STA 360: Homework 3

Samuel Eure

9/23/2018

Math Problem 1

Derive the expected value and variance of a poisson distribution with parameter θ

Expectation

Let

$$(Y_i \mid \theta) \sim^{\text{i.i.d}} \text{Poisson}(\theta)$$

Then, my definition,since the sample space of Y is $0, 1, 2, 3, 4...$,

$$\mathbb{E}[Y \mid \theta] = \sum_{y=0}^{\infty} yP(Y = y \mid \theta) = \sum_{y=0}^{\infty}(y)\left(\frac{\theta^y e^{-\theta}}{y!}\right) = (0)\left(\frac{\theta^0 e^{-\theta}}{0!}\right)\sum_{y=1}^{\infty}(y)\left(\frac{\theta^y e^{-\theta}}{y!}\right) = \sum_{y=1}^{\infty}(y)\left(\frac{\theta^y e^{-\theta}}{y!}\right)$$
$$= \sum_{y=1}^{\infty}\left(\frac{\theta^y e^{-\theta}}{(y-1)!}\right)$$

now, if we let $x + 1 = y \Rightarrow$,

$$\mathbb{E}[Y \mid \theta] = \sum_{x=0}^{\infty}\left(\frac{\theta^{x+1} e^{-\theta}}{(x+1-1)!}\right) = e^{-\theta}\theta\sum_{x=0}^{\infty}\left(\frac{\theta^x}{x!}\right)$$

By the taylor expansion of e^x, it is known that

$$\sum_{x=0}^{\infty}\left(\frac{\theta^x}{x!}\right) = e^{\theta}$$

Thus,

$$\mathbb{E}[Y \mid \theta] = e^{-\theta}e^{\theta}\theta = \theta$$

Which is the parameter of the Poisson distribution.

Variance

By definition

$$Var(Y \mid \theta) = \mathbb{E}[Y^2 \mid \theta] - \left(\mathbb{E}[Y \mid \theta]\right)^2 = \mathbb{E}[Y^2 \mid \theta] - \theta^2$$

Thus, all that is left to do is find $\mathbb{E}[Y \mid \theta])^2$

$$\mathbb{E}[Y \mid \theta])^2 = \sum_{y=0}^{\infty} y^2 P(Y = y \mid \theta) = \sum_{y=0}^{\infty} y^2 \frac{\theta^y e^{-\theta}}{y!}$$
$$= e^{-\theta} \sum_{y=0}^{\infty} y^2 \frac{\theta^y}{y!} = e^{-\theta} \sum_{y=1}^{\infty} y^2 \frac{\theta^y}{y!} = e^{-\theta} \sum_{y=1}^{\infty} y \frac{\theta^y}{(y-1)!}$$

Since the first term of the summation is, again, zero. Furthermore, if we let $v + 1 = y$,

$$\mathbb{E}[Y \mid \theta])^2 = e^{-\theta} \sum_{y=1}^{\infty} y \frac{\theta^y}{(y-1)!} = e^{-\theta} \sum_{v=0}^{\infty} (v+1) \frac{\theta^{v+1}}{v!} = e^{-\theta} \Bigg[\sum_{v=0}^{\infty} (v) \frac{\theta^{v+1}}{v!} + \sum_{v=0}^{\infty} (1) \frac{\theta^{v+1}}{v!} \Bigg] = e^{-\theta}\theta \Bigg[\sum_{v=0}^{\infty} (v) \frac{\theta^{v}}{v!} + \sum_{v=0}^{\infty} \frac{\theta^{v}}{v!} \Bigg]$$
$$= e^{-\theta}\theta \Bigg[\sum_{v=0}^{\infty} (v) \frac{\theta^{v}}{v!} + e^{\theta}] = e^{-\theta}\theta \Bigg[\sum_{v=1}^{\infty} (v) \frac{\theta^{v}}{v!} + e^{\theta} \Bigg] = e^{-\theta}\theta \Bigg[\sum_{v=1}^{\infty} \frac{\theta^{v}}{(v-1)!} + e^{\theta} \Bigg]$$

from the Taylor expansion used above and since the first term of the remaining summation will be zero. Now, if we let z+1 = v,

$$\mathbb{E}[Y \mid \theta])^2 = e^{-\theta}\theta \Bigg[\sum_{v=1}^{\infty} \frac{\theta^{v}}{(v-1)!} + e^{\theta} \Bigg] = e^{-\theta}\theta \Bigg[\sum_{z=0}^{\infty} \frac{\theta^{z+1}}{(z)!} + e^{\theta} \Bigg] = e^{-\theta}\theta \Bigg[\theta \sum_{z=0}^{\infty} \frac{\theta^{z}}{(z)!} + e^{\theta} \Bigg]$$
$$= e^{-\theta}\theta \Bigg[\theta e^{\theta} + e^{\theta} \Bigg] = \theta^2 + \theta$$

Thus

$$Var(Y \mid \theta) = \mathbb{E}[Y^2 \mid \theta] - \theta^2 = (\theta^2 + \theta) - \theta^2 = \theta$$
$$Var(Y \mid \theta) = \theta$$

Math Problem 2: find $p(\theta \mid n_0, t_0)$ if $\phi = log(\theta)$ given

$$p(\phi \mid n_0, t_0) = e^{-n_0 e^{\phi}} e^{n_0 t_0 \phi}$$

First, using Jeffery's Rule for Priors,

$$p(\phi \mid n_0, t_0) = p(\theta \mid n_0, t_0) \left| \frac{d\theta}{d\phi} \right|$$

First I'll find $\frac{d\theta}{d\phi}$

$$\frac{d\theta}{d\phi} = \frac{1}{d\phi/d\theta} = \frac{1}{\frac{d}{d\theta} log(\theta)} = \frac{1}{\frac{1}{\theta}} = \theta$$

Since θ is non-negative,

$$p(\theta \mid n_0, t_0) = \frac{p(\phi \mid n_0, t_0)}{|d\theta/d\phi|} = \frac{p(\phi \mid n_0, t_0)}{\theta} = \frac{e^{-n_0 e^{\phi}} e^{n_0 t_0 \phi}}{\theta}$$

Remembering that $\phi = log(\theta)$

$$p(\theta \mid n_0, t_0) = \frac{e^{-n_0 e^{log(\theta)}} e^{n_0 t_0 log(\theta)}}{\theta} = \frac{\left(e^{n_0 e^{log(\theta)}}\right)^{-1} e^{log(\theta) n_0 t_0}}{\theta} = \frac{\left(e^{n_0 \theta}\right)^{-1} \theta^{n_0 t_0}}{\theta} = \frac{e^{-n_0 \theta} \theta^{n_0 t_0}}{\theta}$$
$$p(\theta \mid n_0, t_0) = \frac{\theta^{n_0 t_0} e^{-n_0 \theta}}{\theta} = \theta^{n_0 t_0 - 1} e^{-n_0 \theta} = \text{gamma}(n_0 t_0, n_0)$$

Problem 3.3

3.3a

Prior information shows $y_a \sim$ Poisson(12)

```
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)
```

Assume $\theta_A \sim gamma(120, 10)$, $\theta_B \sim gamma(12, 1)$, $p(\theta_A, \theta_B) = p(\theta_A)p((\theta_B)$.

Let $n_A = 10$, the number of samples from A, and, similarly for B, let $n_B = 13$.

Posterior distribution for θ_A can be found using Bayes' Rule, as follows:

$$p(\theta_A \mid \mathbf{y_A}) = \frac{p(\mathbf{y_A} \mid \theta_A) p(\theta_A)}{p(\mathbf{y_A})} \propto p(\mathbf{y_A} \mid \theta_A) p(\theta_A) \propto \left(\Pi_{i=1}^{10} \theta_A^{y_i} e^{-\theta_A}\right) \left(\theta_A^{120-1}\right) \left(e^{-10\theta_A}\right)$$
$$\left(\Pi_{i=1}^{10} \theta_A^{y_i} e^{-\theta_A}\right) \left(\theta_A^{120-1}\right) \left(e^{-10\theta_A}\right) = \left(\theta_A^{120 + \left(\sum_{i=1}^{10} y_i\right) - 1}\right) \left(e^{-(10+n_A)\theta_A}\right)$$

Thus

$$p(\theta_A \mid \mathbf{y_A}) \propto \left(\theta_A^{120 + \left(\sum_{i=1}^{10} y_i\right) - 1}\right) \left(e^{-20\theta_A}\right)$$
$$\Longrightarrow$$

Equation 1

$$p(\theta_A \mid \mathbf{y_A}) = \text{gamma}\left(120 + \sum_{i=1}^{10} y_i, 10 + n_A\right)$$

Since

```r
print(paste("nA       = ", length(YA)))
```

```
## [1] "nA       =  10"
```

```r
print(paste("sum(YA) = ", sum(YA)))
```

```
## [1] "sum(YA) =  117"
```

The distribution of $p(\theta_A \mid \mathbf{y_A})$ is

$$p(\theta_A \mid \mathbf{y_A}) = \text{gamma}\big(237, 20\big)$$

which has a mean of $\frac{237}{20}$ and a variance of $\frac{237}{20^2} = \frac{237}{400}$, thus,

$$\mathbb{E}[\theta_A \mid \mathbf{y_A}] = \frac{237}{20}$$

$$Var[\theta_A \mid \mathbf{y_A}] = \frac{237}{400}$$

Using this information about the distirbution, a 95% convidence interval can be found via R's built in gamma distributions

```r
lowerBound = qgamma(.025, 237, 20)
upperBound = qgamma(1-.025, 237, 20)
print(paste("95% quantile based CI for ThetaA:",
            "[", lowerBound,",",upperBound, "]"))
```

```
## [1] "95% quantile based CI for ThetaA: [ 10.3892381909418 , 13.405448325642 ]"
```

To obtain all the above information for θ_B, I simply refer equation 1 which was derived earlier, and use $\sum_{i=1}^{13} y_i$ for $\mathbf{y_B}$, n_B, and replace the 120 and 10 with 12 and 1 respectively since these numbers came from the prior distribution of θ_A (and should now come from the prior of θ_B). This results in

$$p(\theta_B \mid \mathbf{y_B}) = \text{gamma}\left(12 + \sum_{i=1}^{13} y_i, 1 + n_B\right)$$

```r
print(paste("nB       = ", length(YB)))
```

```
## [1] "nB       =  13"
```

```r
print(paste("sum(YB) = ", sum(YB)))
```

```
## [1] "sum(YB) =  113"
```

$$\Longrightarrow$$

$$p(\theta_B \mid \mathbf{y_B}) = \text{gamma}\big(12 + 113, 1 + 13\big) = \text{gamma}\big(125, 14\big)$$

$$\Longrightarrow$$

$$\mathbb{E}[\theta_B \mid \mathbf{y_B}] = \frac{125}{14}$$

$$Var[\theta_B \mid \mathbf{y_B}] = \frac{125}{196}$$

```
lowerBound = qgamma(.025, 125, 14)
upperBound = qgamma(1-.025, 125, 14)
print(paste("95% quantile based CI for ThetaB:",
            "[", lowerBound,",", upperBound, "]"))
```

```
## [1] "95% quantile based CI for ThetaB: [ 7.4320642194643 , 10.5603081492424 ]"
```

3.3b

Using a new prior for θ_B of gamma$(12n_0, n_0)$, the new posterior distribution for θ_B will be

$$p(\theta_B \mid \mathbf{y_B}) = \text{gamma}\big(12n_0 + 113, n_0 + 13\big)$$

which has an expectation of

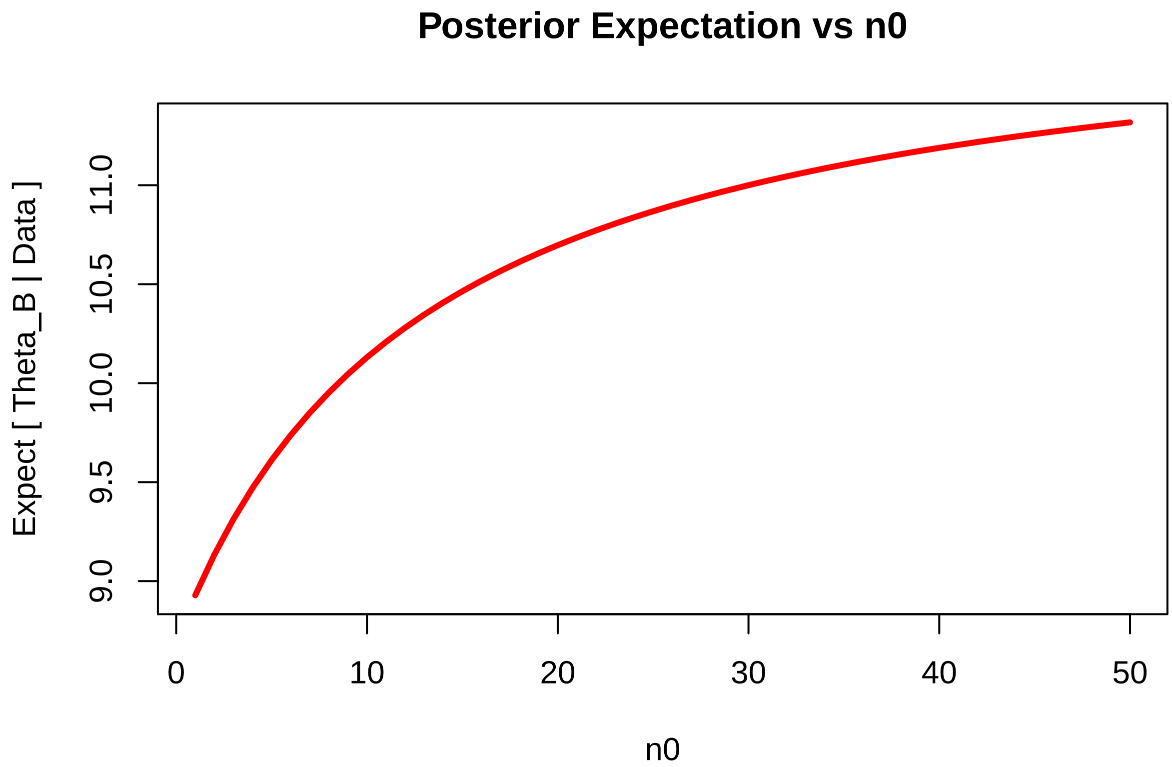
$$\mathbb{E}[\theta_B \mid \mathbf{y_B}] = \frac{12n_0 + 113}{n_0 + 13}$$

based on equation 1 derived above, and where 113 and 13 are the sum of the y_B values and the number of y_B observations. Using $n_0 \in \{1, 2, ..., 50\}$

```
n = 1:50
E = (12*n+sum(YB))/(n+13)
E
```

```
##  [1]  8.928571  9.133333  9.312500  9.470588  9.611111  9.736842  9.850000
##  [8]  9.952381 10.045455 10.130435 10.208333 10.280000 10.346154 10.407407
## [15] 10.464286 10.517241 10.566667 10.612903 10.656250 10.696970 10.735294
## [22] 10.771429 10.805556 10.837838 10.868421 10.897436 10.925000 10.951220
## [29] 10.976190 11.000000 11.022727 11.044444 11.065217 11.085106 11.104167
## [36] 11.122449 11.140000 11.156863 11.173077 11.188679 11.203704 11.218182
## [43] 11.232143 11.245614 11.258621 11.271186 11.283333 11.295082 11.306452
## [50] 11.317460
```

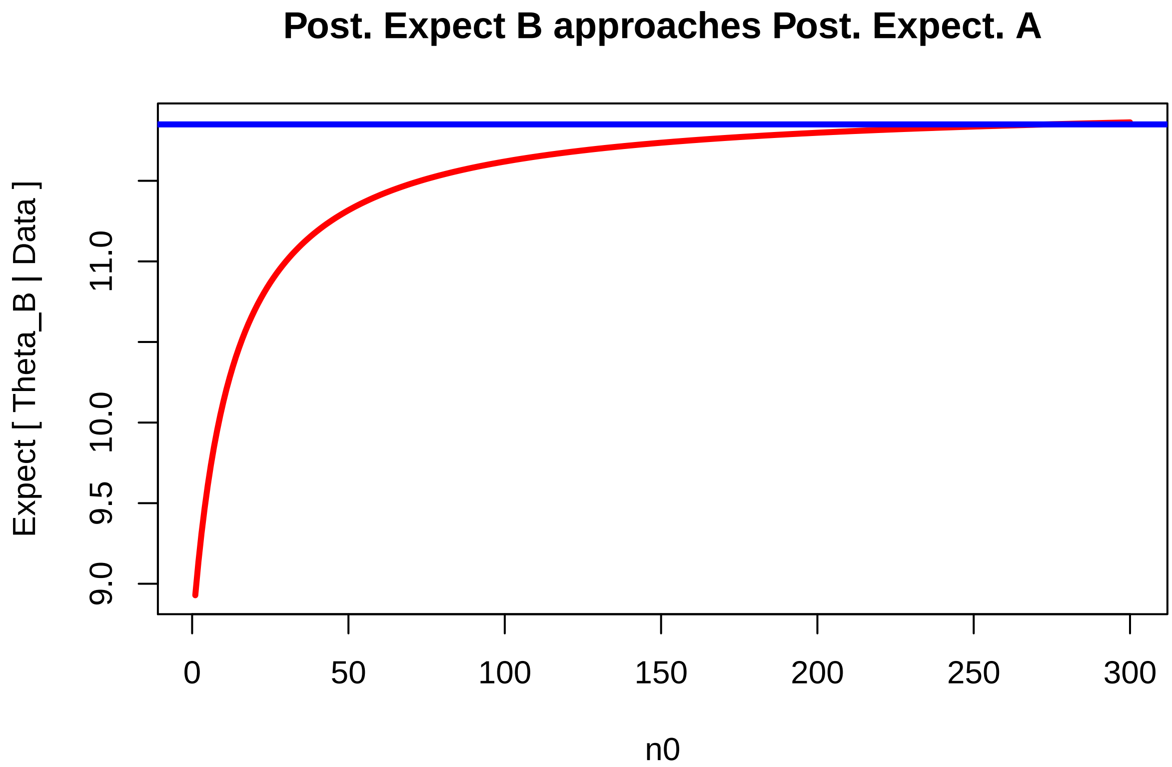
```
plot(n, E, xlab = "n0", ylab = "Expect [ Theta_B | Data ]", type = 'l', lwd = 3, col = "red",
     main = "Posterior Expectation vs n0")
```



Since the data collected from group B ($\mathbf{y_B}$) was larger and had a sample mean lower than that of group A's (8.69 for group B vs. 11.7 for group A), the posterior expectation for θ_B gets pulled down by the samples more severely than the posterior expectation θ_A. This is coupled with the fact that the prior of A is weighted much more heavily than the prior of B (120 vs 12, etc.) which means the posterior expectation of A is less likely to deviate from the prior expectation given a small sample size (since the posterior expectation is proportionally balanced between the prior mean and the sample mean).

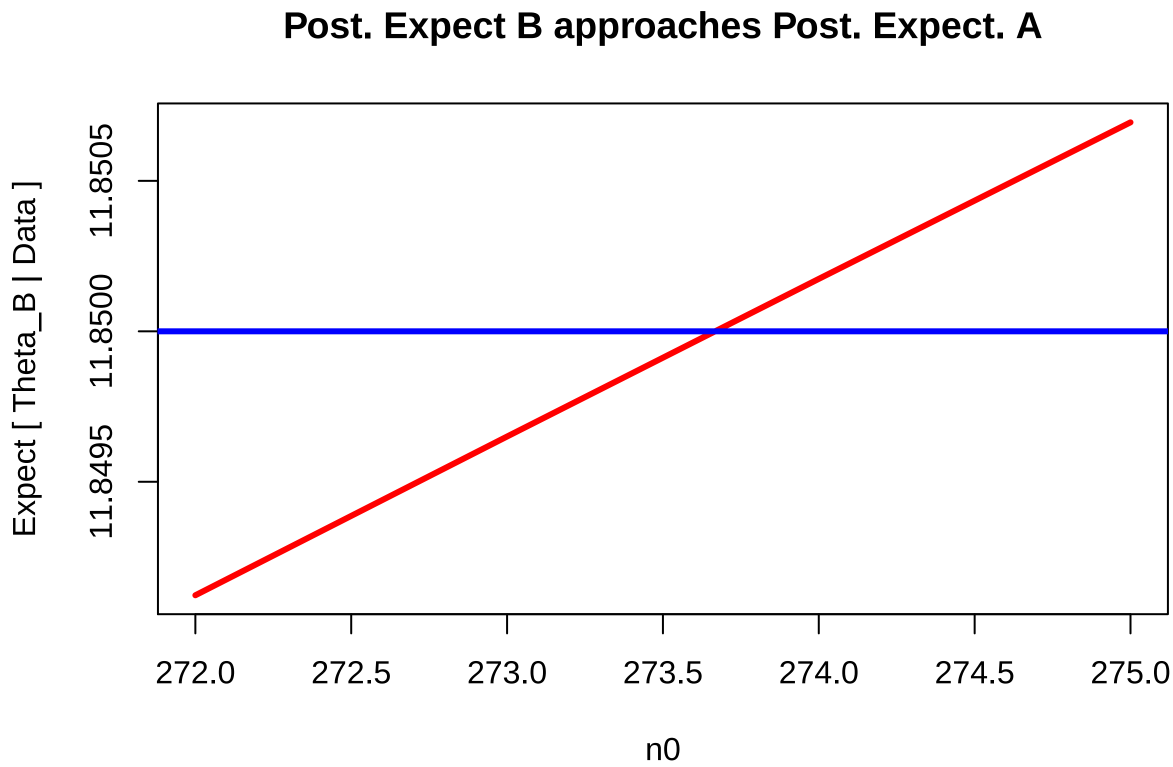
Thus, since the prior expectations are equivalent for θ_B and θ_A, θ_B would need a very heavily weighted prior with $n_0 > (\sim 250)$ in order for the posterior expectation to be similar. This is shown in the graph below. Here, the blue line represents the posterior expectation of θ_A and the red line represents the posterior expectation of θ_B.

```
n = 1:300
E = (12*n+sum(YB))/(n+13)
plot(n, E, xlab = "n0", ylab = "Expect [ Theta_B | Data ]", type = 'l', lwd = 3, col = "red",
     main = "Post. Expect B approaches Post. Expect. A")
abline(h = 237/20, col = 'blue', lwd = 3)
```



Looking at this graph, one may imageing that the posterior expecation of θ_B asymptotically approaches the posterior expectation of θ_A. This of course cannot be true, since the posterior expectation of θ_A is 11.85, which is less than 12, and as $n_0 \to \infty$, the posterior expectation of θ_B approaches 12 (the prior expectation) which is greater than 11.85. In fact, the lines cross at around $n_0 = \sim 273$

```
n = 272:275
E = (12*n+sum(YB))/(n+13)
plot(n, E, xlab = "n0", ylab = "Expect [ Theta_B | Data ]", type = 'l', lwd = 3, col = "red",
     main = "Post. Expect B approaches Post. Expect. A")
abline(h = 237/20, col = 'blue', lwd = 3)
```



3.3c

The problem states that the strain B mice are *related* to the type A mice, and in the problem we even took this knowledge to guess the prior for θ_B. Thus, it seems reasonable to assume that knowing some information about θ_A should provide some information as to what ball park θ_B should be in. This being said, it doesn't make sense to have a complete independence between θ_A and θ_B, and thus having $p(\theta_A, \theta_B) = p(\theta_A)p(\theta_B)$ does not make sense.

Problem 3.9

For $p(\theta)$ to be a conjugate prior to the Galeshore distirubtion, we need

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

which, can be represented as

$$p(\theta \mid y) \propto p(\theta)y^{2a-1}\theta^{c_1}e^{-c_2\theta^2}$$

Thus, if we need our class of priors to be conjugate with the Galeshore(a, θ), we need $p(\theta)$ to be a member of a class which includes terms of θ^{c_1} and $e^{-c_2\theta^2}$ for some constants c_1 and c_2.

Conveniently, we can represent the $p(y \mid a, \theta)$ as

$$p(y \mid a, \theta) = \frac{1}{Z}y^{c_1}\exp\{-c_2y^2\}$$

with $c_1 = 2a - 1$, $c_2 = \theta^2$, and a normalization constant of Z. Thus, a Galeshore Prior is a member of this class of conjugate priors for θ.

Additionally, if we simply redefine θ as $\theta = \phi^{1/2}$, then

$$p(y \mid a, \phi) = \frac{2}{\Gamma(a)} \phi^a y^{2a-1} e^{-\phi y^2}$$

and thus, we could simply find a prior for ϕ, which looks suspiciously like a Gamma, in fact, if

$$p(\phi \mid y) \propto p(\phi) p(y \mid \phi) \propto p(\phi) \phi^{c_1} e^{-c_2 \phi}$$

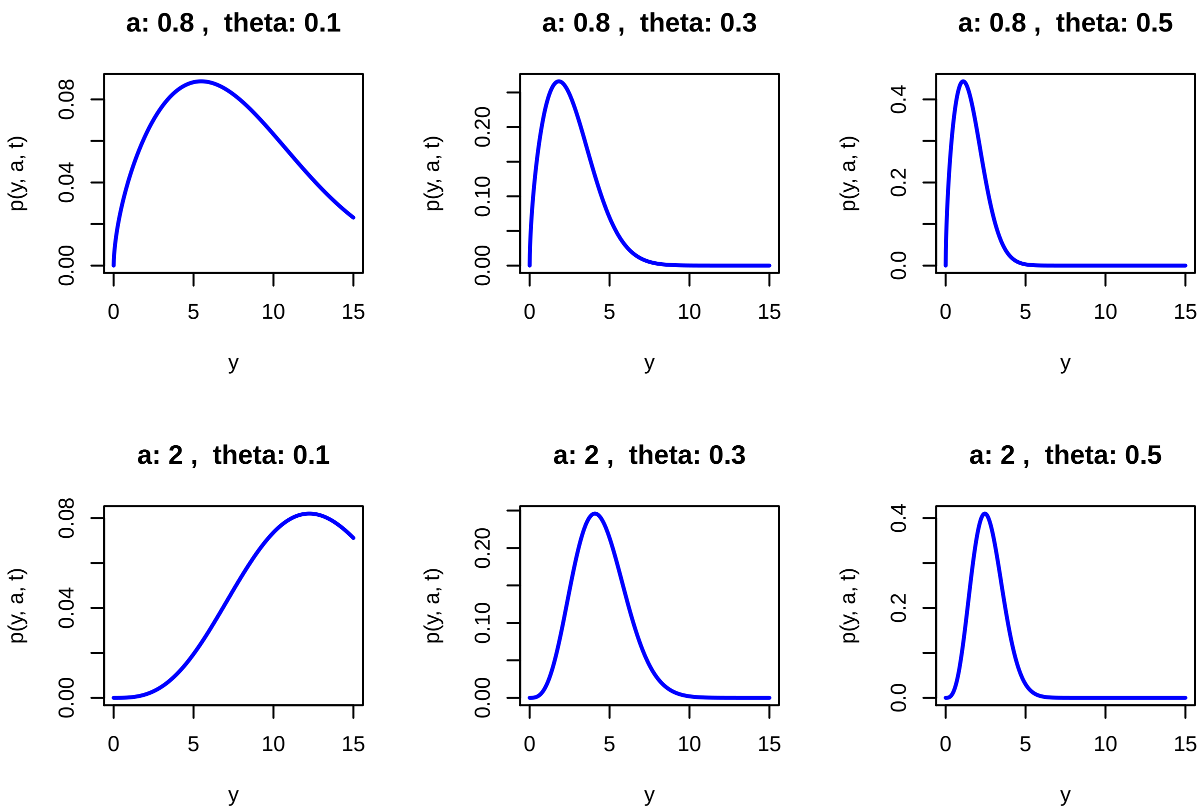
where $c_1 = a$ and $c_2 = y^2$, then if ϕ is simply Gamma (α, β), then

$$p(\phi \mid y) \propto p(\phi) p(y \mid \phi) \propto \phi^{\alpha-1} e^{-\phi\beta} \phi^{c_1} e^{-c_2 \phi} = \phi^{\alpha+c_1-1} e^{-\phi(c_2+\beta)}$$

which results in a $(\phi \mid y) \sim$ Gamma$(\alpha + c_1, c_2 + \beta)$, or, more specifically for this problem, a Gamma $(\alpha + a, \beta + y^2)$.

Now, we've all seen a gamma distribution before, so I'll plot some of the Galenshore priors.

```
par(mfrow = c(2,3))
p = function(y,a,t){
  v = (2/gamma(a))*t^(2*a)*(y^(2*a-1))*exp(-t^2*y^2)
  return(v)
}

y = seq(0,15,by = 0.01)
A = c(.8,2)
Theta = c(.1,.3,.5)

for (a in A){
  for (t in Theta){
    plot(y,p(y,a,t), type = 'l', lwd = 2, col = "blue", main = paste("a:",a,", ","theta:",t))
  }
}
```



3.9b

Using a prior of $(\theta \mid \alpha, \beta) \sim$ Galenshore(α, β), the posterior distribution is

$$p(\theta \mid y_1, y_2, ..., y_n) \propto p(\theta \mid \alpha, \beta)\Pi_{i=1}^n p(Y_i = y_i \mid a, \theta) \propto \left(\theta^{2\alpha-1}e^{-\theta^2\beta^2}\right)\Pi_{i=1}^n \theta^{2a} y_i^{2a-1} e^{-\theta^2 y_i^2}$$

$$\Rightarrow p(\theta \mid y_1, y_2, ..., y_n) \propto \left(\theta^{2\alpha-1}e^{-\theta^2\beta^2}\right)\theta^{(2a)n}\left(\Pi_{i=1}^n y_i\right)^{2a-1} e^{-\theta^2\left(\sum_{i=1}^n y_i^2\right)} \propto \theta^{2(\alpha+an)-1} e^{-\theta^2\left(\beta^2+\sum_{i=1}^n y_i^2\right)}$$

Which means

$$\left(\theta \mid y_1, ..., y_n\right) \sim \text{Galenshore}\left(\alpha + an, \sqrt{\beta^2 + \sum_{i=1}^n y_i^2}\right)$$

Which is a pretty remarkable posterior if you ask me. Additionally, it appears that $\sum_{i=1}^n y_i^2$ is a sufficient statistic for $(y_1, ..., y_n)$ given the data.

3.9c

If we let

$$(\theta_a \mid \alpha_a, \beta_a) \sim \text{Galenshore}(\alpha_a, \beta_a)$$

and

$$(\theta_b \mid \alpha_b, \beta_b) \sim \text{Galenshore}(\alpha_b, \beta_b)$$

Then $\frac{p\left(\theta_a \mid (y_1,...,y_n)\right)}{p\left(\theta_b \mid (y_1,...,y_n)\right)}$ **will be equal to**

$$\frac{p\big(\theta_a \mid (y_1, ..., y_n)\big)}{p\big(\theta_b \mid (y_1, ..., y_n)\big)} = \frac{\frac{p(\theta_a)p(y_1,...,y_n \mid \theta_a)}{p(y_1,...,y_n)}}{\frac{p(\theta_b)p(y_1,...,y_n \mid \theta_b)}{p(y_1,...,y_n)}} = \frac{p(\theta_a)p(y_1, ..., y_n \mid \theta_a)}{p(\theta_b)p(y_1, ..., y_n \mid \theta_b)}$$

$$= \frac{\left(\frac{2}{\Gamma(\alpha_a)}\beta_a^{2\alpha_a}\theta_a^{2\alpha_a-1}e^{-\theta_a^2\beta_a^2}\right)\left(\Pi_{i=1}^n \frac{2}{\Gamma(a)}\theta_a^{2a}y_i^{2a-1}e^{-\theta_a^2 y_i^2}\right)}{\left(\frac{2}{\Gamma(\alpha_b)}\beta_b^{2\alpha_b}\theta_b^{2\alpha_b-1}e^{-\theta_b^2\beta_b^2}\right)\left(\Pi_{i=1}^n \frac{2}{\Gamma(a)}\theta_b^{2a}y_i^{2a-1}e^{-\theta_b^2 y_i^2}\right)}$$

$$= \frac{\left(\frac{2}{\Gamma(\alpha_a)}\beta_a^{2\alpha_a}\theta_a^{2\alpha_a-1}e^{-\theta_a^2\beta_a^2}\right)\left(\left(\frac{2}{\Gamma(a)}\right)^n \left(\theta_a^{2a}\right)^n \left(\Pi_{i=1}^n y_i\right)^{2a-1} e^{-\theta_a^2\left(\sum_{i=1}^n y_i^2\right)}\right)}{\left(\frac{2}{\Gamma(\alpha_b)}\beta_a^{2\alpha_b}\theta_b^{2\alpha_b-1}e^{-\theta_b^2\beta_b^2}\right)\left(\left(\frac{2}{\Gamma(a)}\right)^n \left(\theta_b^{2a}\right)^n \left(\Pi_{i=1}^n y_i\right)^{2a-1} e^{-\theta_b^2\left(\sum_{i=1}^n y_i^2\right)}\right)}$$

$$= \frac{\left(\Gamma(\alpha_b)\beta_a^{2\alpha_a}\theta_a^{2\alpha_a-1}e^{-\theta_a^2\beta_a^2}\right)\left(\left(\theta_a^{2a}\right)^n e^{-\theta_a^2\left(\sum_{i=1}^n y_i^2\right)}\right)}{\left(\Gamma(\alpha_a)\beta_a^{2\alpha_b}\theta_b^{2\alpha_b-1}e^{-\theta_b^2\beta_b^2}\right)\left(\left(\theta_b^{2a}\right)^n e^{-\theta_b^2\left(\sum_{i=1}^n y_i^2\right)}\right)}$$

$$\Rightarrow \frac{p\big(\theta_a \mid (y_1, ..., y_n)\big)}{p\big(\theta_b \mid (y_1, ..., y_n)\big)} = \frac{\Gamma(\alpha_b)}{\Gamma(\alpha_a)}\left(\frac{\beta_a^{\alpha_a}}{\beta_b^{\alpha_b}}\right)^2\left(\frac{\theta_a^{2(\alpha_a+an)-1}}{\theta_b^{2(\alpha_b+an)-1}}\right)\frac{e^{-\theta_a^2(\beta_a^2+\sum_{i=1}^n y_i^2)}}{e^{-\theta_b^2(\beta_b^2+\sum_{i=1}^n y_i^2)}}$$

Thus, $\sum_{i=1}^n y_i^2$ **is a sufficient statistics for the data.**

3.9d

Using the formula for the expected value of a variable Y **taken from a Galenshore**(a, θ) **distribution provided in the book, and using the fact that**

$$\left(\theta \mid y_1, ..., y_n\right) \sim \text{Galenshore}\left(\alpha + an, \sqrt{\beta^2 + \sum_{i=1}^n y_i^2}\right)$$

When $(\theta) \sim$ **Galenshore**(α, β) **the posterior expecation of** θ **is**

$$\mathbb{E}\left[\theta \mid y_1, ..., y_n\right] = \frac{\Gamma(\alpha + an + 1/2)}{\Gamma(\alpha + an)\sqrt{\beta^2 + \sum_{i=1}^n y_i^2}}$$

3.9e

Well, this is going to get messy, but here it goes,

$$p(\tilde{y} \mid y_1, ..., y_n) = \int_0^\infty p(\tilde{y} \mid \theta, y_1, ..., y_n) p(\theta \mid y_1, ..., y_n) d\theta$$

$$= \int_0^\infty \left(\frac{2}{\Gamma(a)} \theta^{2a} \tilde{y}^{2a-1} e^{-\theta^2 \tilde{y}^2} \right) \left(\frac{2}{\Gamma(\alpha + an)} \big(\beta^2 + \sum_{i=1}^n y_i^2\big)^{2(\alpha+an)} \theta^{2(\alpha+an)-1} e^{-\theta^2 \left(\beta^2 + \sum_{i=1}^n y_i^2\right)} \right) d\theta$$

$$= \frac{2}{\Gamma(a)} \frac{2}{\Gamma(\alpha + an)} \tilde{y}^{2a-1} \big(\beta^2 + \sum_{i=1}^n y_i^2\big)^{2(\alpha+an)} \int_0^\infty \left(\theta^{2a} e^{-\theta^2 \tilde{y}^2} \right) \left(\theta^{2(\alpha+an)-1} e^{-\theta^2 \left(\beta^2 + \sum_{i=1}^n y_i^2\right)} \right) d\theta$$

$$= \frac{2}{\Gamma(a)} \frac{2}{\Gamma(\alpha + an)} \tilde{y}^{2a-1} \big(\beta^2 + \sum_{i=1}^n y_i^2\big)^{2(\alpha+an)} \int_0^\infty \left(\theta^{2(\alpha+an)+2a-1} e^{-\theta^2 \left(\tilde{y}^2 + \beta^2 + \sum_{i=1}^n y_i^2\right)} \right) d\theta$$

$$= \frac{2}{\Gamma(a)} \frac{2}{\Gamma(\alpha + an)} \tilde{y}^{2a-1} \big(\beta^2 + \sum_{i=1}^n y_i^2\big)^{2(\alpha+an)} \int_0^\infty \left(\theta^{2(\alpha+a(n+1))-1} e^{-\theta^2 \left(\tilde{y}^2 + \beta^2 + \sum_{i=1}^n y_i^2\right)} \right) d\theta$$

The integral must integrate to $\frac{1}{\textbf{normalization constant}}$ for a Galenshore pdf, thus

$$p(\tilde{y} \mid y_1, ..., y_n) = \frac{2}{\Gamma(a)} \frac{2}{\Gamma(\alpha + an)} \tilde{y}^{2a-1} \big(\beta^2 + \sum_{i=1}^n y_i^2\big)^{2(\alpha+an)} \left(\frac{\Gamma(\alpha + a(n+1))}{2\big(\tilde{y}^2 + \beta^2 + \sum_{i=1}^n y_i^2\big)^{\alpha+a(n+1)}} \right)$$

$$p(\tilde{y} \mid y_1, ..., y_n) = \frac{2\tilde{y}^{2a-1} \big(\beta^2 + \sum_{i=1}^n y_i^2\big)^{2(\alpha+an)}}{\Gamma(a)\Gamma(\alpha + an)} \left(\frac{\Gamma(\alpha + a(n+1))}{\big(\tilde{y}^2 + \beta^2 + \sum_{i=1}^n y_i^2\big)^{\alpha+a(n+1)}} \right)$$

Behold! The posterior predictive density function!

4.1

From problem 3.1e from homework 2, we discovered that the posterior distribution of θ_1 was

$$(\theta_1 \mid y_1, ..., y_n) \sim \text{beta}(1 + \sum_{i=1}^n y_i, 1 + n - \sum_{i=1}^n y_i)$$

thus,

$$(\theta_1 \mid y_1, ..., y_{100}) \sim \text{beta}(1 + 57, 1 + 100 - 57)$$

and

$$(\theta_2 \mid y_1, ..., y_{50}) \sim \text{beta}(1+30, 1+50-30)$$

```
n1 = 100; y1 = 57;
n2 = 50; y2 = 30;
S = 5000;
T1 = rbeta(S, 1+y1,1+n1-y1)
T2 = rbeta(S, 1+y2,1+n2-y2)
solution = mean(T1<T2)
print(paste("Pr(Theta1 < Theta2 |data and priors) =", solution))
```

```
## [1] "Pr(Theta1 < Theta2 |data and priors) = 0.6414"
```

4.2

4.2a

Since $(\theta_A \mid \mathbf{y}_A) \sim$ Gamma$(237, 20)$ and $(\theta_B \mid \mathbf{y}_B) \sim$ Gamma$(125, 14)$, the desired statistic can be calculated by taking 100000 samples from each of the posterior distributions

```
size = 100000
SampleA = rgamma(size, 237, 20)
SampleB = rgamma(size, 125, 14)
print(paste("Pr(ThetaB < ThetaA | yA, yB) = ", mean(SampleB < SampleA)))
```

```
## [1] "Pr(ThetaB < ThetaA | yA, yB) =  0.99568"
```

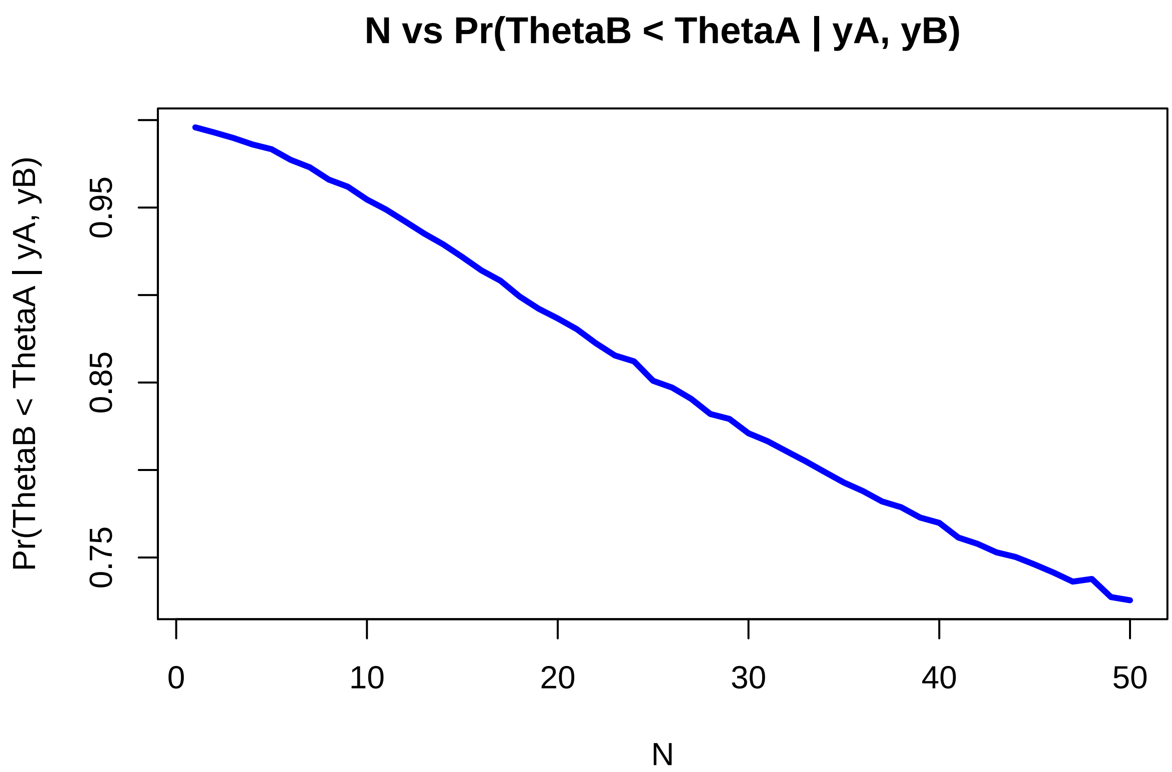
4.2b

```
results = c()
size = 100000
N = 1:50;
for (n in N){
  SampleA = rgamma(size, 237, 20);
  SampleB = rgamma(size, 12*n + 113, n+13);
  solution = mean(SampleB < SampleA);
  results = c(results, solution)
  print(paste("Pr(ThetaB < ThetaA | yA, yB, n=",n,") = ", solution))
}
```

```
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 1 ) =  0.99585"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 2 ) =  0.99291"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 3 ) =  0.98978"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 4 ) =  0.98607"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 5 ) =  0.98336"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 6 ) =  0.97728"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 7 ) =  0.97302"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 8 ) =  0.96598"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 9 ) =  0.9619"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 10 ) =  0.9546"
```

```
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 11 ) =  0.94888"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 12 ) =  0.94206"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 13 ) =  0.93515"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 14 ) =  0.92895"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 15 ) =  0.92171"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 16 ) =  0.91411"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 17 ) =  0.90816"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 18 ) =  0.89919"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 19 ) =  0.89217"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 20 ) =  0.88663"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 21 ) =  0.8805"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 22 ) =  0.87245"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 23 ) =  0.86546"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 24 ) =  0.8621"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 25 ) =  0.85097"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 26 ) =  0.84709"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 27 ) =  0.84066"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 28 ) =  0.83201"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 29 ) =  0.82919"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 30 ) =  0.821"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 31 ) =  0.81646"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 32 ) =  0.81067"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 33 ) =  0.80496"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 34 ) =  0.79885"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 35 ) =  0.79284"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 36 ) =  0.78797"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 37 ) =  0.78203"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 38 ) =  0.77873"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 39 ) =  0.77281"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 40 ) =  0.7698"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 41 ) =  0.7614"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 42 ) =  0.75783"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 43 ) =  0.75294"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 44 ) =  0.7503"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 45 ) =  0.74598"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 46 ) =  0.74133"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 47 ) =  0.7362"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 48 ) =  0.73774"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 49 ) =  0.72737"
## [1] "Pr(ThetaB < ThetaA | yA, yB, n= 50 ) =  0.72555"
```

```r
plot(N, results, main = 'N vs Pr(ThetaB < ThetaA | yA, yB)',
     ylab = 'Pr(ThetaB < ThetaA | yA, yB)', col = 'blue', type = 'l', lwd = 3)
```



As shown in the plot above, $P(\theta_B < \theta_A \mid \mathbf{y}_A, \mathbf{y}_B)$ seems to be almost linearly dependent on the prior sample size of θ_B, which doesn't make it rediculously sensitive to the prior sample size (it isn't exponentially or polynomially dependent), however if one was to make the prior sample size as large as the prior sample size of θ_A, then it can impact the results. But fluxuations of about 10 samples doesn't cause a dramatic change in the outcome.

4.2c

First, I'll draw random θ_A and θ_B from their posterior distributions. Then, I'll sample from the conditional distributions of $\tilde{Y} \mid \theta, y_1, ..., y_n)$.

```
size = 50000;
A1 = 120; B1 = 10;
A2 = 12; B2 = 1;
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);

Y.thetaA = rep(0,size)
Y.thetaB = rep(0,size)
for (s in 1:size){
  ThetaA.y = rgamma(1, A1+sum(YA), B1+length(YA));
  Y.thetaA[s] = rpois(1,ThetaA.y);
  ThetaB.y = rgamma(1, A2+sum(YB), B2+length(YB));
  Y.thetaB[s] = rpois(1, ThetaB.y);
}
print(paste("Pr( YB < YA | ya, yb) = ",mean(Y.thetaB<Y.thetaA)))
```

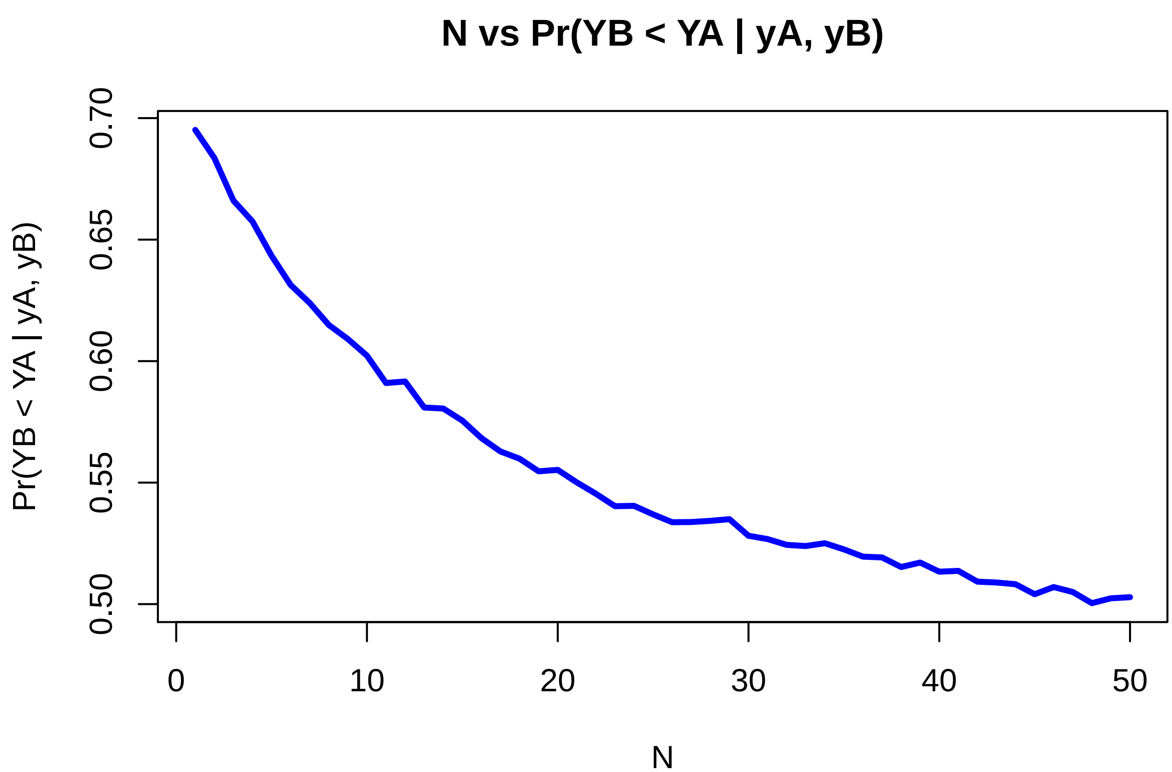
```
## [1] "Pr( YB < YA | ya, yb) =  0.70024"
```

```r
results = c()
size = 50000

A1 = 120; B1 = 10;
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);

N = 1:50;
for (n in N){
  for (s in 1:size){
    ThetaA.y = rgamma(1, A1+sum(YA), B1+length(YA));
    Y.thetaA[s] = rpois(1,ThetaA.y);
    ThetaB.y = rgamma(1, 12*n+sum(YB), n+length(YB));
    Y.thetaB[s] = rpois(1, ThetaB.y);
  }
  solution = mean(Y.thetaB<Y.thetaA);
  results = c(results, solution)
  print(paste("Pr(YB < YA | yA, yB, n=",n,") = ", solution))
}
```

```
## [1] "Pr(YB < YA | yA, yB, n= 1 ) =  0.69514"
## [1] "Pr(YB < YA | yA, yB, n= 2 ) =  0.6836"
## [1] "Pr(YB < YA | yA, yB, n= 3 ) =  0.66608"
## [1] "Pr(YB < YA | yA, yB, n= 4 ) =  0.65744"
## [1] "Pr(YB < YA | yA, yB, n= 5 ) =  0.64334"
## [1] "Pr(YB < YA | yA, yB, n= 6 ) =  0.63134"
## [1] "Pr(YB < YA | yA, yB, n= 7 ) =  0.62392"
## [1] "Pr(YB < YA | yA, yB, n= 8 ) =  0.61488"
## [1] "Pr(YB < YA | yA, yB, n= 9 ) =  0.60908"
## [1] "Pr(YB < YA | yA, yB, n= 10 ) =  0.60226"
## [1] "Pr(YB < YA | yA, yB, n= 11 ) =  0.59098"
## [1] "Pr(YB < YA | yA, yB, n= 12 ) =  0.59162"
## [1] "Pr(YB < YA | yA, yB, n= 13 ) =  0.5809"
## [1] "Pr(YB < YA | yA, yB, n= 14 ) =  0.58048"
## [1] "Pr(YB < YA | yA, yB, n= 15 ) =  0.57552"
## [1] "Pr(YB < YA | yA, yB, n= 16 ) =  0.56832"
## [1] "Pr(YB < YA | yA, yB, n= 17 ) =  0.5628"
## [1] "Pr(YB < YA | yA, yB, n= 18 ) =  0.55984"
## [1] "Pr(YB < YA | yA, yB, n= 19 ) =  0.55466"
## [1] "Pr(YB < YA | yA, yB, n= 20 ) =  0.55524"
## [1] "Pr(YB < YA | yA, yB, n= 21 ) =  0.55008"
## [1] "Pr(YB < YA | yA, yB, n= 22 ) =  0.54542"
## [1] "Pr(YB < YA | yA, yB, n= 23 ) =  0.5403"
## [1] "Pr(YB < YA | yA, yB, n= 24 ) =  0.54044"
## [1] "Pr(YB < YA | yA, yB, n= 25 ) =  0.53692"
## [1] "Pr(YB < YA | yA, yB, n= 26 ) =  0.5337"
## [1] "Pr(YB < YA | yA, yB, n= 27 ) =  0.53378"
## [1] "Pr(YB < YA | yA, yB, n= 28 ) =  0.53426"
## [1] "Pr(YB < YA | yA, yB, n= 29 ) =  0.53496"
## [1] "Pr(YB < YA | yA, yB, n= 30 ) =  0.52812"
## [1] "Pr(YB < YA | yA, yB, n= 31 ) =  0.52676"
## [1] "Pr(YB < YA | yA, yB, n= 32 ) =  0.52438"
## [1] "Pr(YB < YA | yA, yB, n= 33 ) =  0.5239"
## [1] "Pr(YB < YA | yA, yB, n= 34 ) =  0.52506"
```

```
## [1] "Pr(YB < YA | yA, yB, n= 35 ) =  0.52252"
## [1] "Pr(YB < YA | yA, yB, n= 36 ) =  0.51954"
## [1] "Pr(YB < YA | yA, yB, n= 37 ) =  0.51922"
## [1] "Pr(YB < YA | yA, yB, n= 38 ) =  0.51526"
## [1] "Pr(YB < YA | yA, yB, n= 39 ) =  0.5171"
## [1] "Pr(YB < YA | yA, yB, n= 40 ) =  0.51334"
## [1] "Pr(YB < YA | yA, yB, n= 41 ) =  0.51368"
## [1] "Pr(YB < YA | yA, yB, n= 42 ) =  0.50924"
## [1] "Pr(YB < YA | yA, yB, n= 43 ) =  0.5089"
## [1] "Pr(YB < YA | yA, yB, n= 44 ) =  0.50818"
## [1] "Pr(YB < YA | yA, yB, n= 45 ) =  0.50408"
## [1] "Pr(YB < YA | yA, yB, n= 46 ) =  0.50702"
## [1] "Pr(YB < YA | yA, yB, n= 47 ) =  0.50498"
## [1] "Pr(YB < YA | yA, yB, n= 48 ) =  0.5004"
## [1] "Pr(YB < YA | yA, yB, n= 49 ) =  0.5024"
## [1] "Pr(YB < YA | yA, yB, n= 50 ) =  0.50284"
```

```r
plot(N, results, main = 'N vs Pr(YB < YA | yA, yB)',
     ylab = 'Pr(YB < YA | yA, yB)', col = 'blue', type = 'l', lwd = 3)
```



Given my plot above, it appears that this probability is quite sensitive when values of n are low, and then the sensitivity to this value decreases as n gets larger and larger, perhaps eventually reaching a platue.