

SISG Bayes: Key 3

Ken Rice

2026-05-26

Binomial Sampling 2 (Lecture 3)

Q1

Experiment with the priors $\text{Beta}(a, a)$ for the ASE example. In particular, for $a = 2$:

- Obtain a histogram of the posterior probabilities $\text{Pr}(\theta < 0.5|y)$, across genes.

```
ASEdat <- read.table("http://faculty.washington.edu/kenrice/sisgbayes/ASEgene.txt",header=TRUE)
head(ASEdat)
```

```
##      Y      N
## 1   62   107
## 2   33    59
## 3  658  1550
## 4   14    61
## 5   57   153
## 6  218   451
```

```
dim(ASEdat)
```

```
## [1] 4844    2
```

```
ngenes <- nrow(ASEdat)
ASEa <- 2
postprob1 <- postprob2 <- nlogBFR1 <- nlogBFR2 <- pvals <- rep(NA,ngenes)

# non-bayesian p-values (exact test)
for (i in 1:ngenes){
  pvals[i] <- binom.test(ASEdat$Y[i],ASEdat$N[i],
                        p=0.5,alternative="two.sided")$["p.value"]
}

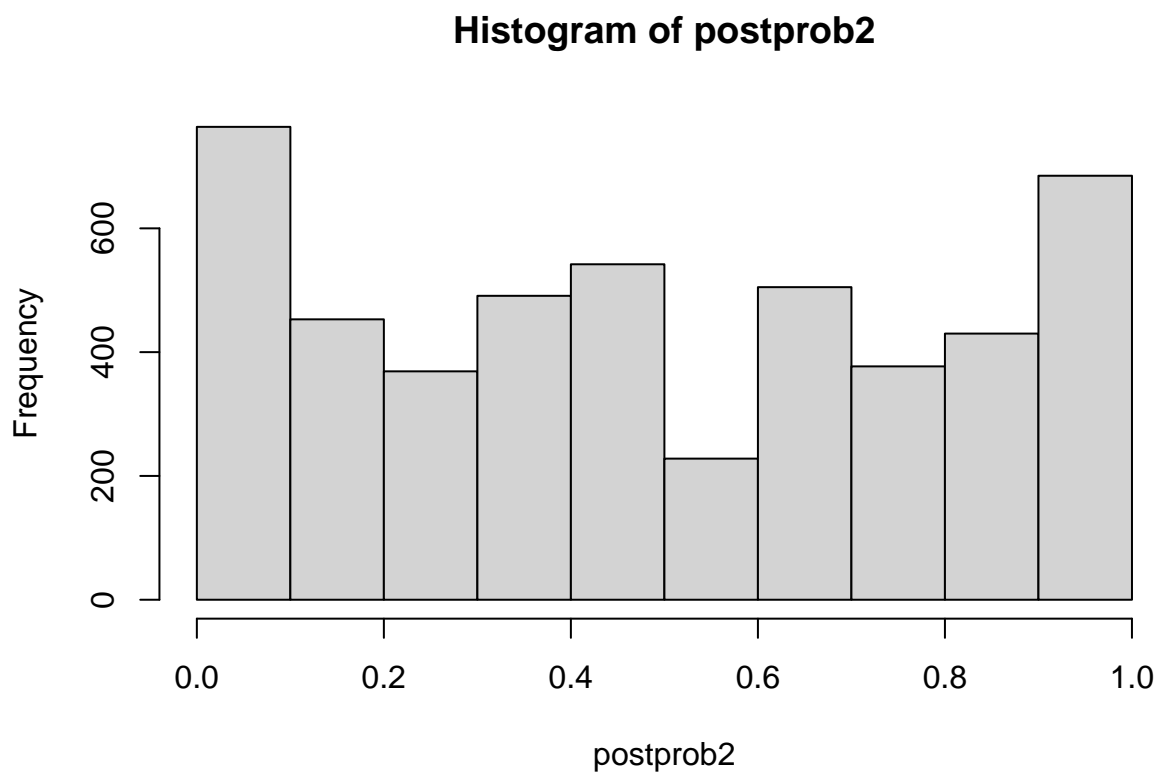
# BF calculations
BFbinomial <- function(N,y,a,b,p0){
  logPrH0 <- lchoose(N,y) + y*log(p0) + (N-y)*log(1-p0)
  logPrH1 <- lchoose(N,y) + lgamma(a+b) - lgamma(a) - lgamma(b) + lgamma(y+a) + lgamma(N-y+b) - lgamma(N)
  logBF <- logPrH0 - logPrH1
  list(logPrH0=logPrH0,logPrH1=logPrH1,logBF=logBF)
}
```

```

}
p0 <- 0.5
for (i in 1:ngenes){
  BFcall1 <- BFbinomial(ASEdat$N[i],ASEdat$Y[i],1,1,p0)
  BFcall2 <- BFbinomial(ASEdat$N[i],ASEdat$Y[i],2,2,p0)
  nlogBFr1[i] <- -BFcall1$logBF
  nlogBFr2[i] <- -BFcall2$logBF
  postprob2[i] <- pbeta(0.5,ASEa+ASEdat$Y[i],ASEa+ASEdat$N[i]-ASEdat$Y[i])
  postprob1[i] <- pbeta(0.5, 1+ASEdat$Y[i], 1+ASEdat$N[i]-ASEdat$Y[i])
}

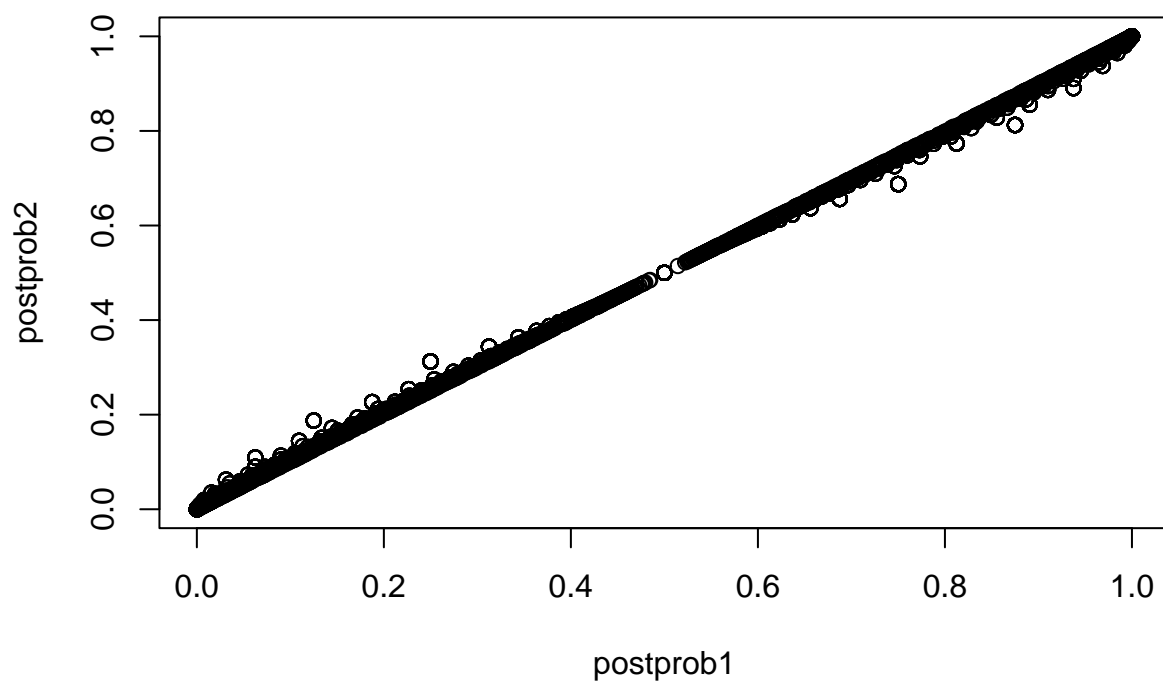
```

```
hist(postprob2)
```



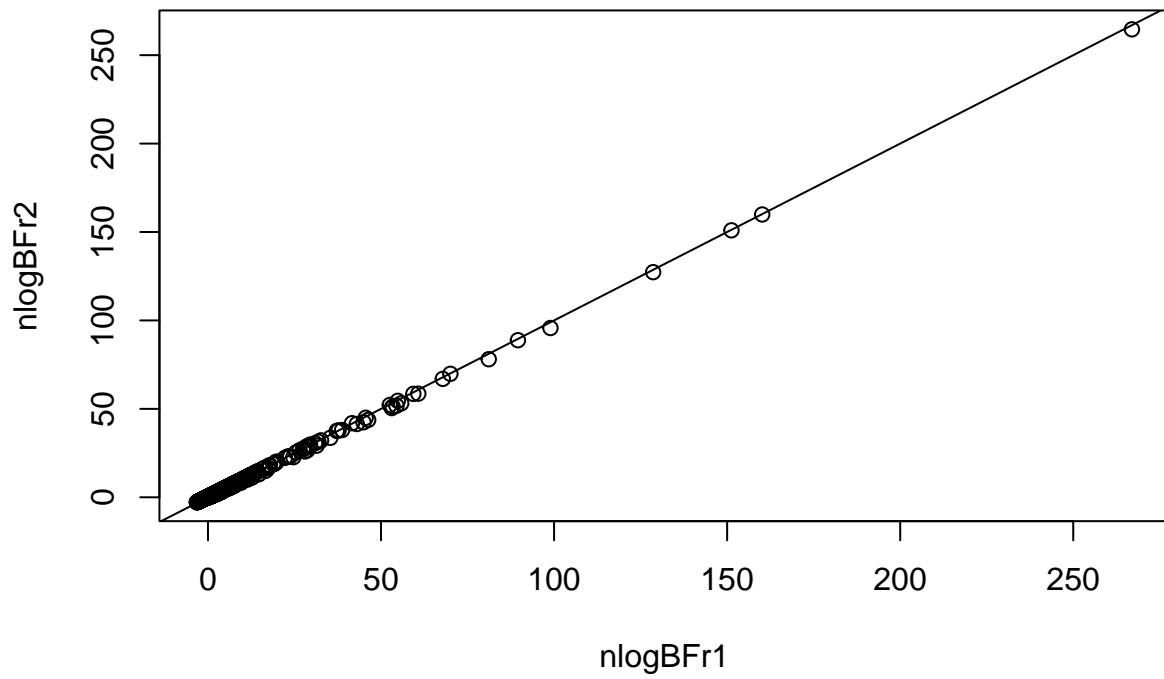
- Plot these posterior probabilities versus the versions under $a = 1$, and comment.

```
plot(postprob1, postprob2)
```

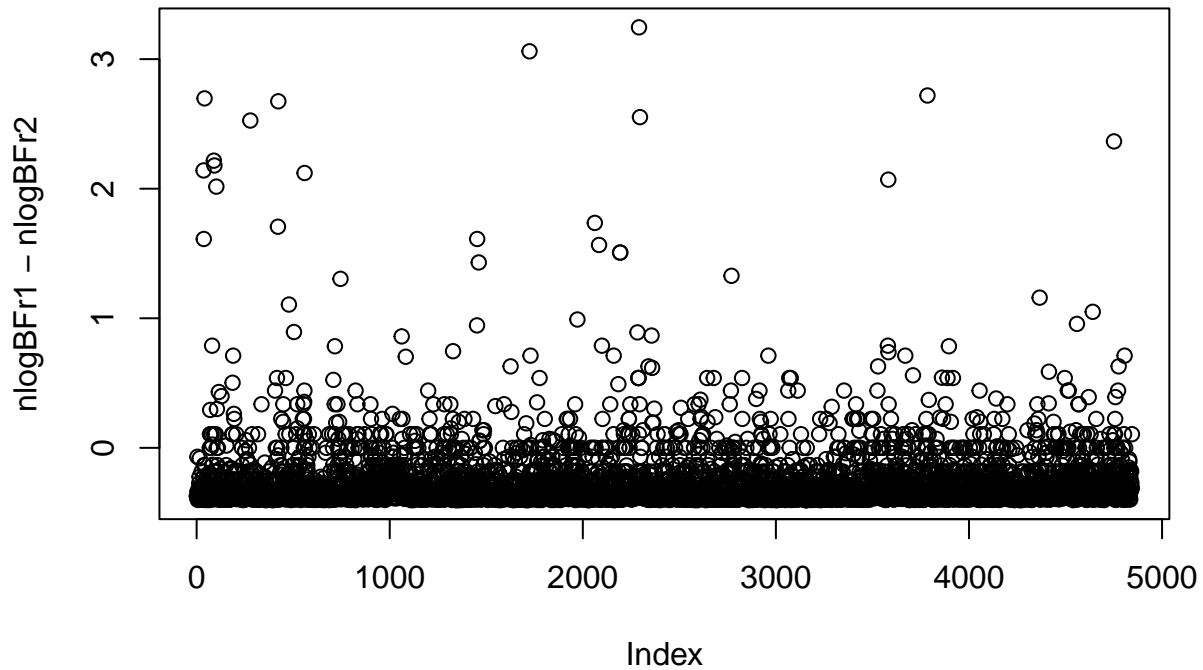


- How sensitive are the (log) Bayes factors to the prior specification?

```
plot(nlogBFr1, nlogBFr2)  
abline(0,1,lty=1)
```



```
plot(nlogBFR1~nlogBFR2)
```



- For how many genes would we reject $H_0 : \theta = 0.5$ if we use a rule of $1/\text{BF} > 150$?

```
p cutoff <- 0.05/ngenes
```

```
cat("log(1/BFr) > log(150) =", sum(nlogBFr1>log(150)), "\n")
```

```
## log(1/BFr) > log(150) = 162
```

```
cat("log(1/BFr) > log(20) =", sum(nlogBFr1>log(20)), "\n")
```

```
## log(1/BFr) > log(20) = 257
```

```
cat("p-values <", p cutoff, sum(pvals<p cutoff), "\n")
```

```
## p-values < 1.032205e-05 111
```

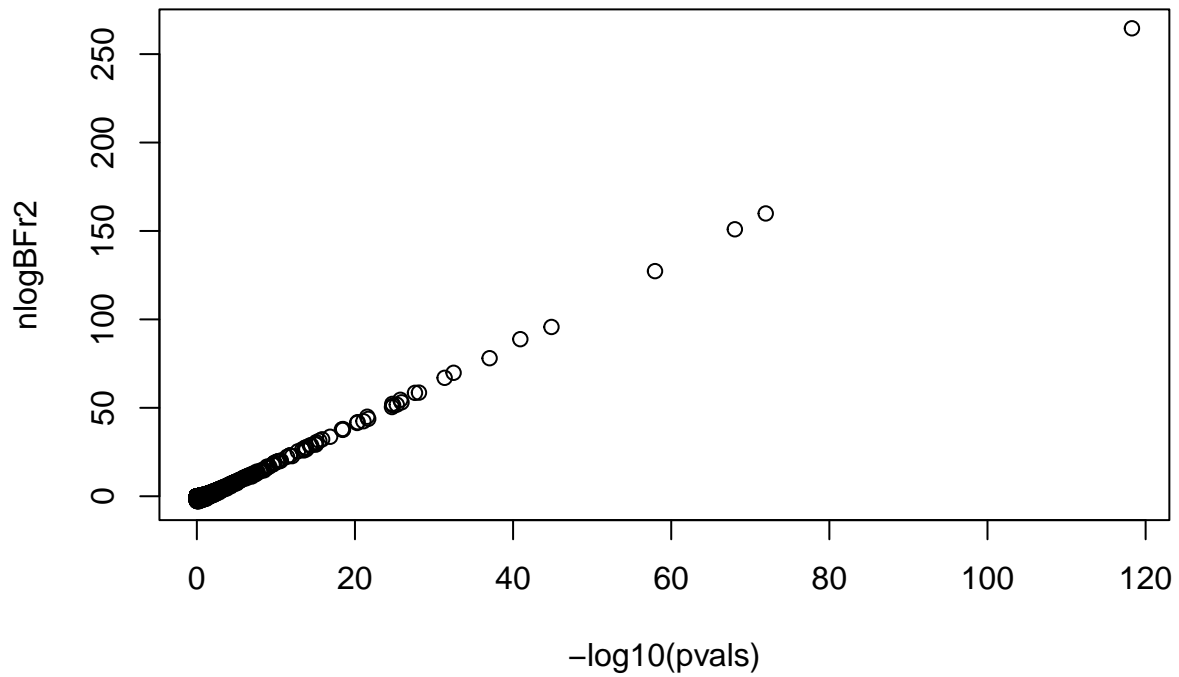
```
cat("postprobs < 0.01 and > 0.99 ", sum(postprob1<0.01), sum(postprob1>0.99), "\n")
```

```
## postprobs < 0.01 and > 0.99 278 242
```

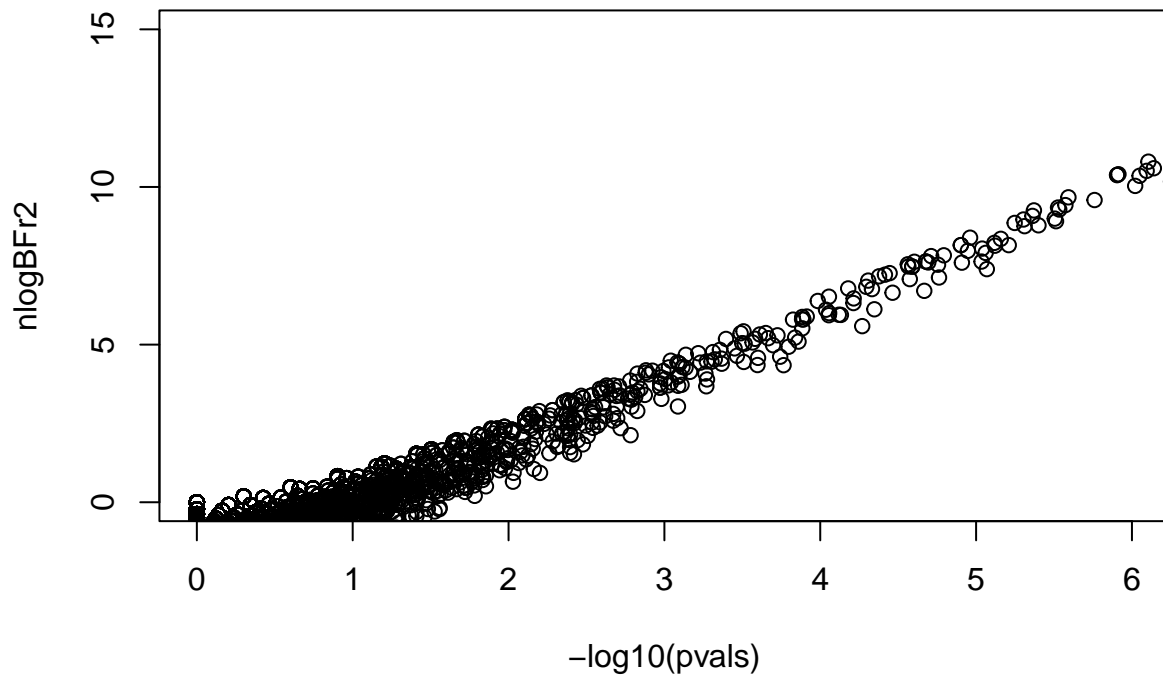
```
cat("postprobs < 0.01 and > 0.99 ",sum(postprob2<0.01),sum(postprob2>0.99),"\n")
```

```
## postprobs < 0.01 and > 0.99 256 222
```

```
plot(-log10(pvals), nlogBFr2)
```



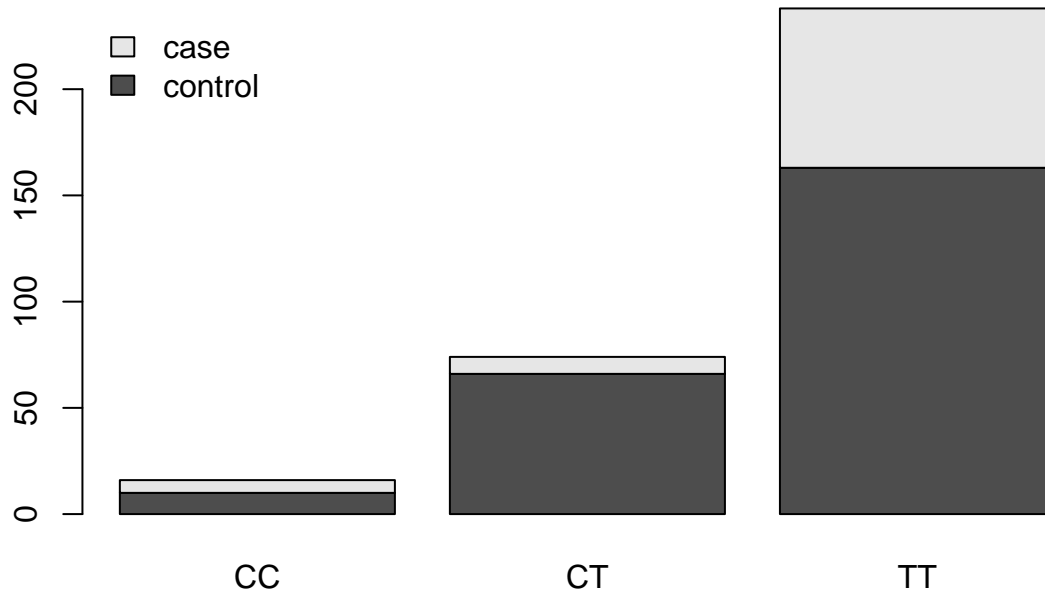
```
plot(-log10(pvals), nlogBFr2, xlim=c(0,6), ylim=c(0,15))
```



Q2

Redo the logistic regression LHON example, but use the less-diffuse prior on (only) the intercept β_0 , not the log odds ratio β_1 . How do the results compare to using the informative prior on β_1 ?

```
# setup the data & likelihood
lhon <- data.frame( y=c(1,1,1,0,0,0), x=c(0,1,2,0,1,2), n=c(6,8,75,10,66,163) )
dmat <- rbind(lhon$n[4:6], lhon$n[1:3])
names(dmat) <- c("CC", "CT", "TT")
barplot(dmat, names.arg=c("CC", "CT", "TT"), legend.text=c("case", "control")[2:1],
        args.legend=list(x="topleft", bty="n"))
```



```

expit <- function(x){exp(x)/(1+exp(x))}
lik <- function(beta){
  probs <- with(lhon, dbinom(y, 1, expit(beta[1] + beta[2]*x)))
  lik <- prod( probs^lhon$n )
}

# function to do the rejection sampling, using different prior
glm1 <- glm( y~x, family="binomial", weights=n, data=lhon)
betahat <- coef(glm1)
do.many <- function(bigB){
  #bigB <- 5
  beta.sample <- matrix(NA,bigB,2)
  i <- 1
  while(i<= bigB){
    beta.try <- c(rnorm(1, 0, sqrt(beta0pvar)), rnorm(1, 0, sqrt(4)))
    u <- runif(1)
    if(u < lik(beta.try)/lik(betahat)){
      beta.sample[i,] <- beta.try
      i <- i+1}
  }
  beta.sample
}

```

Doing informative prior selection:

```

set.seed(4)
#informative prior selection
beta0pvar <- uniroot(function(w){ qnorm(0.975, 0, sqrt(w)) - log(1.5)}, c(0.01, 1))$root
beta.post <- as.data.frame(do.many(100)) # takes a while!
# the informative prior is far from highest values of the likelihood
# so the acceptance probabilities are all low
names(beta.post) <- c("beta0", "beta1")
cbind( apply(beta.post, 2, mean), apply(beta.post, 2, sd))

```

```

##           [,1]      [,2]
## beta0 -0.3514017 0.1623737
## beta1 -0.3098595 0.1188023

```

```

quantile(beta.post[,2], c(0.5, 0.025, 0.975))

```

```

##           50%      2.5%      97.5%
## -0.30483049 -0.52022334 -0.06492171

```

```

par(mar=c(3,3,1.5,1.5)+0.2)
plot(beta1~beta0, data=beta.post, pch=19, col="#0000FF15",
      xlim=c(-3.6,-0.25), ylim=c(-0.5,1.4), xlab="", ylab="")
mtext(side=1, expression(beta[0]), line=2)
mtext(side=2, expression(beta[1]), line=2)

```

```

library("ellipse")

```

```

##
## Attaching package: 'ellipse'

```

```

## The following object is masked from 'package:graphics':

```

```

##
##      pairs

```

```

lines(ellipse(vcov(glm1), centre=coef(glm1)), col="forestgreen", lwd=2)
points(x=coef(glm1)[1], y=coef(glm1)[2], pch=19, col="forestgreen")
lines(ellipse(var(beta.post), centre=colMeans(beta.post)), col="magenta", lwd=2)
points(x=mean(beta.post$beta0), y=mean(beta.post$beta1), pch=19, col="magenta")

```

