Homework 3

Zach White

9/16/2016

Exercise 3.3

```
y.a = c(12,9,12,14,13,13,15,8,15,6)
y.b = c(11,11,10,9,9,8,7,10,6,8,8,9,7)

a.a = 120; a.b = 10
b.a = 12 ; b.b = 1
n.y.a = length(y.a)
n.y.b = length(y.b)
sumy.a = sum(y.a)
sumy.b = sum(y.b)
post.a.a = a.a + sumy.a
post.a.b = n.y.a + a.b

post.b.a = b.a + sumy.b
post.b.b = n.y.b + b.b
```

The closed form of posterior distribution of a poisson sampling model with a gamma prior is $\theta|Y \sim Ga(a + \sum Y, b + n)$

The posterior distribution of $\theta_A|Y_A \sim Ga(237, 20)$, and the posterior distribution of $\theta_B|Y_B \sim Ga(125, 14)$

```
post.mean.a = post.a.a / post.a.b
post.var.a = post.a.a / post.a.b ^2
post.mean.a
```

```
## [1] 11.85
```

```
post.var.a
```

```
## [1] 0.5925
```

```
qgamma(c(.025,.975), post.a.a,post.a.b)
```

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

```
post.mean.b = post.b.b / post.b.b
post.var.b = post.b.b / post.b.b ^2
post.mean.b
```

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

```
post.var.b
```

```
## [1] 0.07142857
```

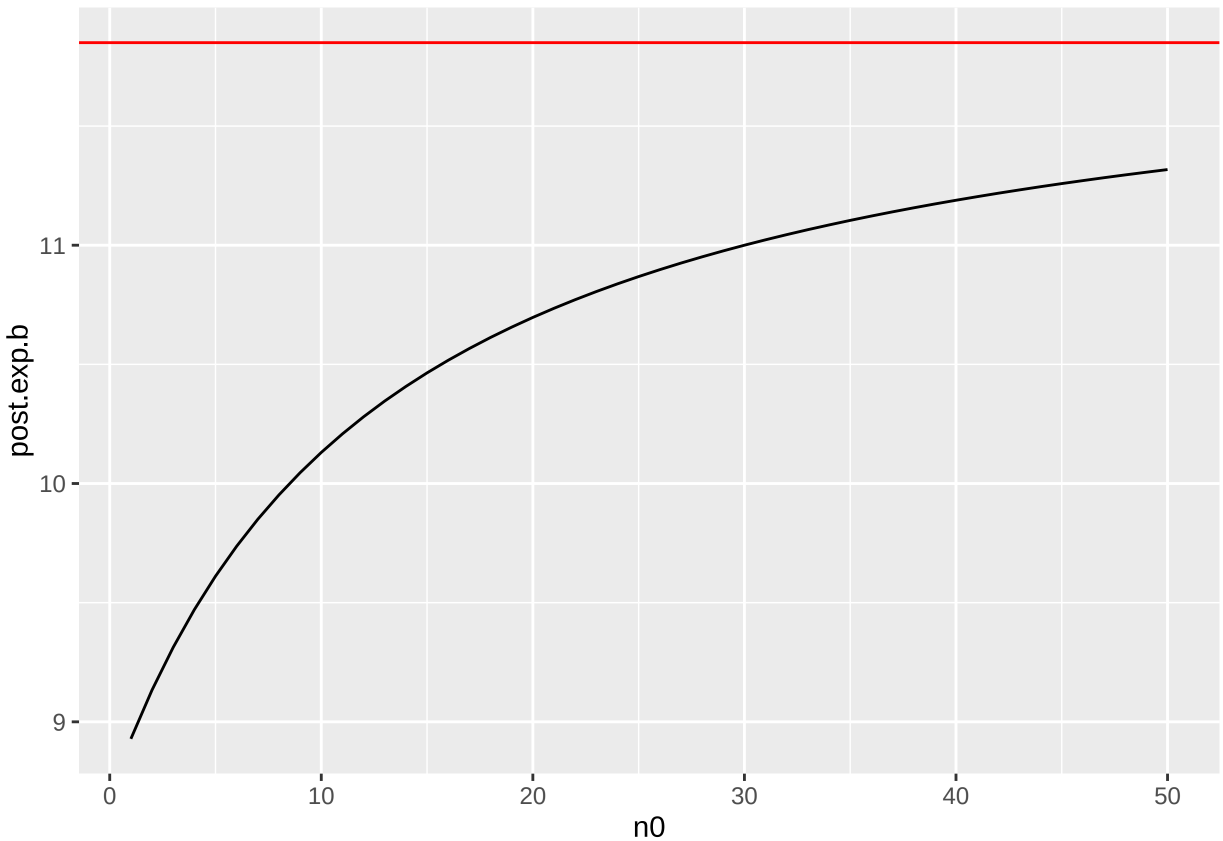
```
qgamma(c(.025,.975), post.b.b,post.b.b)
```

```
## [1] 0.5467093 1.5878854
```

The posterior mean of $\theta = 11.85$

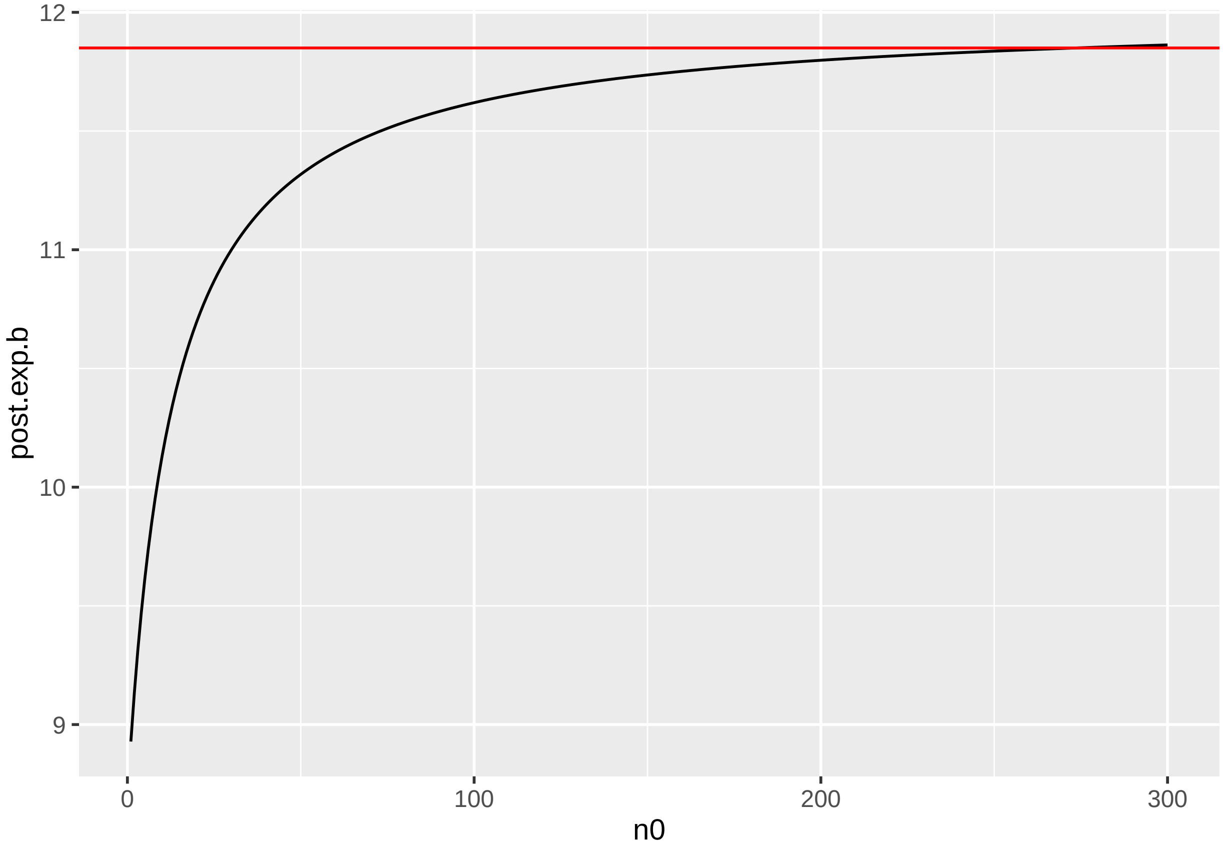
Part B

```
n0 = 1:50
post.b.a.n = 12*n0 + sumy.b
post.b.b.n = n.y.b + n0
post.exp.b = post.b.a.n / post.b.b.n
n0.plot.frame = data.frame(n0,post.exp.b)
plot.exp = ggplot(n0.plot.frame, aes(x= n0, y = post.exp.b))
plot.exp + geom_line() + geom_hline(yintercept = post.mean.a, color = "red")
```



It seems clear that n_0 would need to approach infinity for the posterior expectation of $\theta_B|Y_B$ to approach $E(\theta_A|Y_A)$

```
n0 = 1:300
post.b.a.n = 12*n0 + sumy.b
post.b.b.n = n.y.b + n0
post.exp.b = post.b.a.n / post.b.b.n
n0.plot.frame = data.frame(n0,post.exp.b)
plot.exp = ggplot(n0.plot.frame, aes(x= n0, y = post.exp.b))
plot.exp + geom_line() + geom_hline(yintercept = post.mean.a, color = "red")
```



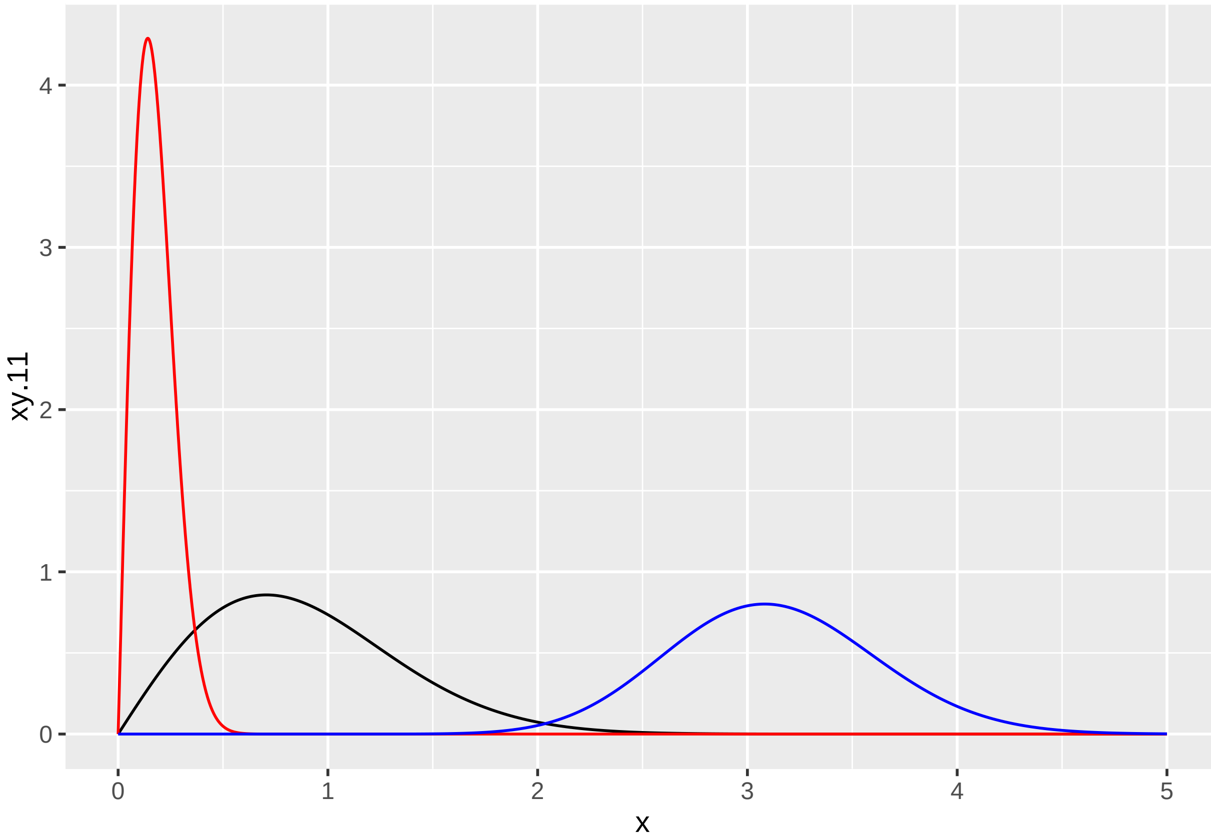
Part C

If knowledge about population A informs us about population B, then it doesn't make sense to have a prior $p(\theta_A, \theta_B) = p(\theta_A) \times p(\theta_B)$ because that prior operates under the assumption that they are independent. A informing B makes that assumption invalid.

Exercise 3.9

```
galenshore = function(y,a,theta){
  (2 / gamma(a)) * theta^(2*a) * y^(2*a - 1) * exp(-theta^2*y^2)
}

x = seq(0,5, by = .001)
xy.11 = galenshore(x,1,1)
xy.15 = galenshore(x,1,5)
xy.101 = galenshore(x,10,1)
galen.data.frame = data.frame(cbind(x,xy.11,xy.15,xy.101))
#curve(galenshore(x,1,1),ylim = c(0,4.5), xlim = c(0,5))
#curve(galenshore(x,1,5), add = TRUE, col = "red")
#curve(galenshore(x,10,5), add = TRUE, col = "blue")

f = ggplot(data = galen.data.frame, aes(x = x), ylim = c(0,5))
f + geom_line(aes(y = xy.11), color = "black") +
```

```r
  geom_line(aes(y = xy.15), color = "red") +
  geom_line(aes(y = xy.101), color = "blue")
```



Exercise 4.1

```r
n.2 = 50
sum.x.2 = 30
n.1 = 100
sum.x.1 = 57
alpha = 1
beta = 1
post.alpha.1 = sum.x.1 + alpha
post.beta.1 = n.1 - sum.x.1 + beta
post.alpha.2 = sum.x.2 + alpha
post.beta.2 = n.2 - sum.x.2 + beta
```

In this case, we assume a uniform prior distribution, which is equivalent to $\theta_2 \sim Ga(1, 1)$. The posterior distribution of $\theta_2 | \sum X \sim Ga$ (31,21).

```r
theta1 = rgamma(5000, post.alpha.1, post.beta.1)
theta2 = rgamma(5000, post.alpha.2, post.beta.2)

theta.both = mean(theta2 > theta1)
theta.both
```

```
## [1] 0.6956
```

Thus $Pr(\theta_A < \theta_B | y_A, y_B) = 0.6956$ in this case.

Exercise 4.2

Part A

Pretty Straightforward.

```
y.a = c(12,9,12,14,13,13,15,8,15,6)
y.b = c(11,11,10,9,9,8,7,10,6,8,8,9,7)

a.a = 120; a.b = 10
b.a = 12 ; b.b = 1
n.y.a = length(y.a)
n.y.b = length(y.b)
sumy.a = sum(y.a)
sumy.b = sum(y.b)
post.a.a = a.a + sumy.a
post.a.b = n.y.a + a.b

post.b.a = b.a + sumy.b
post.b.b = n.y.b + b.b

theta.a.draws = rgamma(10000,post.a.a,post.a.b)
theta.b.draws = rgamma(10000,post.b.a,post.b.b)

prob.diff = mean(theta.b.draws < theta.a.draws)
```

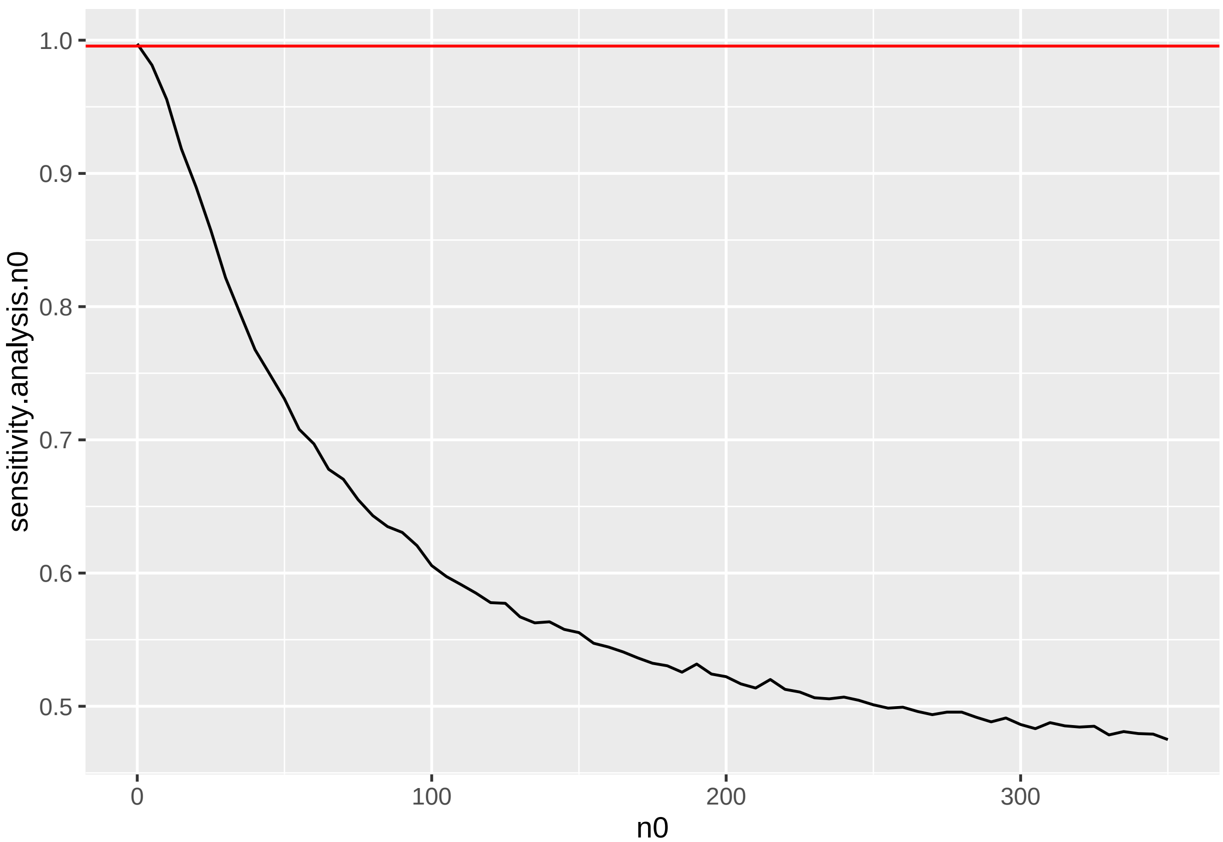
According to this, $P(\theta_B < \theta_A | y_B, y_A) =$ 'r prob.diff', which would indicate that we are quite sure that $\theta_B < \theta_A$.

Part B

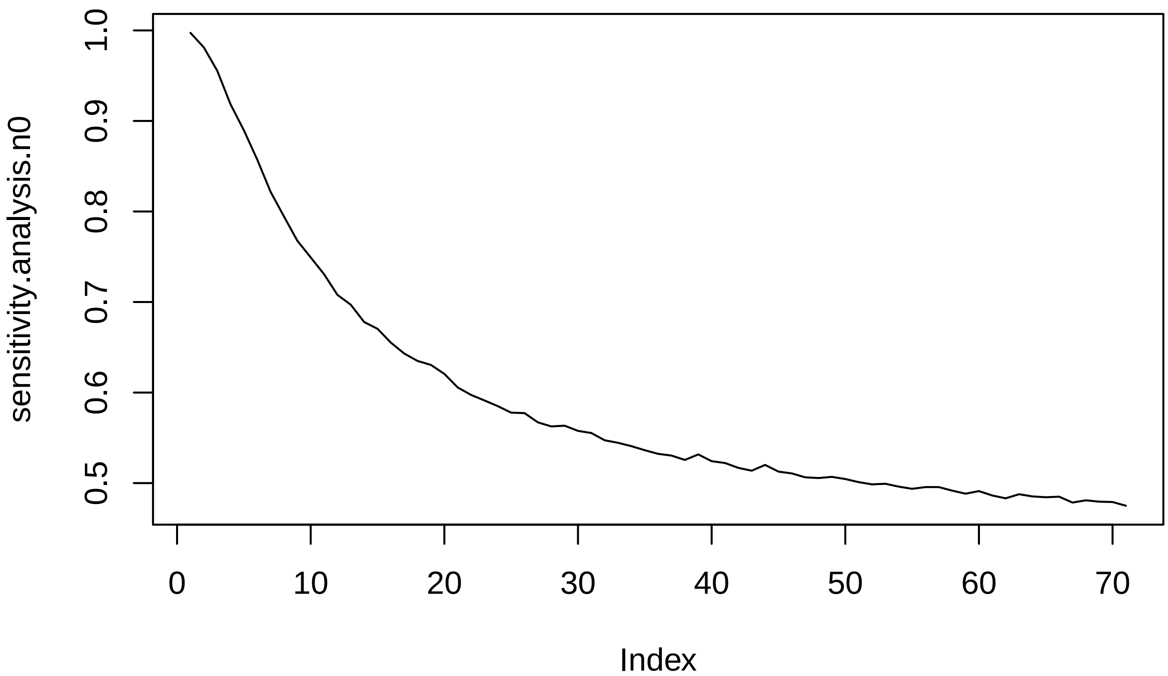
Sensitivity

```
n0 = seq(0,350, by = 5)
post.b.a.n = 12*n0 + sumy.b
post.b.b.n = n.y.b + n0

post.a.a = a.a + sumy.a
post.a.b = n.y.a + a.b

theta.a.draws = rgamma(10000,post.a.a,post.a.b)
theta.b.draws = rgamma(10000,post.b.a,post.b.b)

n.iter = 10000
# This vector is a vector of \theta_B < \theta_A | y_B, y_A for different levels of n_0
sensitivity.analysis.n0 = rep(NA, length(n0))
for(i in 1:length(n0)){
  post.b.a.val = post.b.a.n[i]
  post.b.b.val = post.b.b.n[i]
  theta.b.draws1 = rgamma(n.iter,post.b.a.val,post.b.b.val)
  sensitivity.analysis.n0[i] = mean(theta.b.draws1 < theta.a.draws)
```

```
}

sens.data.frame = as.data.frame(cbind(n0,sensitivity.analysis.n0))

ggplot(sens.data.frame, aes(x = n0)) + geom_line(aes(y = sensitivity.analysis.n0)) +
  geom_hline(yintercept = prob.diff,color = "red")
```




```
plot(sensitivity.analysis.n0, type = "l")
```



Part C.A

Posterior predictive

```
theta.a.draws = rgamma(10000,post.a.a,post.a.b)
theta.b.draws = rgamma(10000,post.b.a,post.b.b)

y.a.draws = rpois(10000,theta.a.draws)
y.b.draws = rpois(10000,theta.b.draws)

post.pred.diff = mean(y.b.draws < y.a.draws)
post.pred.diff
```

```
## [1] 0.6939
```

Through this, we find that $P(\widetilde{Y}_B < \widetilde{Y}_A | Y_A, Y_B) = 0.6939$.

Part C.B

Posterior predictive

```
n0 = seq(0,350, by = 5)
post.b.a.n = 12*n0 + sumy.b
post.b.b.n = n.y.b + n0

post.a.a = a.a + sumy.a
```

```r
post.a.b = n.y.a + a.b

theta.a.draws = rgamma(10000,post.a.a,post.a.b)
y.a.draws = rpois(10000,theta.a.draws)

n.iter = 10000
# This vector is a vector of \theta_B < \theta_A | y_B, y_A for different levels of n_0
sensitivity.pred.n0 = rep(NA, length(n0))
for(i in 1:length(n0)){
  post.b.a.val = post.b.a.n[i]
  post.b.b.val = post.b.b.n[i]
  theta.b.draws1 = rgamma(n.iter,post.b.a.val,post.b.b.val)
  y.b.draws1 = rpois(n.iter,theta.b.draws1)
  sensitivity.pred.n0[i] = mean(y.b.draws1 < y.a.draws)
}
sens.pred.data.frame = as.data.frame(cbind(n0,sensitivity.pred.n0))

ggplot(sens.pred.data.frame, aes(x = n0)) + geom_line(aes(y = sensitivity.pred.n0)) +
  geom_hline(yintercept = post.pred.diff,color = "red")
```