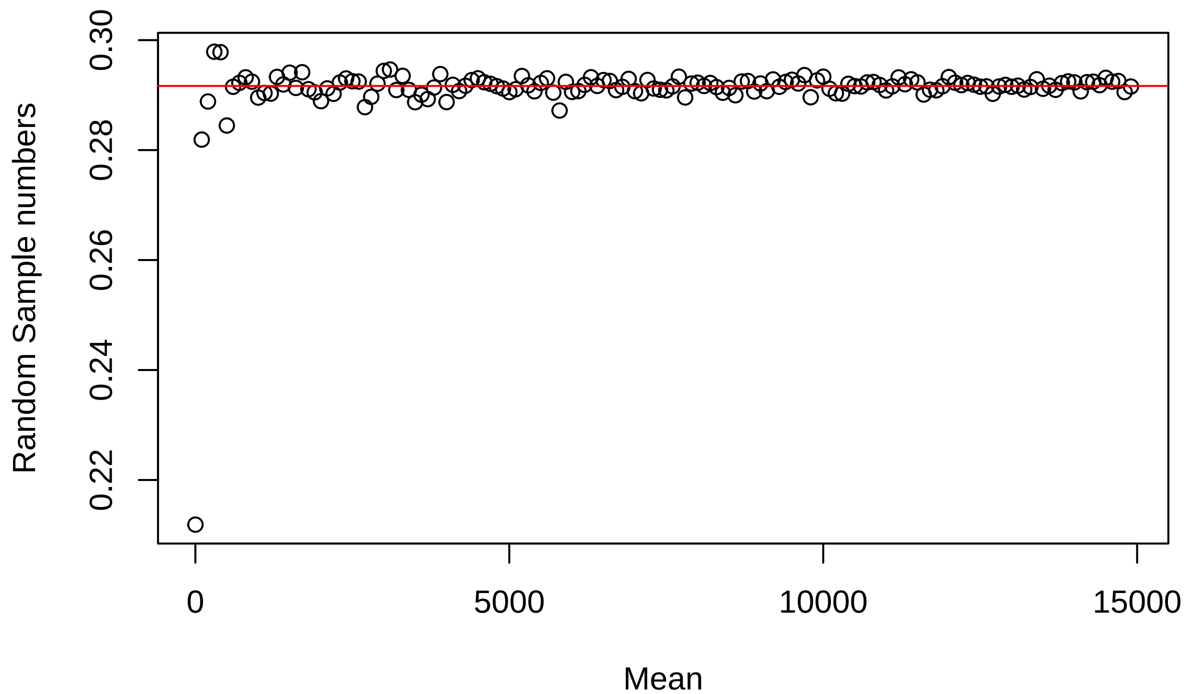
BayesianLab1

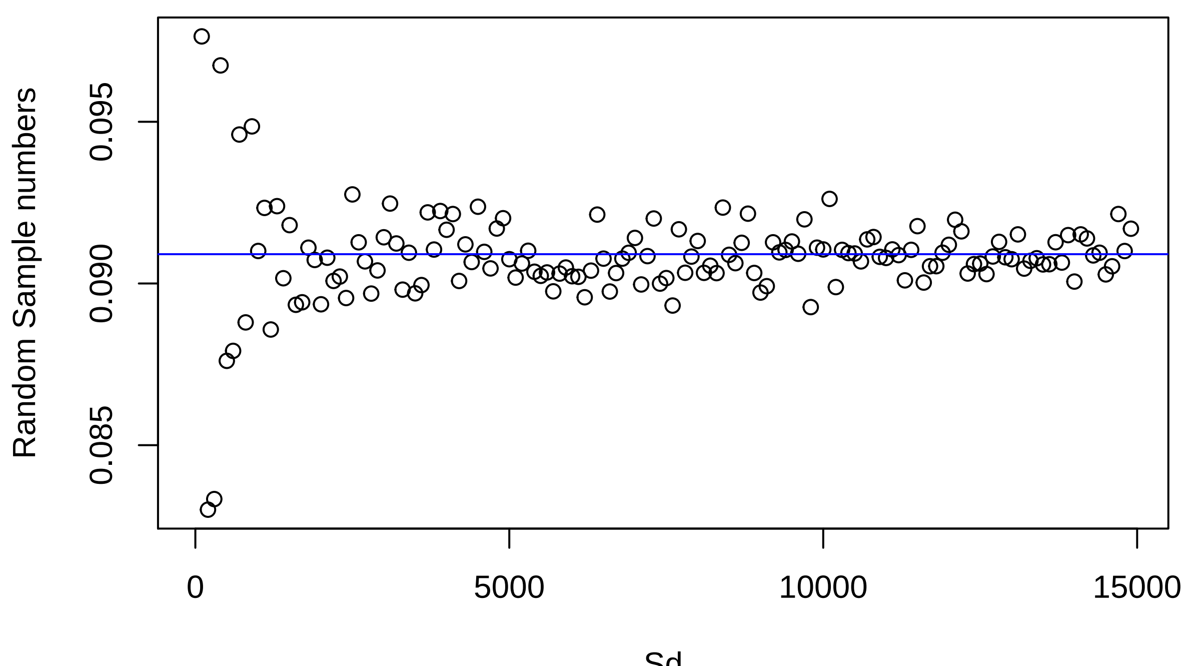
Harshavardhan Subramanian Vinod Kumar Dasari

14/04/2020

1. Bernoulli . . . again

a





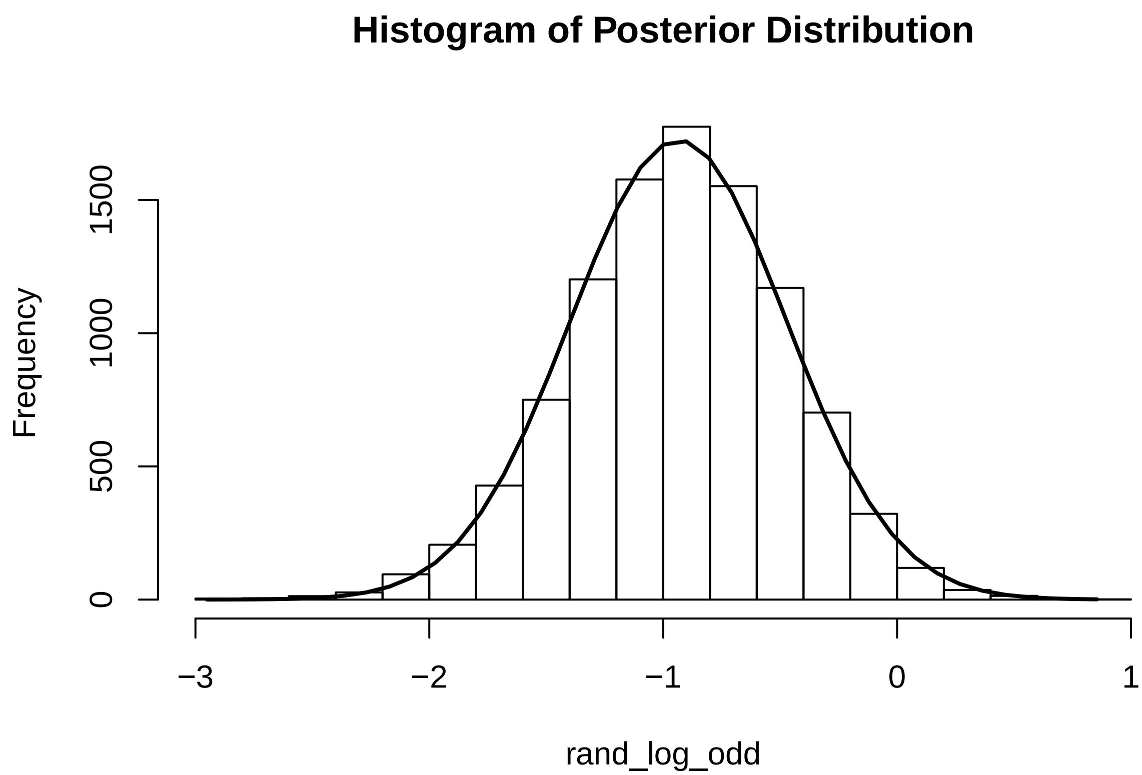
From the above graph it is evident that as the number of random grows, the posterior mean and standard deviation converges to the true values.

b

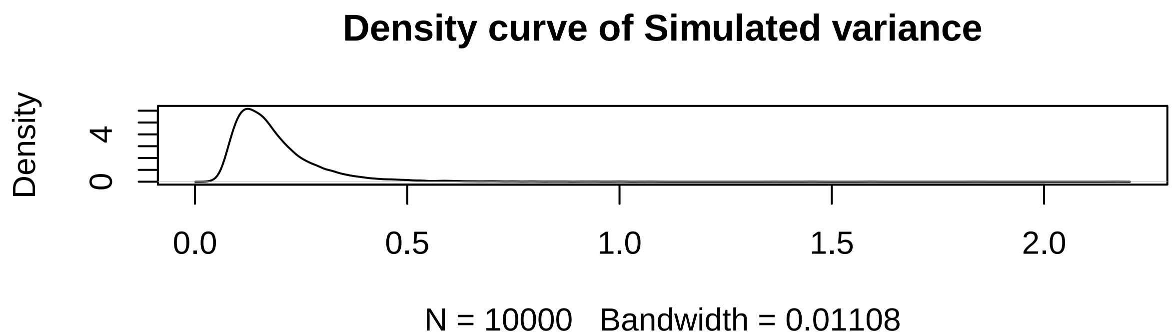
```
## The posterior probability for p(theta > 0.3) using simulation with Ndraws = 10000 is : 0.4341
```

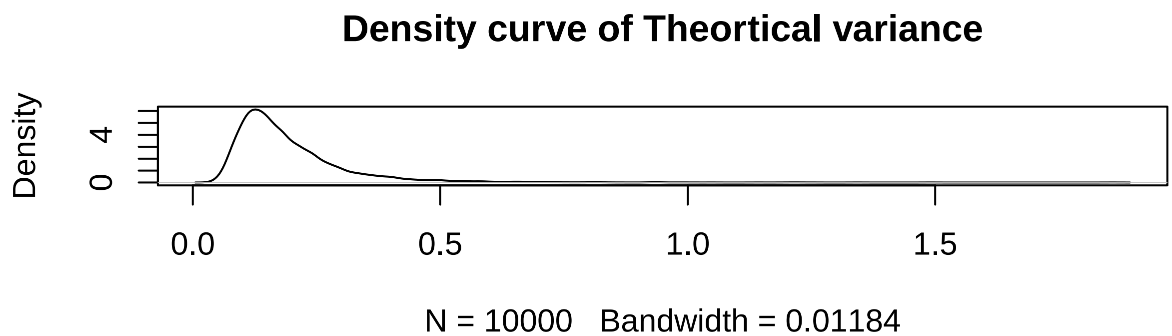
```
## The exact posterior probability for p(theta > 0.3) is : 0.4399472
```

c



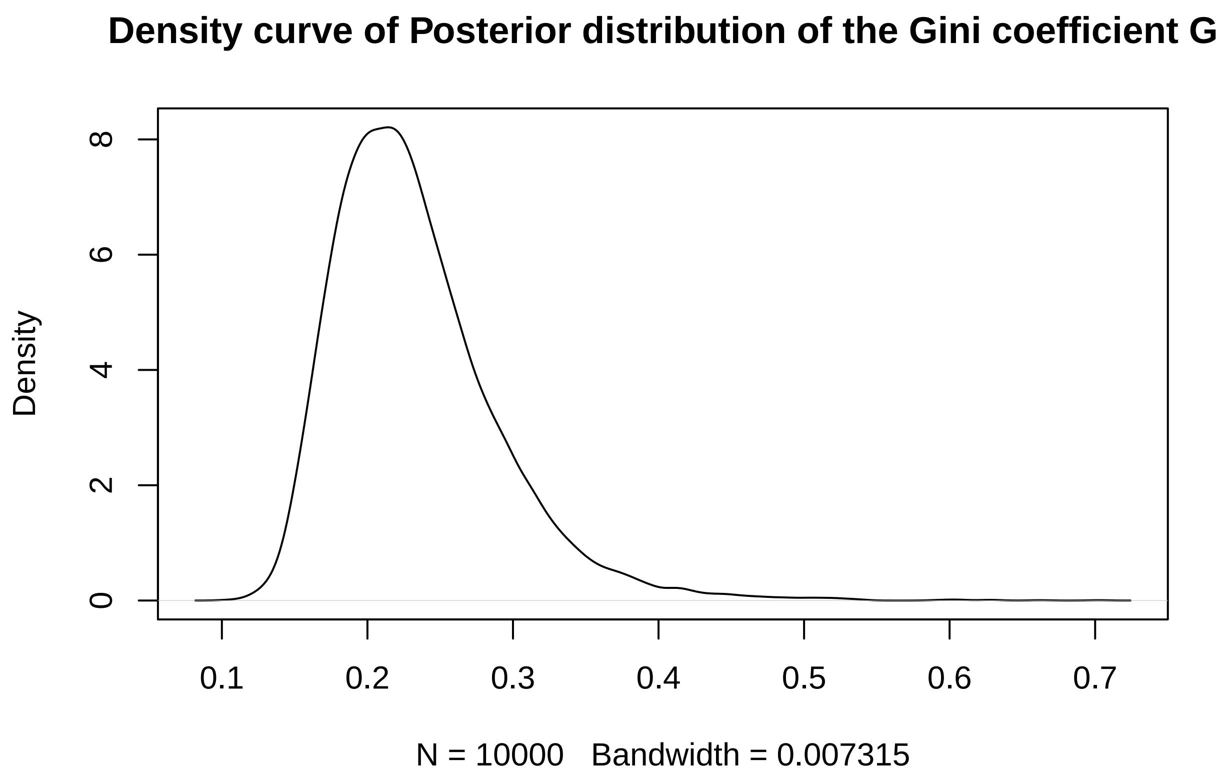
2. Log-normal distribution and the Gini coefficient.





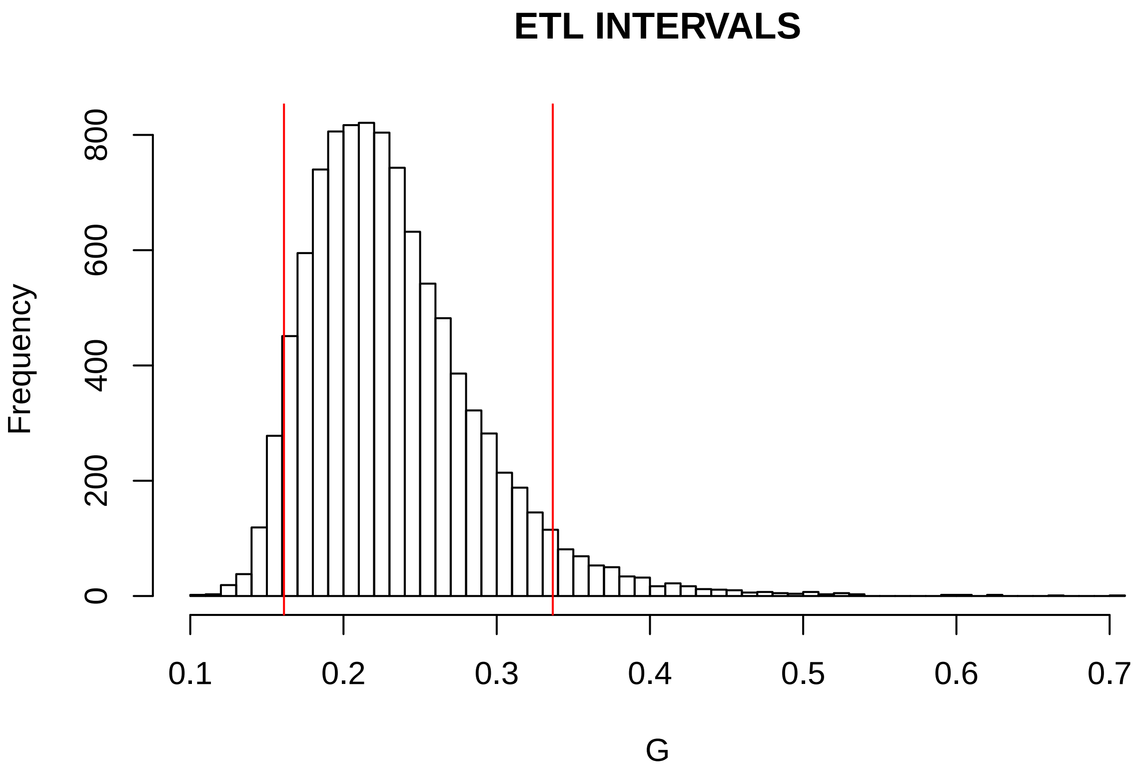
From the density plot it is evident that both simulated and theoretical values pretty much follow same distribution and the mean of both is also very close to each other.

b

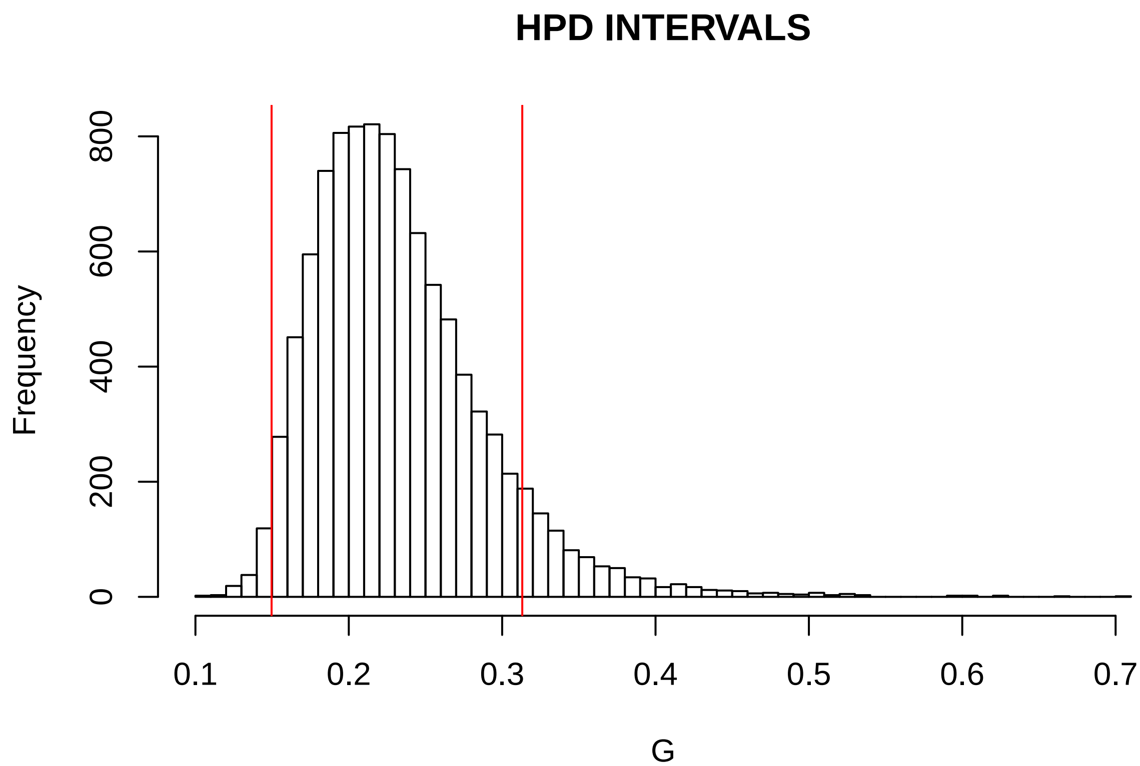


The posterior mean of obtained Gini coefficient is found to be 0.23 which is more closer to 0 than 1, hence the incomes are nearly equal.

c



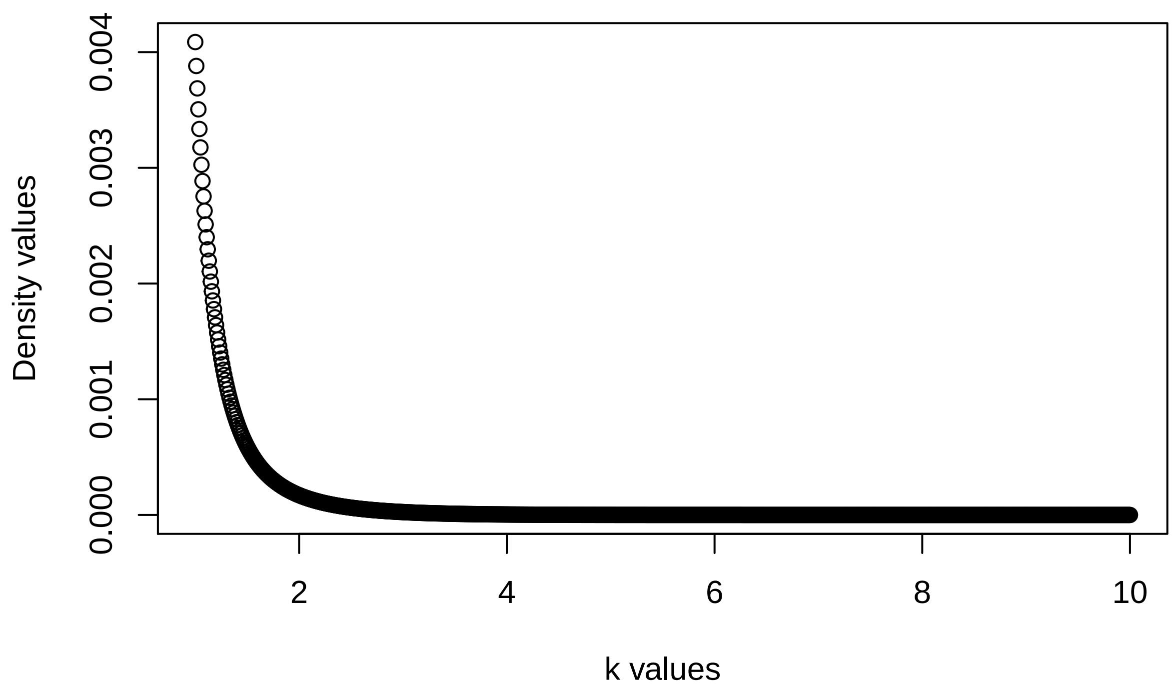

```
## The equal tailed lower and upper credible intervals is found to be
##  0.1611716 and 0.3365902 respectively
```




```
## The Highest Posterior Density lower and upper intervals is found to be
##   0.1496601 and 0.3131126 respectively
```

3. Bayesian inference for the concentration parameter in the von Mises distribution.

a



The above plot depicts the posterior distribution of k for the wind direction data over a fine grid of k values.

b

```
## Approximate posterior mode of k is found to be  0.00408702
```

The approx posterior mode is found from above calculation goes in par with the values of posterior of k values and also from the graph. The value is very close to 0.

Appendix

```
knitr::opts_chunk$set(echo = TRUE,warning = FALSE, message = FALSE)
n <- 20
ao <- 2
bo <- 2
p <- 5
f <- n-p
a_pos <- ao+p
b_pos <- bo+f
actual_mu <- (a_pos)/(a_pos + b_pos)
actual_sd <- sqrt((a_pos * b_pos)/((a_pos + b_pos)^2 * (a_pos+b_pos+1)))
j <- 1
```

```r
post_mu <- numeric()
post_sd <- numeric()
for(i in seq(1,15000,100))
{
  post_rand_num <- rbeta(i,a_pos,b_pos)
  post_mu[j] <- mean(post_rand_num)
  post_sd[j] <- sd(post_rand_num)
  j <- j+1
}

plot(seq(1,15000,100),post_mu,xlab = "Mean",ylab = "Random Sample numbers")
abline(h=actual_mu, col = "red")
plot(seq(1,15000,100),post_sd,xlab = "Sd",ylab = "Random Sample numbers")
abline(h = actual_sd,col="blue")
Ndraws <- 10000
rand_beta <- rbeta(Ndraws,a_pos,b_pos)
rand_prob <- length(which(rand_beta > 0.3)) / length(rand_beta)
cat("The posterior probability for p(theta > 0.3) using simulation with Ndraws = 10000 is :",rand_prob 
prob_beta <- 1 - pbeta(0.3,a_pos,b_pos)
cat("The exact posterior probability for p(theta > 0.3) is :",prob_beta )
rand_log_odd <- log(rand_beta / (1- rand_beta))
h <- hist(rand_log_odd,main = "Histogram of Posterior Distribution")
xfit <- seq(min(rand_log_odd), max(rand_log_odd), length = 40)
yfit <- dnorm(xfit, mean = mean(rand_log_odd), sd = sd(rand_log_odd))
yfit <- yfit * diff(h$mids[1:2]) * length(rand_log_odd)
lines(xfit, yfit, col = "black", lwd = 2)
y <- c(44,25, 45, 52, 30, 63, 19, 50, 34,67)
mu <- 3.7
s2 <- var(y)
n <- length(y)
library(geoR)

## Theoritical calculation
tau_sq <- sum((log(y)-mu)^2)/n
theo_sigma_sq <- rinvchisq(10000,df = n-1, scale = tau_sq)


## Simulation
post_sim <- function(Ndraws,n,tau_sq)
{
  sigma_sq <- numeric()
  for(i in 1:Ndraws)
  {
    X <- rchisq(1,n)
    sigma_sq[i] <- ((n) * tau_sq) / X
    #theta[i] <- rnorm(mu,(sigma_sq/n))
  }
  return(sigma_sq)
}
sigma_sq <- post_sim(10000,n,tau_sq)
par(mfrow = c(2,1))
plot(density(sigma_sq),main = "Density curve of Simulated variance")
plot(density(theo_sigma_sq),main = "Density curve of Theortical variance")
```

```
#hist(theo_sigma_sq,50,main = "Histogram of Theoretical Sigma values")
#hist(sigma_sq,50,main = "Histogram of Simulated Sigma values")
#plot(density(sigma_sq))
#mean(theta)
library(bayestestR)
G <- 2 * pnorm(sqrt(sigma_sq/2),0,1) - 1
plot(density(G),main = "Density curve of Posterior distribution of the Gini coefficient G")
intervals <- quantile(sort(G),probs = c(0.05,0.95))
hist(G, 50, main = "ETL INTERVALS")
abline(v = intervals[1],col = "red")
abline(v = intervals[2],col = "red")
cat("The equal tailed lower and upper credible intervals is found to be \n", intervals[1],"and",interval

#high_pos <- quantile(G_den_val$y,probs = c(0.1))
G_den_val <- density(G)
G_y_sorted <- sort(G_den_val$y,decreasing = TRUE)
G_prob_cdf <- cumsum(G_y_sorted)/sum(G_y_sorted)
G_ind_y <- min(G_y_sorted[which(G_prob_cdf < 0.9)])
#ind_great <- which(G_den_val$y > high_pos)
HPD_range <- range(G_den_val$x[which(G_den_val$y >G_ind_y )])
hist(G,50,main = "HPD INTERVALS")
abline(v = HPD_range[1], col = "red")
abline(v = HPD_range[2], col = "red")


# kern_den_est <- density(G)
# CI_HDI <- ci(G,method = "HDI",ci = 0.90)
#abline(h = quantile(sort(G),probs = c(0.1)))
cat("The Highest Posterior Density lower and upper intervals is found to be\n ", HPD_range[1],"and",HPD_

pos_von_mis <- function(y,mu2,k)
{

  likel <- numeric()
  prior <- exp(-k)
  post_k <- numeric()
  for(i in 1:length(k))
  {
    likel[i] <- 1
    for(j in 1:length(y))
    {
      num <- exp(k[i] * cos(y[j] - mu2))
      deno <- 2 * pi * besselI(k[i],1)
      likel[i] <- likel[i] * num / deno
    }
    post_k[i] <- likel[i] * prior[i]
  }
  #post_k <- prior * likel

  #return(list("post_k" = post_k,"likel" = likel,"prior"=prior))
  return(post_k)
}
y <- c(-2.44,2.14,2.54,1.83,2.02,2.33,-2.79,2.23,2.07,2.02)
```

```r
mu2 <- 2.39
k <- seq(1,10,0.01)
post_k <- pos_von_mis(y,mu2,k)
#post_k <- exp(k * sum(cos(y-mu))-1) / besselI(k,0,expon.scaled = TRUE)
#hist(post_k)
#plot(density(post_k))
plot(x = k, y = post_k, xlab = "k values", ylab = "Density values")
#plot(x = k,y = prior_likel_post$likel, col = "red")
#plot(x= k, y = prior_likel_post$prior, col = "blue")

mode_approx <- post_k[which.max(tabulate(post_k))]
cat("Approximate posterior mode of k is found to be ",mode_approx)
```