Computer Lab 1

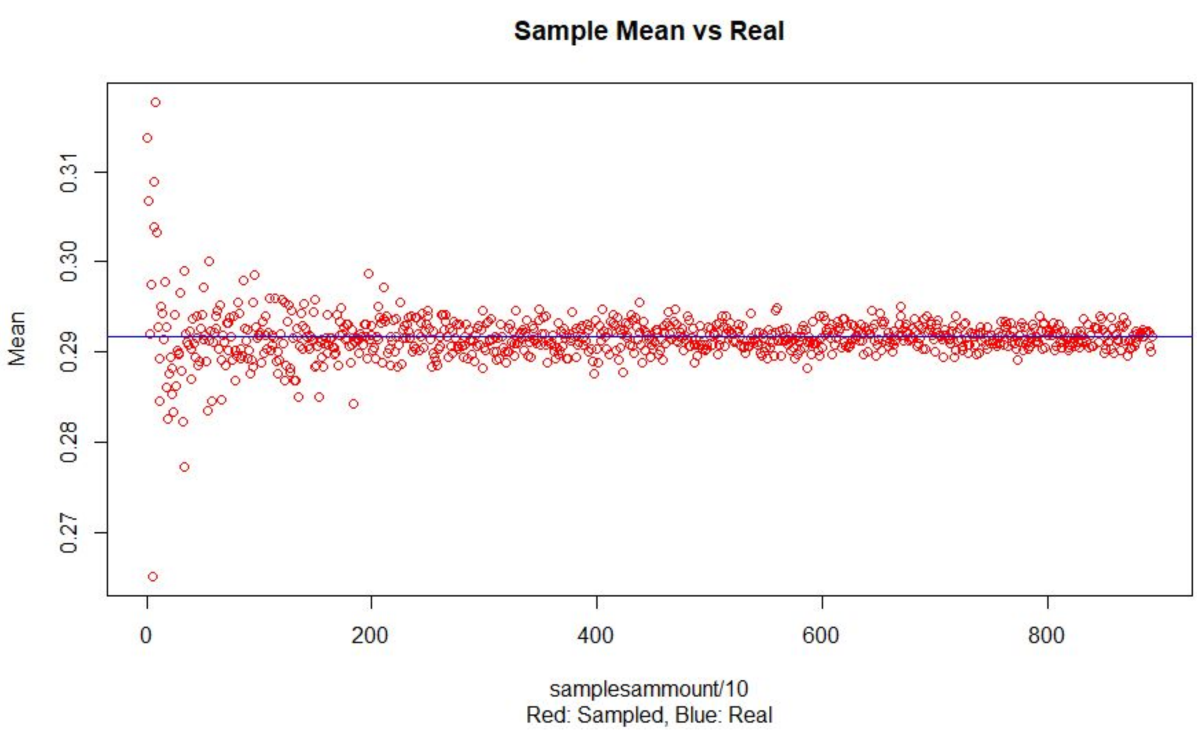
Question 1

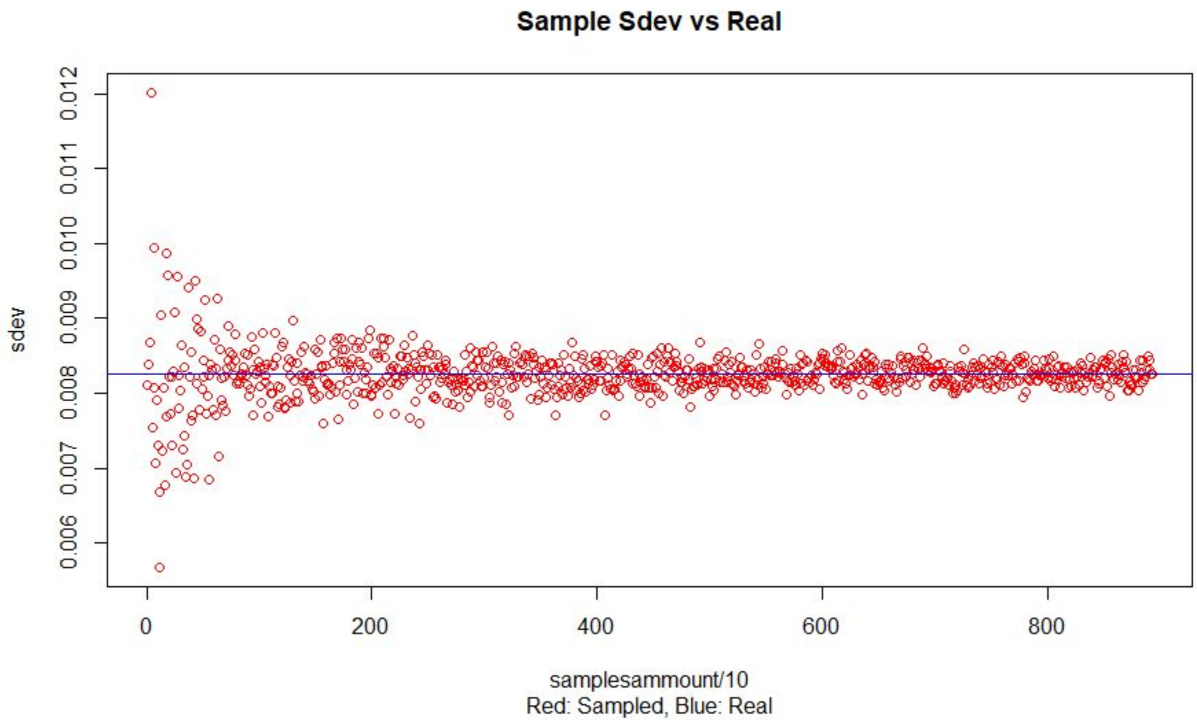
a) **Draw random numbers from the posterior θ|y ~ Beta(α0 + s, β0 + f), y = (y1, . . . , yn), and verify graphically that the posterior mean and standard deviation converges to the true values as the number of random draws grows large.**

We draw samples from the beta distribution with R-function rbeta with parameters given from the assignment. a = 5 + 2 and b = 15 +2.

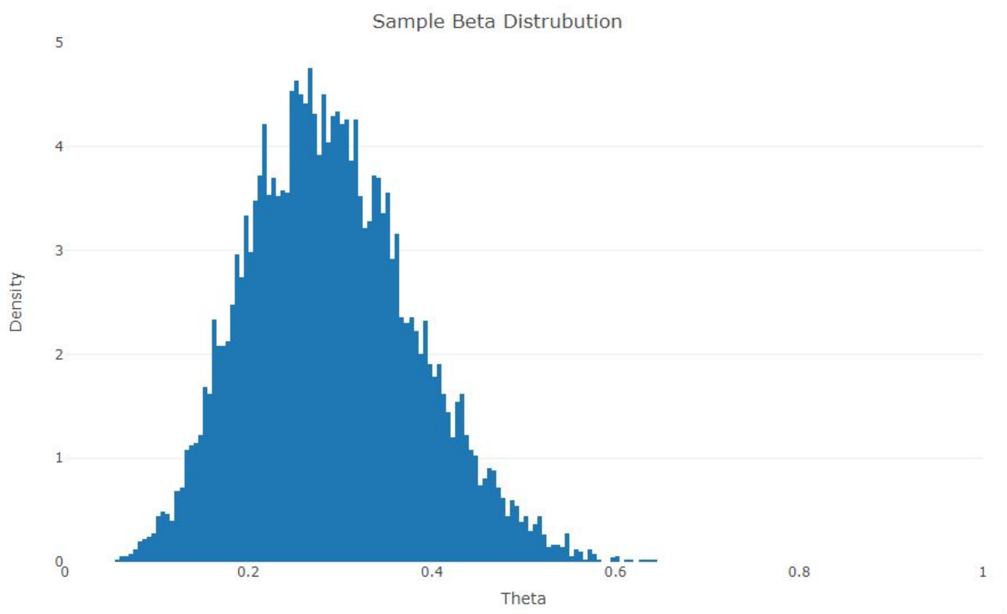
We increase the amount of samples from the distribution until the sample mean and sample variance is within an acceptable threshold from the distribution's expected value and standard deviation. The resulting amount of samples needed varies somewhat, but the result plotted is with 18390 samples. The resulting sample mean and variance is within 0.00001 and 0.00001 difference from the expected values.

We can see from the plots below that the sampled converges mostly after 5000 samples, and after 18390 the sample mean and variance are within our threshold limits.

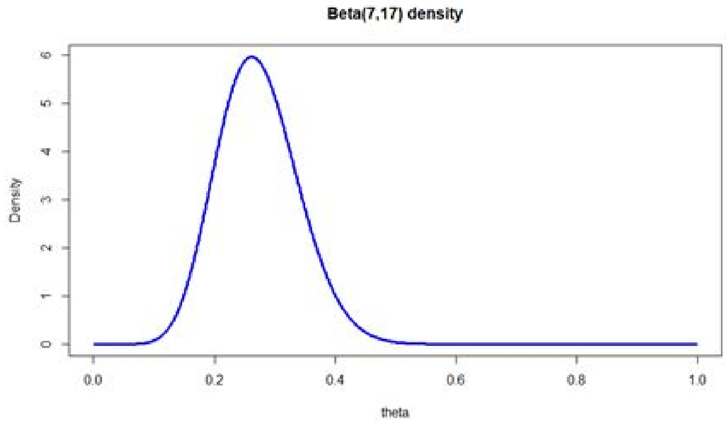




Sample Sdev vs Real
sdev
0.006
0.007
0.008
0.009
0.010
0.011
0.012
0
200
400
600
800
samplesammount/10
Red: Sampled, Blue: Real



Sample Beta Distrubution
Density
0
1
2
3
4
5
0.2
0.4
0.6
0.8
1
Theta



Comparing our sample beta distribution with the real beta(7,17) we can see that the plots are very similar.

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

We draw 10000 samples from the beta distribution in the same manner as a. We then estimate the probability for theta>0.3 by counting the number of observations that are larger than 0.3 in our samples and dividing by the total number of samples.

sampleProb = length(which(posterior>0.3))/length(posterior)

sampleProb ~= 44.26%

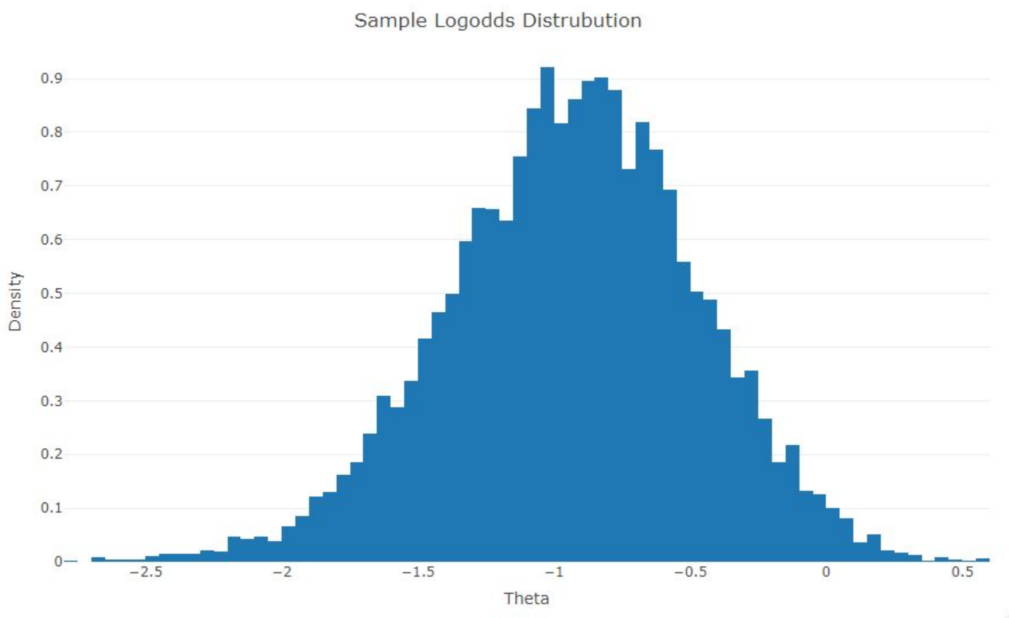
We then compare with the real probability by realProb = pbeta(0.3,a,b,lower.tail = FALSE)

realProb ~= 43.39%

The differences in probability is:

|sampleProb - realProb| = |44,26 - 43,39| = 0,87

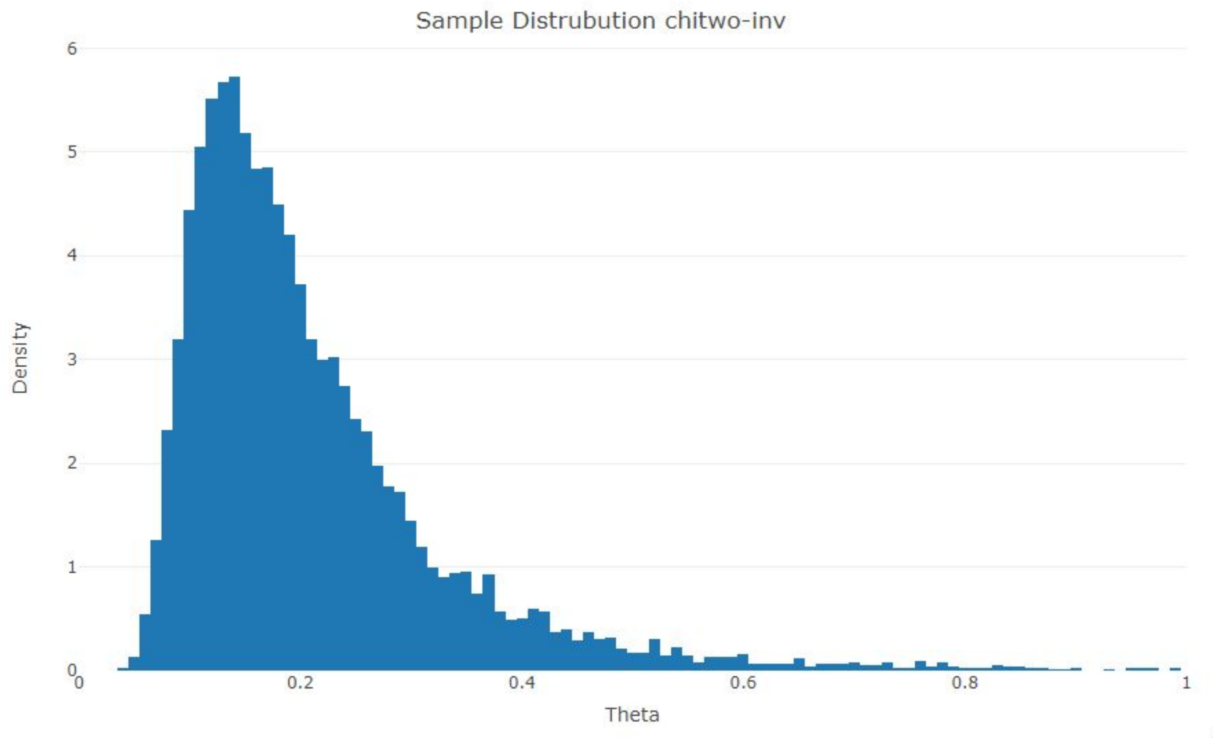
c) **Compute the posterior distribution of the log-odds $\phi = \log \theta\ 1-\theta$ by simulation (nDraws = 10000). [Hint: hist() and density() might come in handy]**

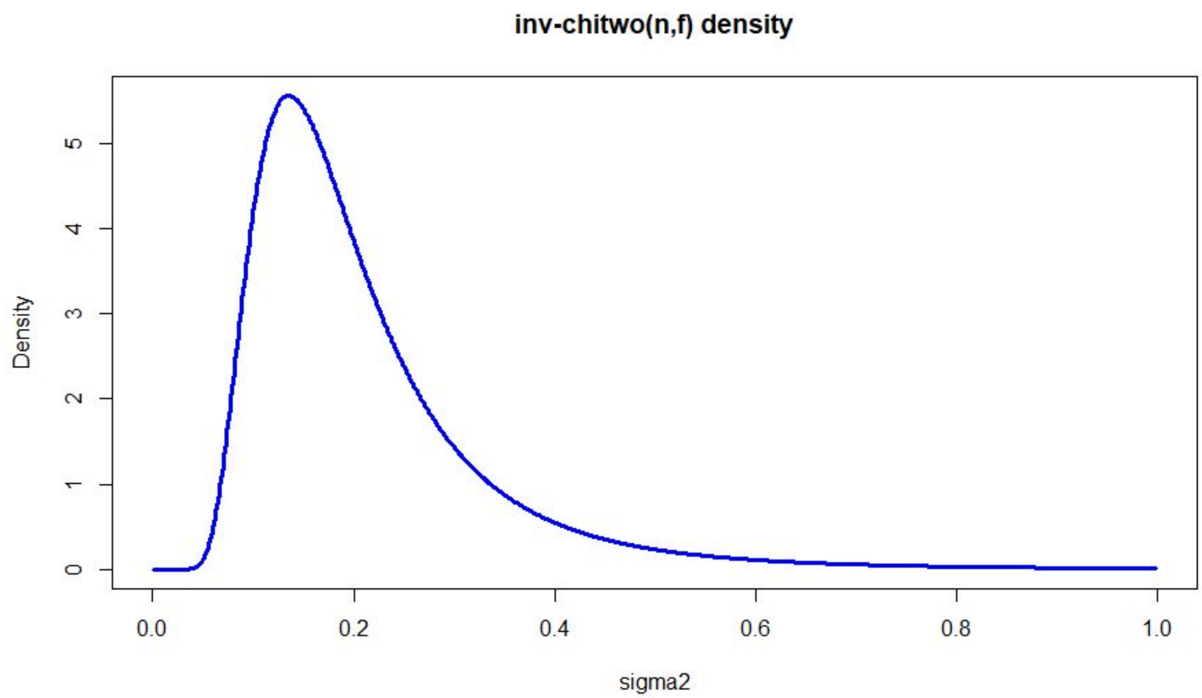


x	y
Min. :-3.14478	Min. :0.0000069
1st Qu.:-2.08994	1st Qu.:0.0039959
Median :-1.03509	Median :0.0702101
Mean :-1.03509	Mean :0.2367704
3rd Qu.: 0.01975	3rd Qu.:0.4469608
Max. : 1.07459	Max. :0.8571856

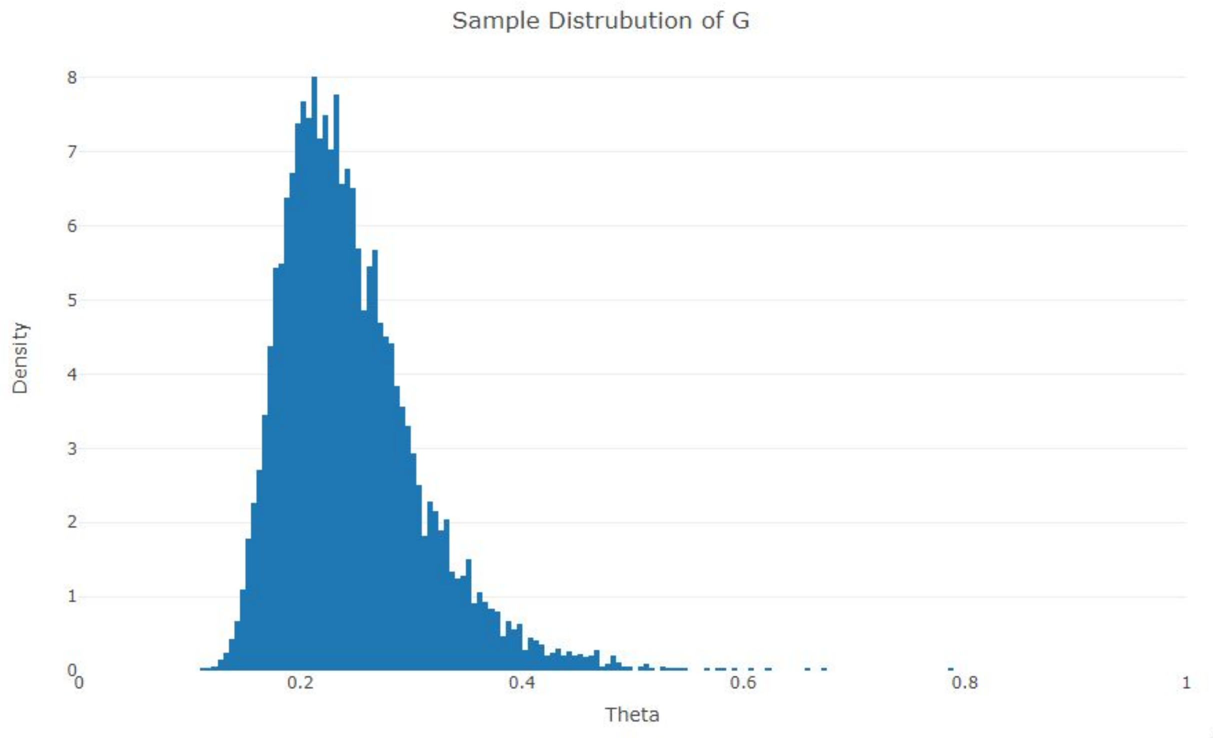
Question 2

a) Simulate 10, 000 draws from the posterior of σ 2 (assuming μ = 3.7) and compare with the theoretical Inv − χ 2 (n, τ 2) posterior distribution.





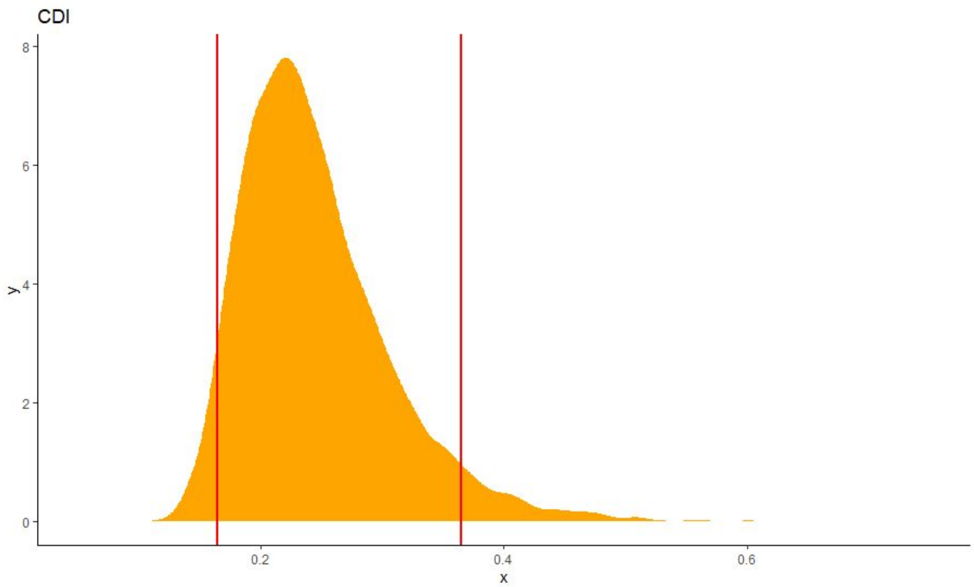
b) Use the posterior draws in a) to compute the posterior distribution of the Gini coefficient G for the current data set.



We compute the sampled G values for each of our sigma-squared observations from the inv-chitwo posterior.
G = 2*pnorm(sqrt(samples)/sqrt(2))-1

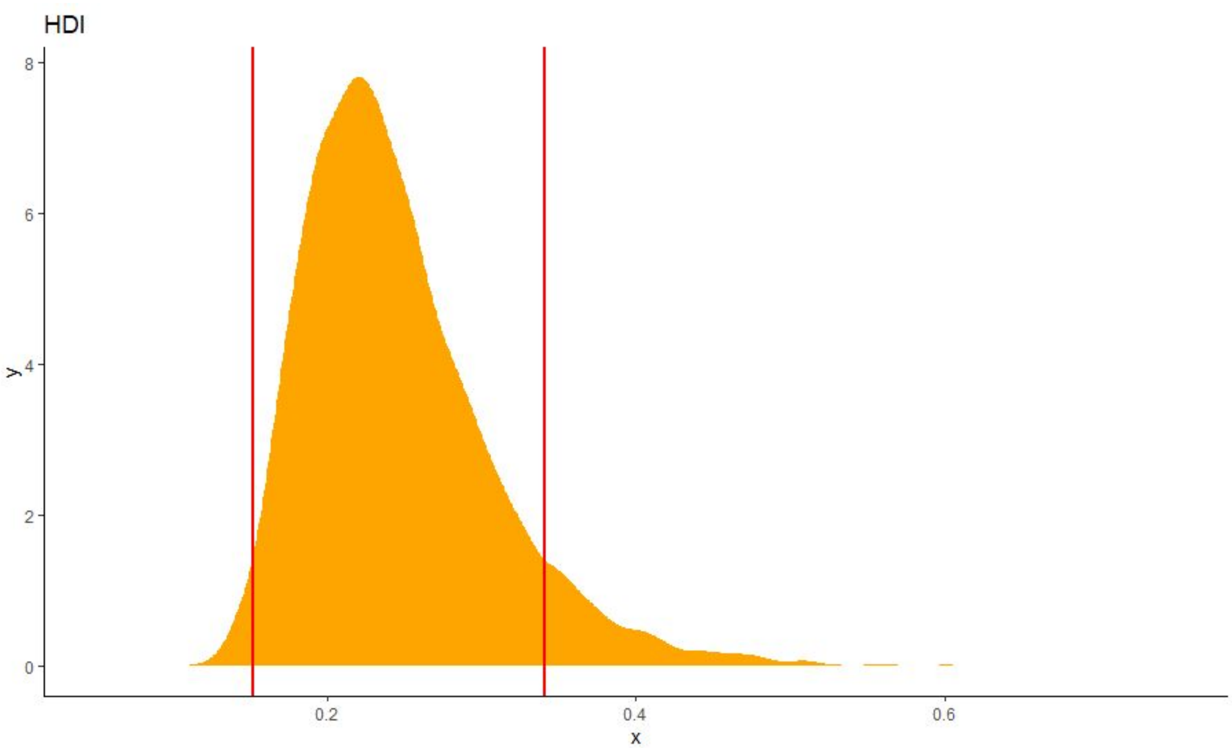
c) **Use the posterior draws from b) to compute a 90% equal tail credible interval for G. Also, do a kernel density estimate of the posterior of G using the density function in R with default settings, and use that kernel density estimate to compute a 90% Highest Posterior Density interval for G. Compare the two intervals**

To compute the Equal Tail Credible interval for G we first use the density function to estimate the probability density over each G value. We then use an iterative function that starts from the mean and includes indices further to the left until the left tail equals 5% of the total mass and vice versa for the right tail.



The CDI spans G = [0.1645958 : 0.3650526]

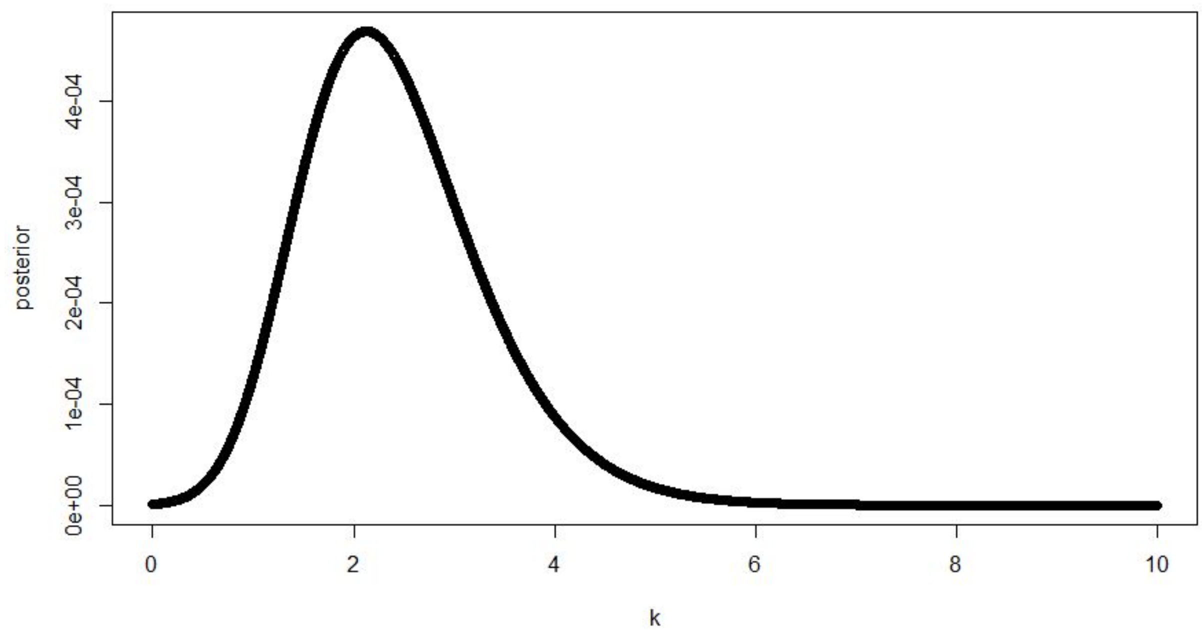
To compute the Highest Posterior Density we constructed our own function that iterates from the G distribution mode and outwards. The next G value to include, starting from the mode, is decided by looking at the closest theta to the left and to the right and including the one with the largest function value. This is done with a while loop until the sum of the function values outside the extreme points of the included Gis equal or above 0,1.



The HDI spans G = [0.1522979 : 0.3404567]

Question 3

a) **Plot the posterior distribution of κ for the wind direction data over a fine grid of κ values**



First we draw many samples of k from 0 to 10. We then calculate the posterior for each of theese k-values. We do this by first calculating the likelihood p(**y**|k) = $\Pi p(y|k)$. Then we multiplicate the likelihood with the prior to get the final posterior value for one specific k according to bayes rule. p(k|**y**) = p(**y**|k)*p(k) where p(y|k) is the von maise distrubtion with mu=2.39 and p(k) being the probability density function of exp(1) at k.

b) **Find the (approximate) posterior mode of κ from the information in a).**

k[which.max(posterior)] = 2.125