtain the mean and standard deviation of Y_2, given $Y_1 = 2$.

```
#pdf("Exercise_3.7-2.pdf",family="Times",height=4.4,width=7)
#par(mar=c(3,3,1,1),mgp=c(1.75,.75,0),oma=c(0,0,.5,0))
py2 <- function(y2){
    choose(278,y2)*beta(y2+3,292-y2)/beta(3,14)
}
y2 <- seq(from = 0, to = 278, by = 1)

# mean
meany2 <- t(y2)%*%py2(y2)
print("The mean is: ")
```

```
## [1] "The mean is: "
```

```
print(meany2)
```

```
##           [,1]
## [1,] 49.05882
```
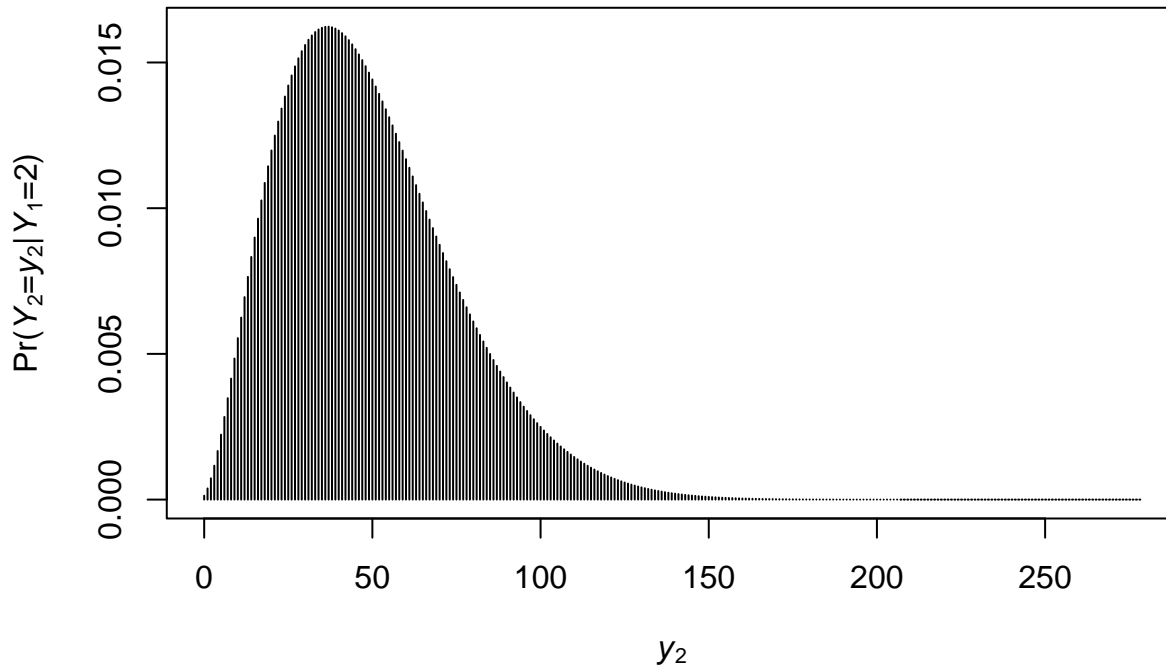
```
# standard deviation
sdy2 <- sqrt(t(y2^2)%*%py2(y2) - (t(y2)%*%py2(y2))^2)
print("The standard deviation is: ")
```

```
## [1] "The standard deviation is: "
```

```
print(sdy2)
```

```
##           [,1]
## [1,] 25.73196
```

```
plot(y2, py2(y2), type="h", ylab=
    expression(paste("Pr(",italic(Y[2]),"=", italic(y[2]),"|", italic(Y[1]), "=2)",sep="")),
    xlab=expression(italic(y[2])))
```

Also, for beta-binomial distribution,

$$\mathbb{E}[Y_2|Y_1=2] = \frac{n\alpha}{\alpha+\beta} = 49.05882$$

$$\text{sd}[Y_2|Y_1=2] = \sqrt{\frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}} = 25.73196$$

Pard d)

The posterior mode and the MLE (maximum likelihood estimate; see Exercise 3.14) of θ, based on data from the pilot study, are both $\hat{\theta} = \frac{2}{15}$. Plot the distribution $\Pr(Y_2 = y_2|\theta = \hat{\theta})$, and find the mean and standard deviation of Y_2 given $\theta = \hat{\theta}$. Compare these results to the plots and calculations in c) and discuss any differences. Which distribution for Y_2 would you use to make predictions, and why?

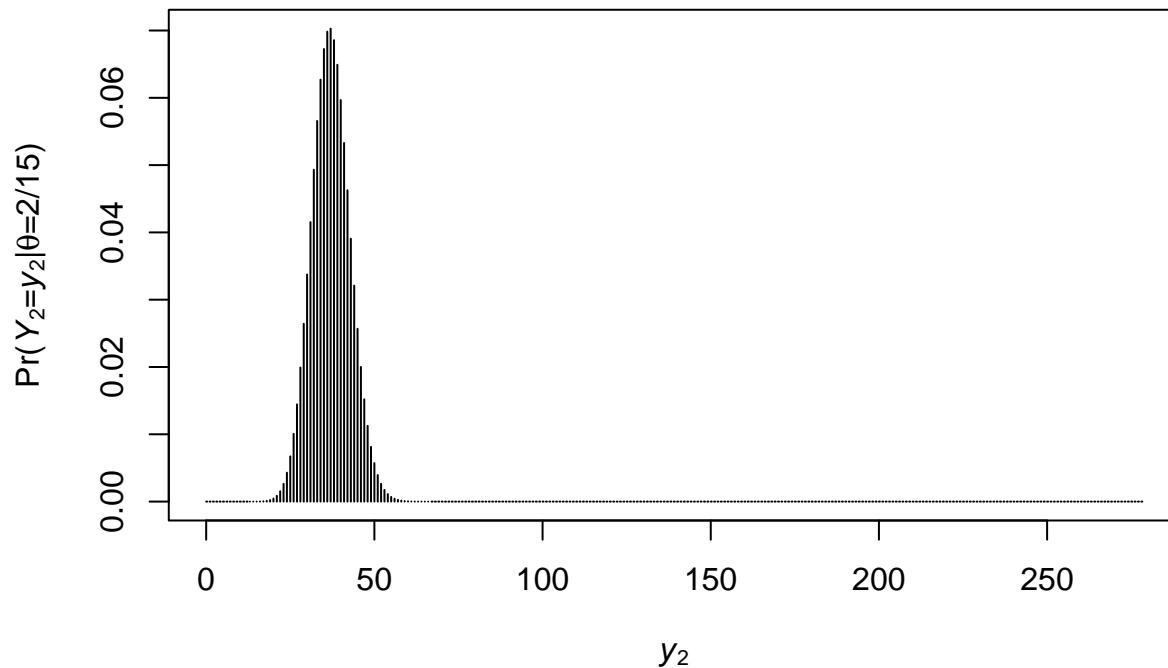$$(Y_2 = y_2|\theta = \hat{\theta} = \frac{2}{15}) \sim \text{Binomial}(278, \frac{2}{15}).$$

$$\Pr(Y_2 = y_2|\theta = \hat{\theta} = \frac{2}{15}) = \binom{278}{y_2}\left(\frac{2}{15}\right)^{y_2}\left(\frac{13}{15}\right)^{278-y_2}.$$

$\mathbb{E}(Y_2 = y_2|\theta = \hat{\theta} = \frac{2}{15}) = 278 \times \frac{2}{15} = 37.067$

$\text{sd}(Y_2 = y_2|\theta = \hat{\theta} = \frac{2}{15}) = \sqrt{278 \times \frac{2}{15} \times \frac{13}{15}} = 5.6678$

```
#pdf("Exercise_3.7-3.pdf",family="Times",height=4.4,width=7)
#par(mar=c(3,3,1,1),mgp=c(1.75,.75,0),oma=c(0,0,.5,0))
y2 <- seq(from = 0, to = 278, by = 1)

plot(y2, dbinom(y2, size=278, prob=2/15), type="h", ylab=
     expression(paste("Pr(",italic(Y[2]),"=", italic(y[2]),"|", theta, "=2/15)",sep="")),
     xlab=expression(italic(y[2])))
```

I will use d) to make predictions because the variance is smaller.

Problem3 Exercise 4.8

More posterior predictive checks:

Let θ_A and θ_B be the average number of children of men in their 30s with and without bachelor's degrees, respectively.

Part a)

Using a Poisson sampling model, a gamma(2,1) prior for each θ and the data in the files `menchild30bach.dat` and `menchild30nobach.dat`, obtain 5,000 samples of $\tilde{Y}_A$ and $\tilde{Y}_B$ from the posterior predictive distribution of the two samples. Plot the Monte Carlo approximations to these two posterior predictive distributions.

We know from chapter 3 that in a poisson model, the posterior distribution for θ given data is a Gamma distribution:

$$\theta \mid y_1, \ldots, y_n \sim \mathrm{Gamma}(a + \sum_i y_i, b + n)$$

```
#pdf("Exercise_4.8.pdf",family="Times",height=3.5,width=7)
#par(mar=c(3,3,1,1),mgp=c(1.75,.75,0))
#par(mfrow=c(1,2))

menchild30bach = scan("menchild30bach.dat")
menchild30nobach = scan('menchild30nobach.dat')

a = 2
b = 1
N = 5000
```
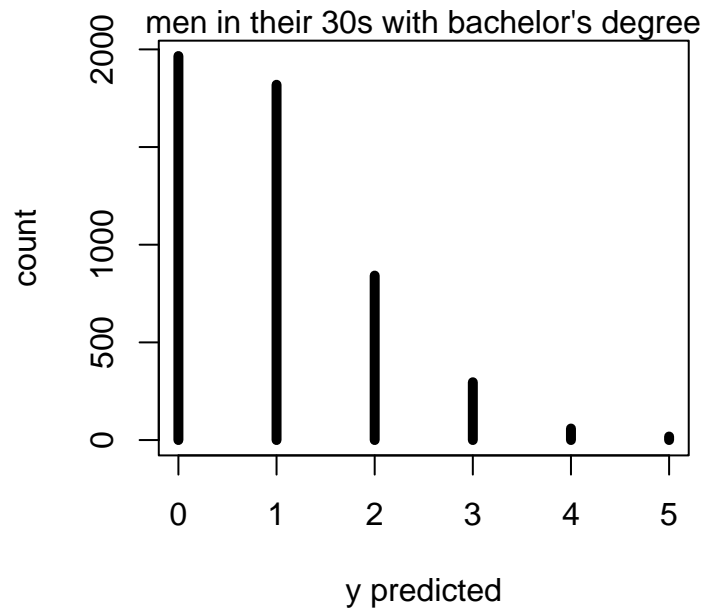
```
set.seed(123)
thetaA.prior = rgamma(N, a + sum(menchild30bach), b + length(menchild30bach))
thetaB.prior = rgamma(N, a + sum(menchild30nobach), b + length(menchild30nobach))

ynewA = rpois(N, thetaA.prior)
ynewB = rpois(N, thetaB.prior)

plot(table(ynewA), type="h", xlab="y predicted", ylab="count", lwd=5)
mtext("men in their 30s with bachelor's degree",side=3)
```
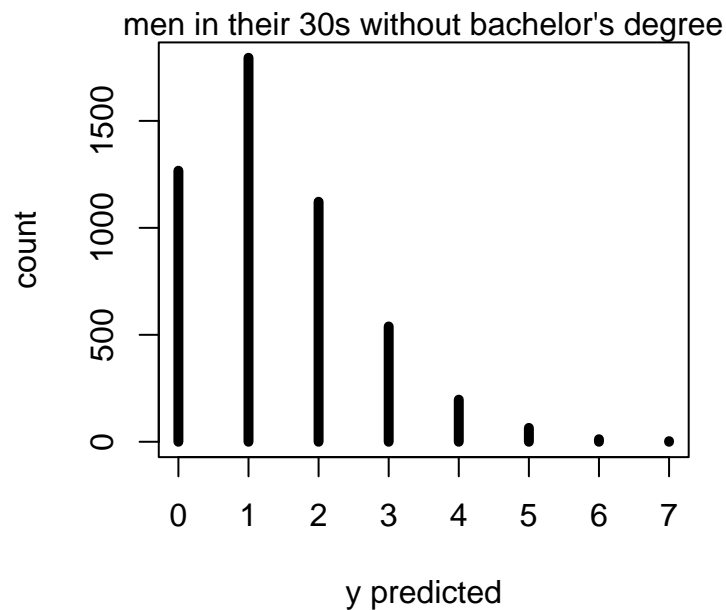


```
plot(table(ynewB), type="h", xlab="y predicted", ylab="count", lwd=5)
mtext("men in their 30s without bachelor's degree",side=3)
```

```
#dev.off()
```

Part b)

Find 95% quantile-based posterior confidence intervals for $\theta_B - \theta_A$ and $Y_B - Y_A$. Describe in words the differences between the two populations using these quantities and the plots in a), along with any other results that may be of interest to you.

$\theta_B - \theta_A$ is:

```
theta.diff = thetaB.prior - thetaA.prior
print(quantile(theta.diff, c(0.025, 0.975)))
```

```
##      2.5%     97.5%
## 0.1559684 0.7359099
```

$Y_B - Y_A$ is:

```
ynew.diff = ynewB - ynewA
print(quantile(ynew.diff, c(0.025, 0.975)))
```

```
##  2.5% 97.5%
##   -3     4
```

Since the lower bound of the confidence interval for $\theta_B - \theta_A$ is positive, we believe with at least approximately 95% probability that the true posterior mean θ_B is greater than θ_A. However, we cannot draw the same conclusion for the difference between two individuals sampled from this distribution, due to the increased uncertainty in the posterior predictive distribution.

Part c)

Obtain the empirical distribution of the data in group B. Compare this to the Poisson distribution with mean $\hat{\theta} = 1.4$. Do you think the Poisson model is a good fit? Why or why not?

```
# empirical distribution of the data in group B
ecdf<-as.vector((table(menchild30nobach))/sum(table(menchild30nobach)))
Nchildren = length(table(menchild30nobach))
plot(1:Nchildren +.05,
     ecdf,
     type="h",lwd=5,
     xlab="number of children",
     ylab="Probability",
     col="gray")
points(1:Nchildren-.05,
       dpois(1:Nchildren, 1.4),
       lwd=5,col="black",type="h")
legend(3.25,.31,
       legend=c(
        expression(paste("Poisson distribution when ",
                  italic(theta),"=1.4",sep="")),
        "empirical distribution"),
       lwd=c(2,2),col=c("black","gray"),bty="n",cex=.75)
```
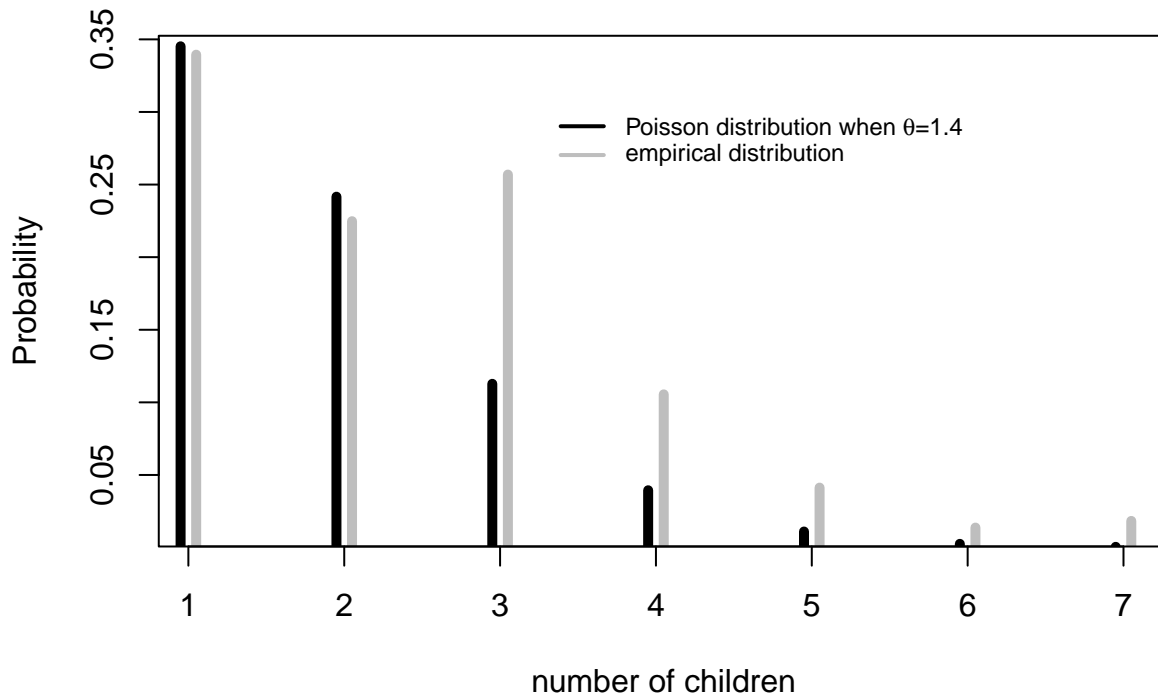
From the plot, the probability of having 3 or 4 children in empirical distribution is two times the probability of poisson distribution when $\theta = 1.4$. So the poisson model is not a good fit.

Part d)

For each of the 5,000 θ_B-values you sampled, sample $n_B = 218$ Poisson random variables and count the number of 0s and the number of 1s in each of the 5,000 simulated datasets. You should now have two sequences of length 5,000 each, one sequence counting the number of people having zero children for each of the 5,000 posterior predictive datasets, the other counting the number of people with one child. Plot the two sequences against one another (one on the x-axis, one on the y-axis). Add to the plot a point marking how many people in the observed dataset had zero children and one child. Using this plot, describe the adequacy of the Poisson model.

Model checking, by computing a statistic that may be of interest:

```r
par(mar=c(3,3,1,1),mgp=c(1.75,.75,0),oma=c(0,0,.5,0))
Zero_One <- lapply(thetaB.prior, function(theta) {
  nB <- rpois(218, theta)
  n0s <- sum(nB == 0)
  n1s <- sum(nB == 1)
  c(n0s, n1s)
})

zeros <- sapply(Zero_One, function(t) t[1])
ones <- sapply(Zero_One, function(t) t[2])

plot(zeros, ones, xlab = "Zeros", ylab = "Ones", main = "Scatter Plot", col = rgb(0, 0, 0, 0.5))

points(sum(menchild30nobach == 0), sum(menchild30nobach == 1), col = 'red', pch = 17)
```

Scatter Plot
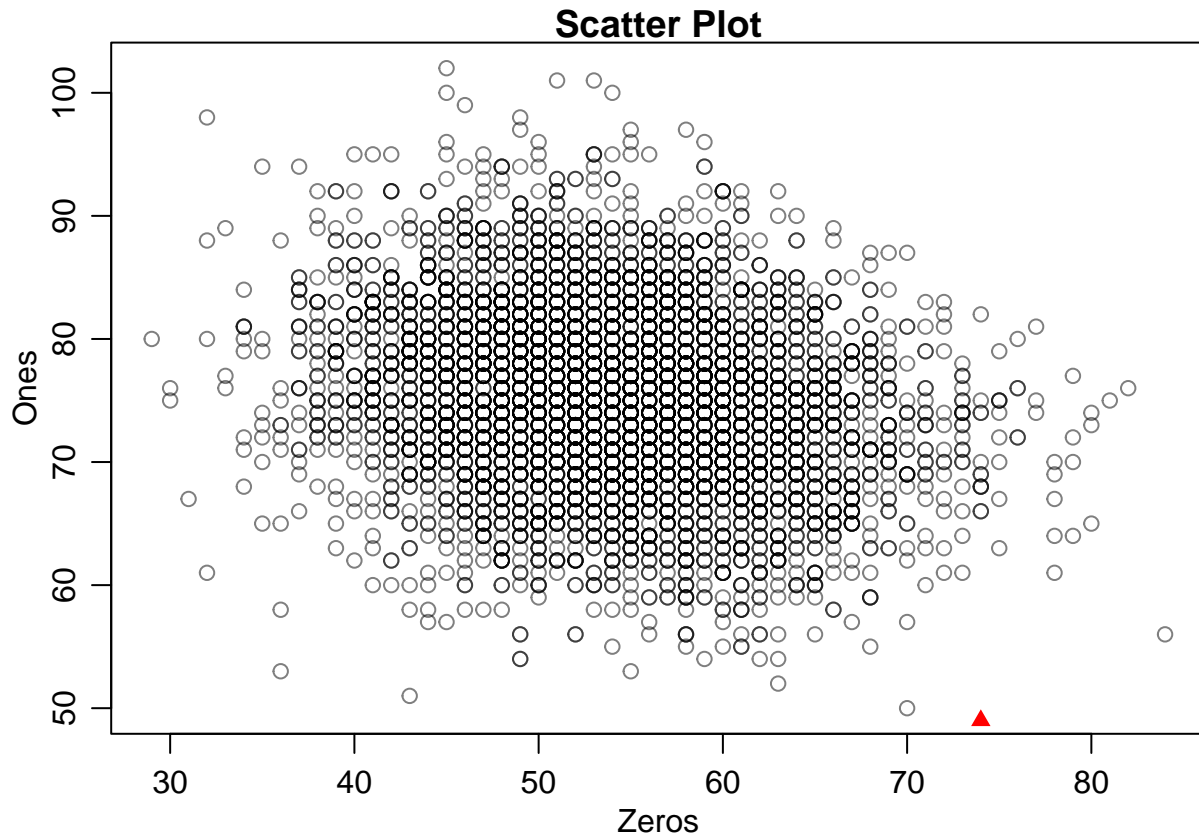
The observed statistic marked in solid triangle appears noticeably unusual when compared to sample datasets generated from our Poisson model with $\theta = 1.4$, particularly regarding the count of ones observed. This strongly suggests that the fit of our Poisson model is less than optimal.

Problem 4 Exercise 5.2

Sensitivity analysis: Thirty-two students in a science classroom were randomly assigned to one of two study methods, A and B, so that $n_A = n_B = 16$ students were assigned to each method. After several weeks of study, students were examined on the course material with an exam designed to give an average score of 75 with a standard deviation of 10. The scores for the two groups are summarized by $\{\bar{y}_A = 75.2, s_a = 7.3\}$ and $\{\bar{y}_B = 77.5, s_b = 8.1\}$. Consider independent, conjugate normal prior distributions for each of θ_A and θ_B, with $\mu_0 = 75$ and $\sigma_0^2 = 100$ for both groups. For each $(\kappa_0, \nu_0) \in \{(1, 1), (2, 2), (4, 4), (8, 8), (16, 16), (32, 32)\}$ (or more values), obtain $\Pr(\theta_A < \theta_B | \mathbf{y}_A, \mathbf{y}_B)$ via Monte Carlo sampling. Plot this probability as a function of (κ_0, ν_0). Describe how you might use this plot to convey the evidence that $\theta_A < \theta_B$ to people of a variety of prior opinions.

```
# prior

k0nu0 = c(1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024)
mu0 = 75
s20 = 100

# data
n.A = n.B = 16
ybar.A = 75.2
s2.A = 7.3^2
ybar.B = 77.5
```

```r
s2.B = 8.1^2


# define the function to calculate difference
# between posterior mean
MC.diff <- function(p) {
  # p is the common parameter for k0 and nu0

  # MC sample size
  NSAMPLE = 100000

  # posterior inference
  kn.A <- p + n.A
  nun.A <- p + n.A
  mun.A <- (p * mu0 + n.A * ybar.A) / kn.A
  s2n.A <- (p * s20 + (n.A - 1) * s2.A + ((p * n.A) * (ybar.A - mu0)^2 / kn.A)) / nun.A

  # MC
  s2.A.mc = 1 / rgamma(NSAMPLE, nun.A / 2, rate=s2n.A * nun.A / 2)
  theta.A.mc = rnorm(NSAMPLE, mun.A, sqrt(s2.A.mc))

  # posterior inference
  kn.B <- p + n.B
  nun.B <- p + n.B
  mun.B <- (p * mu0 + n.B * ybar.B) / kn.B
  s2n.B <- (p * s20 + (n.B - 1) * s2.B + ((p * n.B) * (ybar.B - mu0)^2 / kn.B)) / nun.B

  # MC
  s2.B.mc = 1 / rgamma(NSAMPLE, nun.B / 2, rate=s2n.B * nun.B / 2)
  theta.B.mc = rnorm(NSAMPLE, mun.B, sqrt(s2.B.mc))

  mean(theta.A.mc < theta.B.mc)
}

set.seed(1)
prob = sapply(k0nu0, MC.diff)

plot(k0nu0,prob,type = 'l',
     xlab=expression(paste("prior belief=",kappa[0],"=",nu[0],sep="")),
     ylab=expression(paste("Pr(",theta[A],"<",theta[B],"|",y[A],",",y[B],")",sep=""))
)
points(k0nu0,prob)
```
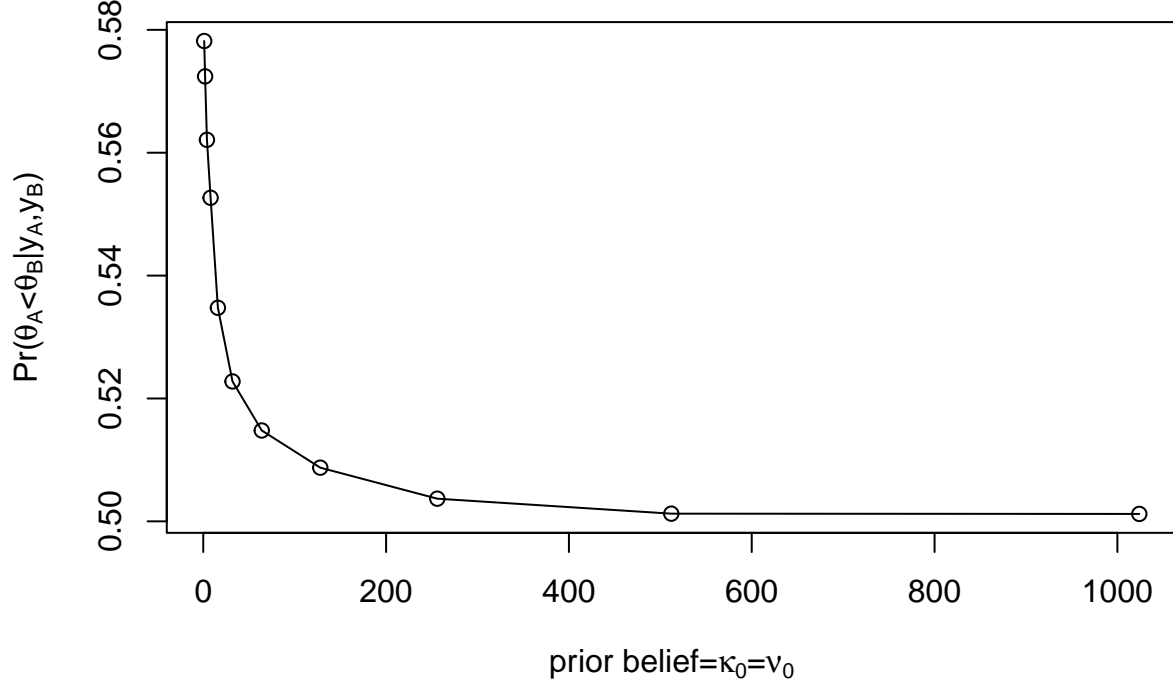
From the Monte Carlo sampling result we can see that there is a weak evidence that $\theta_A < \theta_B$. The probability that $\theta_A < \theta_B$ decreases as the prior belief (which is quantified by the "prior sample size" $\nu_0 = \kappa_0$) gets stronger (when $\nu_0 = \kappa_0$ increases). The probability of $\theta_A < \theta_B$ approach to 0.5 when the prior belief is extremely strong.

Problem 5 Exercise 6.1

Poisson population comparisons: 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 θ_A and θ_B as $\theta_A = \theta$, $\theta_B = \theta \times \gamma$. In this parameterization, γ represents the relative rate $\frac{\theta_B}{\theta_A}$. Let $\theta \sim$ gamma (a_θ, b_θ) and let $\gamma \sim$ gamma (a_γ, b_γ).

Part a)

Are θ_A and θ_B independent or dependent under this prior distribution? In what situations is such a joint prior distribution justified?

$$
\begin{aligned}
\text{Cov}(\theta_A, \theta_B) &= E\left[\theta_A \theta_B\right] - E[\theta_A]E[\theta_B] \\
&= E[\theta^2 \gamma] - E[\theta]E[\theta\gamma] \\
&= E\left[\theta^2\right]E\left[\gamma\right] - E[\theta]E[\theta]E[\gamma] \\
&= E\left[\theta^2\right]E\left[\gamma\right] - E[\theta]^2 E[\gamma] \\
&= \left(E[\theta^2] - E[\theta]^2\right)E[\gamma] \\
&= \text{Var}(\theta)E[\gamma] > 0
\end{aligned}
$$

$\therefore \theta_A$ and θ_B are NOT independent.

We can use this prior if we have known that θ_B is some product of θ_A plus random Gamma-distributed noise.

Part b)

Obtain the form of the full conditional distribution of θ given $\mathbf{y}_A, \mathbf{y}_B$ and γ.

First the joint posterior distribution

$$
\begin{aligned}
p(\theta, \gamma \mid \boldsymbol{y}_A, \boldsymbol{y}_B) &\propto p(\theta, \gamma) \times p(\boldsymbol{y}_A, \boldsymbol{y}_B \mid \theta, \gamma) \\
&= p(\theta) \times p(\gamma) \times p(\boldsymbol{y}_A \mid \theta) \times p(\boldsymbol{y}_B \mid \theta, \gamma) \\
&\propto \left(\theta^{a_\theta - 1} e^{-b_\theta \theta}\right) \times \left(\gamma^{a_\gamma - 1} e^{-b_\gamma \gamma}\right) \times \left(\prod_{i=1}^{n_A} \theta^{y_{A_i}} e^{-\theta}\right) \times \left(\prod_{i=1}^{n_B} (\gamma\theta)^{y_{B_i}} e^{-\gamma\theta}\right) \\
&= \left(\theta^{a_\theta - 1} e^{-b_\theta \theta}\right) \times \left(\gamma^{a_\gamma - 1} e^{-b_\gamma \gamma}\right) \times \left(\theta^{\sum_{i=1}^{n_A} y_{A_i}} e^{-n_A \theta}\right) \times \left((\gamma\theta)^{\sum_{i=1}^{n_B} y_{B_i}} e^{-n_B \gamma\theta}\right) \\
&= \left(\theta^{a_\theta - 1} e^{-b_\theta \theta}\right) \times \left(\gamma^{a_\gamma - 1} e^{-b_\gamma \gamma}\right) \times \left(\theta^{n_A \bar{y}_A} e^{-n_A \theta}\right) \times \left((\gamma\theta)^{n_B \bar{y}_B} e^{-n_B \gamma\theta}\right)
\end{aligned}
$$

$\therefore$

$$
\begin{aligned}
p(\theta, \mid \boldsymbol{y}_A, \boldsymbol{y}_B, \gamma) &\propto \left(\theta^{a_\theta - 1} e^{-b_\theta \theta}\right) \times \left(\gamma^{a_\gamma - 1} e^{-b_\gamma \gamma}\right) \times \left(\theta^{n_A \bar{y}_A} e^{-n_A \theta}\right) \times \left((\gamma\theta)^{n_B \bar{y}_B} e^{-n_B \gamma\theta}\right) \\
&\propto \left(\theta^{a_\theta - 1} e^{-b_\theta \theta}\right) \times \left(\theta^{n_A \bar{y}_A} e^{-n_A \theta}\right) \times \left((\gamma\theta)^{n_B \bar{y}_B} e^{-n_B \gamma\theta}\right) \\
&\propto \theta^{a_\theta + n_A \bar{y}_A + n_B \bar{y}_B - 1} \exp\left(-(b_\theta + n_A + n_B\gamma)\theta\right) \\
&\propto \mathrm{dgamma}\left(a_\theta + n_A \bar{y}_A + n_B \bar{y}_B, b_\theta + n_A + n_B\gamma\right)
\end{aligned}
$$

Part c)

Obtain the form of the full conditional distribution of γ given $\mathbf{y}_A, \mathbf{y}_B$, and θ.

$$
\begin{aligned}
p(\gamma, \mid \boldsymbol{y}_A, \boldsymbol{y}_B, \theta) &\propto \left(\theta^{a_\theta - 1} e^{-b_\theta \theta}\right) \times \left(\gamma^{a_\gamma - 1} e^{-b_\gamma \gamma}\right) \times \left(\theta^{n_A \bar{y}_A} e^{-n_A \theta}\right) \times \left((\gamma\theta)^{n_B \bar{y}_B} e^{-n_B \gamma\theta}\right) \\
&\propto \left(\gamma^{a_\gamma - 1} e^{-b_\gamma \gamma}\right) \times \left((\gamma\theta)^{n_B \bar{y}_B} e^{-n_B \gamma\theta}\right) \\
&\propto \left(\gamma^{a_\gamma - 1} e^{-b_\gamma \gamma}\right) \times \left(\gamma^{n_B \bar{y}_B} e^{-n_B \gamma\theta}\right) \\
&\propto \gamma^{a_\gamma + n_B \bar{y}_B - 1} \exp\left(-(b_\gamma + n_B\theta)\gamma\right) \\
&\propto \mathrm{dgamma}\left(a_\gamma + n_B \bar{y}_B, b_\gamma + n_B\theta\right)
\end{aligned}
$$

Part 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 $E\left[\theta_B - \theta_A\right] \mid \mathbf{y}_A, \mathbf{y}_B$. Describe the effects of the prior distribution for γ on the results.

```
Y_a <- scan('menchild30bach.dat')
Y_b <- scan('menchild30nobach.dat')
n_a <- length(Y_a)
n_b <- length(Y_b)
ybar_a <- mean(Y_a)
ybar_b <- mean(Y_b)

a_theta <- 2
b_theta <- 1

S <- 5000
```

```r
ab_gamma <- c(8, 16, 32, 64, 128)

theta_diff <- sapply(ab_gamma, function(abg) {
  a_gamma <- b_gamma <- abg

  THETA <- numeric(S)
  GAMMA <- numeric(S)

  # Starting values
  theta <- ybar_a
  gamma <- ybar_a / ybar_b  # Relative rate \theta_B / \theta_A

  for (s in 1:S) {
    # Sample theta
    theta <- rgamma(
      1,
      a_theta + n_a * ybar_a + n_b * ybar_b,
      b_theta + n_a + n_b * gamma
    )

    # Sample gamma
    gamma <- rgamma(
      1,
      a_gamma + n_b * ybar_b,
      b_gamma + n_b * theta
    )

    THETA[s] <- theta
    GAMMA[s] <- gamma
  }

  # Reconstruct \theta_A, \theta_B
  THETA_A <- THETA
  THETA_B <- THETA * GAMMA

  mean(THETA_B - THETA_A)
})

plot(ab_gamma, theta_diff, type = "p", xlab = expression(paste(a[gamma], ", ", b[gamma], sep=" ")), ylab
lines(ab_gamma, theta_diff, type = "l")
```
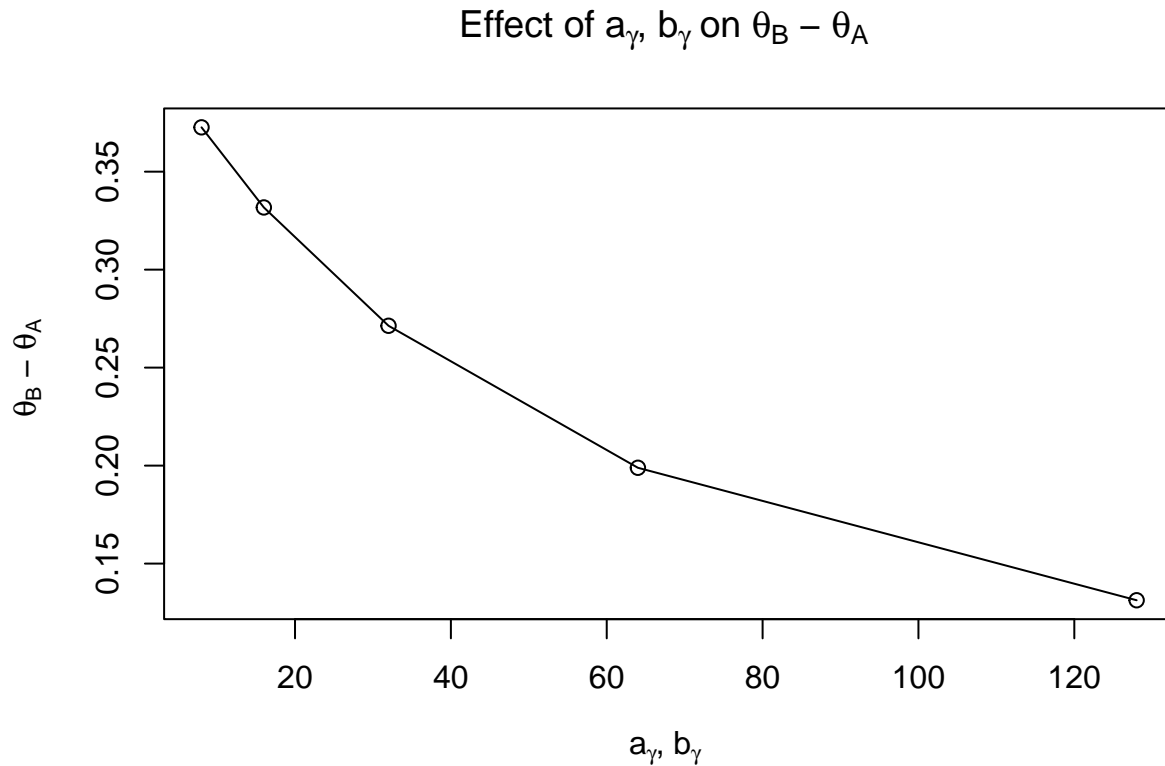
Effect of a_γ, b_γ on $\theta_B - \theta_A$

The effects of the prior distribution for *gamma* on the results are determined by the equality of a_γ and b_γ. This equality centers the gamma distribution around 1, with its magnitude reflecting the intensity of our conviction regarding *gamma* (the proportion θ_B/θ_A) being 1. Consequently, as our confidence in this belief strengthens, the average posterior disparity between θ_B and θ_A diminishes.

Problem 6 Exercise 7.5

Imputation: The file interexp.dat contains data from an experiment that was interrupted before all the data could be gathered. Of interest was the difference in reaction times of experimental subjects when they were given stimulus A versus stimulus B. Each subject is tested under one of the two stimuli on their first day of participation in the study, and is tested under the other stimulus at some later date. Unfortunately the experiment was interrupted before it was finished, leaving the researchers with 26 subjects with both A and B responses, 15 subjects with only A responses and 17 subjects with only B responses.

Part a)

Calculate empirical estimates of θ_A, θ_B, ρ, σ_A^2, σ_B^2 from the data using the commands mean, cor and var. Use all the A responses to get $\hat\theta_A$ and $\hat\sigma_A^2$, and use all the B responses to get $\hat\theta_B$ and $\hat\sigma_B^2$. Use only the complete data cases to get ρ.

```r
interexp = as.data.frame(read.table("interexp.dat", header=TRUE))
no_NA_index <- which(!is.na(interexp$yA)&!is.na(interexp$yB), TRUE)

interexp.no_NA <- interexp[no_NA_index,]
thetaA = mean(interexp$yA, na.rm=TRUE)
thetaB = mean(interexp$yB, na.rm=TRUE)
rho = cor(interexp.no_NA$yA,interexp.no_NA$yB)
sigmaA2 = var(interexp$yA, na.rm=TRUE)
sigmaB2 = var(interexp$yB, na.rm=TRUE)
```

$\hat{\theta}_A$	$\hat{\theta}_B$	$\hat{\rho}$	$\hat{\sigma}_A^2$	$\hat{\sigma}_B^2$
24.20049	24.80535	0.6164509	4.0928	4.691578

Part b)

For each person i with only an A response, impute a B response as

$$\hat{y}_{i,B} = \hat{\theta}_B + (y_{i,A} - \hat{\theta}_A)\hat{\rho}\sqrt{\hat{\sigma}_B^2/\hat{\sigma}_A^2}.$$

For each person i with only an B response, impute a A response as

$$\hat{y}_{i,A} = \hat{\theta}_A + (y_{i,B} - \hat{\theta}_B)\hat{\rho}\sqrt{\hat{\sigma}_A^2/\hat{\sigma}_B^2}.$$

You now have two "observations" for each individual. Do a paired sample t-test and obtain a 95% confidence interval for $\theta_A - \theta_B$.

```
interexp.partb = interexp

for(i in 1:nrow(interexp.partb)){
  if(is.na(interexp.partb[i,]$yA)){
    interexp.partb[i,]$yA <- thetaA +
        (interexp.partb[i,]$yB - thetaB) *
        rho * sqrt(sigmaA2 / sigmaB2)
  }else if(is.na(interexp.partb[i,]$yB)){
    interexp.partb[i,]$yB <- thetaB +
        (interexp.partb[i,]$yA - thetaA) *
        rho * sqrt(sigmaB2 / sigmaA2)
  }
}

t.test(x=interexp.partb$yA, y=interexp.partb$yB, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  interexp.partb$yA and interexp.partb$yB
## t = -3.2807, df = 57, p-value = 0.00177
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##   -0.9850730 -0.2383347
## sample estimates:
## mean difference
##      -0.6117038
```

The 95% CI for $\theta_A - \theta_B$ is $[-0.9850730, -0.2383347]$

Part c)

Using either Jeffreys' prior or a unit information prior distribution for the parameters, implement a Gibbs sampler that approximates the joint distribution of the parameters and the missing data. Compute a posterior mean for $\theta_A - \theta_B$ as well as a 95% posterior confidence interval for $\theta_A - \theta_B$. Compare these results with the results from b) and discuss.

```r
#### Simulate multivariate normal vector
rmvnorm<- function(n,mu,Sigma)
{
  p<-length(mu)
  res<-matrix(0,nrow=n,ncol=p)
  if( n>0 & p>0 ) {
    E<-matrix(rnorm(n*p),n,p)
    res<-t(t(E%*%chol(Sigma)) +c(mu))
  }
  res
}


#### Simulate from the Wishart distribution
rwish<-function(n,nu0,S0)
{
  sS0 <- chol(S0)
  S<-array( dim=c( dim(S0),n ) )
  for(i in 1:n)
  {
    Z <- matrix(rnorm(nu0 * dim(S0)[1]), nu0, dim(S0)[1]) %*% sS0
    S[,,i]<- t(Z)%*%Z
  }
  S[,,1:n]
}


interexp.partc = interexp
n<-dim(interexp.partc)[1]
p<-dim(interexp.partc)[2]
O<-!is.na(interexp.partc)

## prior parameters
p<-dim(interexp.partc)[2]
mu0<-c(24.20049,24.80535)
sd0<-(mu0/2)
L0<-matrix(.1,p,p); diag(L0)<-1 ; L0<-L0*outer(sd0,sd0)
nu0<-p+2 ; S0<-L0
###

### starting values
Sigma<-S0
interexp.partc.full<-interexp.partc
for(j in 1:p)
{
  interexp.partc.full[is.na(interexp.partc.full[,j]),j]<-mean(interexp.partc.full[,j],na.rm=TRUE)
}
###


### Gibbs sampler
THETA<-SIGMA<-interexp.partc.MISS<-NULL
set.seed(1)

for(s in 1:1000)
```

```
{

  ###update theta
  ybar<-apply(interexp.partc.full,2,mean)
  Ln<-solve( solve(L0) + n*solve(Sigma) )
  mun<-Ln%*%( solve(L0)%*%mu0 + n*solve(Sigma)%*%ybar )
  theta<-rmvnorm(1,mun,Ln)
  ###

  ###update Sigma
  Sn<- S0 + ( t(interexp.partc.full)-c(theta) )%*%t( t(interexp.partc.full)-c(theta) )
# Sigma<-rinvwish(1,nu0+n,solve(Sn))
  Sigma<-solve( rwish(1, nu0+n, solve(Sn)) )
  ###

  ###update missing data
  for(i in 1:n)
  {
    b <- ( O[i,]==FALSE )
    a <- ( O[i,]==TRUE )
    iSa<- solve(Sigma[a,a])
    beta.j <- Sigma[b,a]%*%iSa
    s2.j   <- Sigma[b,b] - Sigma[b,a]%*%iSa%*%Sigma[a,b]
    theta.j<- theta[b] + beta.j%*%(t(interexp.partc.full[i,a])-theta[a])
    interexp.partc.full[i,b] <- rmvnorm(1,theta.j,s2.j )
  }

  ### save results
  THETA<-rbind(THETA,theta) ; SIGMA<-rbind(SIGMA,c(Sigma))
  interexp.partc.MISS<-rbind(interexp.partc.MISS, interexp.partc.full[O==FALSE] )
  ###

  #cat(s,theta,"\n")
}
```

```
mean(THETA[,1]-THETA[,2])
```

```
## [1] -0.6023525
```

```
quantile(THETA[,1]-THETA[,2], prob=c(.025,.5,.975) )
```

```
##      2.5%       50%     97.5%
## -1.6243078 -0.6145060  0.4248431
```

The posterior mean for $\theta_A - \theta_B$ is -0.6023525, 95% posterior confidence interval for $\theta_A - \theta_B$ is $[-1.6243078, 0.4248431]$. Compared to the CI from part b), this is wider and also contains 0 which means we cannot say there is a significant increase in average reaction times when switching from stimulus A to stimulus B.

Problem 7

Choose a conjugate prior setting and investigate the performance of credible intervals (let's say at the 80% level) in this setting by doing the following Demonstrate that for some combination of little information in the data (i.e., a small sample size) and non-negligible information in the prior (e.g., not a uniform distribution as the prior), the frequentist coverage will be wrong.

Part(a)

(a) More specifically, repeatedly generate datasets for a true parameter value in the tail of the prior, and verify that the proportion of 80% credible intervals containing this true value is less than the nominal 80%. Similarly, show that higher-than-nominal coverage obtains for a true value 'in the middle' of the prior.

I choose Beta distribution as the prior with $\alpha = 2$ $\beta = 5$, for a binomial model, investigating the proportion of 80% credible intervals containing $\theta_{tail} = 0.7$ when the true parameter (the θ) is at the tail of the prior, and $\theta_{middle} = 2/7$ when the true parameter is at the middle of the prior.

```r
calc_credible_interval <- function(alpha, beta, data, alpha_level = 0.8) {
  q_low <- qbeta((1 - alpha_level) / 2, alpha, beta)
  q_high <- qbeta(1 - (1 - alpha_level) / 2, alpha, beta)
  interval <- c(q_low, q_high)
  proportion_inside <- mean(data >= interval[1] & data <= interval[2])
  return(list(interval = interval, proportion_inside = proportion_inside))
}


# Function to simulate data and calculate credible interval coverage
simulate_coverage <- function(n_simulations, true_theta, alpha, beta, sample_size) {
  coverage <- numeric(n_simulations)

  for (i in 1:n_simulations) {
    # Simulate data from the model
    data <- rbinom(sample_size, 1, true_theta)

    # Calculate posterior parameters
    alpha_posterior <-  alpha + sum(data)
    beta_posterior <- beta + sample_size - sum(data)

    # Calculate credible interval
    credible_interval <- calc_credible_interval(alpha_posterior, beta_posterior, data)$interval

    # Check if true_theta is inside the credible interval
    coverage[i] <- true_theta >= credible_interval[1] & true_theta <= credible_interval[2]
  }

  return(mean(coverage))
}


set.seed(1)
# Set parameters
alpha <- 2
beta <- 5
true_theta_tail <- 0.7
true_theta_middle <- alpha / (alpha+beta)
sample_size <- 10
n_simulations <- 1000

# Simulate coverage for true parameter in the tail of the prior
coverage_tail <- simulate_coverage(n_simulations, true_theta_tail, alpha, beta, sample_size)

# Simulate coverage for true parameter in the middle of the prior
coverage_middle <- simulate_coverage(n_simulations, true_theta_middle, alpha, beta, sample_size)
```

```
# Output results
cat("Coverage for true parameter in the tail of the prior:", coverage_tail, "\n")
```

Coverage for true parameter in the tail of the prior: 0.381

```
cat("Coverage for true parameter in the middle of the prior:", coverage_middle, "\n")
```

## Coverage for true parameter in the middle of the prior: 0.918

Part(b)

Now take the setting you chose in part (a), and investigate a different kind of coverage. Repeatedly sample a (parameter, dataset) pair by sampling a parameter value from the prior distribution and then sampling a dataset from the model using this parameter value. What do you find about the proportion of 80% credible intervals containing the parameter value in this kind of repeated sampling? (You could regard this kind of repeated sampling as corresponding to a series of studies of different phenomena, so that the true parameter value is different for each study.)

```
set.seed(1)
# Set parameters
alpha <- 2
beta <- 5
true_theta <- rbeta(1000, alpha, beta)
sample_size <- 10
n_simulations <- 1000

cov_result = NULL
for (i in 1:1000){
cov_result[i] <- simulate_coverage(n_simulations, true_theta[i] ,alpha, beta, sample_size)
}

plot(true_theta, cov_result)
```
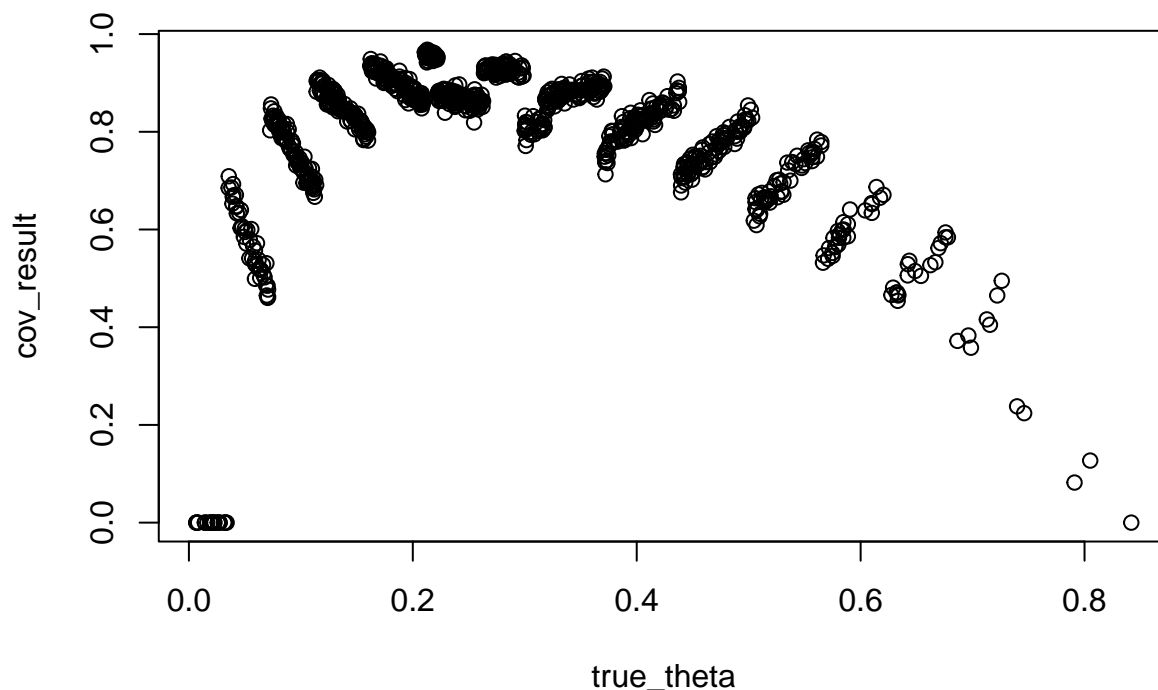
We can see that the proportion of credible intervals containing the true parameter value is consistently lower than 80% for a true value in the tail of the prior, it suggests that the intervals are narrower than they should be. On the other hand, the proportion of credible intervals containing the true parameter value is consistently higher than 80% for a true value in the middle of the prior, it suggests that the intervals are wider than they should be. Hence, the frequentist coverage is not correct.

Problem 8

Consider the following situation. The observable data consists of a single observation $Y \sim N(\theta, 1)$. The investigator will use a prior distribution of the form $\theta \sim N(0, \tau^2)$, and estimate θ by $E(\theta|Y)$. What is the Bayes risk of this estimator when 'nature's prior' is $\theta \sim N(0, k^2)$, i.e., what is the average mean-squared error of the investigator's estimator with respect to 'nature's' distribution over the parameter space? Fix k^2 and plot the Bayes risk as the investigator's choice of τ^2 varies. Does the pattern you see conform with the basic tenets of decision theory? Is there any support for the notion that an 'overly-flat' prior is less damaging than an 'overly-concentrated' prior?

The *Bayes risk* of $\delta()$ with respect to the 'prior distribution' $\theta \sim \pi$ is

$$r(\pi, \delta) = E_\pi\{R(\theta, \delta)\}$$
$$= \int R(\theta, \delta) d\Pi(\theta)$$

The risk function is given by:

$$R(\theta, \delta) = E_\theta\{L(\theta, \delta(Y))\}$$
$$= \int L(\theta, \delta(y)) p(y|\theta) dy$$

$$\hat{\theta} = E(\theta|Y)$$
$$= \frac{\frac{1}{\tau_0}\mu_0 + \frac{n}{\theta\sigma^2}\overline{y}}{\frac{1}{\tau_0} + \frac{n}{\sigma^2}} \qquad \text{(Chapter 5)} \qquad \mu_0 = 0, \tau_0 = \tau, n = \sigma^2 = 1,$$
$$= \frac{\overline{y}}{\frac{1}{\tau} + 1} = \frac{\tau\overline{y}}{\tau + 1}$$

$$L(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^2$$
$$= \left(\frac{\tau\overline{y} - (\tau + 1)\theta}{\tau + 1}\right)^2$$
$$= \left(\frac{\tau y - (\tau + 1)\theta}{\tau + 1}\right)^2$$

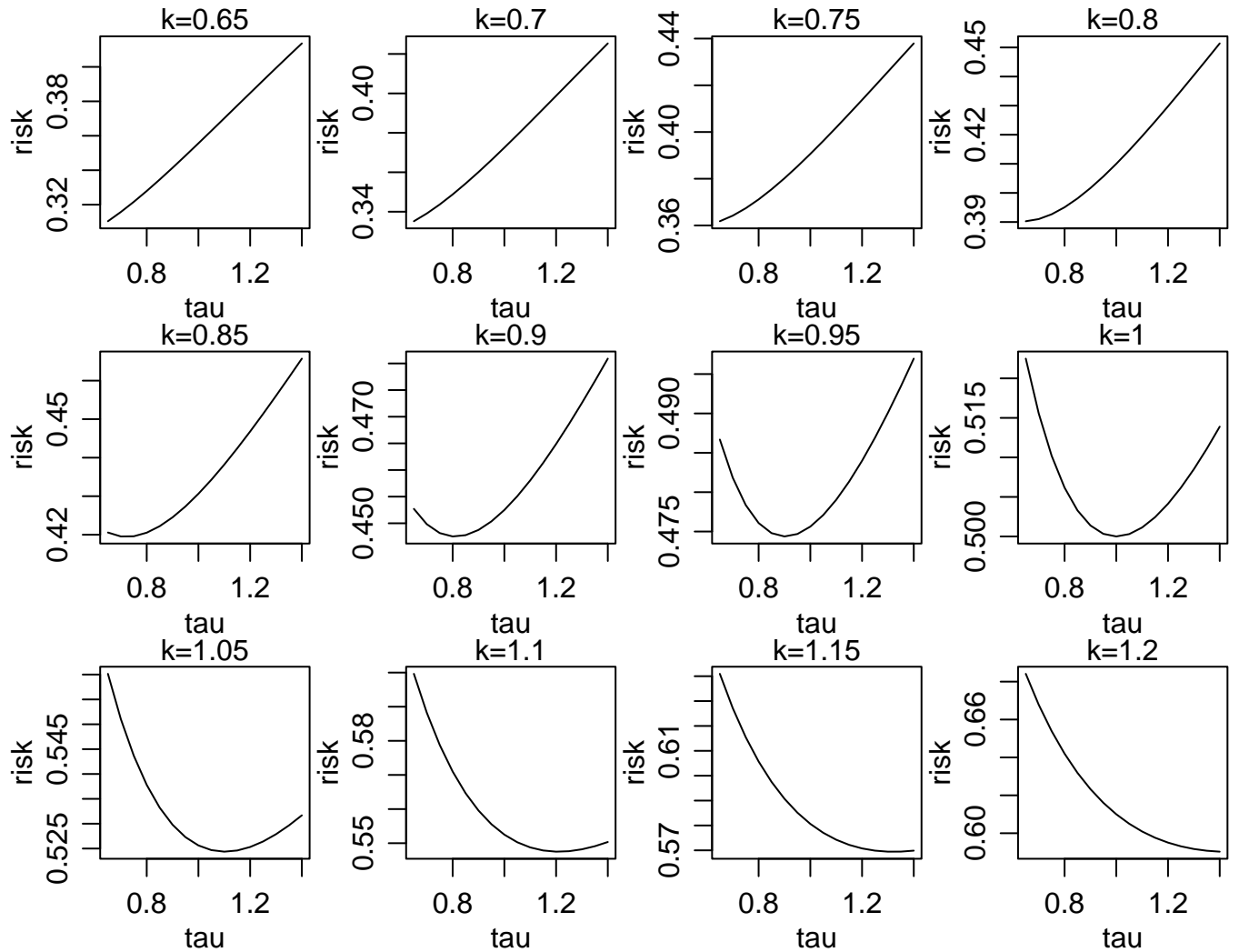$$R(\theta, \hat{\theta}) = \int_{-\infty}^{\infty} \left(\frac{\tau y}{\tau + 1} - \theta\right)^2 \frac{1}{\sqrt{2\pi}} e^{\frac{-(y-\theta)^2}{2}} dy$$
$$= \left(\frac{\tau}{\tau + 1}\right)^2 \cdot E(Y^2) - 2\left(\frac{\tau}{\tau + 1}\right)\theta \cdot E(Y) + \theta^2$$
$$= \left(\frac{\tau}{\tau + 1}\right)^2 \cdot (1 + \theta^2) - 2\left(\frac{\tau}{\tau + 1}\right)\theta^2 + \theta^2$$
$$= \left(\frac{\tau - \theta}{\tau + 1}\right)^2$$

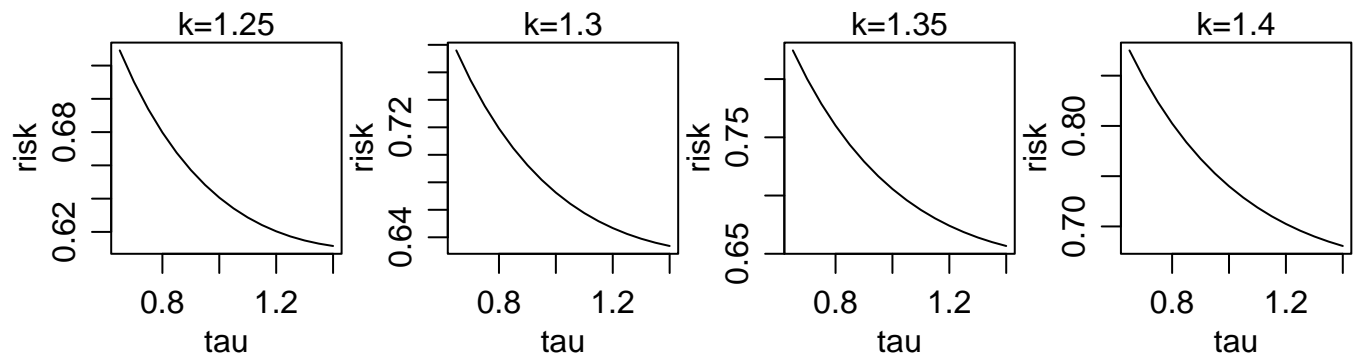$$r\left(N(0,k^2), N(0,\tau^2)\right) = \int_{-\infty}^{\infty} \left(\frac{\tau - \theta}{\tau + 1}\right)^2 \frac{1}{k\sqrt{2\pi}} e^{-\frac{\theta^2}{2k^2}} d\theta$$

$$= \left(\frac{1}{\tau+1}\right)^2 \int_{-\infty}^{\infty} \left(\tau^2 - 2\tau\theta + \theta^2\right) \frac{1}{k\sqrt{2\pi}} e^{-\frac{\theta^2}{2k^2}} d\theta$$

$$= \left(\frac{\tau}{\tau+1}\right)^2 - \left(\frac{1}{\tau+1}\right)^2 \cdot 2\tau \cdot E(\theta) + \left(\frac{1}{\tau+1}\right)^2 \cdot E(\theta^2)$$

$$= \left(\frac{\tau}{\tau+1}\right)^2 + \left(\frac{k}{\tau+1}\right)^2$$

```
par(mar=c(3,3,1,1),mgp=c(1.75,.75,0),oma=c(0,0,.5,0))
k <- seq(0.65, 1.4, by=0.05)
tau <- seq(0.65, 1.4, by=0.05)
for (i in k){
risk=(tau^2+i^2)/(tau+1)^2
plot(tau,risk,type="l")
mtext(paste("k=", i, sep=""))
}
```

From the graph, we can see that when $\tau > k$, the risk decreases when increasing τ (picking a flat prior). Also when k increases, the risk increases, so 'overly-flat' prior is less damaging than 'overly-concentrated' prior.