Lab 1 JAE Bayes

Roi Naveiro y David Ríos Insua

22/06/2021

Birth rates

We compare number of children of 40 year old women with and without college degree.

```
load("data/gss.RData")
```

We show empirical distributions of both groups.

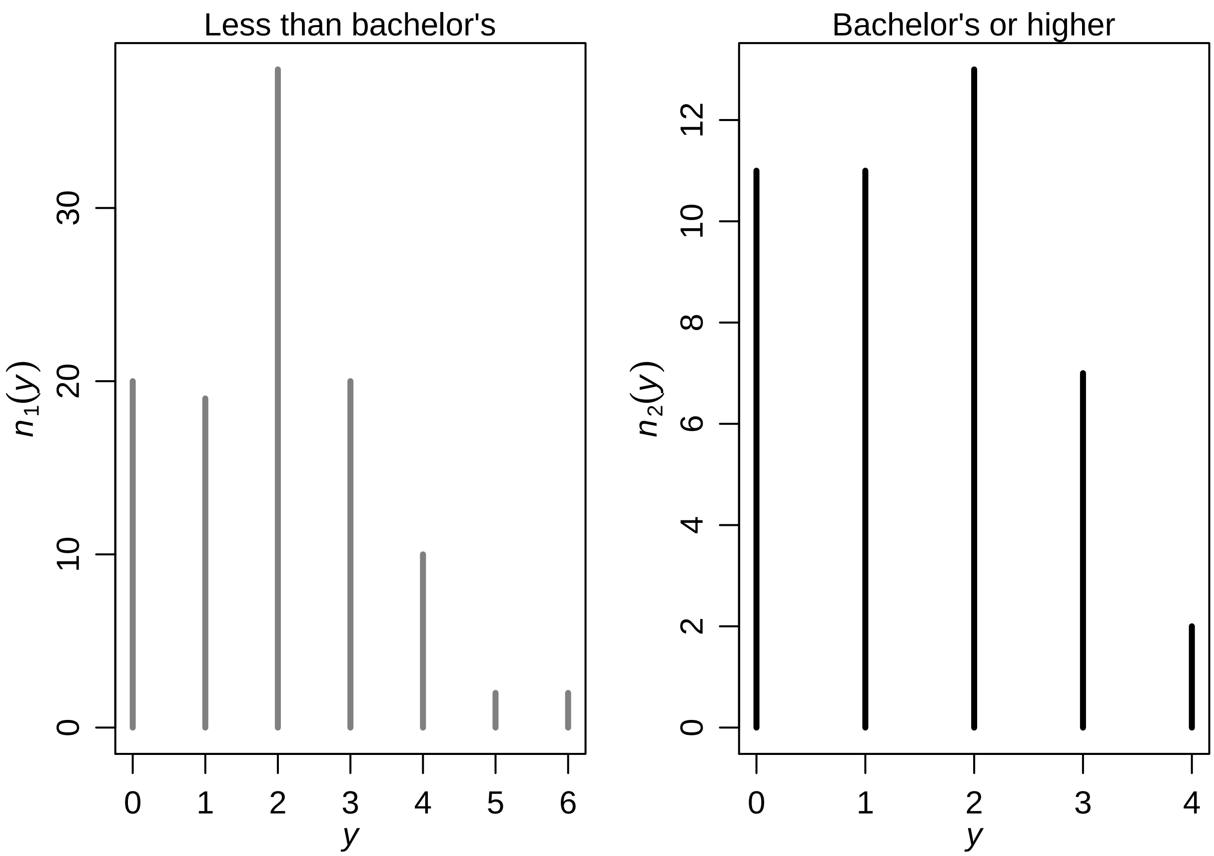
```
#### Fig 3.9
y2<-gss$CHILDS[gss$FEMALE==1 &  gss$YEAR>=1990  & gss$AGE==40 & gss$DEG>=3 ]
y1<-gss$CHILDS[gss$FEMALE==1 &  gss$YEAR>=1990  & gss$AGE==40 & gss$DEG<3 ]

y2<-y2[!is.na(y2)]
y1<-y1[!is.na(y1)]

par(mar=c(3,3,1,1),mgp=c(1.75,.75,0))
par(mfrow=c(1,2))

set.seed(1)
n1<-length(y1) ; n2<-length(y2)
s1<-sum(y1)
s2<-sum(y2)

par(mfrow=c(1,2),mar=c(3,3,1,1),mgp=c(1.75,.75,0))
plot(table(y1), type="h",xlab=expression(italic(y)),ylab=expression(italic(n[1](y))),col=gray(.5) ,lwd=3
mtext("Less than bachelor's",side=3)
plot(table(y2), type="h",xlab=expression(italic(y)),ylab=expression(italic(n[2](y))),col=gray(0),lwd=3)
mtext("Bachelor's or higher",side=3,lwd=3)
```



We use a poisson sampling model for each group

$$Y_{1,1} \ldots, Y_{n_1,1} \sim \text{Poisson}(\theta_1)$$

$$Y_{1,2} \ldots, Y_{n_2,2} \sim \text{Poisson}(\theta_2)$$

and a common gamma prior $\theta_1, \theta_2 \sim \text{gamma}(a = 2, b = 1)$. We can calculate posterior means, modes, and intervals

```
a <- 2 ; b <- 1
n1 <- 111 ; sy1 <- 217
n2 <- 44 ;  sy2 <- 66

print( paste0( "Posterior mean women without college degree: ", ( a+sy1 ) / ( b+n1 ) ))
```

```
## [1] "Posterior mean women without college degree: 1.95535714285714"
```

```
print( paste0( "Posterior mode women without college degree: ", ( a+sy1 - 1)/(b+n1 ) ))
```

```
## [1] "Posterior mode women without college degree: 1.94642857142857"
```

```
print( "Posterior 95% CI women without college degree: " )
```

```
## [1] "Posterior 95% CI women without college degree: "
```

```
print( qgamma( c ( 0.025 , 0.975 ) , a+sy1 , b+n1 ))
```

```
## [1] 1.704943 2.222679
```

```r
print( paste0( "Posterior mean women with college degree: ", ( a+sy2 ) / ( b+n2 ) ))
```

```
## [1] "Posterior mean women with college degree: 1.51111111111111"
```

```r
print(paste0( "Posterior mode women with college degree: ", ( a+sy2 - 1)/(b+n2 ) ))
```

```
## [1] "Posterior mode women with college degree: 1.48888888888889"
```

```r
print( "Posterior 95% CI women with college degree: " )
```

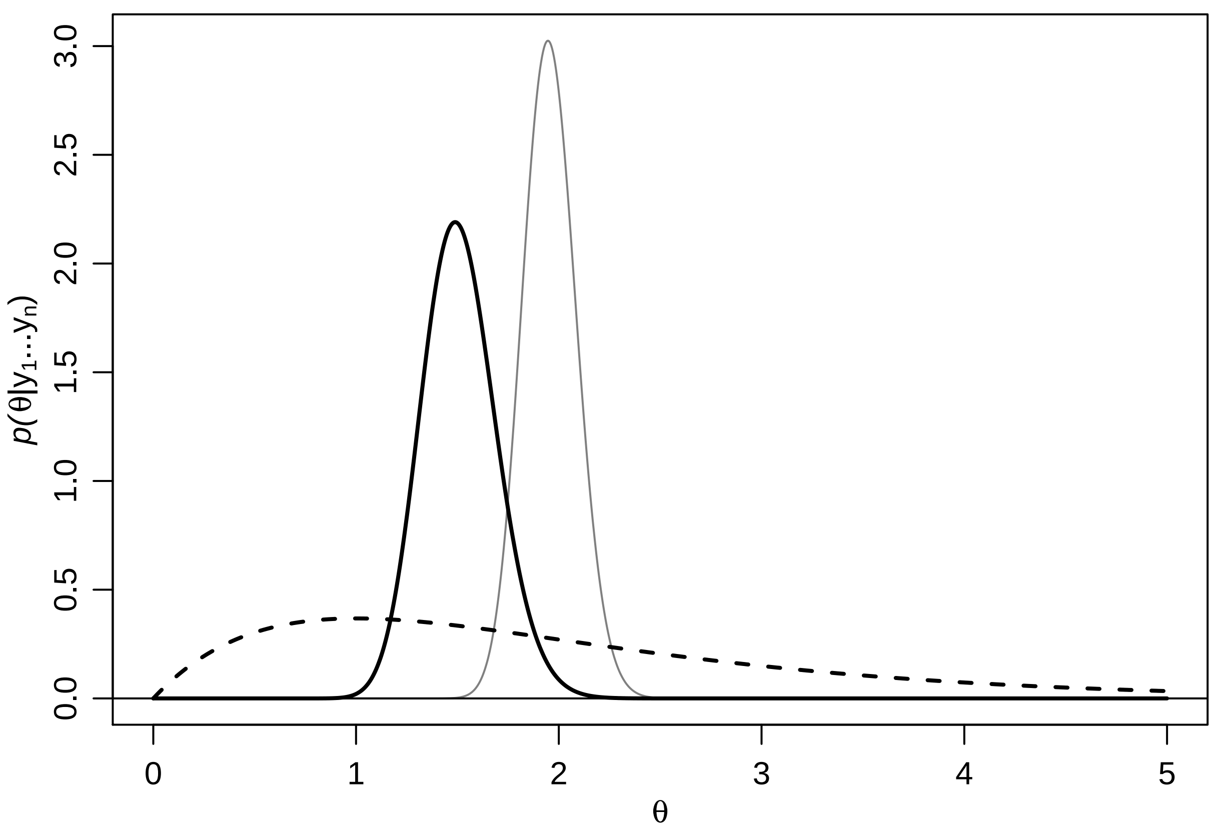
```
## [1] "Posterior 95% CI women with college degree: "
```

```r
print( qgamma( c ( 0.025 , 0.975 ) , a+sy2 , b+n2 ))
```

```
## [1] 1.173437 1.890836
```

Posterior densities

```r
par(mar=c(3,3,1,1),mgp=c(1.75,.75,0))
#par(mfrow=c(1,2))
a<-2
b<-1
xtheta<-seq(0,5,length=1000)
plot(xtheta,dgamma(xtheta,a+s1,b+n1),type="l",col=gray(.5),xlab=expression(theta),
  ylab=expression(paste(italic("p("),theta,"|",y[1],"...",y[n],")",sep="")))
lines(xtheta,dgamma(xtheta,a+s2,b+n2),col=gray(0),lwd=2)
lines(xtheta,dgamma(xtheta,a,b),type="l",lty=2,lwd=2)
abline(h=0,col="black")
```



Strong evidence for $\theta_1 > \theta_2$. What is the posterior probability of $\theta_1 > \theta_2$?

```
theta_1_samples <- rgamma(100000, a+sy1, b+n1)
theta_2_samples <- rgamma(100000, a+sy2, b+n2)

mean(theta_1_samples > theta_2_samples)
```

```
## [1] 0.97278
```

The predictive is neg binomial

```
y<-(0:12)
plot(y-.1, dnbinom(y, size=(a+s1), mu=(a+s1)/(b+n1)) , col=gray(.5) ,type="h",
ylab=expression(paste(italic("p("),y[n+1],"|",y[1],"...",y[n],")",sep="")),
   xlab=expression(italic(y[n+1])),ylim=c(0,.35),lwd=3)
points(y+.1, dnbinom(y, size=(a+s2), mu=(a+s2)/(b+n2)) , col=gray(0) ,type="h",lwd=3)
legend(1,.375,legend=c("Less than bachelor's","Bachelor's or higher"),bty="n",
  lwd=c(3,3),col=c(gray(.5),gray(0)))
```