Pstat 115 HW2

Kalvin Goode

2018/10/20

(1)

(a)

We see that $E(A) = \frac{120}{10} = 12 = \frac{12}{1} = E(B)$. So we expect to have same average of incidence of cancer. However, we see that $Var(A) = \frac{120}{10^2} = 1.2 < 12 = \frac{12}{1^2} = Var(B)$. Therefore, we are more certain on A because it has lower variance.

(b)

```
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)

cat("Posterior distribution for A is gamma(",120+sum(y_A),",",10+length(y_A),")")
```

```
## Posterior distribution for A is gamma( 237 , 20 )
```

```
cat("Posterior distribution for B is gamma(",12+sum(y_B),",",1+length(y_B),")")
```

```
## Posterior distribution for B is gamma( 125 , 14 )
```

```
cat("Posterior mean for A is",(120+sum(y_A))/(10+length(y_A)))
```

```
## Posterior mean for A is 11.85
```

```
cat("Posterior mean for B is",(12+sum(y_B))/(1+length(y_B)))
```

```
## Posterior mean for B is 8.928571
```

```
cat("Posterior variance for A is",(120+sum(y_A))/(10+length(y_A))^2)
```

```
## Posterior variance for A is 0.5925
```

```
cat("Posterior variance for B is",(12+sum(y_B))/(1+length(y_B))^2)
```

```
## Posterior variance for B is 0.6377551
```

```
cat("The 95% quantile-based credible intervals for theta_A is (",qgamma(0.025,120+sum(y_A),10+length(y_
```

```
## The 95% quantile-based credible intervals for theta_A is ( 10.38924 , 13.40545 )
```

```
cat("The 95% quantile-based credible intervals for theta_B is (",qgamma(0.025,12+sum(y_B),1+length(y_B)
```

```
## The 95% quantile-based credible intervals for theta_B is ( 7.432064 , 10.56031 )
```

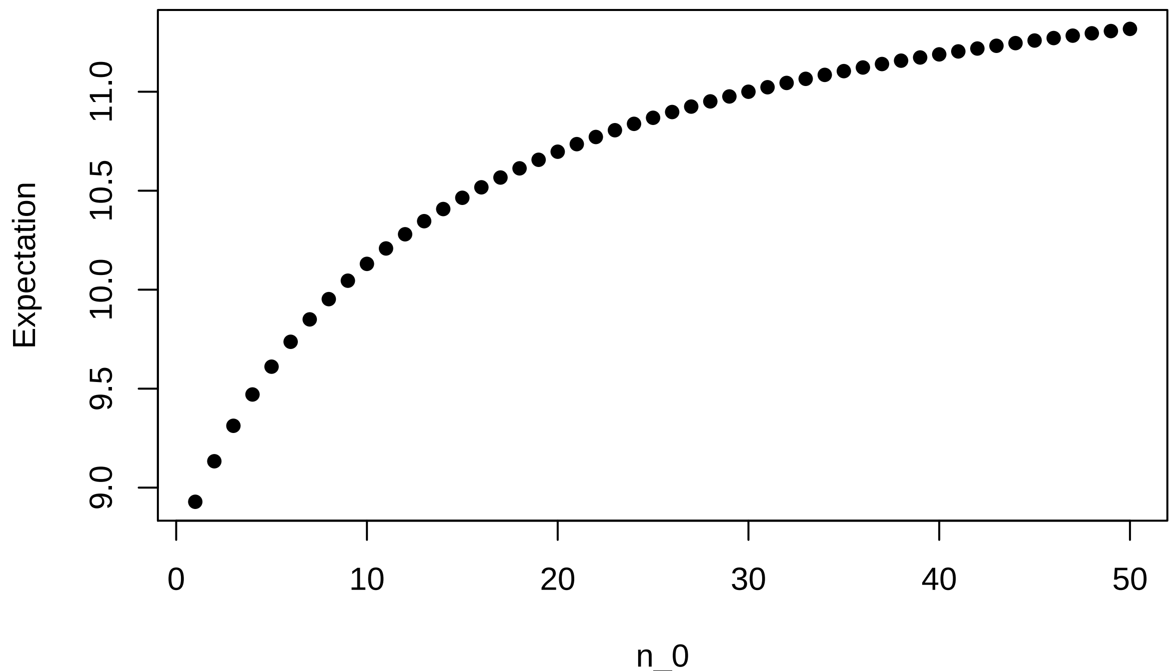
(c)

```
n_0=c(1:50)
a=12*n_0+sum(y_B)
b=n_0+length(y_B)
a/b
```

```
##  [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(a/b,xlab="n_0",ylab="Expectation",pch=16)
```



(d)

According to the 95% quantile-based credible intervals for θ_A and θ_B, we see that these two groups are not independent and are related. Therefore, it does not make sense to have $p(\theta_A), \theta_B) = p(\theta_A) \times p(\theta_B)$

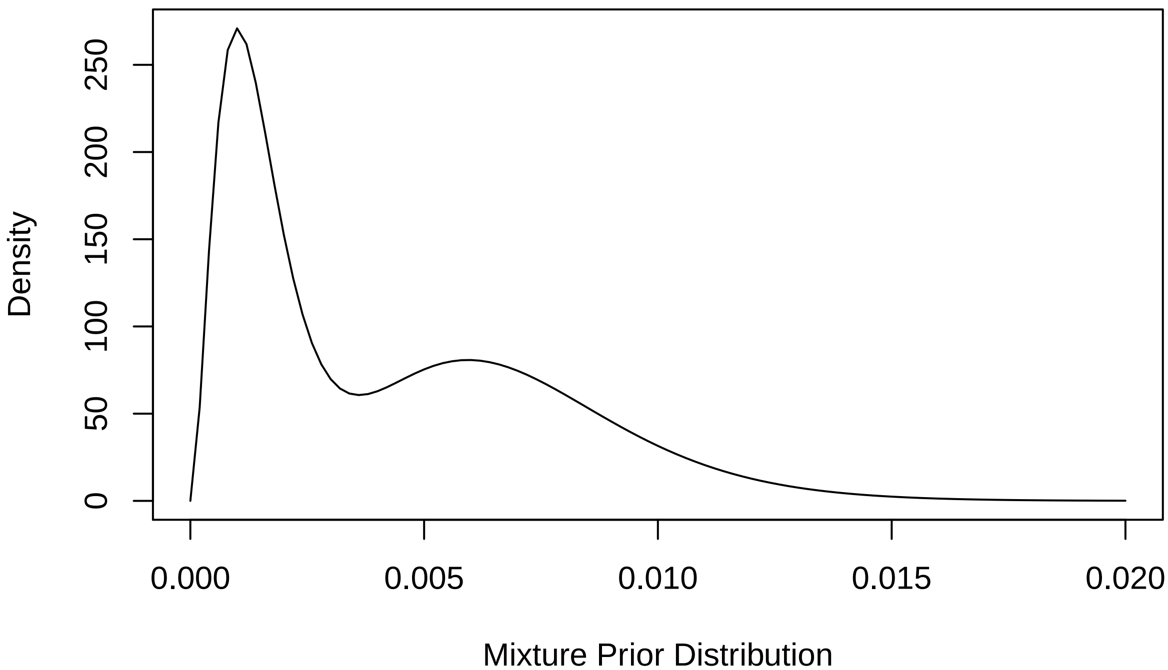
(2)

(a)

The first expert has mean=3/2000 and variance=$3/2000^2$, and second expert has mean=7/1000 and variance=$7/1000^2$. The first expert has more confidence because the vaiance is smaller than that of second expert.

(b)

```
curve((dgamma(x,3,2000)+dgamma(x,7,1000))/2,0,0.02,xlab="Mixture Prior Distribution", ylab="Density")
```



(c)

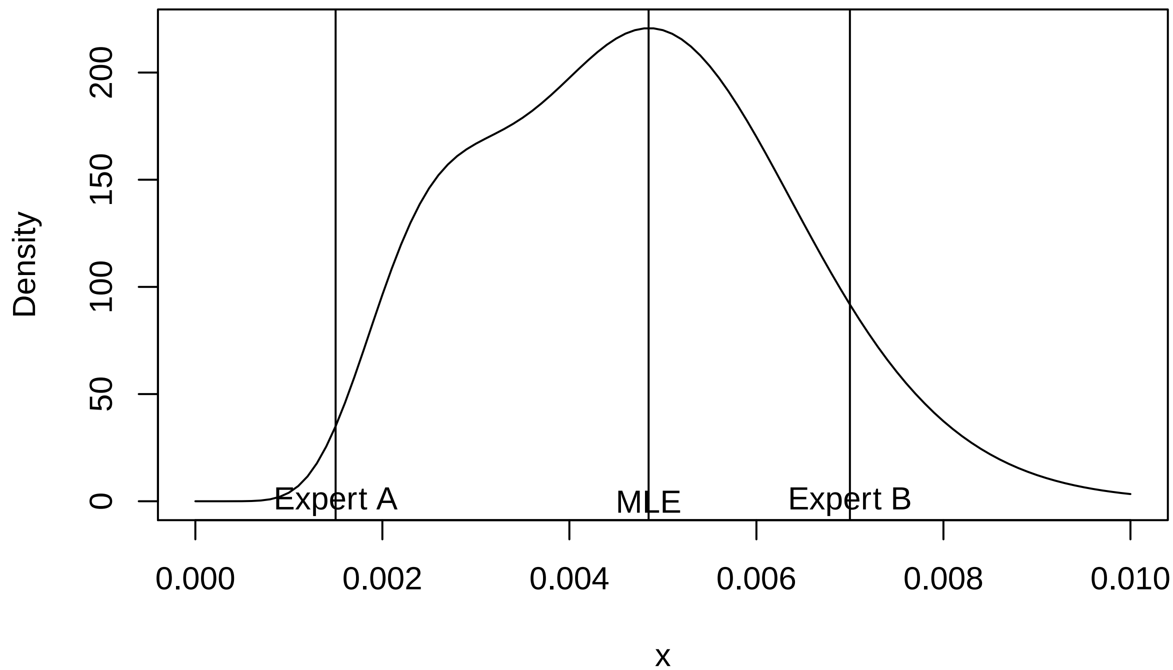
Let c_1 and c_2 be constants. By conjugation, we have $c_1 gamma(x;(3+8),(2000+1767)) = Lgamma(x;3,2000)$ and $c_2 gamma(x;(7+8),(1000+1767)) = Lgamma(x;7,1000)$. We know that $gamma(x;a,b) = \frac{x^{a-1}}{\Gamma(x)b^a}exp(\frac{-x}{b})$. Thus, $c_1 = \frac{11!\times 2000^3}{3767^{11}\times 3!}$ and $c_2 = \frac{15!\times 1000^7}{2767^{15}\times 7!}$ And the posterior distribution up to a proportionality is $\frac{1}{2}(c_1 gamma(x;11,3767) + c_2 gamma(x;15,2767))$

(d)

We have $p(\lambda|y) = \frac{L(\lambda;y)p(\lambda)}{K} = \frac{c_1 gamma(x;11,3767)+c_2 gamma(x;15,2767)}{c_1+c_2}$. Notice that, from (c), we have $\frac{c_1}{c_1+c_2} = 0.287$ and $\frac{c_1}{c_1+c_2} = 0.712$. Thus, $p(\lambda|y) \sim 0.287 gamma(11,3767) + 0.712 gamma(15,2767)$

(e)

```
y<-function(x)
{
  fx<-0.28766*dgamma(x,11,3767)+0.71233*dgamma(x,15,2767)
}
curve(0.28766*dgamma(x,11,3767)+0.71233*dgamma(x,15,2767),0,0.01,xlab="x",ylab="Density")
abline(v=3/2000)
abline(v=7/1000)
MLE=optimize(y,c(0,0.01),maximum=TRUE)
text(3/2000, 0, labels='Expert A')
text(7/1000, 0, labels='Expert B')
text(MLE, 0, labels='MLE')
abline(v=MLE)
```




```r
cat("The MLE is approximately")
```

```
## The MLE is approximately
```

```r
MLE
```

```
## $maximum
## [1] 0.004847024
##
## $objective
## [1] 220.7259
```