Key2

Jon Wakefield

2022-07-20

Bayes for Genetics 2021 SISG

Prob theory and Binomial Sampling (Lecture 2)

Q1

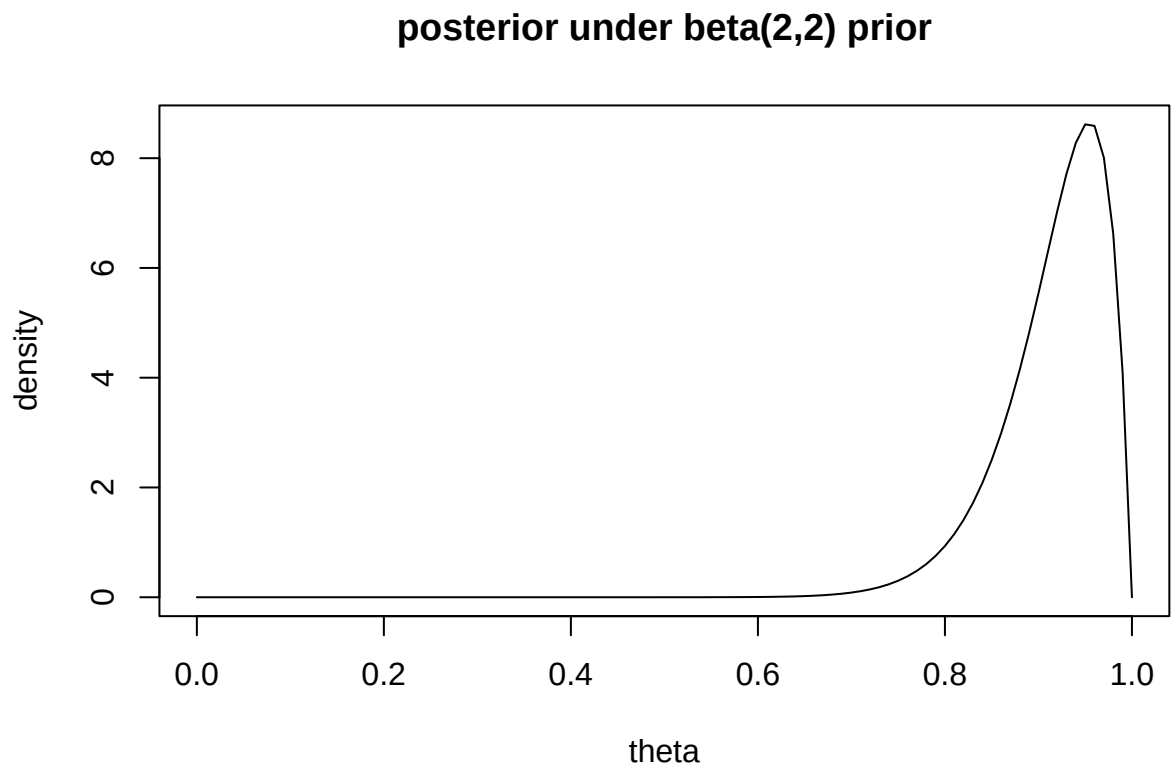
Suppose we observe data with N = 20, y = 20 and we assume a binomial likelihood with probability theta.

- What are the MLE and standard error of the MLE?

$$MLE = \hat{\theta} = 20/20 = 1 \text{ and estimated } se = 0$$

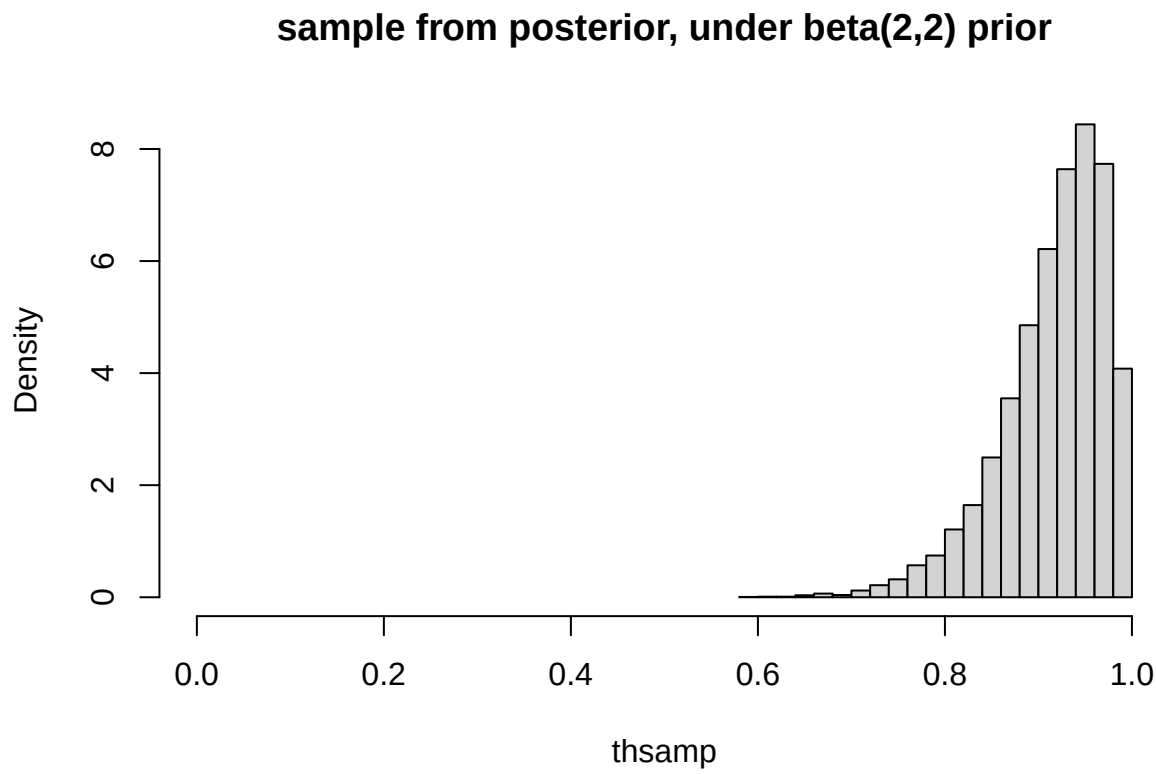
- Plot the posterior distribution under a Beta(2,2) prior

```
y <- n <- 20
a <- b <- 2
theta <- seq(0,1,.01)
plot(dbeta(theta,y+a,n-y+b)~theta,type="l", ylab="density", main="posterior under beta(2,2) prior")
```



- Simulate from the posterior and draw a histogram of the posterior samples.

```
set.seed(4)
thsamp <- rbeta(10000,y+a,n-y+b)
hist(thsamp,xlim=c(0,1), main="sample from posterior, under beta(2,2) prior", freq=FALSE)
```



- What is the posterior median, and give a 90% credible interval for θ: evaluate these quantities in two ways, one via the `qbeta` function and the other via sampling.

```
qbeta(p=c(0.05,.5,.95),y+a,n-y+b)  # exact answer
```

```
## [1] 0.8097960 0.9280943 0.9843259
```

```
quantile(thsamp,p=c(0.05,.5,0.95)) # using posterior sample
```

```
##        5%       50%       95%
## 0.8066491 0.9277860 0.9846921
```

- What is the MLE for the odds $\theta/(1-\theta)$?

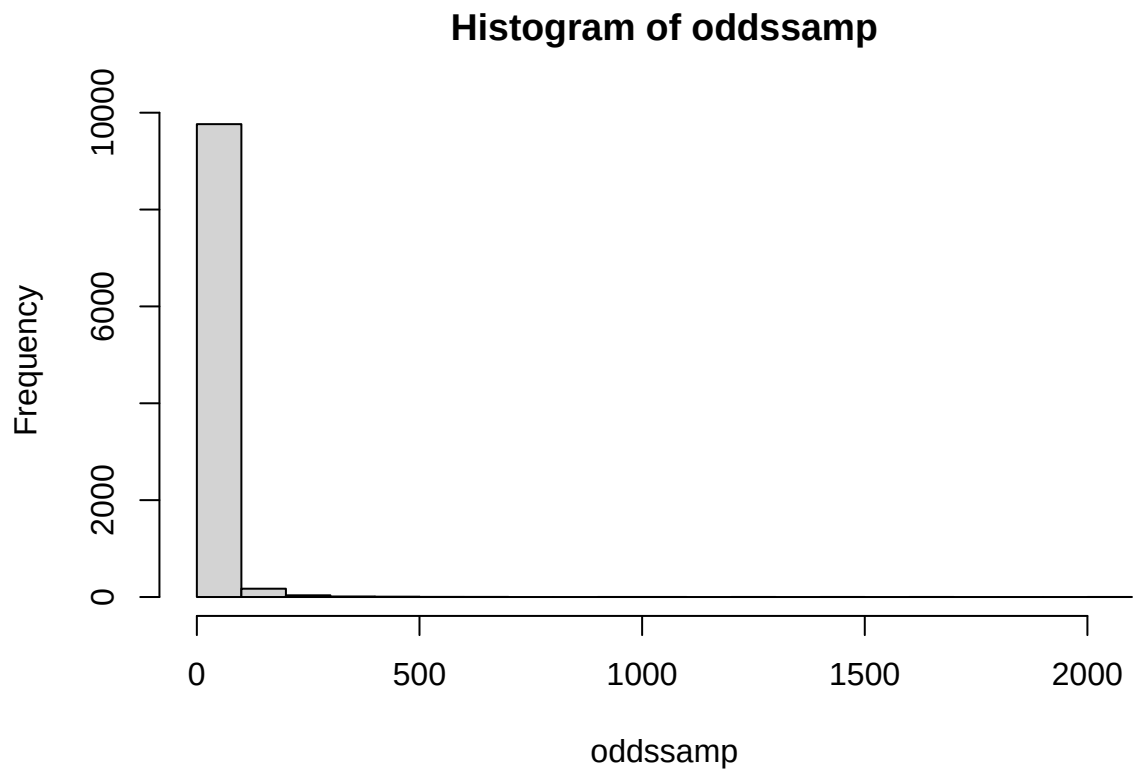
MLE for odds is undefined

- What is the posterior probability that the odds are greater than 100?

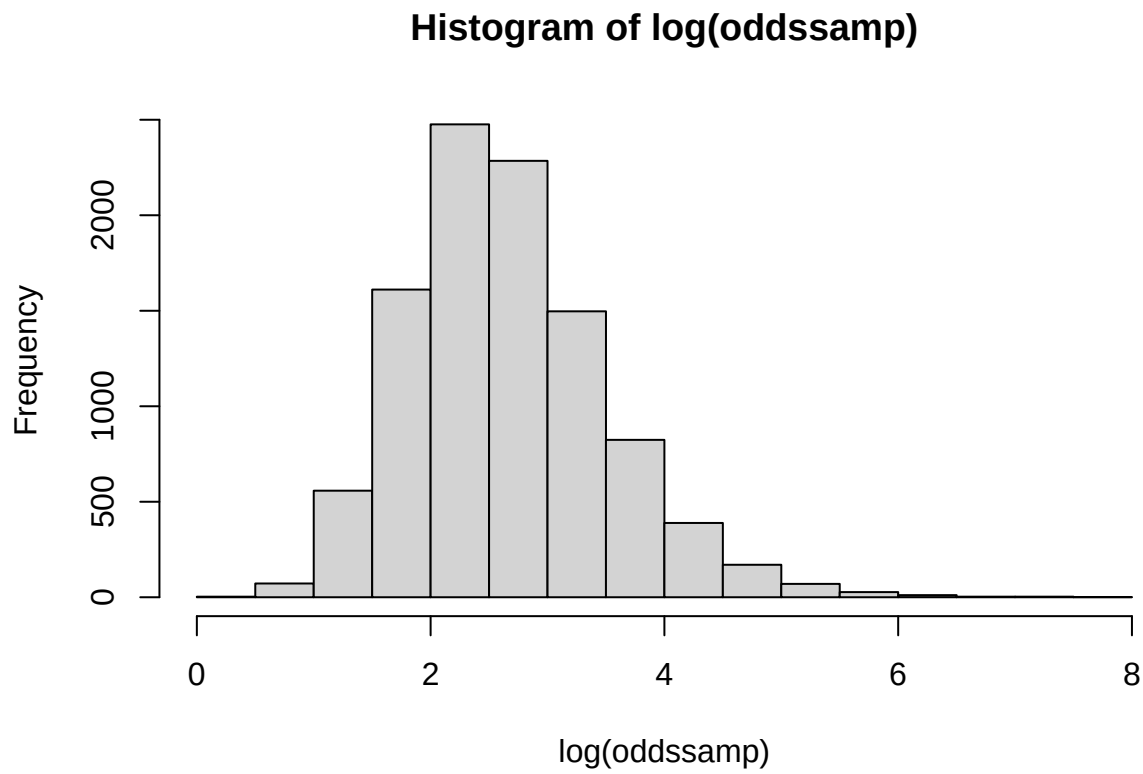
```
oddssamp <- thsamp/(1-thsamp)
mean(oddssamp>100) # proportion of sample with value > 100
```

```
## [1] 0.0236
```

```r
hist(oddssamp) # long right tail!
```



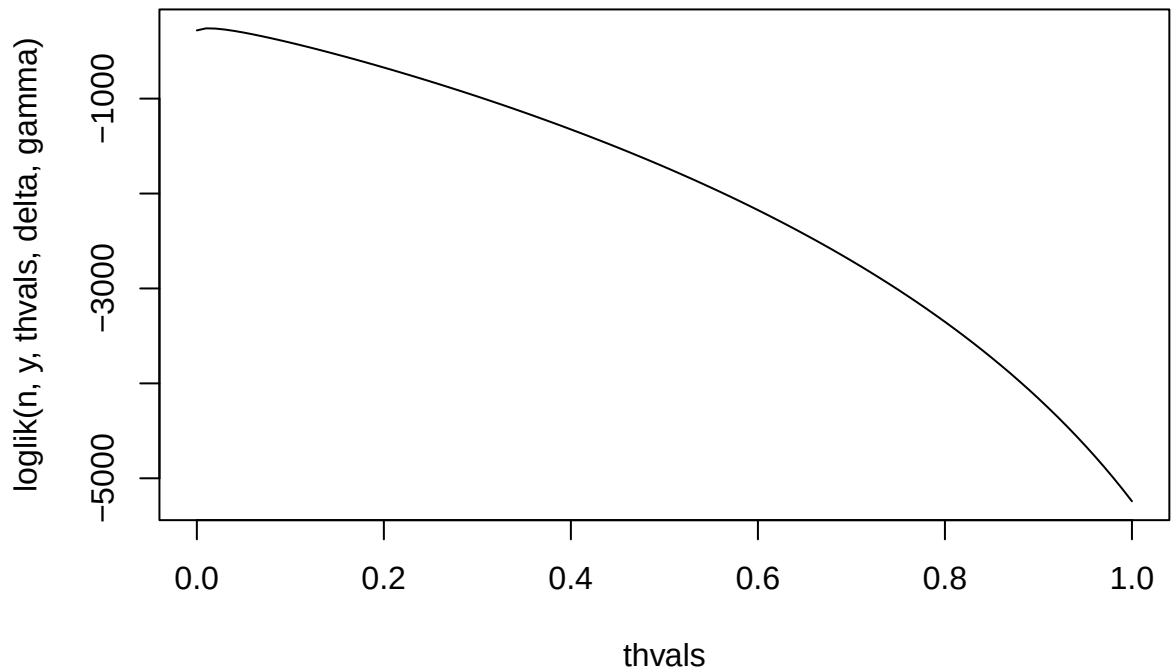

```r
hist(log(oddssamp))
```



#Q2 ##Redo the seroprevalence example with a Beta(1,9) prior on the prevalence.

```
y <- 50; n <- 3300
delta <- 0.8; gamma <- 0.995
A <- delta + gamma - 1; B <- 1-gamma
MLE <- (y - n*B)/(n*A)
loglik <- function(n,y,prev,delta,gamma){
  A <- delta + gamma - 1; B <- 1-gamma
  p <- prev*A + B
  loglik <- y*log(p) + (n-y)*log(1-p)
  loglik
}
maxl <- loglik(n,y,MLE,delta,gamma)
nsim <- 1000
success <- 0
a <- 1
b <- 9

thvals <- seq(0,1,l=101)
plot( loglik(n,y,thvals,delta, gamma)~thvals, type="l")
```




```
post <- rep(NA, nsim)
set.seed(4)
while(success<nsim+1){
  U <- runif(1); theta <- rbeta(1,a,b)
  test <- loglik(n,y,theta,delta,gamma)
  if (log(U) < test - maxl) {
    success <- success + 1
    post[success] <- theta
  }
}
mean(post)
```

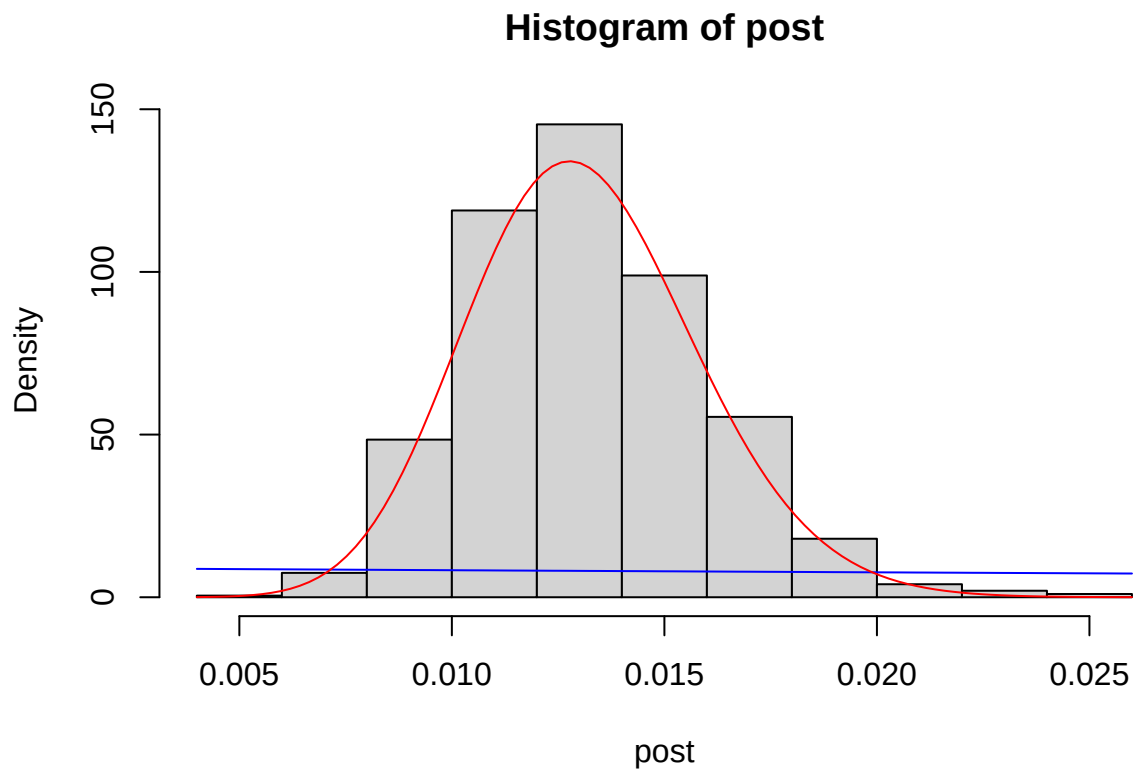
```
## [1] 0.01320074
```

```
quantile(post,p=c(0.05,0.5,0.95))
```

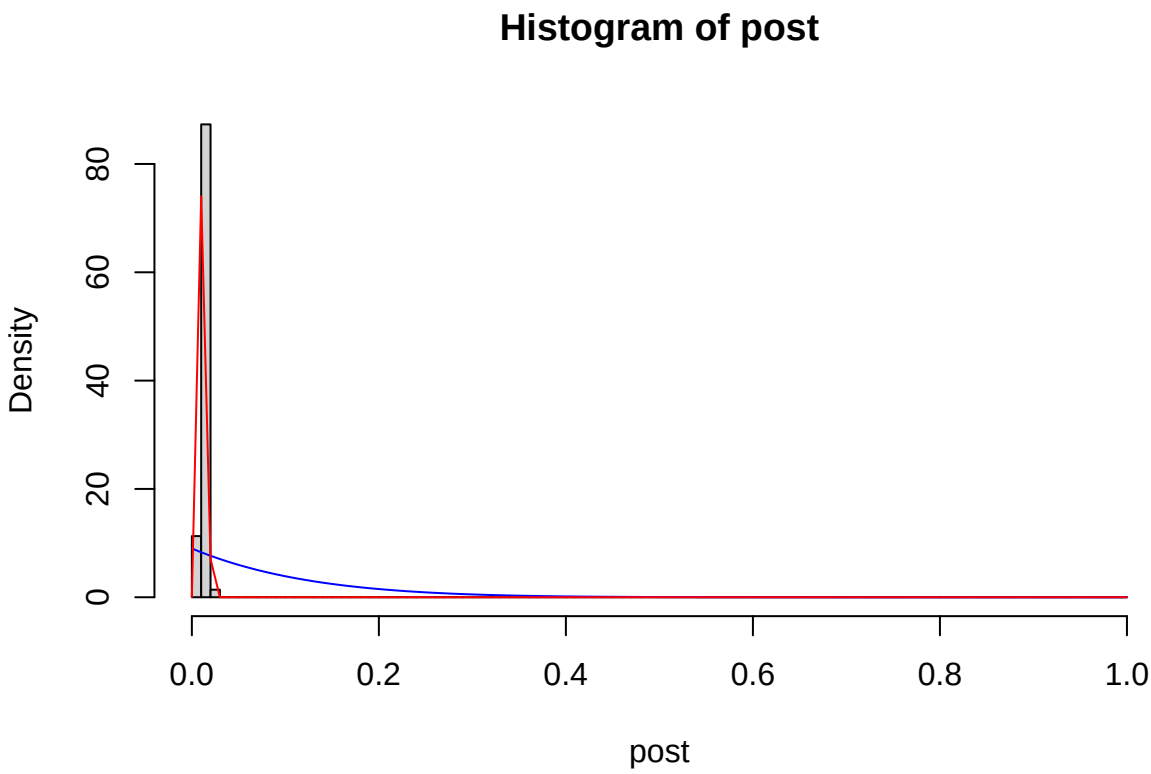
```
##         5%        50%        95%
## 0.00899876 0.01302424 0.01791566
```

```
hist(post, freq=FALSE)

# adding prior and (scaled) likelihood
hist(post, freq=FALSE)
curve(dbeta(x,a,b), col="blue", add=TRUE)
curve( exp(loglik(n,y,x,delta, gamma)+264), add=TRUE, col="red")
```




```
# adding prior and (scaled) likelihood, full range
hist(post, freq=FALSE, breaks=seq(0,1,l=101))
curve(dbeta(x,a,b), col="blue", add=TRUE)
curve( exp(loglik(n,y,x,delta, gamma)+264), add=TRUE, col="red")
```


Histogram of post
Density
0
20
40
60
80
0.0
0.2
0.4
0.6
0.8
1.0
post