Bayesian Learning Lab1

Weng Hang Wong

4/8/2020

1. Bernoulli ... again.

Let $y_1, ..., y_n|\theta \sim Bern(\theta)$ and assume that you have obtained a sample with s = 5 successes in n = 20 trials. Assume a Beta(α_0,β_0) prior for θ and let $\alpha_0 = \beta_0 = 2$

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

Verification

To verify praphically that the posterior mean and standard deviation, meaning that it is needed to use the mean and variance from the Beta Probability Density Function.

Mean of Beta PDF: $\frac{\alpha}{\alpha+\beta}$

Variance of Beta PDF: $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Bernouli Distruibution

The probability density function of Bernouli distribution is:

$$p(s) = \binom{n}{s}\theta^s(1-\theta)^{y-s}$$

The likelihood function is:

$$p(y_1, ...y_n|\theta) = \binom{n}{s}\theta^s(1-\theta)^{y-s}$$

The prior is given Beta(α_0,β_0):

$$p(\theta) = \frac{\Gamma(\alpha_0 + \beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)}\theta^{\alpha_0-1}(1-\theta)^{\beta_0-1}$$

We have:

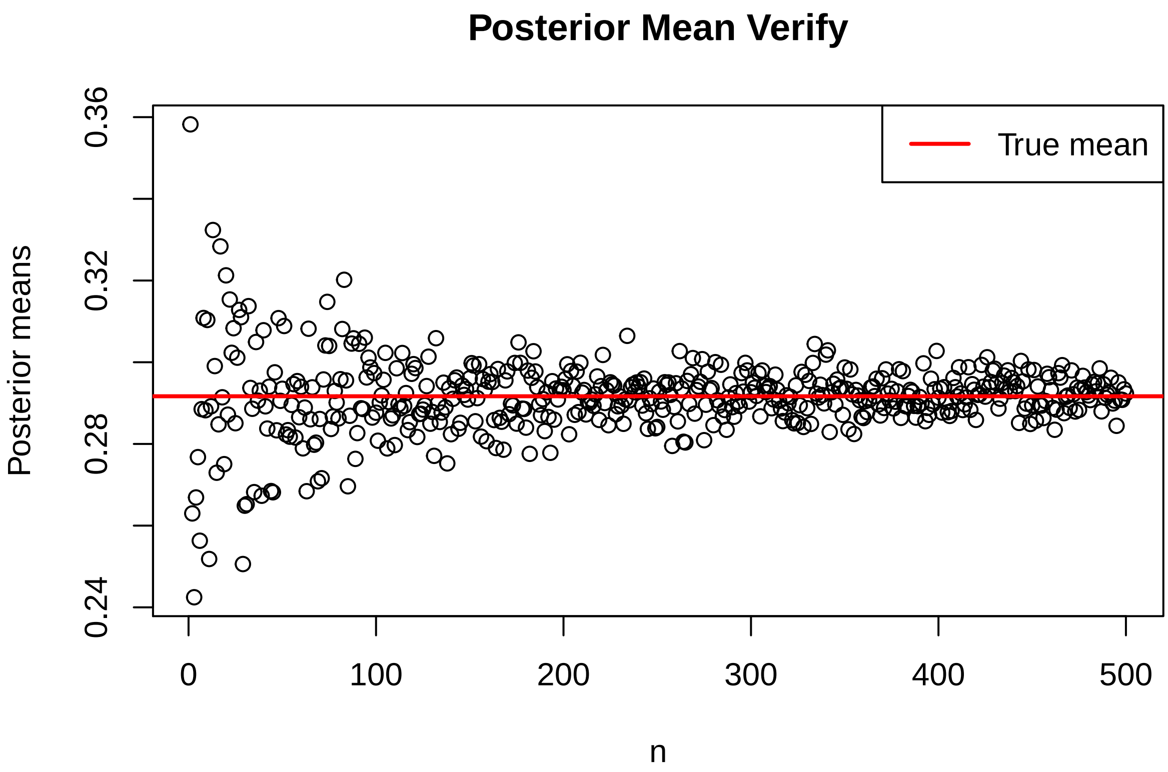
Posterior$\propto$likelihood*prior

$$\theta|y \sim Beta(\alpha_0 + s, \beta_0 + f)$$

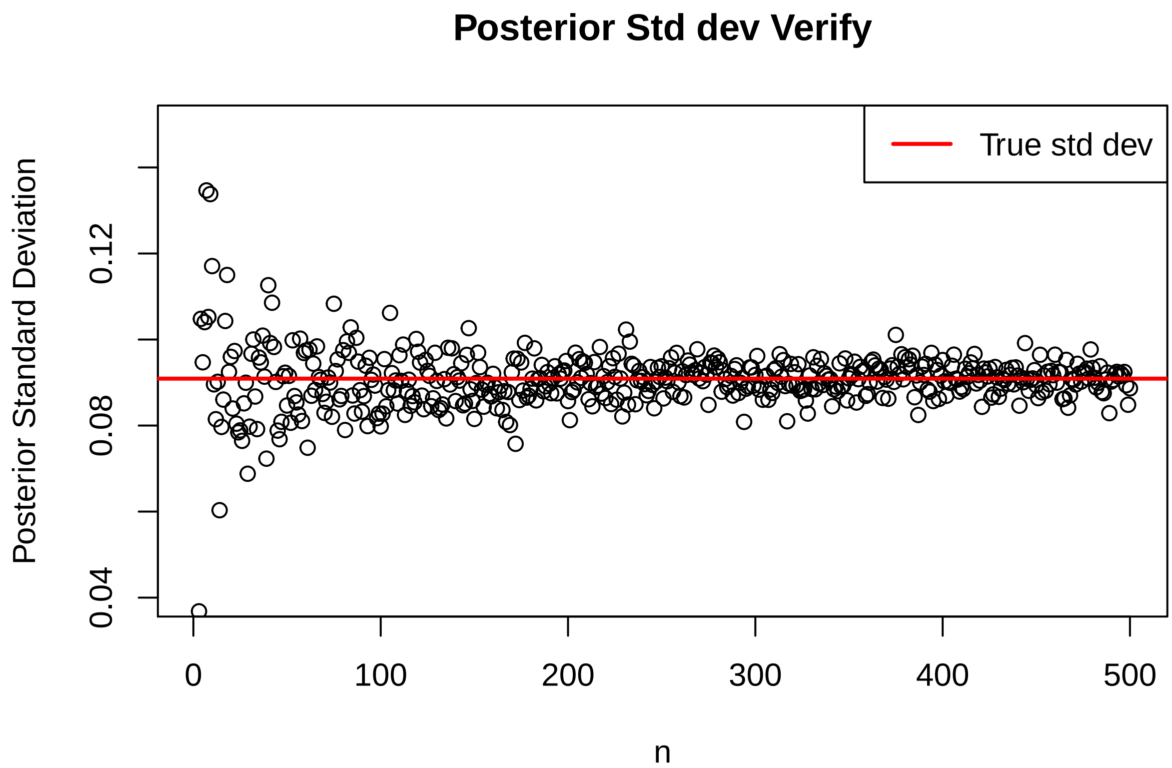
From the above given posterior, we know that $\alpha = \alpha_0 + s = 7$, $\beta = \beta_0 + f = 17$. By using the alpha and beta we have, it is possible to simulate the random samples from the rbeta() function. With the growing number of random draws(1:500), we can plot the samples and investigate how the mean and standard deviation distributed.

```
## [1] 0.09090593
```

```
## The true mean is: 0.2916667
```




```
## The true Standard deviation is: 0.09090593
```



According to the two figures above, it is successfully verified the posterior mean and standard devaiations are converaged to the true value.

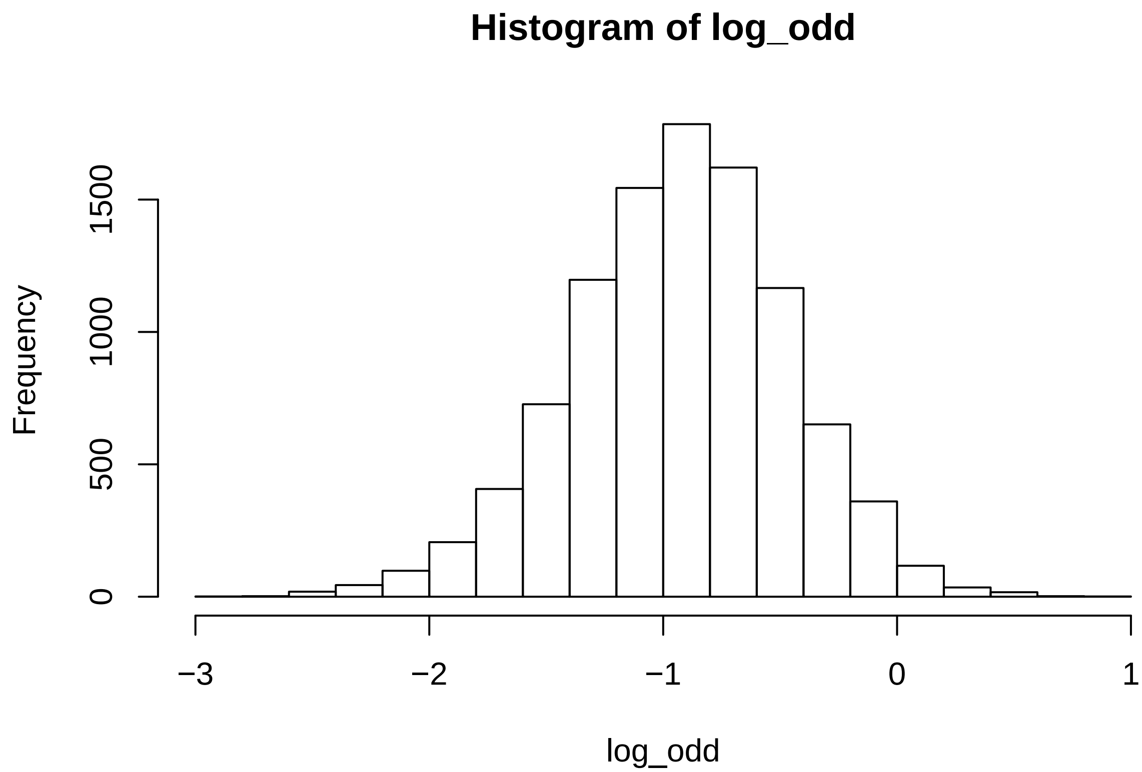
(b) Use simulation (nDraws = 10000) to compute the posterior probability $Pr(\theta > 0.3|y)$ and compare with the exact value [Hint: pbeta()].

From the simulation of nDraws = 10000, the posterior probability $Pr(\theta > 0.3|y)$ is 0.4392, and it is very close to the true probability computed from the pbeta() function, which is 0.4399472.

```
## The simulation posterior probability (> 0.3) is:  0.4392
```

```
## The true probability is : 0.4399472
```

(c) Compute the posterior distribution of the log-odds $\phi = log\frac{\theta}{(1-\theta)}$ (nDraws = 10000). [Hint: hist() and density() might come in handy]




```
##
## Call:
##  density.default(x = log_odd)
##
## Data: log_odd (10000 obs.);  Bandwidth 'bw' = 0.06469
##
##        x                 y
##  Min.   :-3.17277   Min.   :0.0000071
##  1st Qu.:-2.08423   1st Qu.:0.0027689
##  Median :-0.99569   Median :0.0560336
##  Mean   :-0.99569   Mean   :0.2294409
##  3rd Qu.: 0.09285   3rd Qu.:0.4063222
##  Max.   : 1.18139   Max.   :0.8987902
```

2 Log-normal distribution and the Gini coefficient.

Assume that you have asked 10 randomly selected persons about their monthly income (in thousands Swedish Krona) and obtained the following ten observations: 44, 25, 45, 52, 30, 63, 19, 50, 34 and 67. A common model for non-negative continuous variables is the log-normal distribution. The log-normal distribution

$log(N(\mu, \sigma^2))$ has density function:

$$p(y|\mu, \sigma^2) = \frac{1}{y * \sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(log(y)-\mu)^2}$$

For y>0, $\mu > 0 and \sigma^2 > 0$. The log-normal distribution is related to the normal distribution as follows: if $y \ logN(\mu, \sigma^2)$ then $log(y) \ N(\mu, \sigma^2)$. Let $y_1, ..., y_n|\mu, \sigma^2 \ ^{iid} log(N(\mu, \sigma^2))$, where $\mu = 3.7$ is assumed to be known but σ^2 is unknown with non-informative prior $p(\sigma^2) \propto \frac{1}{\sigma^2}$. The posterior for the σ^2 is the $Inv - \chi^2(n, \tau^2)$ distribution, where

$$\tau^2 = \frac{\sum_{i=1}^{n}(log(y_i) - \mu)^2}{n}$$

a.

Simulate 10,000 draws from the posterior of σ^2 (assuming $\mu = 3.7$) and compare with the theoretical $Inv - \chi^2(n, \tau^2)$ posterior distribution.

We worked with the log of the data, assuming that is approximately normally distributed with $N(\mu, \sigma^2)$ ### (a)

Appendix

```
## Draw random numbers from the posterior
set.seed(12345)
beta_sample = function(n,alpha, beta){
  sample = rbeta(n, alpha, beta)
  mean_sam = mean(sample)
  sd_sam = sd(sample)
  return(list("Mean"=mean_sam,"sd"= sd_sam))
}

alpha=7
beta=17

true_mean = alpha/(alpha+beta)
true_sd =  sqrt((alpha*beta) / (((alpha+beta)^2)*(alpha+beta+1)))
true_sd

##verify of n =500
n=500
veri_means=c()
veri_sd = c()
for(i in 1:n){
  veri_means[i] = beta_sample(i, alpha, beta)$Mean
  veri_sd[i]=beta_sample(i, alpha, beta)$sd
}

cat("The true mean is:", true_mean)
plot( veri_means, main="Posterior Mean Verify", xlab="n", ylab="Posterior means")
abline(h=true_mean, col="red", lwd=2)
legend("topright",legend="True mean", lty=1,col="red",lwd=2)
```

```
cat("The true Standard deviation is:", true_sd)
plot(veri_sd, main="Posterior Std dev Verify", xlab="n", ylab="Posterior Standard Deviation",ylim=c(0.0
abline(h=true_sd, col="red", lwd=2)
legend("topright",legend="True std dev", lty=1,col="red",lwd=2)


set.seed(12345)
nDraws=10000
beta_sample2 = rbeta(nDraws, alpha, beta)

sample_prob = sum(beta_sample2 > 0.3)/nDraws
real_prob = pbeta(q=0.3,alpha, beta, lower.tail = FALSE)


cat("The simulation posterior probability (> 0.3) is: ",sample_prob)
cat("The true probability is :", real_prob)


log_odd = log(beta_sample2/(1-beta_sample2))

hist(log_odd)
density(log_odd)
#log_odd
```