STA 601/360 Homework 3

Yifei Wang

18 September, 2019

Exec 1. Derive the mean and variance of a Poisson distribution with parameter θ from first principles.

$$\begin{aligned}
E[Y \mid \theta] &= \sum_{y \geq 0} y p(y \mid \theta) \\
&= \sum_{y \geq 0} y \frac{\theta^y e^{-\theta}}{y!} \\
&= \sum_{y \geq 1} \frac{\theta^y e^{-\theta}}{(y-1)!} \\
&= \theta \sum_{y \geq 1} \frac{\theta^{y-1} e^{-\theta}}{(y-1)!} \\
&= \theta \sum_{y \geq 0} \frac{\theta^y e^{-\theta}}{(y)!} \\
&= \theta
\end{aligned}$$

$$\begin{aligned}
E[Y^2 \mid \theta] &= \sum_{y \geq 0} y^2 p(y \mid \theta) \\
&= \sum_{y \geq 0} y^2 \frac{\theta^y e^{-\theta}}{y!} \\
&= \sum_{y \geq 1} y\theta \frac{\theta^{y-1} e^{-\theta}}{(y-1)!} \\
&= \theta \sum_{y \geq 0} (y+1) \frac{\theta^y e^{-\theta}}{y!} \\
&= \theta \left[\sum_{y \geq 0} y \frac{\theta^y e^{-\theta}}{y!} + \sum_{y \geq 0} \frac{\theta^y e^{-\theta}}{y!} \right] \\
&= \theta \left[\sum_{y \geq 1} \theta \frac{\theta^{y-1} e^{-\theta}}{(y-1)!} + 1 \right] \\
&= \theta \left[\theta \sum_{y \geq 0} \frac{\theta^y e^{-\theta}}{y!} + 1 \right] \\
&= \theta \left[\theta + 1 \right] \\
&= \theta^2 + \theta
\end{aligned}$$

$$Mean(Poi(Y \mid \theta)) = E[Y \mid \theta] = \theta$$

$$
\begin{aligned}
Var(Poi(Y \mid \theta)) &= E[(Y - E[Y \mid \theta])^2 \mid \theta] \\
&= E[Y^2 \mid \theta] - 2E[YE[Y \mid \theta] \mid \theta] + E^2[Y \mid \theta] \\
&= E[Y^2 \mid \theta] - 2E^2[Y \mid \theta] + E^2[Y \mid \theta] \\
&= E[Y^2 \mid \theta] - E^2[Y \mid \theta] \\
&= \theta^2 + \theta - \theta^2 \\
&= \theta
\end{aligned}
$$

Exec 2. Hoff 3.3.

(a). Posterior distributions, means, variances and 95% quantile-based confidence interval.

According to the proof in class and the text books, if

$$
\begin{aligned}
P(Y \mid \theta) &\sim Poisson(\theta) \\
P(\theta) &\sim gamma(a, b)
\end{aligned}
$$

we have

$$
\begin{aligned}
P(\theta \mid Y) &\sim gamma(a + \sum_{i=1}^{n} Y_i, b + n) \\
E[\theta \mid Y] &= \frac{a}{b} \\
Var[\theta \mid Y] &= \frac{a}{b^2}
\end{aligned}
$$

Therefore we have,

$$
\begin{aligned}
P(\theta_A \mid y_A) &\sim gamma(237, 20) \\
E[\theta_A \mid y_A] &= \frac{237}{20} \\
Var[\theta_A \mid y_A] &= \frac{237}{20 \times 20} = \frac{237}{400}
\end{aligned}
$$

$$
\begin{aligned}
P(\theta_B \mid y_B) &\sim gamma(125, 14) \\
E[\theta_B \mid y_B] &= \frac{125}{14} \\
Var[\theta_B \mid y_B] &= \frac{125}{14 \times 14} = \frac{125}{196}
\end{aligned}
$$

```
q <- c(0.025, 0.975)
a <- 237
b <- 20
qgamma(q, a, b)
```

```
## [1] 10.38924 13.40545
```

```
q <- c(0.025, 0.975)
a <- 125
b <- 14
qgamma(q, a, b)
```

```
## [1]  7.432064 10.560308
```

Thus, the 95% quantile-based confidence interval is $[10.38924, 13.40545]$ for θ_A and $[7.432064, 10.560308]$ for θ_B

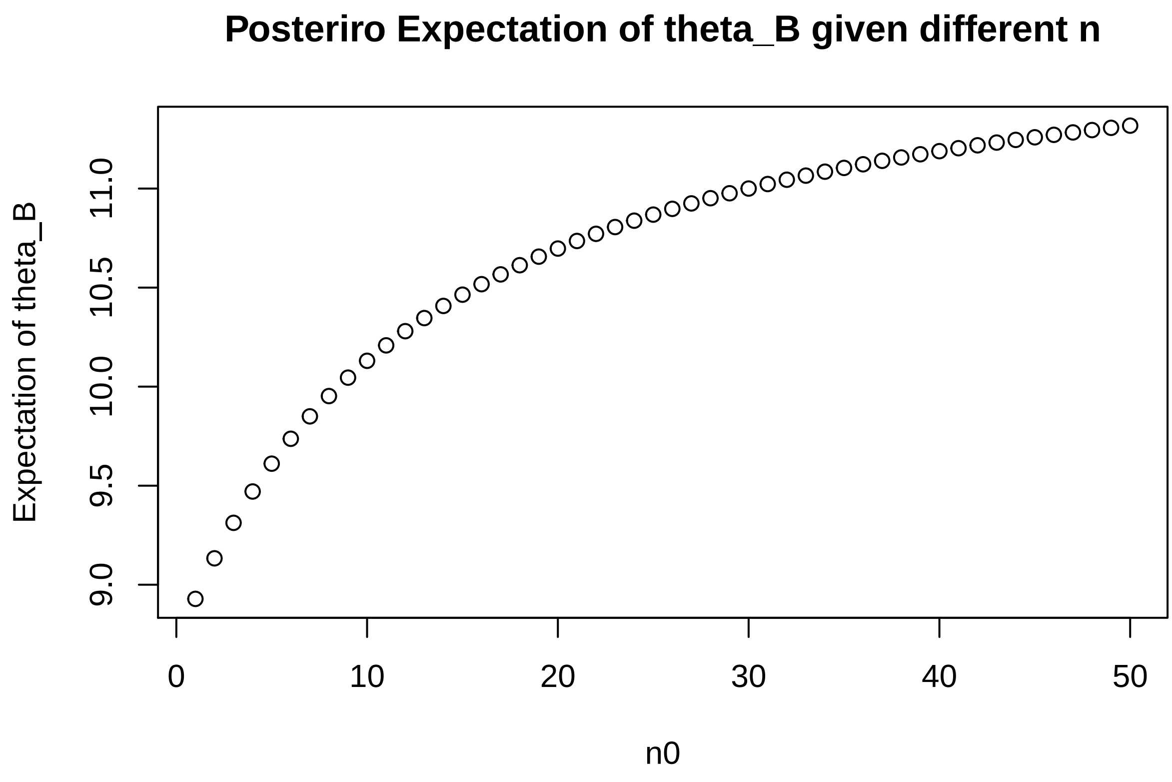
(b) Posterior expectation of θ_B

As we mentioned before, given a posterior $gamma(a, b)$, the mean would be $\frac{a}{b}$ and the variance is $\frac{a}{b^2}$.

```
n_B <- 13
sum_y_B <- 113

n0 <- seq(1, 50)
prior_a <- 12 * n0
prior_b <- n0
post_a <- prior_a + sum_y_B
post_b <- prior_b + n_B
E_post_theta <- post_a / post_b

plot(n0, E_post_theta, xlab = "n0", ylab = "Expectation of theta_B",
     main = "Posteriro Expectation of theta_B given different n")
```



We know the posterior mean of θ_A is 11.85 in this case, which is a weighted average of prior mean (12) and sample mean (11.7). The posteriro mean of θ_B is also the weighted average of prior mean (12) and sample mean (8.69). In order to make the sample mean to be close to 11.85, we need a large n_0, but not to lare to get over 11.85, which is 274 in this case.

(c) Does it make sense to have $p(\theta_A, \theta_B) = p(\theta_A)\, p(\theta_B)$?

This highly depends on the background domain knowledge. If study shows that type A mice and type B mice are fundamentally different on the tumor counts case, we could assume that $p(\theta_A, \theta_B) = p(\theta_A)\, p(\theta_B)$, which means Type A mice and Type B mice are independent on the number of tumor counts.

However, if studys show that there is some relationship betwwen the tumor counts of Type A mice and Type B mice, we might be more careful before we make the assumption of $p(\theta_A, \theta_B) = p(\theta_A)\, p(\theta_B)$. Because this means that the prior beliefs of tumor rate among Type A mice is independent of the prior beliefs of tumor rate among Type B mice. If we know they are related, our prior beliefs about the tumor rates in type A and B mice should not be independent.

Exec 3. Hoff 3.9.

(a). Identify a class of conjugate prior densities for θ

$$\begin{aligned} p(\theta \mid y) &\propto p(y \mid \theta) p(\theta) \\ &= p(\theta)\, \frac{2}{\Gamma(a)} \theta^{2a} y^{2a-1} e^{-\theta^2 y^2} \\ &\propto p(\theta)\, \theta^{2a} e^{-\theta^2} \end{aligned}$$

Therefore our prior should include terms like $\theta^{c_1} e^{-c_2 \theta^2}$. If we have,

$$p(\theta) = \text{Galenshore}(a_0, \theta_0) = \frac{2}{\Gamma(a_0)} {\theta_0}^{2a_0} \theta^{2a_0 - 1} e^{-\theta^2 {\theta_0}^2}$$

Then we will have

$$p(\theta \mid y) \propto \theta^{2a+2a_0-1} e^{-(\theta_0^2+1)\theta^2}$$

which is also a Galenshore distribution with parameter $a + a_0$ and $\sqrt{\theta_0^2 + 1}$.

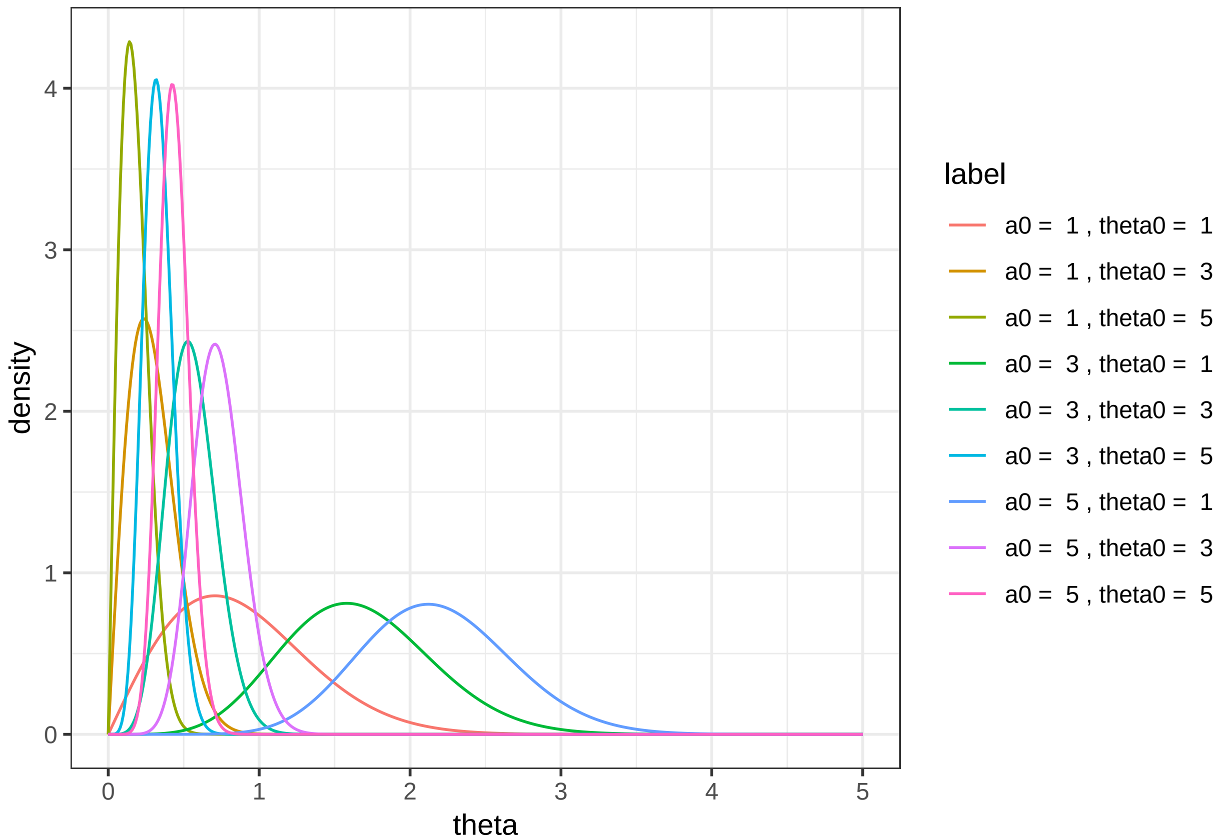
```
Galenshore <- function(x, a, theta) {
  return((2/gamma(a)) * theta^(2*a) * x^(2*a - 1) * exp(-(theta^2) * (x^2)))
}

theta <- seq(0, 5, 0.01)
df <- data.frame()
a0 <- seq(1, 5, 2)
theta0 <- seq(1, 5, 2)

for (a0_ in a0){
  for (theta0_ in theta0) {
    df = rbind(df, data.frame(theta = theta,
                              density = Galenshore(theta, a0_, theta0_),
                              label = paste("a0 = ", a0_, ", theta0 = ", theta0_)))
  }
}

ggplot(df, aes(x = theta,
               y = density,
               group = label,
```

```r
                color = label)) +
  geom_line()
```



(b). Find the posterior distribution of θ given $Y_1, \ldots, Y_n$, using a prior from your conjugate class.

Since we only care about θ, we could let a be a constant in this case. If we have a prior

$$p(\theta) = \text{Galenshore}(a_0, \theta_0) = \frac{2}{\Gamma(a_0)} {\theta_0}^{2a_0} \theta^{2a_0 - 1} e^{-{\theta_0}^2 \theta^2}$$

Then the posterior will be

$$\begin{aligned} p(\theta \mid Y_1, \ldots, Y_n) &\propto p(Y_1, \ldots, Y_n \mid \theta)\, p(\theta) \\ &= \frac{2}{\Gamma(a_0)} {\theta_0}^{2a_0} \theta^{2a_0-1} e^{-{\theta_0}^2 \theta^2} \prod_{i=1}^{n} \frac{2}{\Gamma(a)} \theta^{2a} Y_i^{2a-1} e^{-\theta^2 {Y_i}^2} \\ &\propto \theta^{2a_0-1} e^{-{\theta_0}^2\theta^2} \theta^{2na} e^{-\theta^2 \sum_{i=1}^n {Y_i}^2} \\ &= \theta^{2na+2a_0-1} e^{-({\theta_0}^2 + \sum_{i=1}^n {Y_i}^2)\theta^2} \end{aligned}$$

After setting the integration of θ on this posterior to be 1, we will find that

$$p(\theta \mid Y_1, \ldots, Y_n) \sim \text{Galenshore}(na + a_0, \sqrt{{\theta_0}^2 + \sum_{i=1}^{n} {Y_i}^2})$$

(c). Write down $p(\theta_a \mid Y_1, \ldots, Y_n)/p(\theta_b \mid Y_1, \ldots, Y_n)$ and simplify. Identify a sufficient statistic

As we have proved in part (b)

$$\begin{aligned}
\frac{p(\theta_a \mid Y_1, \ldots, Y_n)}{p(\theta_b \mid Y_1, \ldots, Y_n)} &= \frac{{\theta_a}^{2na+2a_0-1} e^{-({\theta_0}^2 + \sum_{i=1}^n {Y_i}^2){\theta_a}^2}}{{\theta_b}^{2na+2a_0-1} e^{-({\theta_0}^2 + \sum_{i=1}^n {Y_i}^2){\theta_b}^2}} \\
&= \left(\frac{\theta_a}{\theta_b}\right)^{2na+2a_0-1} e^{-({\theta_a}^2 - {\theta_b}^2)({\theta_0}^2 + \sum_{i=1}^n {Y_i}^2)}
\end{aligned}$$

We can see that this formular of θ_a and θ_b only depends on $Y_1, \ldots, Y_n$ through $\sum_{i=1}^n {Y_i}^2$. Thus, $\sum_{i=1}^n {Y_i}^2$ is a sufficient statistics for θ and $p(Y_1, \ldots, Y_n \mid \theta)$.

(d). Determine $E[\theta \mid y_1, \ldots, y_n]$.

According to the given formula for the mean of Galenshore distribution, and the posterior distribution for θ in part (b), we have,

$$E[\theta \mid y_1, \ldots, y_n] = \frac{\Gamma(na + a_0 + \frac{1}{2})}{\Gamma(na + a_0)\sqrt{{\theta_0}^2 + \sum_{i=1}^n {Y_i}^2}}$$

(e). Determine the form of the posterior predictive density $p(\tilde{y} \mid y_1, \ldots, y_n)$

$$\begin{aligned}
p(\tilde{y} \mid y_1, \ldots, y_n) &= \int_0^\infty p(\tilde{y} \mid \theta, y_1, \ldots, y_n)\, p(\theta \mid y_1, \ldots, y_n)\, d\theta \\
&= \int_0^\infty p(\tilde{y} \mid \theta)\, p(\theta \mid y_1, \ldots, y_n)\, d\theta \\
&= \int_0^\infty \frac{2}{\Gamma(a)} \theta^{2a} \tilde{y}^{2a-1} e^{-\theta^2 \tilde{y}^2} \frac{2}{\Gamma(na + a_0)} \left({\theta_0}^2 + \sum_{i=1}^n {Y_i}^2\right)^{na+a_0} \theta^{2na+2a_0-1} e^{-\theta^2({\theta_0}^2 + \sum_{i=1}^n {Y_i}^2)}\, d\theta \\
&= \frac{2}{\Gamma(a)} \tilde{y}^{2a-1} \frac{2\big({\theta_0}^2 + \sum_{i=1}^n {Y_i}^2\big)^{na+a_0}}{\Gamma(na + a_0)} \int_0^\infty \theta^{2(na+a+a_0)-1} e^{-\theta^2(\tilde{y}^2 + {\theta_0}^2 + \sum_{i=1}^n {Y_i}^2)}\, d\theta
\end{aligned}$$

According to the kernal trick, we have

$$\begin{aligned}
1 &= \int_0^\infty \frac{2}{\Gamma(a)} \theta^{2a} y^{2a-1} e^{-\theta^2 y^2}\, dy \\
\frac{\Gamma(a)}{2\,\theta^{2a}} &= \int_0^\infty y^{2a-1} e^{-\theta^2 y^2}\, dy
\end{aligned}$$

Using this formula, we have,

$$\begin{aligned}
p(\tilde{y} \mid y_1, \ldots, y_n) &= \frac{2}{\Gamma(a)} \tilde{y}^{2a-1} \frac{2\big({\theta_0}^2 + \sum_{i=1}^n {Y_i}^2\big)^{na+a_0}}{\Gamma(na + a_0)} \int_0^\infty \theta^{2(na+a+a_0)-1} e^{-\theta^2(\tilde{y}^2 + {\theta_0}^2 + \sum_{i=1}^n {Y_i}^2)}\, d\theta \\
&= \frac{2}{\Gamma(a)} \tilde{y}^{2a-1} \frac{2\big({\theta_0}^2 + \sum_{i=1}^n {Y_i}^2\big)^{na+a_0}}{\Gamma(na + a_0)} \frac{\Gamma(na + a + a_0)}{2\,(\tilde{y}^2 + {\theta_0}^2 + \sum_{i=1}^n {Y_i}^2)^{(na+a+a_0)}} \\
&= \frac{2\Gamma(na + a + a_0)}{\Gamma(a)\Gamma(na + a_0)} \frac{\big({\theta_0}^2 + \sum_{i=1}^n {Y_i}^2\big)^{na+a_0}}{(\tilde{y}^2 + {\theta_0}^2 + \sum_{i=1}^n {Y_i}^2)^{(na+a+a_0)}} \tilde{y}^{2a-1}
\end{aligned}$$

Exec 4. Hoff 4.1. Posterior comparisons: Reconsider the sample survey in Exercise 3.1. Suppose you are interested in comparing the rate of support in that county to the rate in another county. Suppose that a survey of sample size 50 was done in the second county, and the total number of people in the sample who supported the policy was 30. Identify the posterior distribution of θ_2 assuming a uniform prior. Sample 5,000 values of each of θ_1 and θ_2 from their posterior distributions and estimate $Pr(\theta_1 < \theta 2 \mid \text{the data and prior})$.

According to our previous work, we know the posterior distribution for the first county is $Beta(58, 44)$.

For the second county, given a uniform ($Beta(1, 1)$) prior, we know that the posterior will be $Beta(31, 21)$.

```
a1 <- 57 + 1
b1 <- 100 - 57 + 1
a2 <- 30 + 1
b2 <- 50 - 30 + 1
theta1 <- rbeta(5000, a1, b1)
theta2 <- rbeta(5000, a2, b2)

Pr1 <- sum(theta1 < theta2) / 5000
Pr1
```

```
## [1] 0.6398
```

The estimated probability of $Pr(\theta_1 < \theta 2 \mid \text{the data and prior})$ is 0.6398 in this case.

Exec 5. Hoff 4.2. Tumor count comparisons: Reconsider the tumor count data in Exercise 3.3:

(a). For the prior distribution given in part (a) of that exercise, obtain $Pr(\theta_B < \theta_A \mid y_A, y_B)$ via Monte Carlo sampling.

From the previous execrise we know that

$$\theta_A \sim Gamma(120 + 117, 10 + 10)$$
$$\theta_B \sim Gamma(12 + 113, 1 + 13)$$

```
a1 <- 120 + 117
b1 <- 10 + 10
a2 <- 12 + 113
b2 <- 1 + 13
N <- 10000

theta_A <- rgamma(N, a1, b1)
theta_B <- rgamma(N, a2, b2)
Pr2 <- sum(theta_B < theta_A) / N
Pr2
```

```
## [1] 0.9958
```

The estimated probability of $Pr(\theta_B < \theta_A \mid y_A, y_B)$ via Monte Carlo sampling method is 0.9958 in this case.

(b). For a range of values of n_0, obtain $Pr(\theta_B < \theta_A \mid y_A, y_B)$ for $\theta_A \sim Gamma(120, 10)$ and $\theta_B \sim Gamma(12 \times n_0, n_0)$. Describe how sensitive the conclusions about the event $\{\theta_B < \theta_A\}$ are to the prior distribution on θ_B.

From the previous execrise we know that

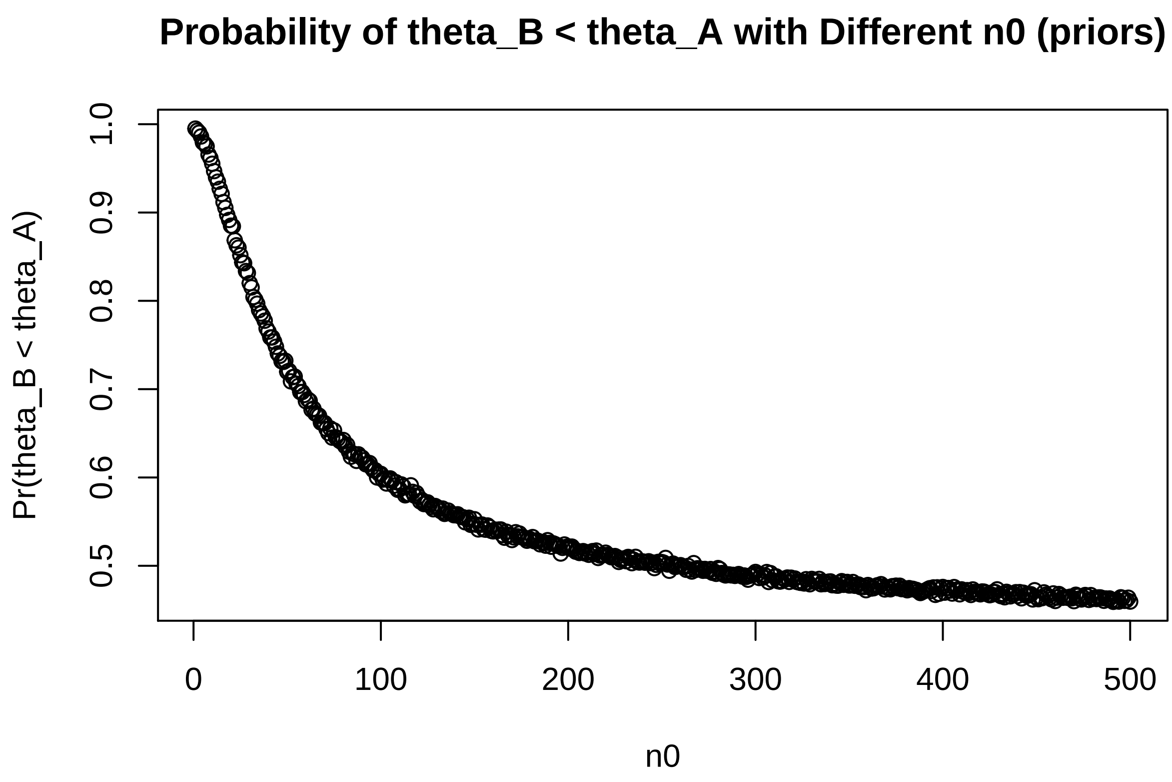
$$p(\theta_A \mid y_A) \sim Gamma(120 + 117, 10 + 10)$$
$$p(\theta_B \mid y_B) \sim Gamma(12 \times n_0 + 113, n_0 + 13)$$

```r
N <- 10000

a1 <- 120 + 117
b1 <- 10 + 10
theta_A <- rgamma(N, a1, b1)

Pr3 <- c()
for (n0 in seq(1, 500)) {
  a2 <- 12 * n0 + 113
  b2 <- n0 + 13
  theta_B <- rgamma(N, a2, b2)
  Pr3[n0] <- sum(theta_B < theta_A) / N
}

plot(seq(1, 500), Pr3, xlab = "n0", ylab = "Pr(theta_B < theta_A)",
     main = "Probability of theta_B < theta_A with Different n0 (priors)")
```



We know that $p(\theta_A \mid y_A) \sim Gamma(237, 20)$, with mean to be 11.85, and $p(\theta_B \mid y_B) \sim Gamma(12 \times n_0 + 113, n_0 + 13)$, with mean to be $\lim_{n_0 \to \infty} \frac{113+12n_0}{13+n_0} = 12$. As shown in the book and previous execrise, the posterior is shifting towards prior as n_0, the prior sample size, grows to 500, because there are only 13 samples in the given data. Thus, $p(\theta_B \mid y_B)$, as well as $Pr(\theta_B < \theta_A \mid y_A, y_B)$, are sensitive to n_0 when $n_0 < 100$, because the ratio of $\frac{n_0}{13}$ changes rapidly in this range. When $n_0 > 100$, this ratio changes slower and slower as n_0 linearly increases, and the posterior $p(\theta_B \mid y_B)$, as well as $Pr(\theta_B < \theta_A \mid y_A, y_B)$ become less and less sensitive to n_0.

(c). Repeat parts (a) and (b), replacing the event $\{\theta_B < \theta_A\}$ with the event $\{\tilde{Y}_B < \tilde{Y}_A\}$, where $\tilde{Y}_A$ and $\tilde{Y}_B$ are samples from the posterior predictive distribution.

```
a1 <- 120 + 117
b1 <- 10 + 10
a2 <- 12 + 113
b2 <- 1 + 13
N <- 10000

theta_A <- rgamma(N, a1, b1)
theta_B <- rgamma(N, a2, b2)
pre_Y_A <- rpois(N, theta_A)
pre_Y_B <- rpois(N, theta_B)

Pr4 <- mean(pre_Y_B < pre_Y_A)
Pr4
```

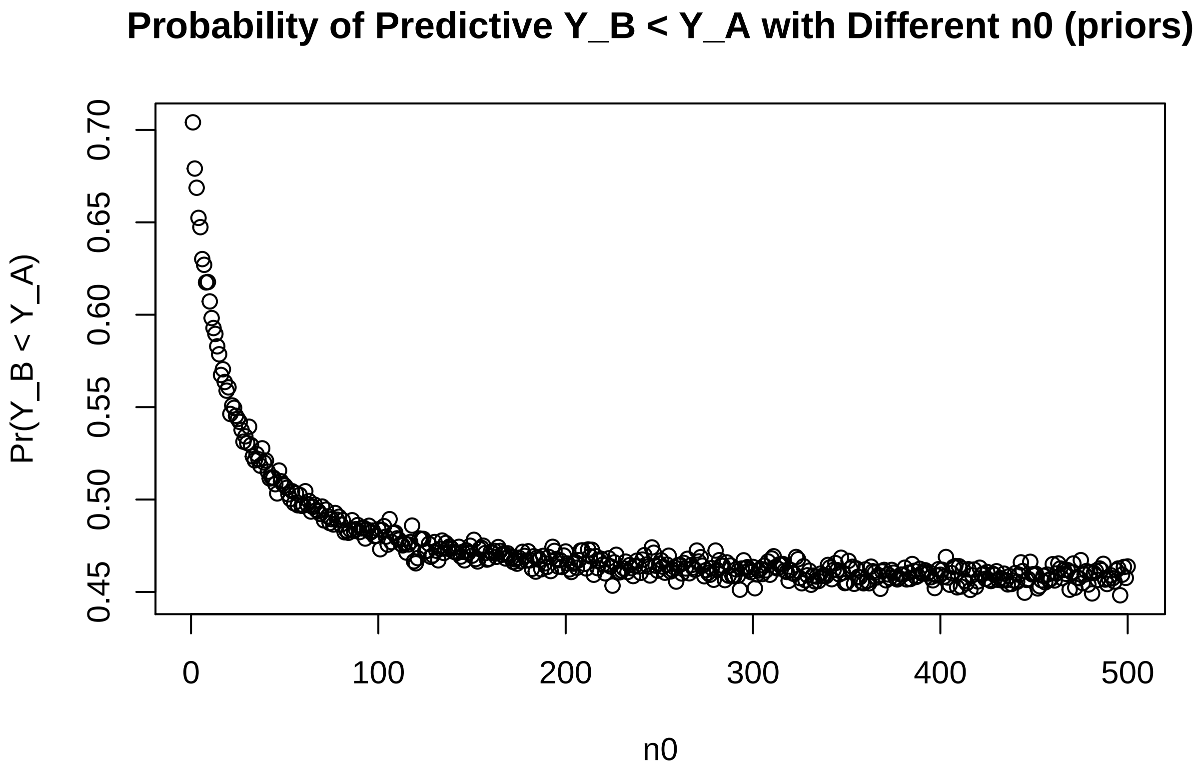
```
## [1] 0.7006
```

```
N <- 10000

a1 <- 120 + 117
b1 <- 10 + 10
theta_A <- rgamma(N, a1, b1)
pre_Y_A <- rpois(N, theta_A)

Pr5 <- c()
for (n0 in seq(1, 500)) {
  a2 <- 12 * n0 + 113
  b2 <- n0 + 13
  theta_B <- rgamma(N, a2, b2)
  pre_Y_B <- rpois(N, theta_B)
  Pr5[n0] <- mean(pre_Y_B < pre_Y_A)
}

plot(seq(1, 500), Pr5, xlab = "n0", ylab = "Pr(Y_B < Y_A)",
     main = "Probability of Predictive Y_B < Y_A with Different n0 (priors)")
```


Probability of Predictive Y_B < Y_A with Different n0 (priors)
Pr(Y_B < Y_A)
0.45
0.50
0.55
0.60
0.65
0.70
0
100
200
300
400
500
n0