Key 3

Jon Wakefield

2022-07-21

Q1

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

- Obtain a histogram of the posterior probabilities $\Pr(\theta < 0.5|y)$, across genes.
- Plot these posterior probabilities versus the versions under $a = 1$, and comment.
- How sensitive are the (log) Bayes factors to the prior specification?
- For how many genes would we reject $H_0 : \theta = 0.5$ if we use a rule of $1/BF > 150$?

```
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)
```

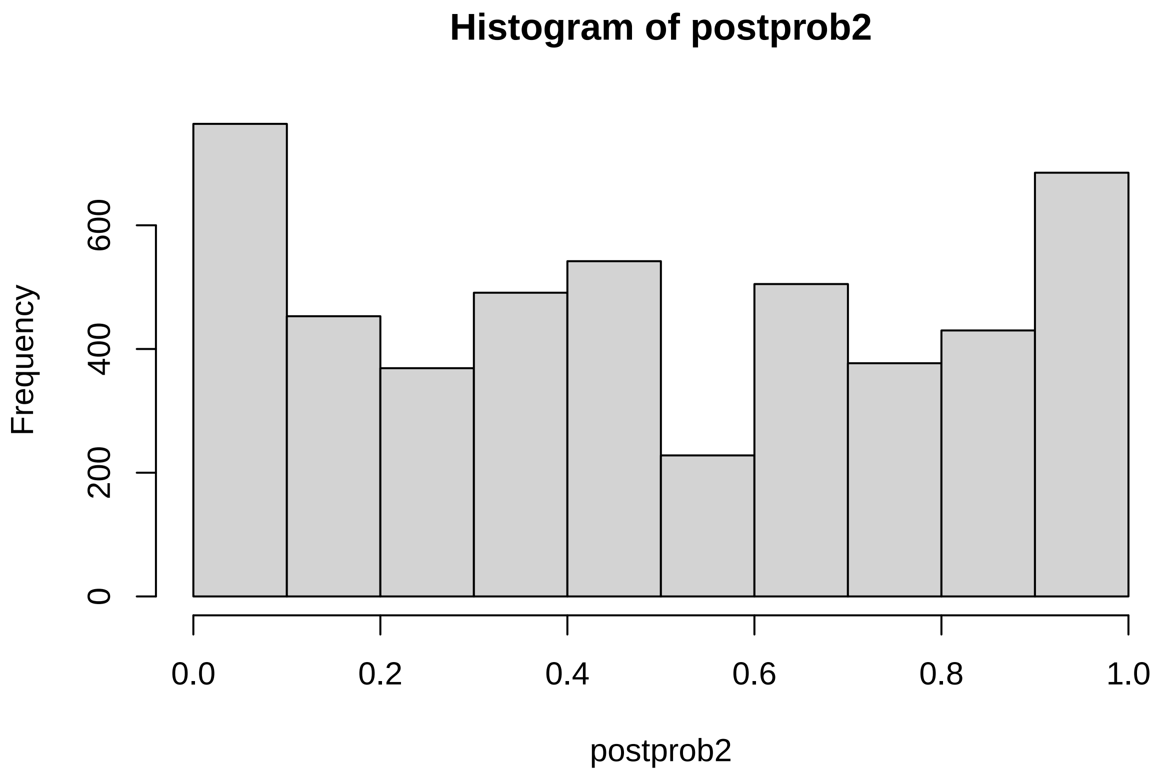
```
## [1] 4844    2
```

```
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) + gamma(a+b) - lgamma(a) - lgamma(b) + lgamma(y+a) + lgamma(N-y+b) -lgamma(N+a
  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])
}
```

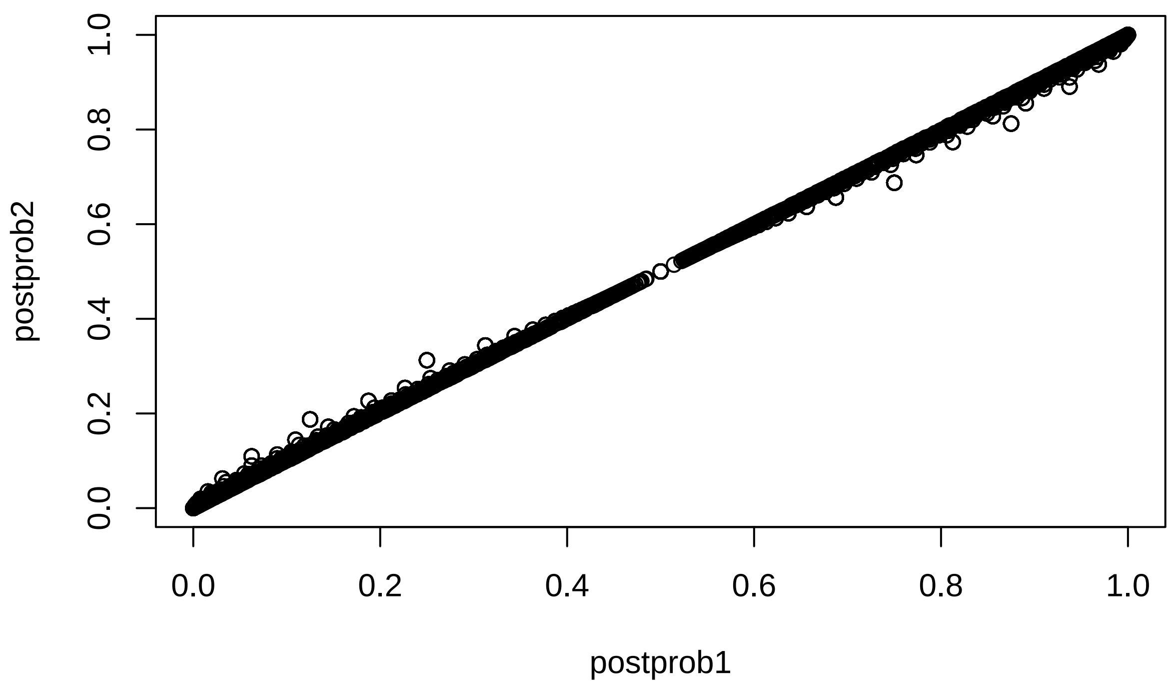
- Obtain a histogram of the posterior probabilities $\Pr(\theta < 0.5|y)$, across genes.

```
hist(postprob2)
```



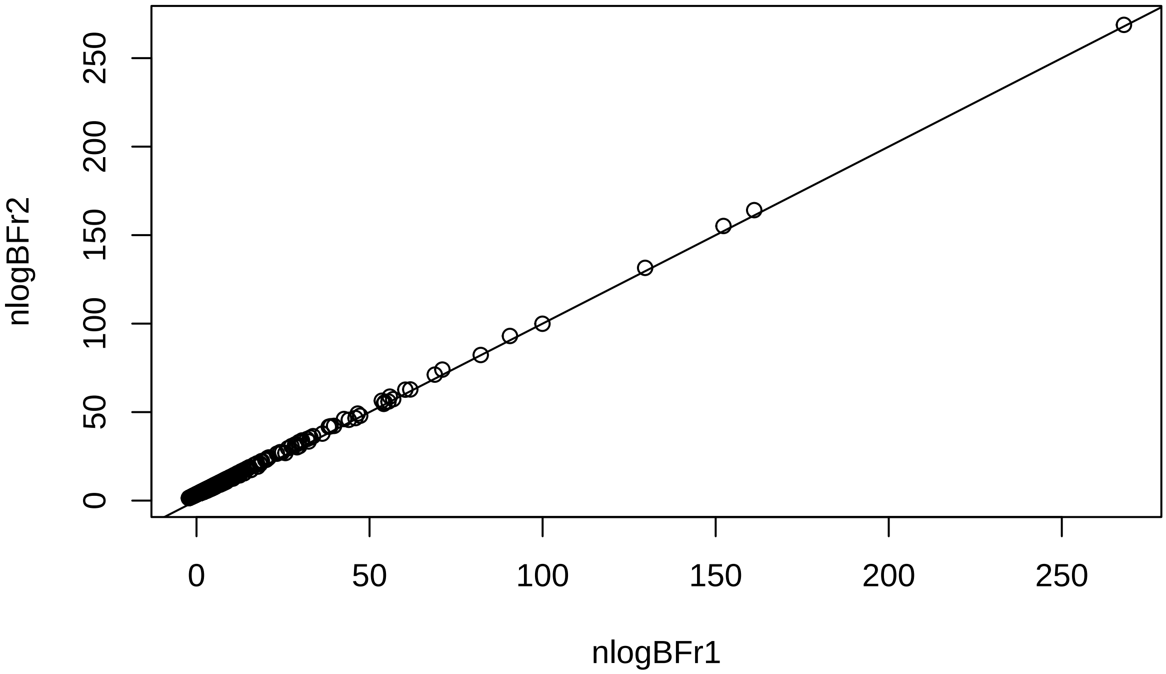
- Plot these posterior probabilities versus the versions under $a = 1$, and comment.

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

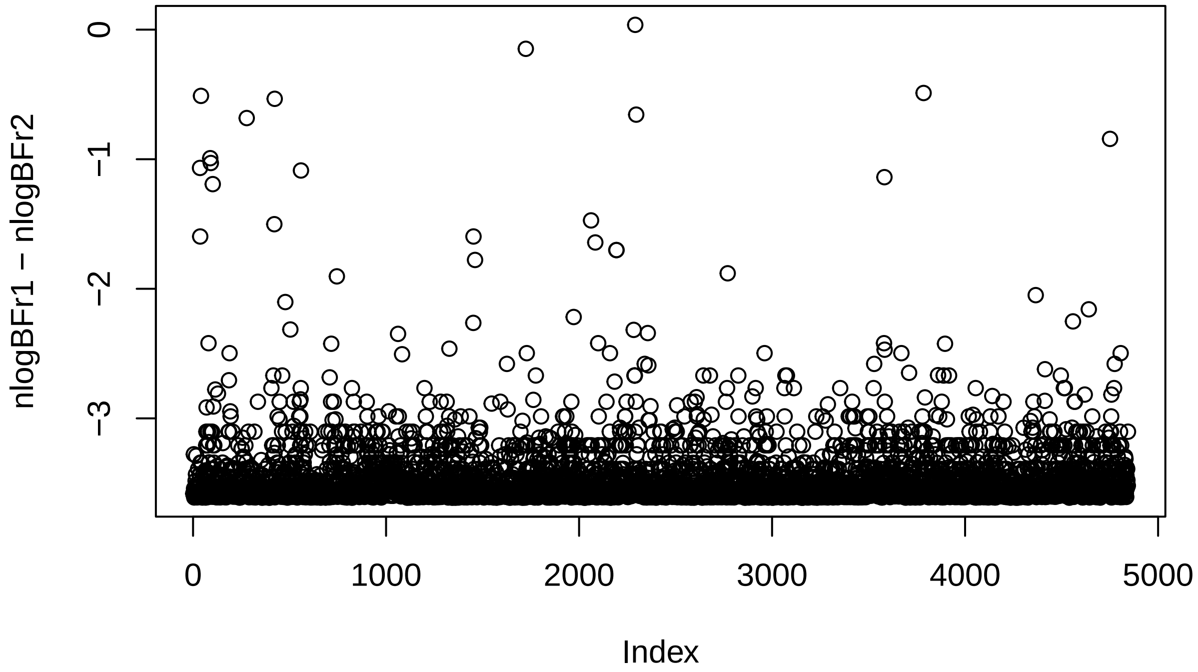


- How sensitive are the (log) Bayes factors to the prior specification?

```
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 $1/\mathrm{BF} > 150$?

```
pcutoff <- 0.05/ngenes

cat("log(1/BFr) > log(150) =",sum(nlogBFr2>log(150)),"\n")
```

```
## log(1/BFr) > log(150) = 640
```

```
cat("log(1/BFr) > log(20) =",sum(nlogBFr2>log(20)),"\n")
```

```
## log(1/BFr) > log(20) = 3826
```

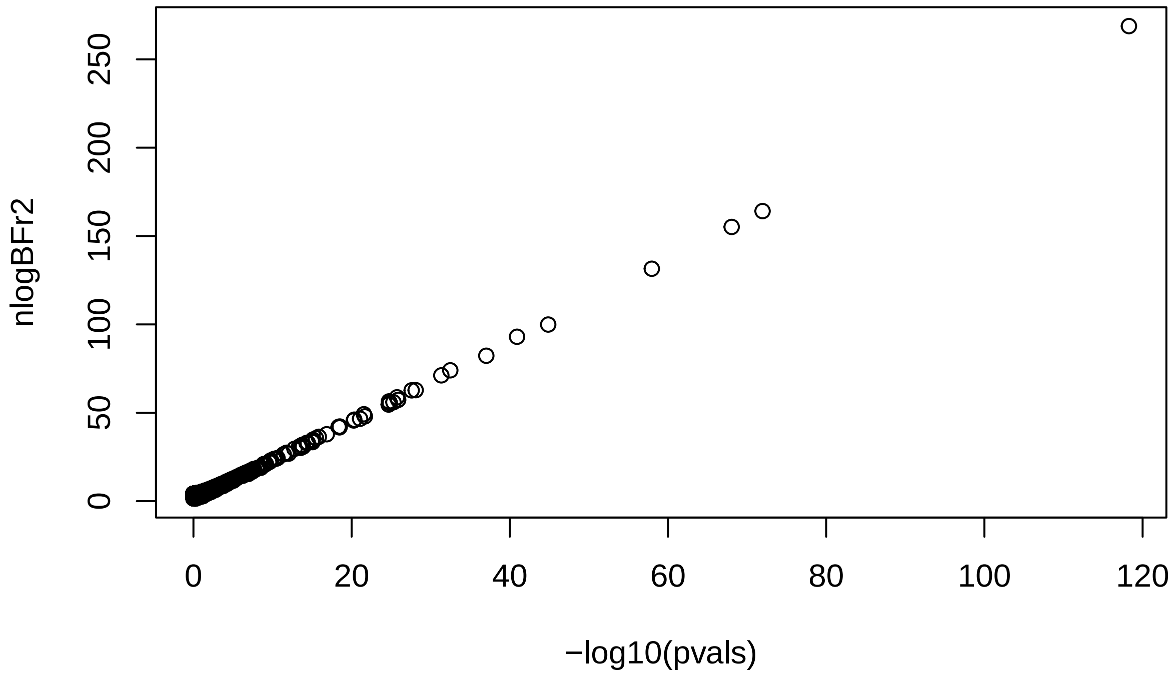
```
cat("p-values >",pcutoff,sum(pvals<pcutoff),"\n")
```

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

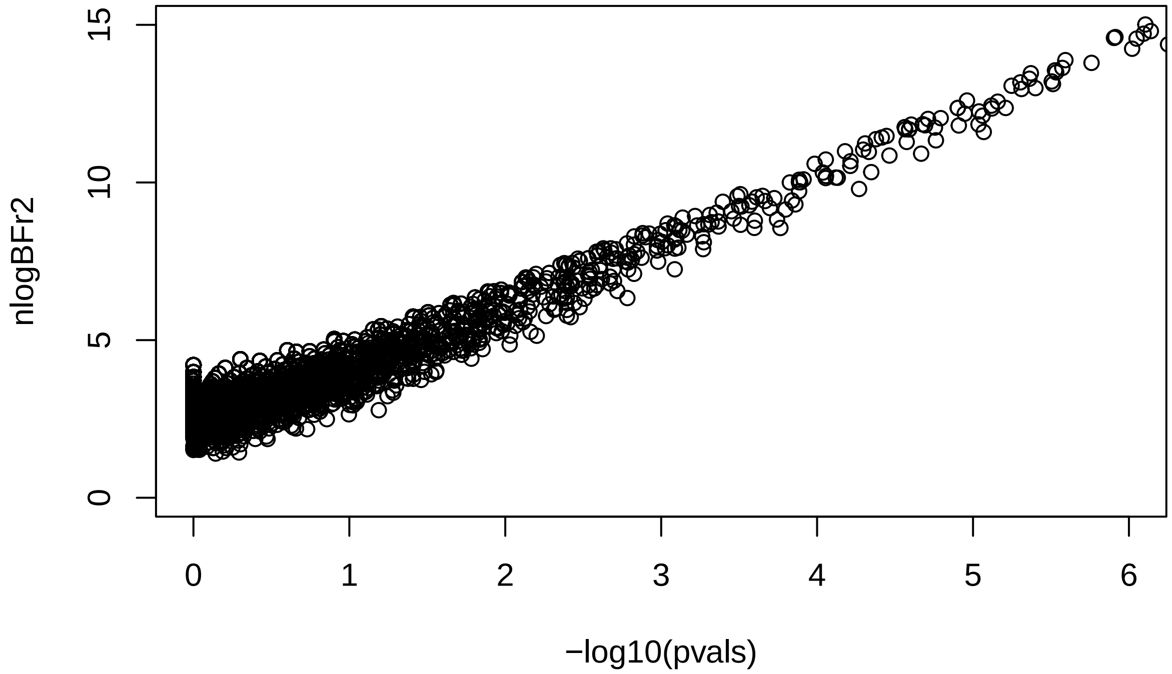
```
cat("postprobs < 0.01 and > 0.99 ",sum(postprob2<0.01),sum(postprob2>0.99),"\n")
```

```
## 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 birthweight example, using the default priors, but with the smoking variable only in the model.

- Compare with a frequentist smoking only example with the `glm` function.

```
# installing INLA, commands from INLA's site:
# install.packages("INLA",repos=c(getOption("repos"),INLA="https://inla.r-inla-download.org/R/stable"),

# installing brinla, for the dataset:
#devtools::install_github('julianfaraway/brinla')
#remotes::install_github('julianfaraway/brinla')
library("brinla")
```

```
## Loading required package: INLA
```

```
## Loading required package: Matrix
```

```
## Loading required package: foreach
```

```
## Warning: package 'foreach' was built under R version 4.2.1
```

```
## Loading required package: parallel
```

```
## Loading required package: sp
```

```
## Warning: package 'sp' was built under R version 4.2.1
```

```
## This is INLA_22.05.07 built 2022-05-07 09:52:03 UTC.
##  - See www.r-inla.org/contact-us for how to get help.
```

```
data(lowbwt)

# do the analysis
library("INLA")
bwt.inla <- inla(LOW ~ SMOKE, data = lowbwt, family = "binomial", Ntrials = 1,
                 control.compute = list(dic = TRUE, cpo = TRUE))
bwt.inla$summary.fixed
```

```
##                   mean        sd  0.025quant   0.5quant 0.975quant mode
## (Intercept) -1.0870194 0.2146983 -1.52115710 -1.0827790 -0.6786275   NA
## SMOKE1       0.7078681 0.3196038  0.08272406  0.7066084  1.3371714   NA
##                      kld
## (Intercept) 2.622894e-12
## SMOKE1      7.372548e-05
```

```
bwt.glm <- glm(LOW ~ SMOKE, data = lowbwt, family = "binomial")
coef(summary(bwt.glm))
```

```
##               Estimate Std. Error   z value     Pr(>|z|)
## (Intercept) -1.0870515  0.2147338 -5.062322 4.141802e-07
## SMOKE1       0.7040592  0.3196423  2.202647 2.761962e-02
```