

TDDE07 - Lab 1

Pontus Svensson (ponsv690) & Kristian Sikiric (krisi211)

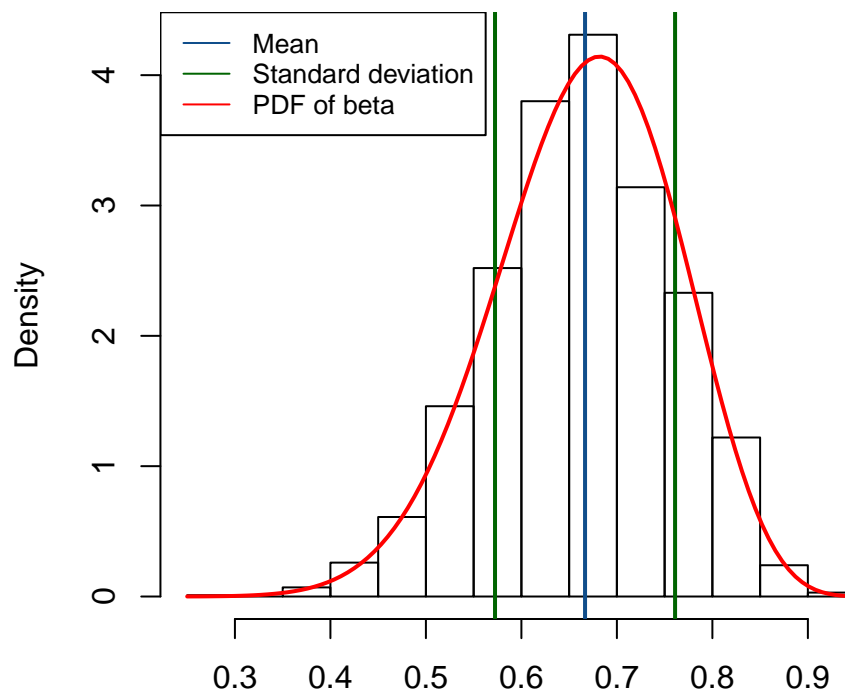
Assignment 1

Bernoulli ... again.

In this assignment we assumed to have obtained a Bernoulli distributed sample of 14 successes and 6 failures, 20 trials in total. We assumed a $Beta(\alpha, \beta)$ prior with $\alpha = \beta = 2$.

For the first part of the assignment, random numbers from the posterior Beta distribution were drawn. In the plot below, the histogram is the randomly sampled beta distribution, the blue line is the true mean, the green lines are the standard deviation and the red line is the PDF of the beta distribution. When the number of draws increases, the mean of the histogram converges to the true mean, and the same for the standard deviation. The plot shows a histogram after 2000 draws.

Histogram of beta distribution

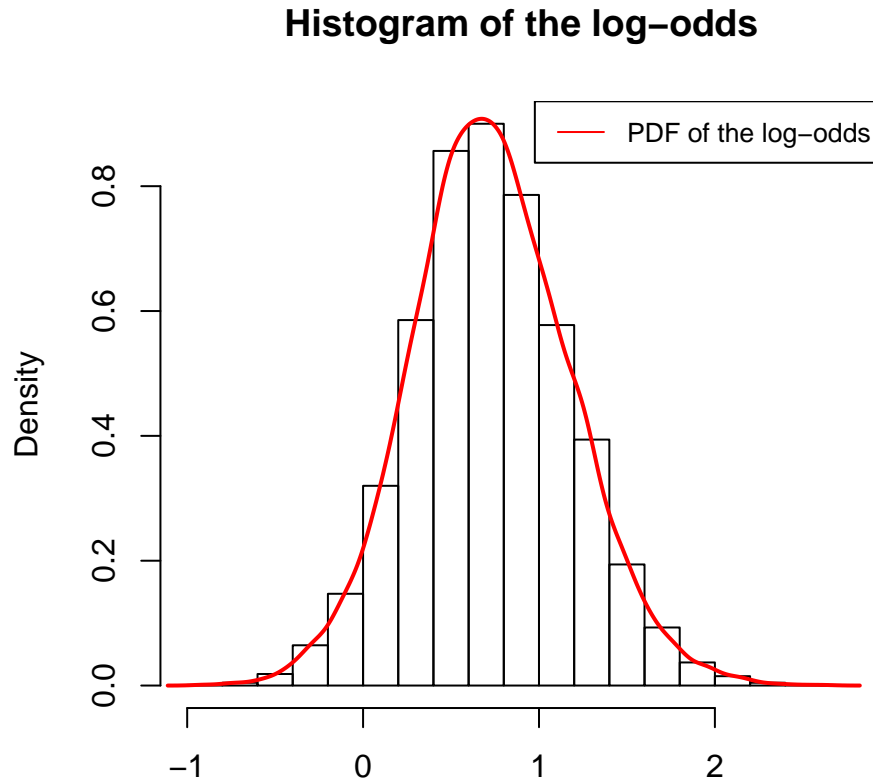


For the second part of the assignment the posterior probability were to be calculated by simulation by 10000 draws and compared to the exact value. The results are shown below. When the draws are increased the simulated probability converges to the exact value.

```
## [1] "Simulated probability: 0.47"
```

```
## [1] "Exact probability: 0.397268082810898"
```

For the final part of the assignment, the posterior log-odds distribution were to be computed by simulation with 10000 draws. The result is shown below.



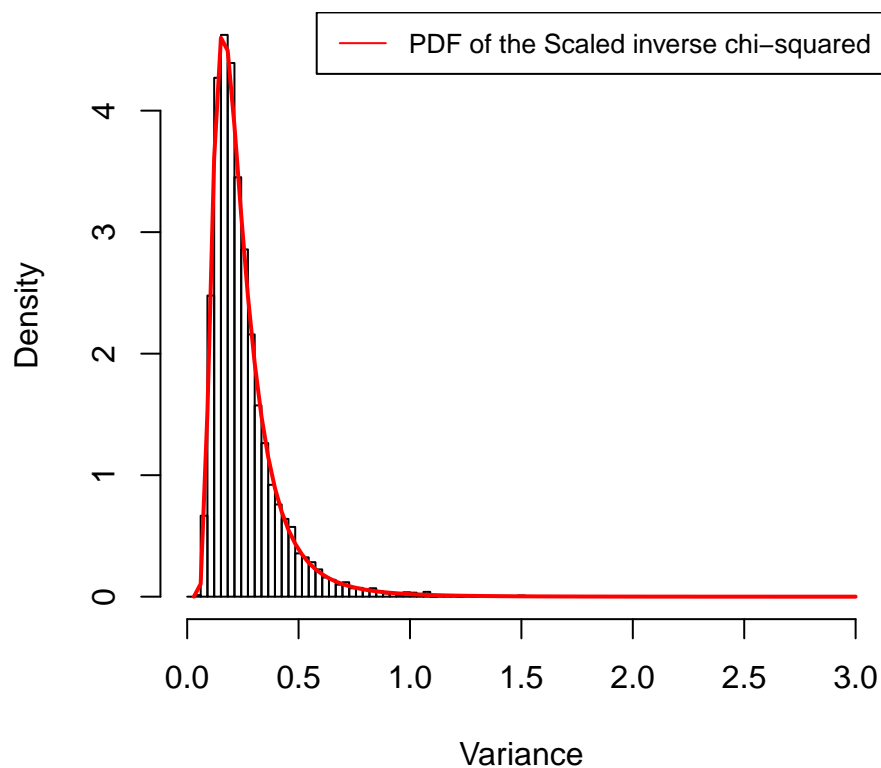
Assignment 2

Log-normal distribution and the Gini coefficient

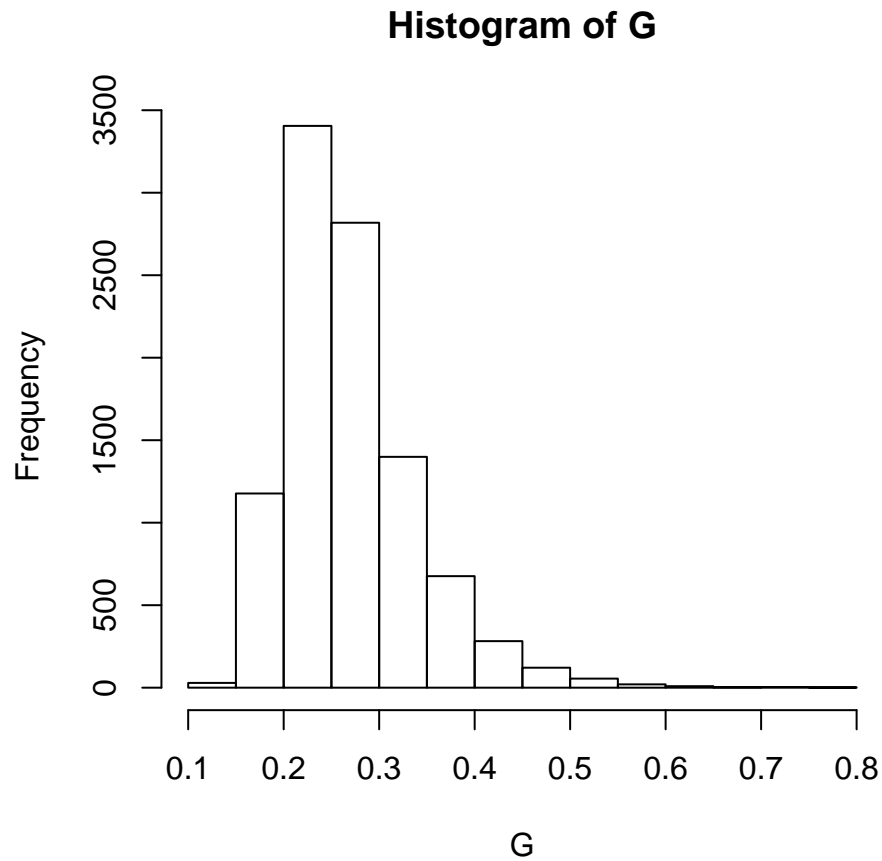
In this assignment ten observations of monthly incomes were given. This kind of data is said to be normally distributed with $\mu = 3.5$ and unknown variance σ^2 with prior $p(\sigma^2) \propto 1/\sigma^2$. The posterior for σ^2 is the $Inv - \chi^2(n, \tau^2)$ distribution.

10000 draws were simulated from the posterior of σ^2 , which can be seen in the histogram below. The red line in the histogram is the theoretical scaled inverse chi-squared distribution.

Histogram of variance



By using the posterior draws in the previous task the posterior distribution of the Gini coefficient G for the current data set was calculated. It was assumed that $G = 2\Phi(\sigma/\sqrt{2}) - 1$ when incomes follow a $\log \mathcal{N}(\mu, \sigma^2)$. The results are shown in the histogram below.



Now the 95% equal tail credible interval and the 95% Highest Posterior Density interval for G was calculated and plotted. The intervals ended up being very similar, which is resonable since the PDF is quite symmetric.

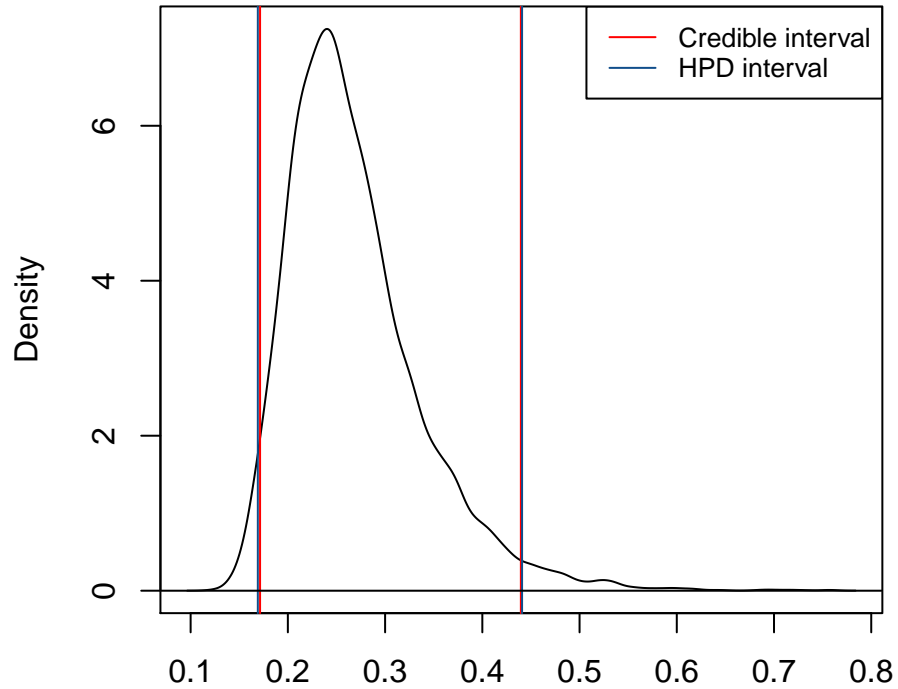
```
## [1] "Lower threshold of the credible interval: 0.171413594743392"
```

```
## [1] "Upper threshold of the credible interval: 0.43964333604288"
```

```
## [1] "Lower threshold of the highest posterior density interval: 0.16909501185074"
```

```
## [1] "Upper threshold of the highest posterior density interval: 0.440858656923549"
```

PDF of G



Assignment 3

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

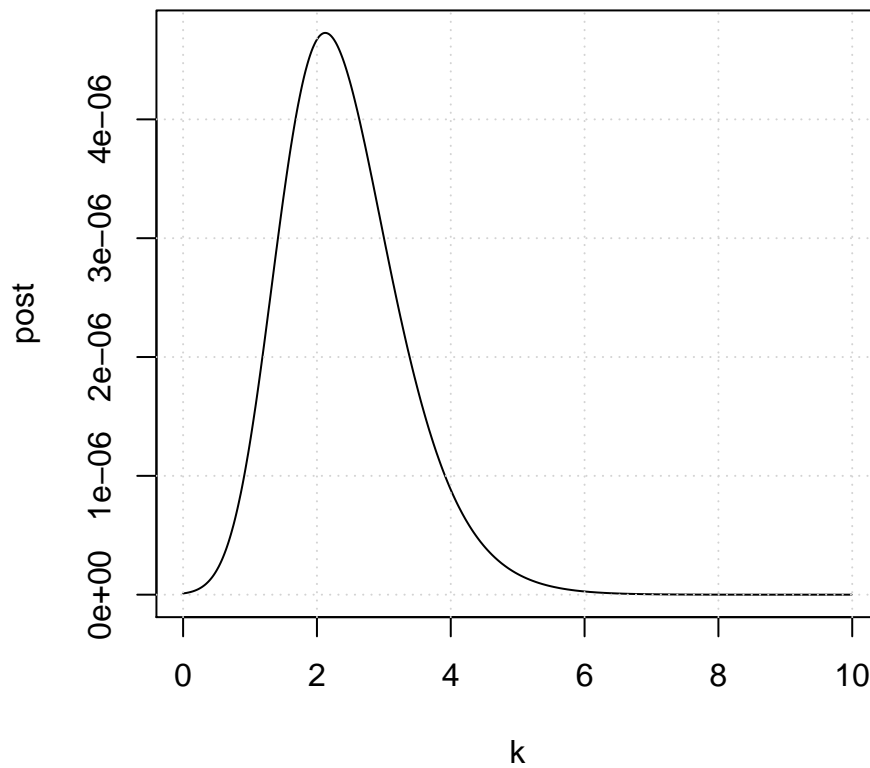
In this exercise we were given a set y of measured wind directions converted into radians.

$$y = (-2.44, 2.14, 2.54, 1.83, 2.02, 2.33, -2.79, 2.23, 2.07, 2.02)$$

These data points were assumed to be independent observations following the von Mises distribution.

$$p(y|\mu, \kappa) = \frac{\exp[\kappa \cos(y - \mu)]}{2\pi I_0(\kappa)}$$

We assumed that $\mu = 2.39$ and that κ is exponentially distributed. We were now supposed to plot the posterior of κ , this was done by multiplying the prior of κ with the likelihood function described above. The following plot is the result of this.



From this plot, we can approximate the mode to be around 2.

Appendix

Assignment 1

```
### 1 a)
set.seed(235)

# Variables
s = 14
n = 20
a0 = b0 = 2
f = n - s
a_n = a0 + s
b_n = b0 + f

rand.samples = rbeta(n*100, a_n, b_n)

# True values
mean = (a_n) / (a_n+b_n)
```

```

variance = ((a_n) * (b_n)) / (((a_n) + (b_n))^2 * ((a_n) + (b_n) + 1))
stdev = variance^0.5

# Plots
hist(rand.samples,freq=FALSE,main = "Histogram of beta distribution")
abline(v=mean, col = "blue")
abline(v=mean-stdev, col = "green")
abline(v=mean+stdev, col = "green")
x = seq(0,1,length=100)
curve(dbeta(x,a0+s,b0+f),add=TRUE,col="red")

### 1 b)
set.seed(235)

nDraws = 10000
rand.samples = rbeta(nDraws, a_n, b_n)
p_calc = 100*sum(rand.samples<0.4)/length(rand.samples)
p_true = 100*pbeta(0.4, a0+s,b0+f)

### 1 c)

prior = beta(a0, b0)
hist(log(rand.samples/(1-rand.samples)), freq=FALSE)
lines(density(log(rand.samples/(1-rand.samples))))

```

Assignment 2

```

### 2 a)
set.seed(235)

draws = 10000
income = c(14, 25, 45, 25, 30, 33, 19, 50, 34, 67)
n = 10
v = 9

y = log(income)
tau2=sum((y-3.5)^2)/10
tau = tau2**(0.5)

var<-((n-1)*tau2)/rchisq(draws,n-1)
hist(var,breaks = seq(0,3,length=100),freq=FALSE, xlab="Variance", main="Histogram of variance")
theta = seq(0,3,length=100)
p = ((v/2)^(v/2))*(tau^v)*(theta^-((v/2)+1))*exp((-v*tau2)/(2*theta))
p = p/gamma(v/2)
lines(theta,p,col="red", lwd=2)
legend("topright", col = c("red"), legend=c("PDF of the Scaled inverse chi-squared"),lty=1)

### 2 b)
set.seed(235)

G = 2*pnorm(var**0.5/2**0.5)-1
hist(G)

```

```

### 2 c)
set.seed(235)

q = quantile(G, probs=c(0.025, 0.975))
print(paste("Lower threshold of the credible interval: ", q[1]))
print(paste("Upper threshold of the credible interval: ", q[2]))
newG = G[which(G > q[1])]
newG = newG[which(newG < q[2])]
plot(density(G), main="PDF of G", xlab="")
legend("topright", col = c("red", "dodgerblue4"), legend=c("Credible interval", "HPD interval"), lty=1)
abline(v=q[1], col="red")
abline(v=q[2], col="red")

dx <- density(G)
dn <- cumsum(dx$y)/sum(dx$y)
li <- which(dn>=0.025)[1]
ui <- which(dn>=0.975)[1]
dx$x[c(li, ui)]
print(paste("Lower threshold of the highest posterior density interval: ", dx$x[c(li, ui)][1]))
print(paste("Upper threshold of the highest posterior density interval: ", dx$x[c(li, ui)][2]))
abline(v = 0.1690950, col="dodgerblue4")
abline(v = 0.4408587, col="dodgerblue4")
abline(h=0)

```

Assignment 3

```

set.seed(235)

data = c(-2.44, 2.14, 2.54, 1.83, 2.02, 2.33, -2.79, 2.23, 2.07, 2.02)

mu = 2.39
k = seq(0.001, 10, 0.01)

calc_posterior = function(k){
  prior = dexp(k)
  likelihood = exp(k * cos(data - mu))/(2 * pi * besseli(k, 0))
  post = prod(likelihood) * prior

  return(post)
}

post = sapply(k, calc_posterior)

plot(k, post, type = "l")
grid()

```