732A91 - Lab 1

Joris van Doorn || Weng Hang Wong

09 April 2020

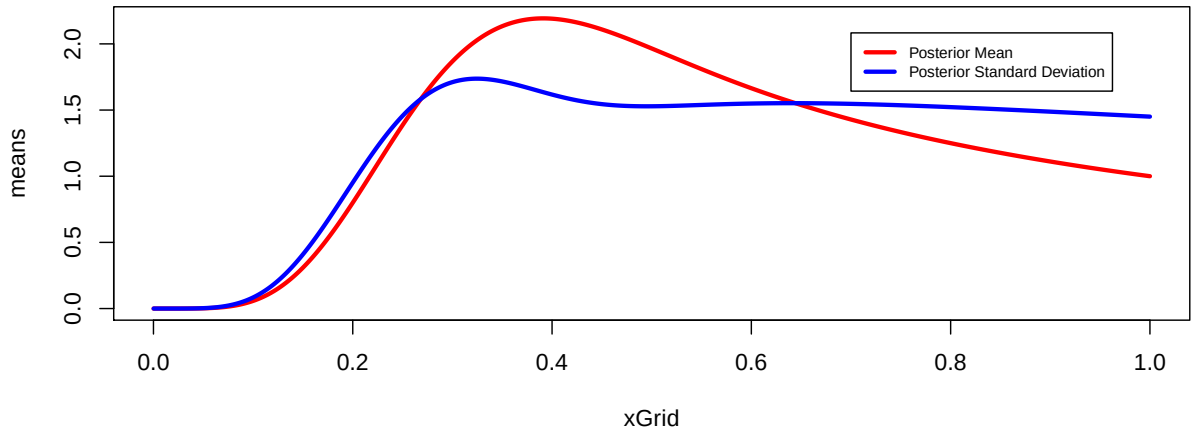
1. Bernoulli ... again.

Let $y_1, ..., y_n|\theta$ $Bern(\theta)$, and assume that you have obtained a sample with $s = 5$ successes in $n = 20$ trials. Assume a $Beta(\alpha_0, \beta_0$ prior for θ and let $\alpha_0 = \beta_0 = 2$

a.

Draw random numbers from the posterior $\theta|y$ $Beta(\alpha_0 + s, \beta_0 + f), y = (y_1, ..., y_n)$, and verify graphically that the posterior mean and standard deviation converges to the true values as the number of random draws grows large.

```
## Number of Draws:  10000
## Posterior Mean:  1
## Posterior Standard Deviation:  1.450244
```



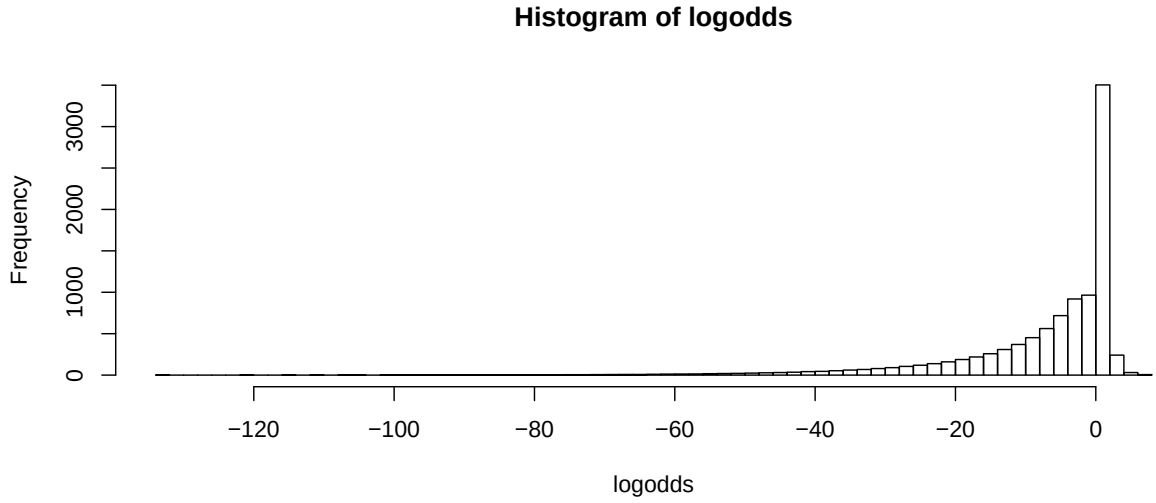
b.

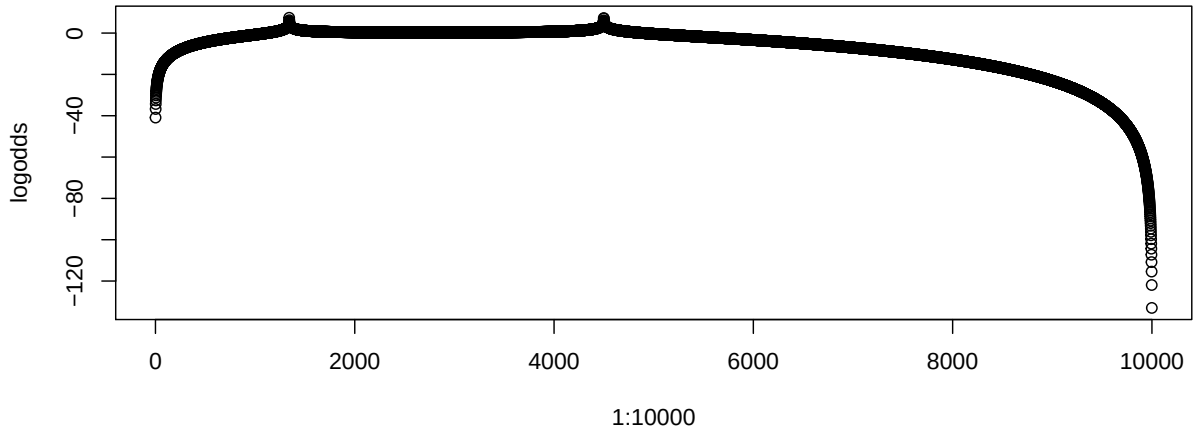
Use simulation (nDraws = 10000) to compute the posterior probability $Pr(\theta > 0.3|y)$ and compare with the exact value

```
## The probability of theta being larger than 0.3 is:  0.7613
```

c.

Compute the posterior distribution of the log-odds ϕ






```
##
## Call:
##  density.default(x = logodds)
##
## Data: logodds (10000 obs.);  Bandwidth 'bw' = 1.154
##
##        x                 y
##  Min.   :-136.47   Min.   :0.000e+00
##  1st Qu.: -99.63   1st Qu.:4.515e-05
##  Median : -62.80   Median :3.945e-04
##  Mean   : -62.80   Mean   :6.783e-03
##  3rd Qu.: -25.96   3rd Qu.:3.620e-03
##  Max.   :  10.88   Max.   :1.309e-01
```

Appendix

```r
knitr::opts_chunk$set(echo = TRUE)
knitr::opts_chunk$set(fig.width=9, fig.height = 4.1)
library(tidyverse)
library(dplyr)
library(knitr)
RNGversion("3.6.2")
set.seed(12345)
# ----------------------------------------------------------
# 1a

a = b = 2
n = 20
s = 5
nDraws = 10000
#xGrid <- seq(0.001, 0.999, by=0.001)
#posterior = dbeta(xGrid, a+s, b+(n-s))

means <- c()
sds <- c()

set.seed(12345)

for(i in 1:nDraws){
  xGrid <- seq(1/nDraws, i/nDraws, by=1/nDraws)
  posterior = dbeta(xGrid, a+s, b+(n-s))
  means[i] <- mean(posterior)
  sds[i] <- sd(posterior)

  #at("\nNumber of Draws: ", i , "\nMean: ", mean(posterior), "\nStandard Deviation: ", sd(posterior))
}

cat("Number of Draws: ", nDraws , "\nPosterior Mean: ", means[nDraws], "\nPosterior Standard Deviation:

plot(xGrid, means, type = 'l', lwd = 3, col = "red")
lines(xGrid, sds, lwd = 3, col = "blue")
legend(x = max(xGrid)*0.70, y = 0.95*max(means), legend = c("Posterior Mean", "Posterior Standard Devia
# ----------------------------------------------------------
# 1b

xGrid <- seq(1/nDraws, nDraws/nDraws, by=1/nDraws)
posterior = pbeta(xGrid, a+s, b+(n-s)) # Ask for the difference between pbeta & dbeta

prob_03 <- posterior[posterior > 0.3]
prob <- length(prob_03)/nDraws

cat("The probability of theta being larger than 0.3 is: ", prob)
# ----------------------------------------------------------
# 1c

xGrid <- seq(1/nDraws, nDraws/nDraws, by=1/nDraws)
posterior = dbeta(xGrid, a+s, b+(n-s))
```

```r
logodds <- c()

for(i in 1:length(posterior)){
  logodds[i] <- log(abs(posterior[i]/(1-posterior[i])))
}

hist(logodds, breaks = 100)
plot(1:10000,logodds)
density(logodds)
```