Exam file

Jacob Lundberg, Sebastian Callh

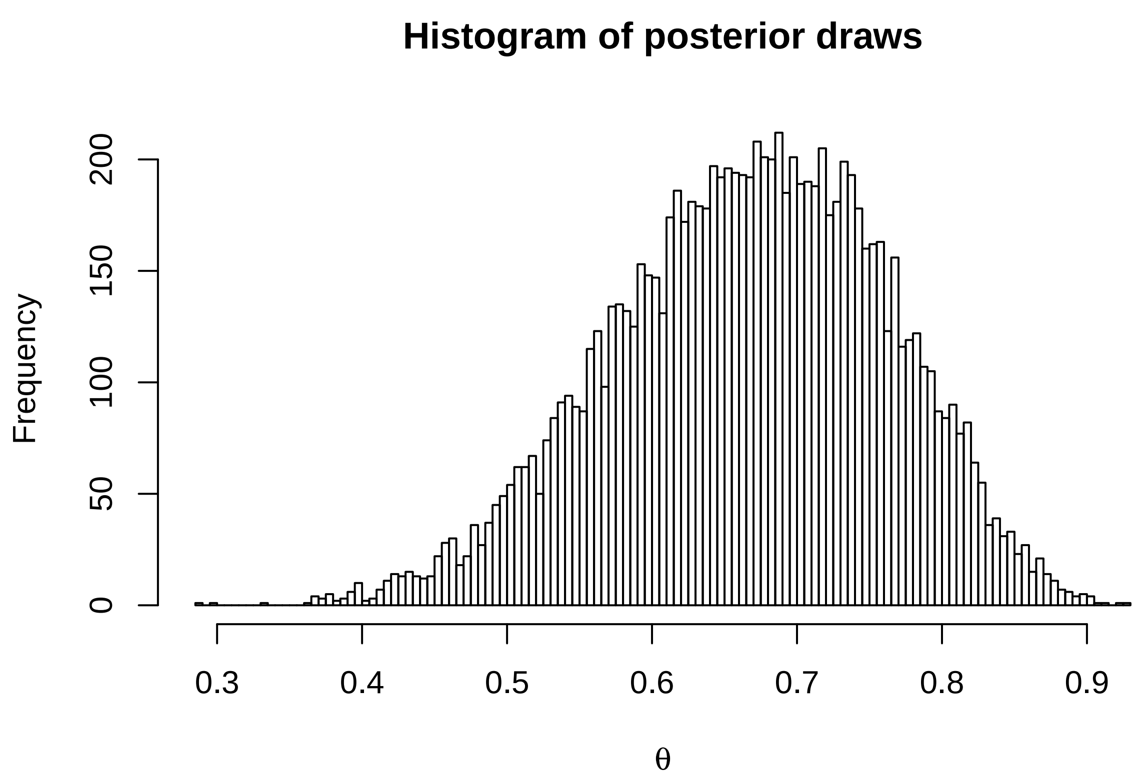
25 maj 2018

——- Lab 1 ——-

1. Bernoulli . . . again

a)

The mean of a Bernoulli distribution is given by $p = \frac{s}{n} = \frac{14}{20} \approx 0.7$ which is very close to the mean of the simulated posterior distribution plotted below.



The posterior mean and the true mean are close to each other and increasing the number of draws will let them come arbitrarily close.

```
## [1] "Mean of posterior draws 0.666574"
```

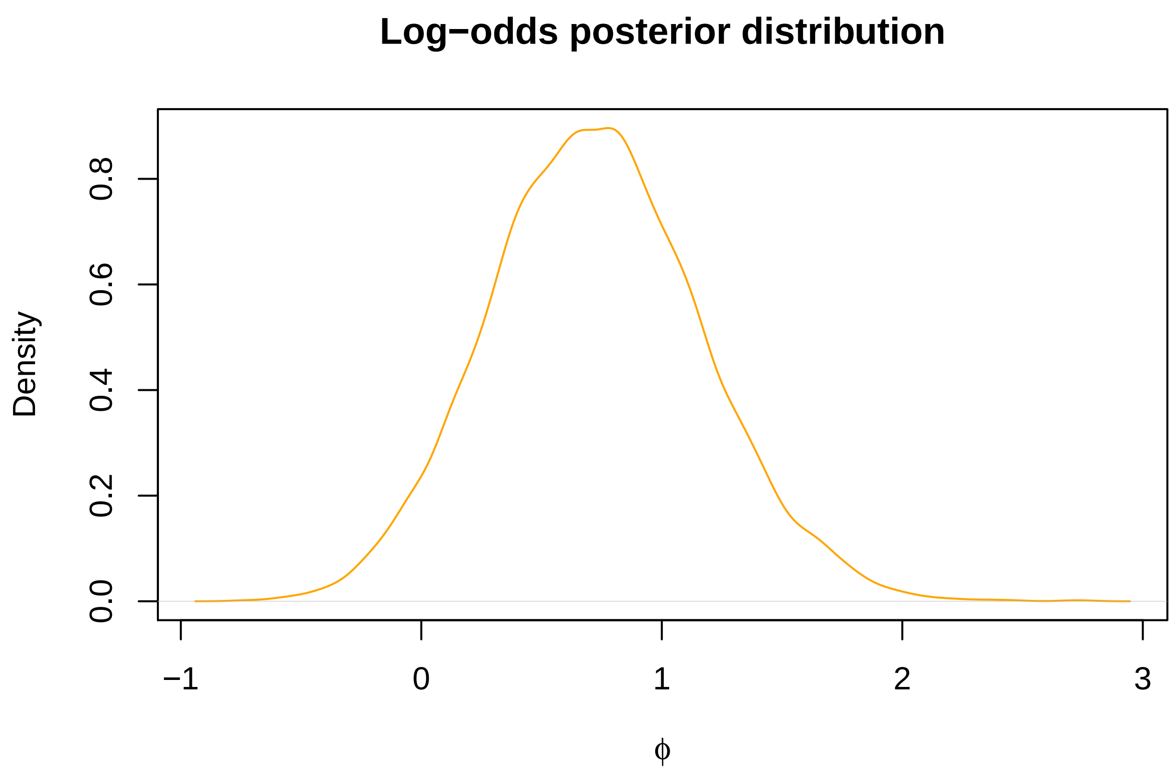
b)

The posterior probability and the true probability for $P(\theta < 0.4)$ lie very close to each other. As with the mean value, increasing the number of draws will let the posterior probability approach the true probability.

```
## [1] "Simulated probability of theta < 0.4: 0.003000"
## [1] "Exact probability of theta < 0.4: 0.003973"
```

c)

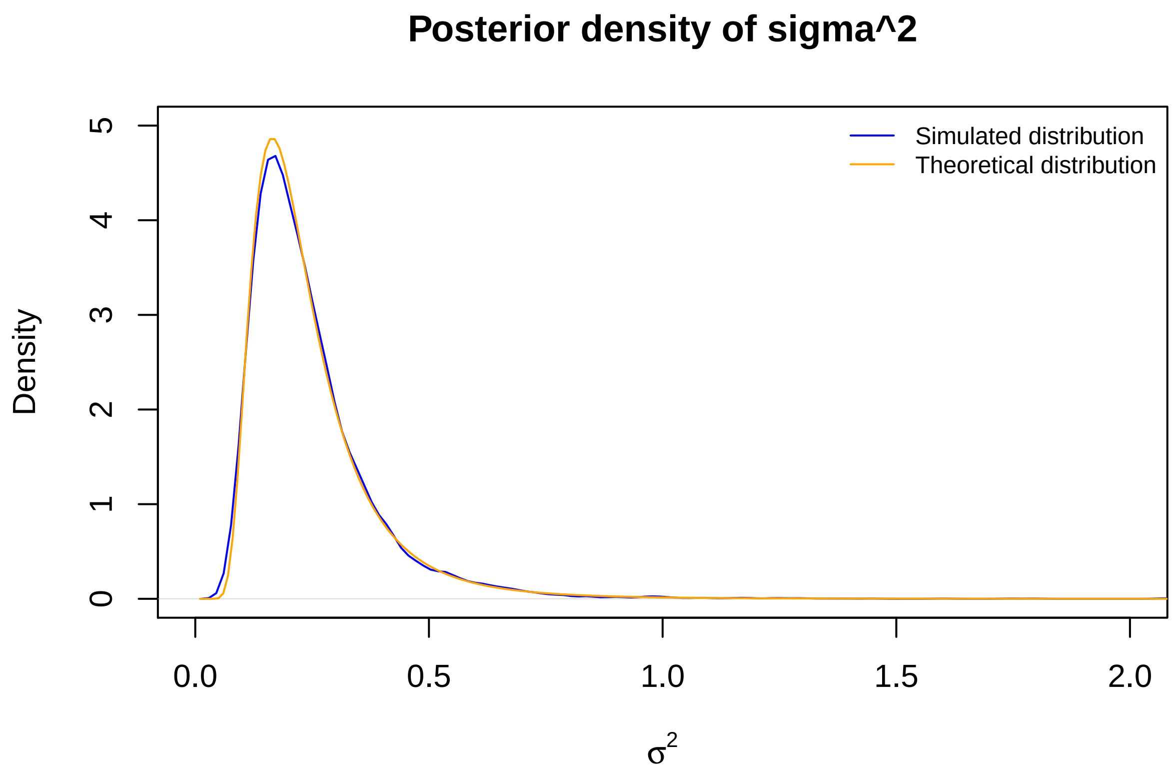
The log-odds posterior distribution can be seen in the plot below.



2. Log-normal distribution and the Gini coefficient.

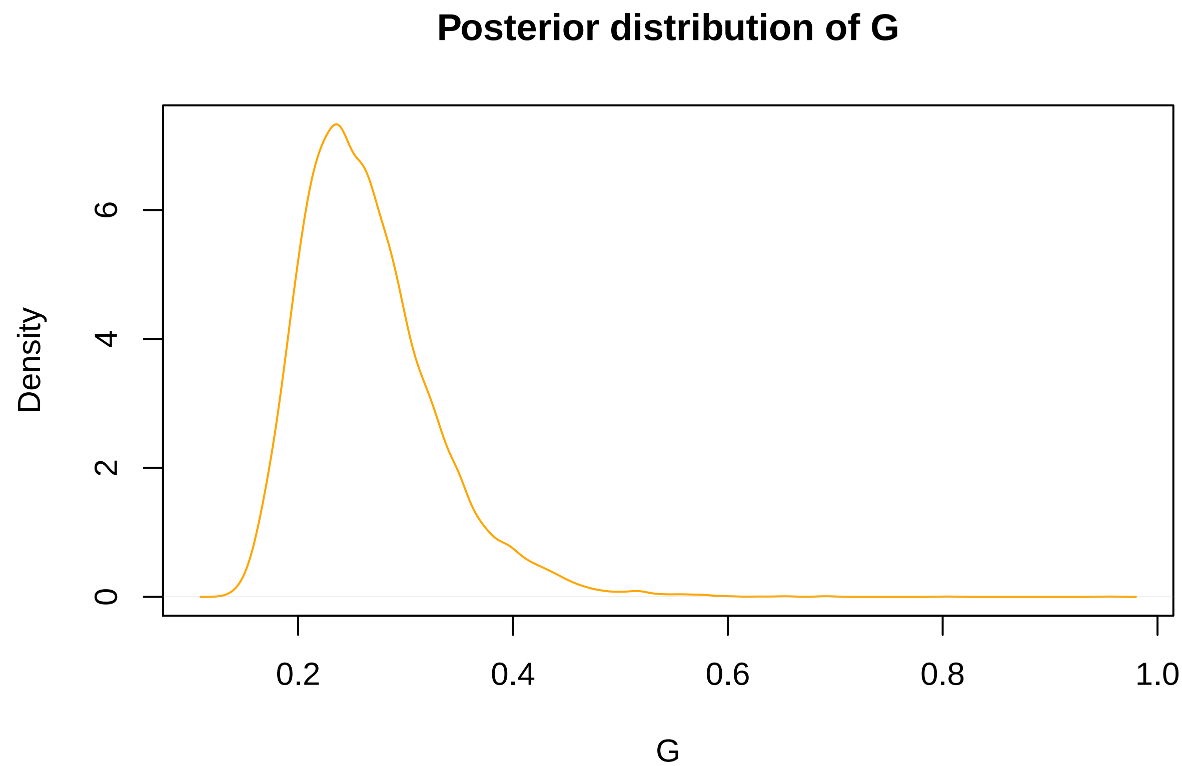
a)

Simulating from the posterior and plotting the approximated density together with the theoretical density reveals that the approximation is quite good, which can be seen in the plot below.



b)

Using the samples from a) and computing the posterior distribution of the Gini coefficient G results in the plot below.



c)

Computing the 95% credible interval results in

```
## [1] "Credible interval lower bound: 0.173441"
```

```
## [1] "Credible interval upper bound: 0.418847"
```

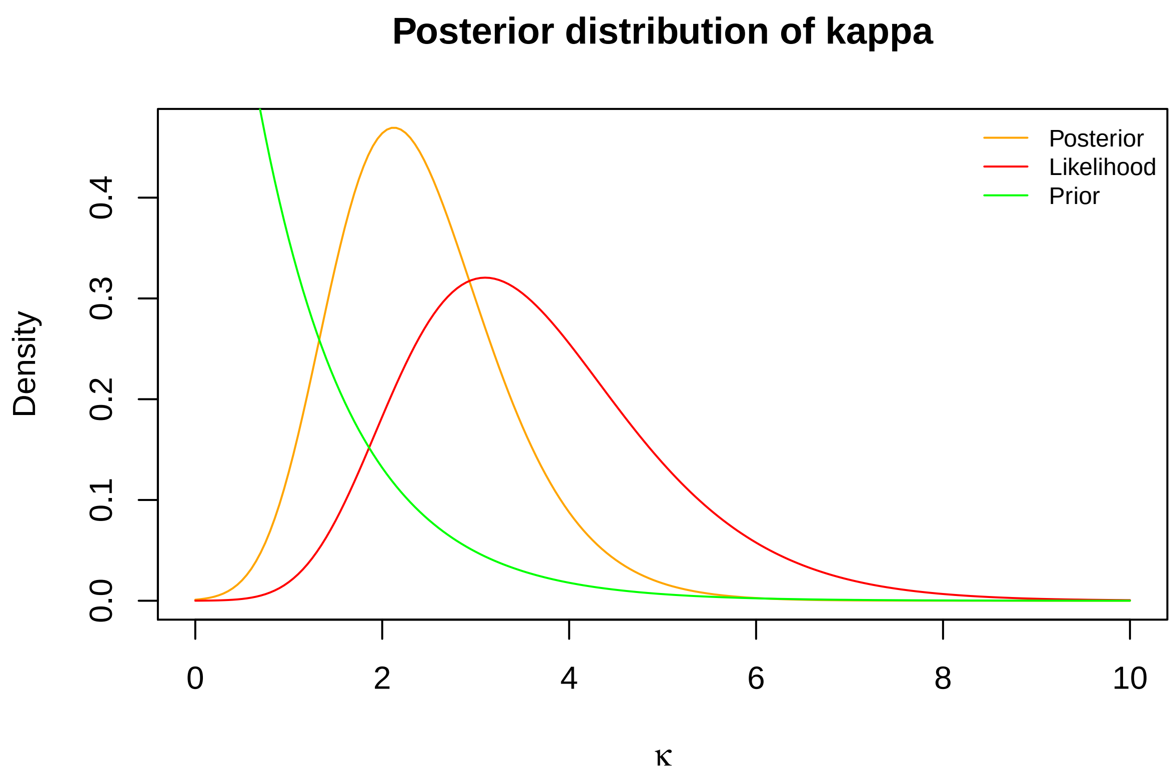
and computing the Highest Posterior Density interval results in

```
## [1] "HPD lower bound: 0.160149"
```

```
## [1] "HPD upper bound: 0.391934"
```

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

Plotting the posterior distribution of κ together with the prior and likelihood results in the plot below.



Which has it's mode at

```
## [1] "Posterior mode: 2.10"
```

```r
# 1
# a) Show that posterior converges to 14/20 approx. 0.7
alpha0 <- 2
beta0  <- 2
nDraws <- 10000
n      <- 20
s      <- 14
f      <- n - s
post   <- rbeta(nDraws, alpha0 + s, beta0 + f)
hist(post,
     breaks = 100,
     main   = "Histogram of posterior draws",
     xlab   = expression(theta))

sprintf("Mean of posterior draws %f", mean(post))

# b) Simulation to compute posterior probability P(theta < 0.4 | y)
thetas <- rbeta(nDraws, alpha0 +s , beta0 + f)
sprintf("Simulated probability of theta < 0.4: %f", length(thetas[thetas<0.4]) / length(thetas))  # Sim
sprintf("Exact probability of theta < 0.4: %f", pbeta(0.4, alpha0 + s, beta0 + f))                # Exa

# c) Log odds posterior distribution
phi <- sapply(thetas, function(theta) { log(theta / (1 - theta)) })
phi.dens <- density(phi)
plot(phi.dens,
     main = "Log-odds posterior distribution",
     ylab = "Density",
     xlab = expression(phi),
     col  = "orange")

# 2
incomes <- c(14, 25, 45, 25, 30, 33, 19, 50, 34, 67)
mu      <- 3.5
n       <- length(incomes)
tau2    <- sum((log(incomes) - mu)^2) / n
nDraws  <- 10000
dinvgamma <- function(x, a, b) {
  (b^a)/gamma(a) * x^(-a-1) * exp(-b/x)
}

dinvchisq2 <- function(x, n, tau2) {
  a <- n / 2
  b <- n*tau2 / 2
  dinvgamma(x, a, b)
}

# a) Simulate posterior draws
X           <- rchisq(nDraws, n)
sigma2      <- n*tau2/X
sigma2.dens <- density(sigma2)

plot(sigma2.dens, col = "blue",
     main = "Posterior density of sigma^2",
```

```r
     xlab = expression(sigma^2),
     ylim = c(0,5),
     xlim = c(0,2))

# Theoretical sigmas
delta  <- 0.01
grid   <- seq(0.01, 2.5, delta)
sigmas <- dinvchisq2(grid, n, tau2)
lines(grid, sigmas/(sum(sigmas)*delta), type = 'l', col = "orange")
legend("topright",
       c("Simulated distribution", "Theoretical distribution"),
       col = c("blue", "orange"),
       lty = 1,
       bty='n',
       cex=.75)

# b) Gini coefficient posterior
G      <- 2*pnorm(sqrt(sigma2/2)) - 1
#G.hist <- hist(G, breaks = 100, freq = FALSE, xlim = 0:1)

# c) G Credible interval
G.slice <- sort(G)[(0.025*nDraws):(0.975*nDraws)]
cred.lower.bound <- min(G.slice)
cred.upper.bound <- max(G.slice)
G.dens <- density(G)
plot(G.dens,
     main = "Posterior distribution of G",
     xlab = "G",
     col  = "orange")

sprintf("Credible interval lower bound: %f", cred.lower.bound)
sprintf("Credible interval upper bound: %f", cred.upper.bound)

# c) G Highest Posterior Density (HPD)
i      <- order(G.dens$y, decreasing = TRUE)
csum   <- cumsum(G.dens$y[i])
m      <- length(csum[csum < sum(G.dens$y)*0.95])
H      <- G.dens$x[i][1:m]
hpd.lower.bound <- min(H)
hpd.upper.bound <- max(H)
sprintf("HPD lower bound: %f", hpd.lower.bound)
sprintf("HPD upper bound: %f", hpd.upper.bound)


# 3

Von.Mises <- function (y, mu, k){
  exp(k*cos(y - mu)) / (2*pi*besselI(k, 0))
}

Y           <- c(-2.44, 2.14, 2.54, 1.83, 2.02, 2.33, -2.79, 2.23, 2.07, 2.02)
mu          <- 2.39
lambda      <- 1
```

```r
delta      <- 0.05
kappas     <- seq(0, 10, delta)
likelihood <- sapply(kappas, function (k) { prod(Von.Mises(Y, mu, k)) })
prior      <- dexp(kappas, lambda)
posterior  <- likelihood * prior

plot(kappas, posterior/(sum(posterior)*delta),
     main = "Posterior distribution of kappa",
     xlab = expression(kappa),
     ylab = "Density",
     col  = "orange",
     type = 'l')
lines(kappas, likelihood/(sum(likelihood)*delta), col ="red")
lines(kappas, prior/(sum(prior)*delta), col ="green")
legend("topright",
       c("Posterior", "Likelihood", "Prior"),
       lty=1,
       col=c("orange", "red", "green"),
       bty='n',
       cex=.75)

sprintf("Posterior mode: %.2f", kappas[which.max(posterior)])
```

——- Lab 2 ——-