Assignment 2

Terry Liu 1630005038

```
#setup
library(LearnBayes)
library(lattice)
```

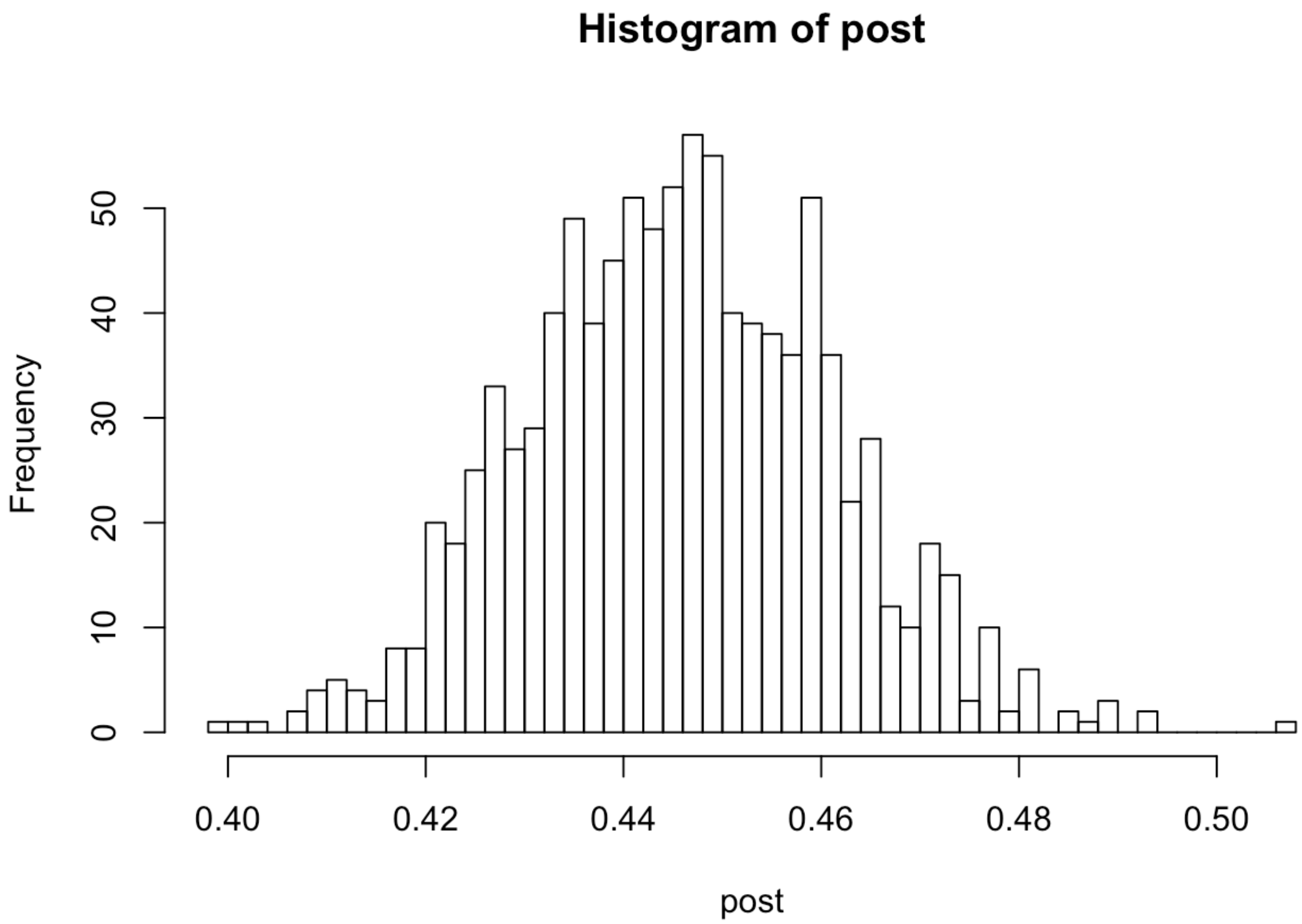
Question One

Reproduce the results of example Probability of a girl birth given placenta previa in page 37 of textbook by R

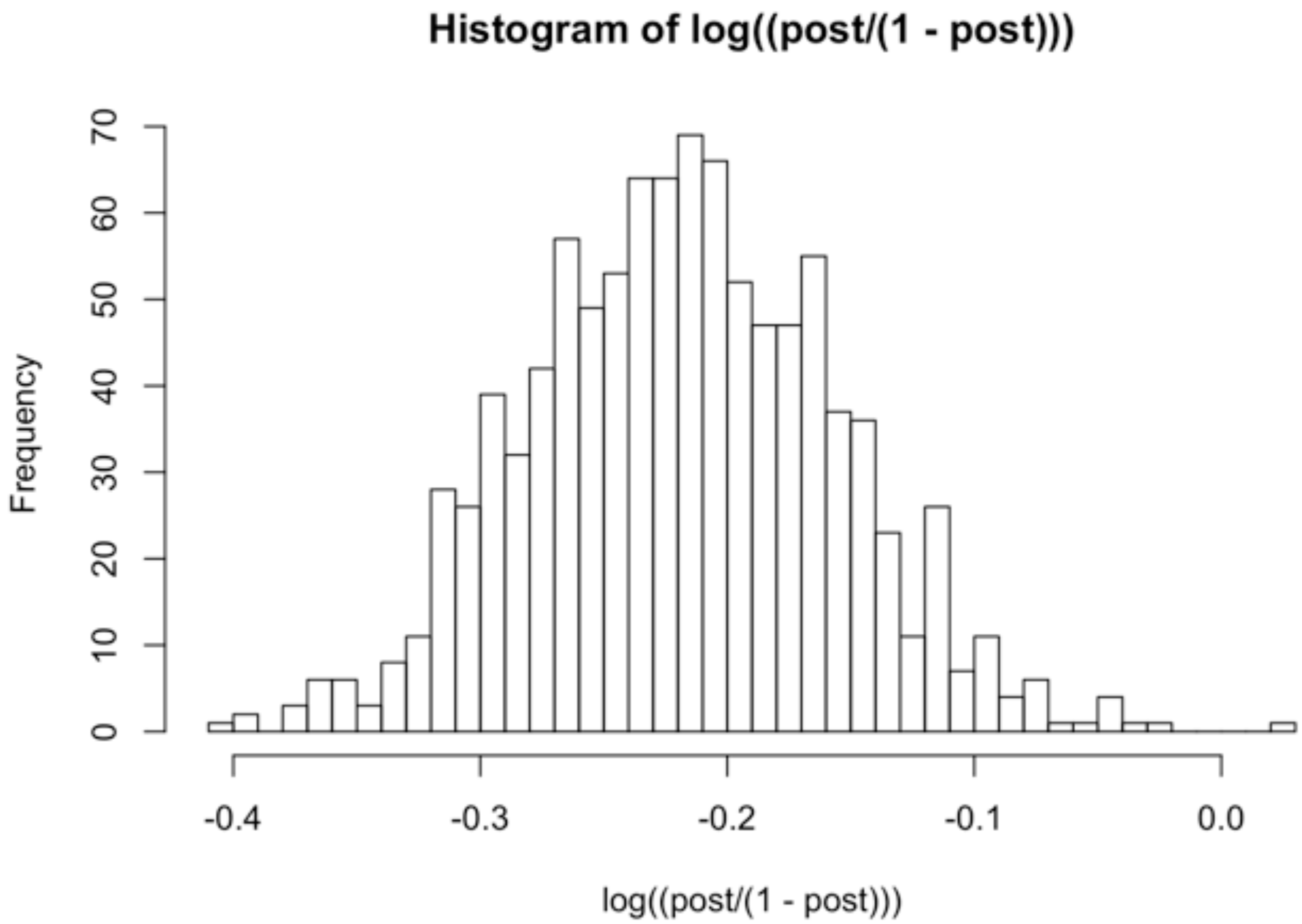
```
#Question One
par(mfrow<-c(1,3))
```

```
## NULL
```

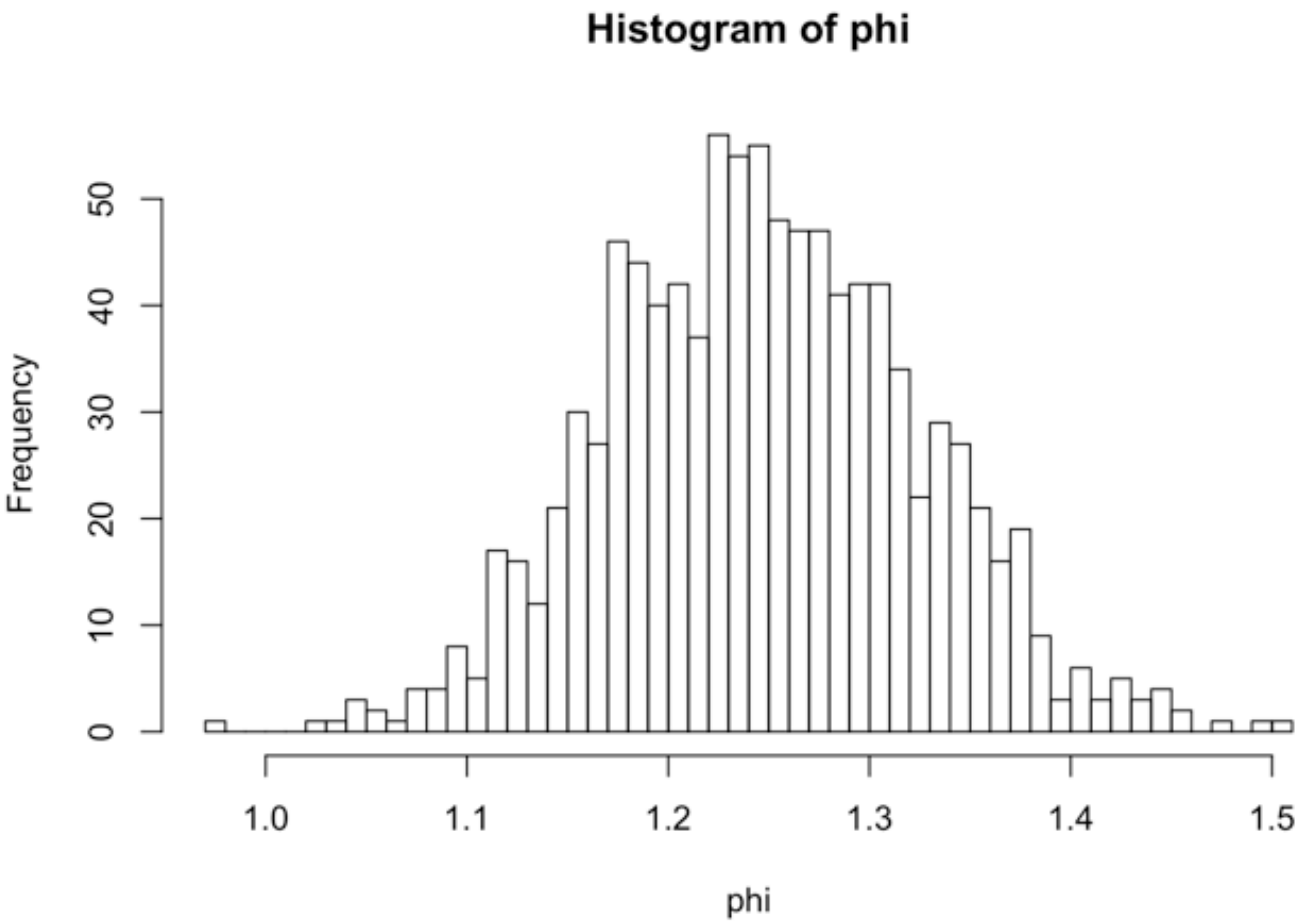
```
a <- 438
b <- 544
post <- rbeta(1000,a,b)
hist(post,breaks<-50)
```




```
hist(log((post/(1-post))),breaks<-50)
```




```
phi <- (1-post)/post
hist(phi,breaks<-50)
```




```
E <- 0.5
SUM <- 2
a <- E*SUM
b <- SUM-a
s <- 438
f <- 544
poster <- rbeta(1000,a+s,b+f)
quantile(poster,c(0.5,0.5))
```

```
##       50%       50%
## 0.4469159 0.4469159
```

```
quantile(poster,c(0.025,0.975))
```

```
##      2.5%     97.5%
## 0.4157754 0.4772843
```

Question Two

Discrete data: Table 2.2 gives the number of fatal accidents and deaths on scheduled airline flights per year over a ten-year period. We use these data as a numerical example for fitting discrete data models.

a. Assume that the numbers of fatal accidents in each year are independent with a Poisson(θ) distribution. Set a prior distribution for θ and determine the posterior distribution based on the data from 1976 through 1985. Under this model, give a 95% predictive interval for the number of fatal accidents in 1986. You can use the normal approximation to the gamma and Poisson or compute using simulation.

b. Assume that the numbers of fatal accidents in each year follow independent Poisson distributions with a constant rate and an exposure in each year proportional to the number of passenger miles flown. Set a prior distribution for θ and determine the posterior distribution based on the data for 1976–1985. (Estimate the number of passenger miles flown in each year by dividing the appropriate columns of Table 2.2 and ignoring round-off errors.) Give a 95% predictive interval for the number of fatal accidents in 1986 under the assumption that 8 × 1011 passenger miles are flown that year.

c. Repeat (a) above, replacing ‘fatal accidents’ with ‘passenger deaths.’

d. Repeat (b) above, replacing ‘fatal accidents’ with ‘passenger deaths.’

e. In which of the cases (a)–(d) above does the Poisson model seem more or less reasonable? Why? Discuss based on general principles, without specific reference to the numbers in Table 2.2. Incidentally, in 1986, there were 22 fatal accidents, 546 passenger deaths, and a death rate of 0.06 per 100 million miles flown. We return to this example in Exercises 3.12, 6.2, 6.3, and 8.14.

Part a

```
#Part a
y <- c(24,25,31,31,22,21,26,20,16,22)
ybar <-mean(y)
n <- 10
a <- 0
b <- 0
theta <- rgamma(1000,n*ybar+a,n+b)
yhat <- rpois(1000,theta)
quantile(yhat,c(0.025,0.975))
```

```
##  2.5% 97.5%
##    15    35
```

```
sort(yhat)[c(25,976)]
```

```
## [1] 15 35
```

```
GE <- ybar
GVar <- sqrt(ybar*n)/n
E <- GE
Var <- GE+GVar**2
Sd <- sqrt(Var)
round(c(E-1.96*Sd,E+1.96*Sd))
```

```
## [1] 14 34
```

Part b

```
Pd <- c(734,516,754,877,814,362,764,809,223,1066)
Dr <- c(0.19,0.12,0.15,0.16,0.14,0.06,0.13,0.13,0.03,0.15)
z <- (Pd/Dr)*100000000
zbar <- mean(z)
n <- 10
a <- 0
b <- 0
theta2 <- rgamma(1000,n*ybar,n*zbar)
yhat2 <- rpois(1000,theta2*8e11)
quantile(yhat2,c(0.025,0.975))
```

```
##  2.5% 97.5%
##    22    47
```

```
sort(yhat2)[c(25,976)]
```

```
## [1] 22 47
```

```
GE2 <- (n*ybar)/(zbar*n)
GVar2 <- (ybar*n)/((zbar*n)**2)
E2 <- 8e11*GE2
Var2 <- 8e11*GE2+(8e11**2)*(GVar2**2)
Sd2 <- sqrt(Var2)
round(c(E2-1.96*Sd2,E2+1.96*Sd2))
```

```
## [1] 22 45
```

Part c

```
Pd<-c(734,516,754,877,814,362,764,809,223,1066)
xbar<-mean(Pd)
n<-10
a<-0
b<-0
theta<-rgamma(1000,n*xbar+a,n+b)
xhat<-rpois(1000,theta)
round(quantile(xhat,c(0.025,0.975)))
```

```
##  2.5% 97.5%
##   638   747
```

```
sort(xhat)[c(25,976)]
```

```
## [1] 638 748
```

```
GE3<-xbar
GVar3<-sqrt(xbar*n)/n
E3<-GE3
Var3<-GE3+GVar3**2
Sd3<-sqrt(Var3)
round(c(E3-1.96*Sd3,E3+1.96*Sd3))
```

```
## [1] 638 746
```

Part d

```
Pd< - c(734,516,754,877,814,362,764,809,223,1066)
```

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

```
Dr <- c(0.19,0.12,0.15,0.16,0.14,0.06,0.13,0.13,0.03,0.15)
w <- (Pd/Dr)*100000000
wbar <- mean(w)
n <- 10
a <- 0
b <- 0
theta4 <- rgamma(1000,n*xbar,n*wbar)
what<- rpois(1000,theta4*8e11)
quantile(what,c(0.025,0.975))
```

```
##  2.5% 97.5%
##   903  1034
```

```
sort(what)[c(25,976)]
```

```
## [1]  903 1034
```

```
GE4 <- (n*xbar)/(wbar*n)
GVar4 <- (xbar*n)/((wbar*n)**2)
E4 <- 8e11*GE4
Var4 <- 8e11*GE4+(8e11**2)*(GVar4**2)
Sd4 <- sqrt(Var4)
round(c(E4-1.96*Sd4,E4+1.96*Sd4))
```

```
## [1]  907 1029
```