Lab 3

Martin Gustafsson (margu424) and Axel Gard (axega544)

2020-10-13

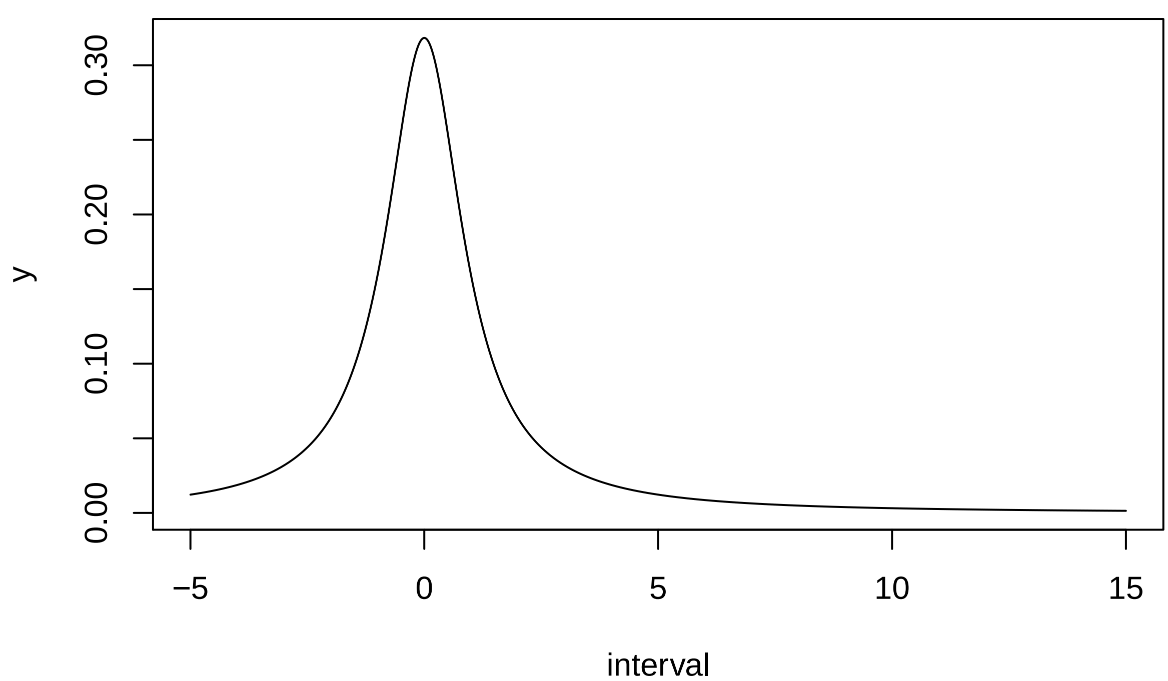
3.1.1

Visualisera posteriorn

(1)

Visualisera din prior exakt (d.v.s. använd dt()) över intervallet [5, 15]

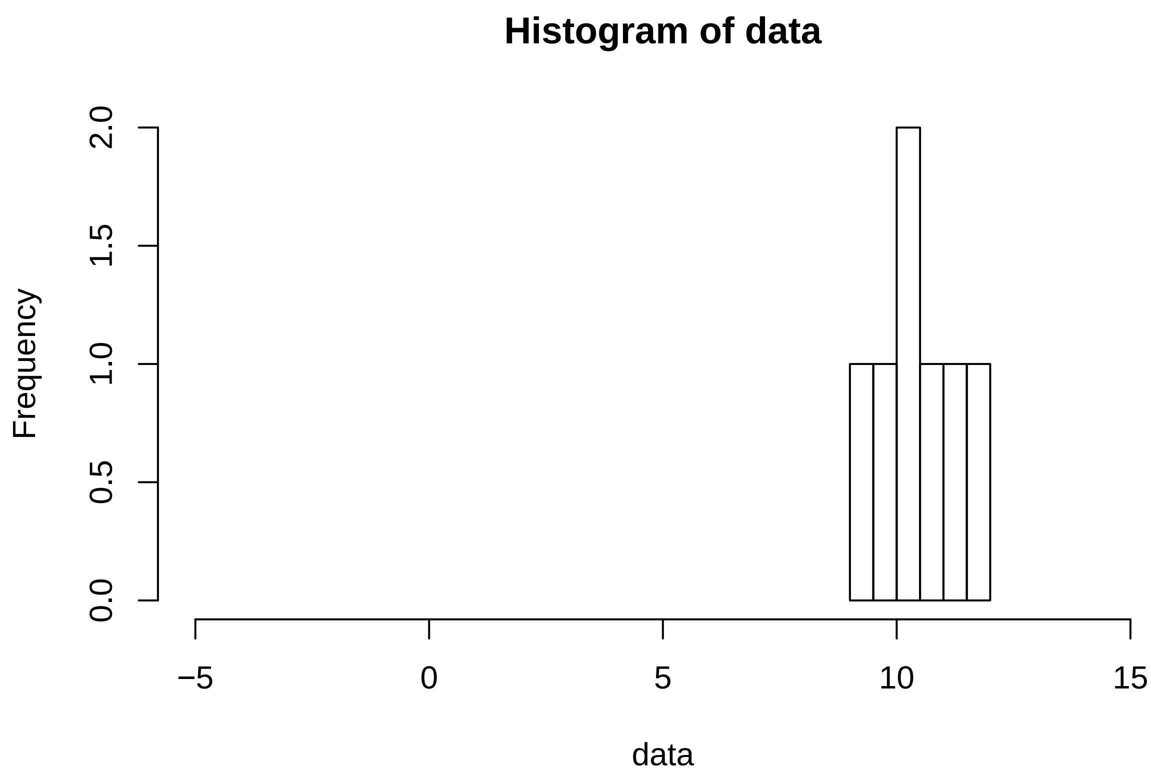
```
sigma <- 1
interval <- seq(-5, 15, 0.01)
y <- dt(x = interval, df = 1)
plot(interval, y, type="l")
```



(2)

Visualisera dessa som ett histogram på intervallet [-5, 15]

```r
data <- c(11.3710, 9.4353, 10.3631, 10.6329, 10.4043, 9.8939, 11.5115)
hist(data, xlim = c(-5, 15))
```



(3)

Skapa en funktion för log likelihooden för µ som du kallar normal_log_likelihood(mu, data)

$$\ln \mathcal{L}(\mu, \sigma^2) = \sum_{i=1}^{n} \ln f(x_i \mid \mu, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2.$$

Sigmaˆ2 = 1 ger då:

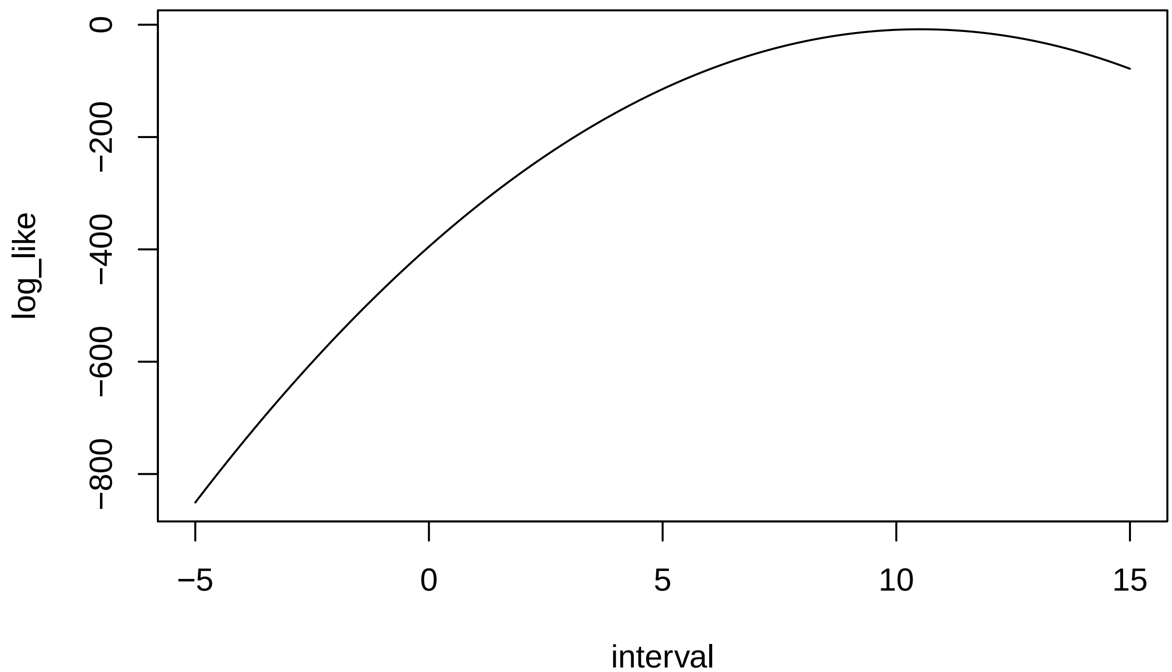
$$\ln \mathcal{L}(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^2$$

```r
normal_log_likelihood <- function(mu, data) {
    return((-length(data)/2) * log(2*pi) - (1/2) * sum((data - mu)^2))
}
llik <- normal_log_likelihood(5, data)
print(round(llik, 1))
```

```
## [1] -114.6
```

```r
log_like <- c()
for (mu in interval) {
    log_like <- c(log_like, normal_log_likelihood(mu, data))
```

```r
}
plot(interval, log_like, type="l")
```



(4)

Härled den proportionella posteriorn för mu

$$p(\theta \mid y) \propto p(y \mid \theta) * p(\theta)$$

Eftersom det är parametern mu vi vill ska posteriorn för så byter vi ut theta mot mu.

$$p(\mu \mid y) \propto p(y \mid \mu) * p(\mu)$$

Likelihood funktionen:

$$p(y \mid \mu) = (2\pi\sigma^2)^{\frac{-n}{2}} * \exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right)$$

Sigmaˆ2 = 1 ger:

$$p(y \mid \mu) = (2\pi)^{\frac{-n}{2}} * \exp\left(\frac{-1}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right)$$

Vi kan förkorta bort termer som inte innehåller mu:

$$p(y \mid \mu) = \exp\left(\frac{-1}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right)$$

Prior funktionen:

$$p(\mu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{\mu^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

v = 1 ger:

$$p(\mu) = \frac{\Gamma(1)}{\sqrt{\pi}\,\Gamma(\frac{1}{2})\,(1+\mu^2)} = \frac{1}{\sqrt{\pi}\,\Gamma(\frac{1}{2})\,(1+\mu^2)}$$

Posteriorn blir då:

$$p(\mu \mid y) = (\exp\left(\frac{-1}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right))\left(\frac{1}{\sqrt{\pi}\,\Gamma(\frac{1}{2})\,(1+\mu^2)}\right) =$$

$$= \frac{\exp\left(\frac{-1}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right)}{\sqrt{\pi}\,\Gamma(\frac{1}{2})\,(1+\mu^2)} = \frac{1}{\sqrt{\pi}\,\Gamma(\frac{1}{2})} * \frac{\exp\left(\frac{-1}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right)}{1+\mu^2} = c * \frac{\exp\left(\frac{-1}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right)}{1+\mu^2}$$

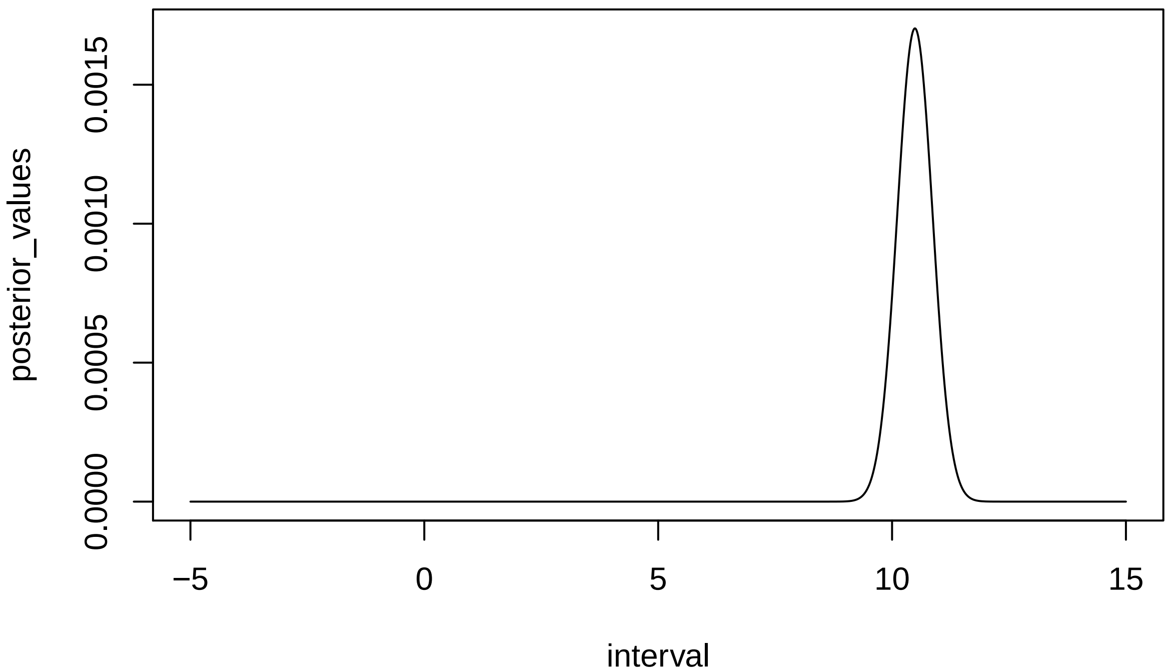
Eftersom vi vill ha den proportionella posteriorn så kan vi förkorta bort konstanten c:

$$p(\mu \mid y) = \frac{\exp\left(\frac{-1}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right)}{1+\mu^2}$$

(5)

Visualisera den onormaliserade posteriorn

```
posterior <- function(mu, data) {
    return(exp((-1/2)*sum((data - mu)^2)) / (1 + mu^2))
}

posterior_values <- c()
for (mu in interval) {
    posterior_values <- c(posterior_values, posterior(mu, data))
}
plot(interval, posterior_values, type="l")
```



3.2.1

Produkt A eller B?

(1)

Materialet behöver du bestämma din prior för p som en beta fördelning. Välj parametrar.

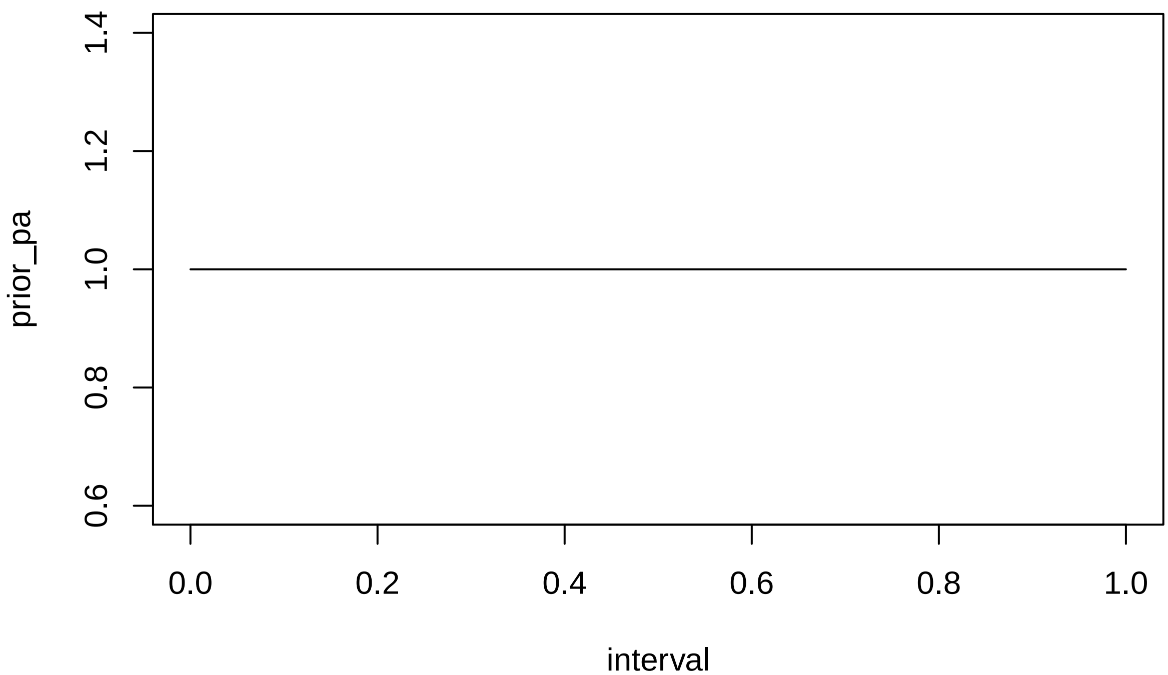
Eftersom vi inte sätt någon data så vill vi skapa en likformig fördelning.

Proir för Pa: Beta (1, 1)

Proir för Pb: Beta (1, 1)

Eftersom de är samma så plottar vi bara ena

```
alpha = 1
beta = 1
interval <- seq(0, 1, 0.01)
prior_pa <- dbeta(x = interval, shape1 = alpha, shape2 = beta)
plot(interval, prior_pa, type="l")
```



(2)

Beräkna den förväntade proportionen för respektive produkt. Vilken produkt har den högsta förväntadepro-portionen intresserade?

	Produkt A	Produkt B
Frågade	13	3
Intresserade	8	2

Posteriorn för Pa = Proir för Pa * Likelihood för binomialfördelning

Binomial ger Likelihood:

$$f(x \mid p) = \binom{n}{k} * p^k * (1-p)^{n-k}$$

Beta prior: Beta(alpha, beta)

$$\pi(p) = \frac{1}{B(\alpha, \beta)} * p^{\alpha-1} * (1-p)^{\beta-1}$$

Posteriorn fås då:

$$\pi(p \mid x) = \frac{f(x \mid p)\pi(p)}{\int f(x \mid p)\pi(p)\, dp} =$$

$$= \frac{\binom{n}{k} p^k (1-p)^{n-k} \frac{1}{B(\alpha,\beta)} p^{\alpha-1} (1-p)^{\beta-1}}{\int \binom{n}{k} p^k (1-p)^{n-k} \frac{1}{B(\alpha,\beta)} p^{\alpha-1} (1-p)^{\beta-1} dp} =$$

$$= \frac{p^{k+\alpha-1} (1-p)^{n-k+\beta-1}}{\int p^{k+\alpha-1} (1-p)^{n-k+\beta-1} dp} =$$

Vi kan bryta ut 1 / integralen till en konstant c.

$$= c * p^{k+\alpha-1} (1-p)^{n-k+\beta-1}$$

Beta(k+alpha, n-k+beta)

Väntevärdet för beta fördelning:

$$\mathbb{E}(x) = \frac{\alpha}{\alpha+\beta}$$

Så för Produkt A: n = 13 k = 8

E(x) = (8+1) / (8+1 + 13-8+1) = 9 / 15 = 0.6

För Produkt B: n = 3 k = 2

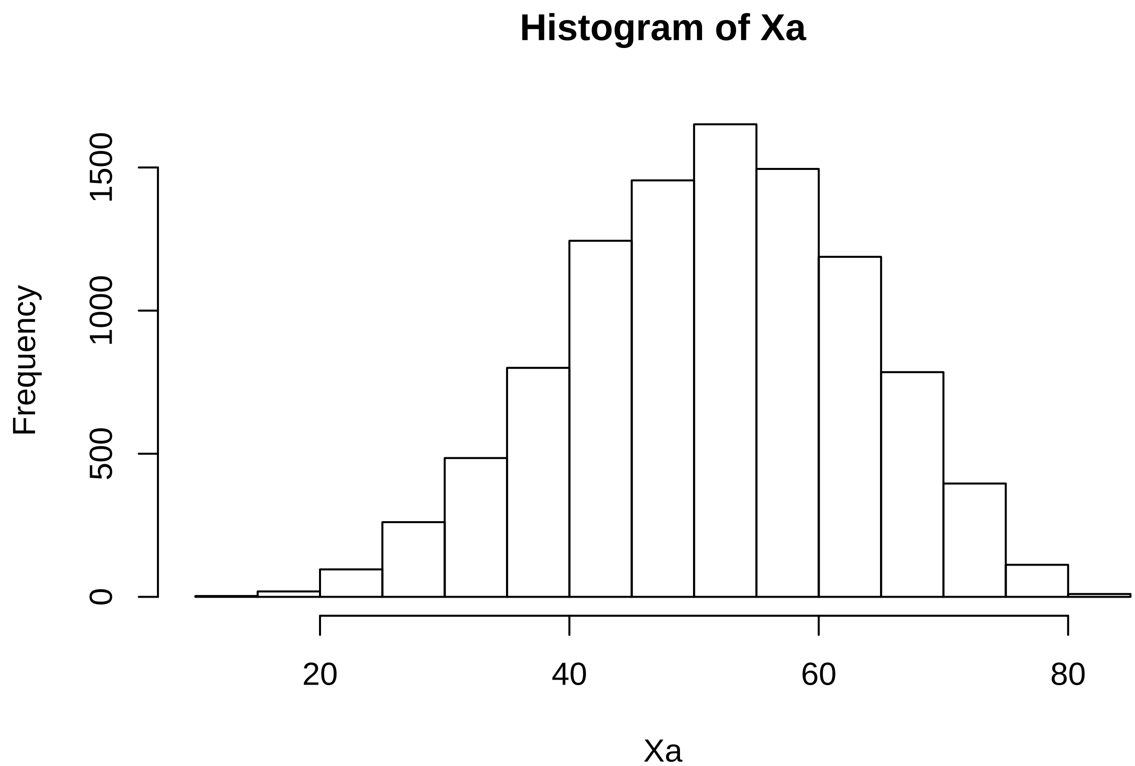
E(x) = (2+1) / (2+1 + 3-2+1) = 3 / 5 = 0.6

Det vi ser här är att båda produkterna har samma förväntade proportion intresserade.

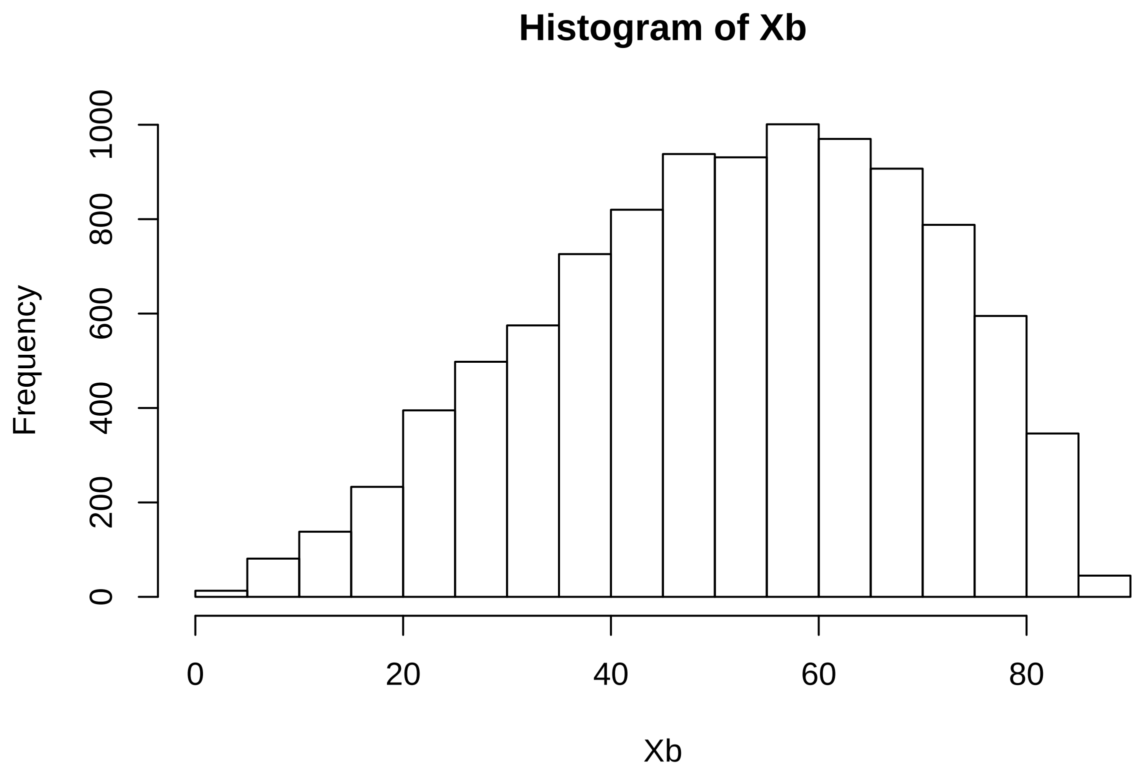
(3)

Använd dina två aposteriorfördelningar för respektive produkt för att simulera hur många intresserade kunder du kan tänkas få för respektive produkt.

```r
n <- 10000
pA <- rbeta(n = n, shape1 = 8+1, shape2 = 13-8+1)
pB <- rbeta(n = n, shape1 = 2+1, shape2 = 3-2+1)
Xa <- rbinom(n = n, size = 87, prob = pA)
Xb <- rbinom(n = n, size = 87, prob = pB)

hist(Xa)
```




```r
hist(Xb)
```



(a) Hur stor är sannolikheten att du får fler än 40 intresserade kunder med respektiveprodukt ?

```r
print(sum(Xa > 40) / n)
```

```
## [1] 0.8336
```

```r
print(sum(Xb > 40) / n)
```

```
## [1] 0.7341
```

(b) Vad är det förväntade antalet intresserade kunder av respektive produkt, dvs E(Xa) och E(XB)?

```r
print(mean(Xa))
```

```
## [1] 52.0384
```

```r
print(mean(Xb))
```

```
## [1] 52.3198
```

3.3.1

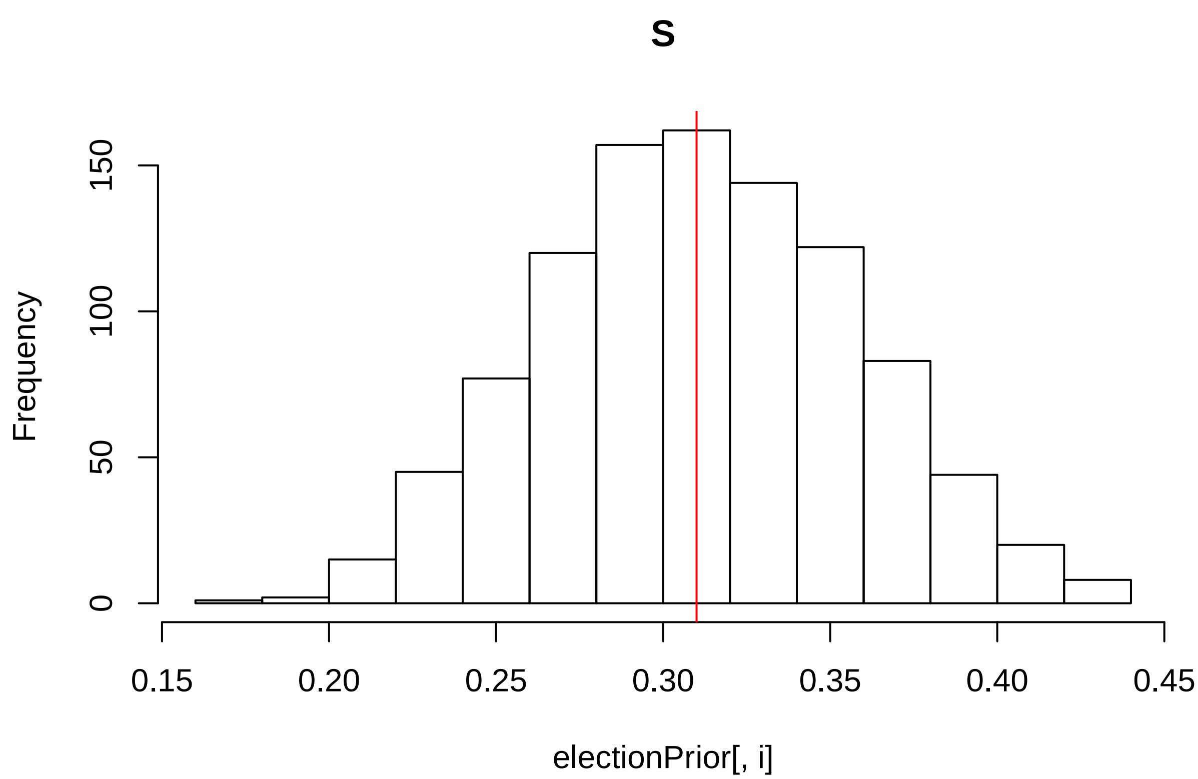
Analys av opinionsundersökningar.

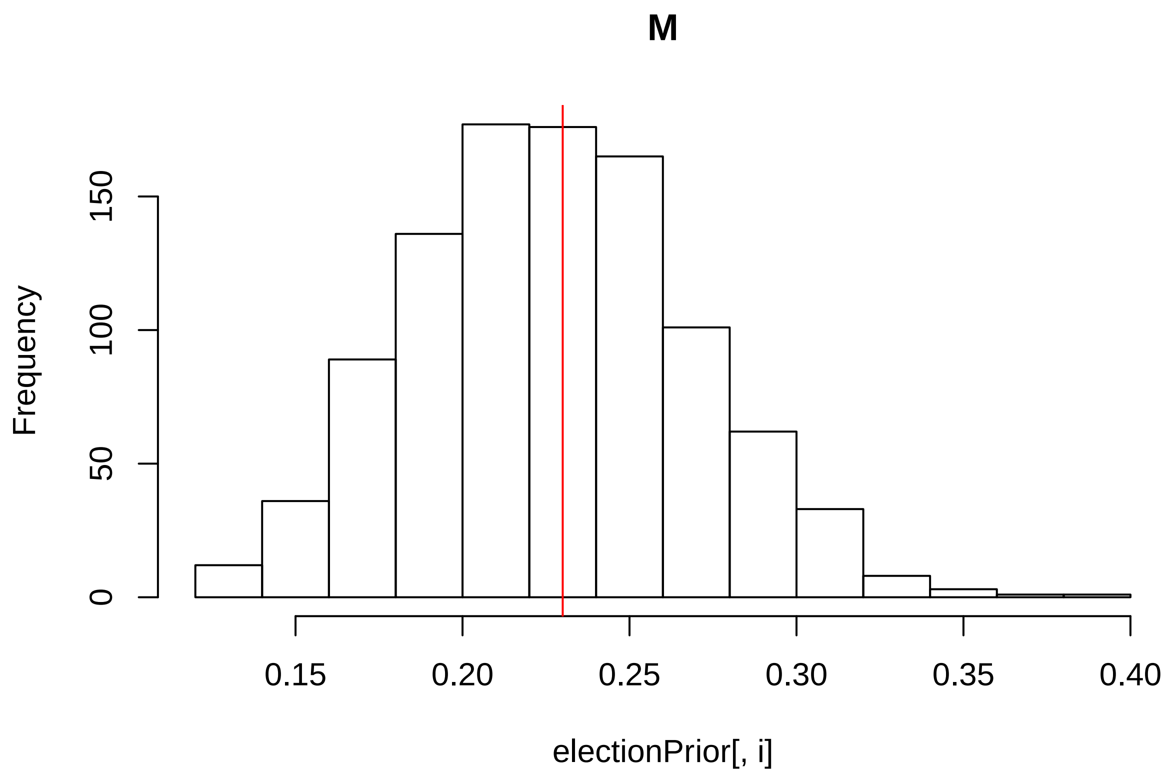
(1)

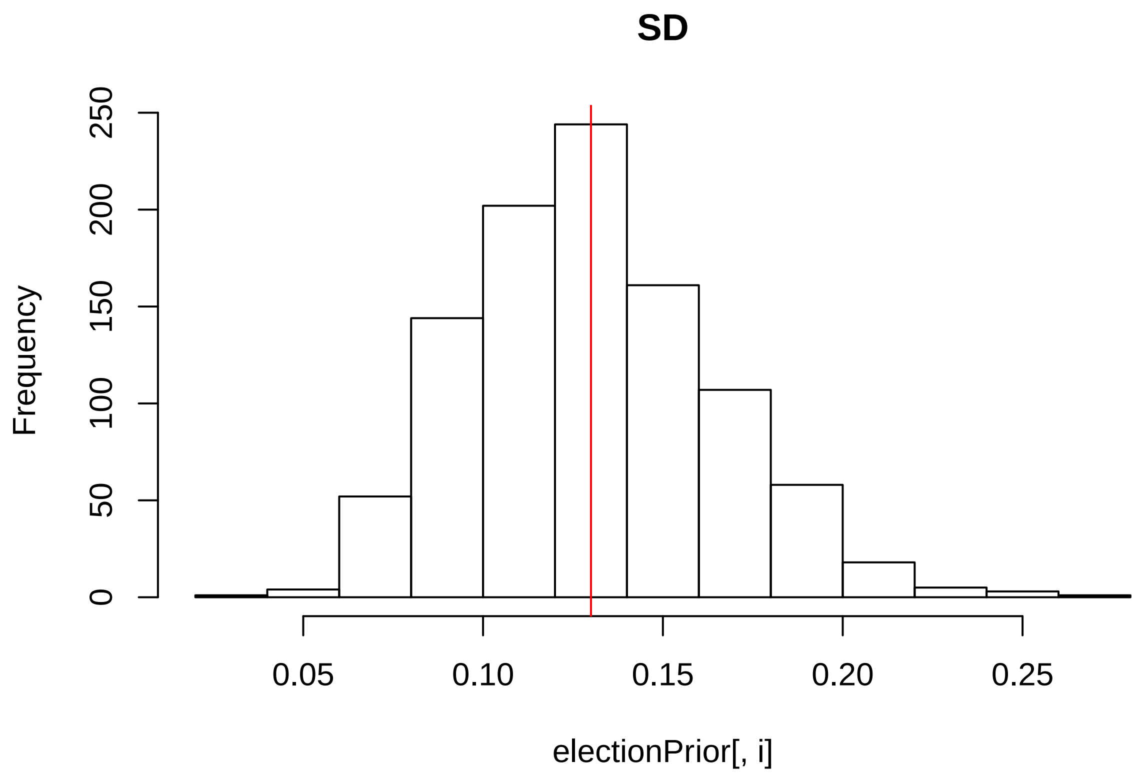
bestämma vår apriorifördelning för de olika partierna.

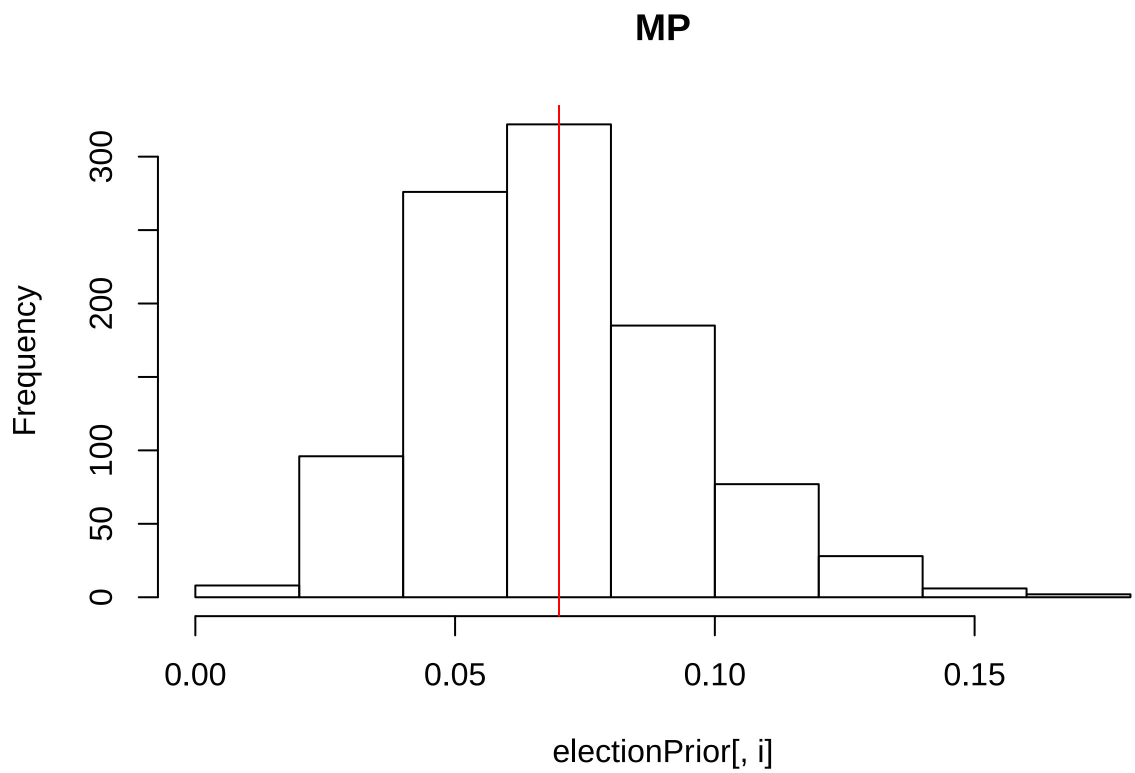
Parti	%
S	31.01
M	23.33
SD	12.86
MP	6.89
C	6.11
V	5.72
FP	5.42
KD	4.57
FI	3.12

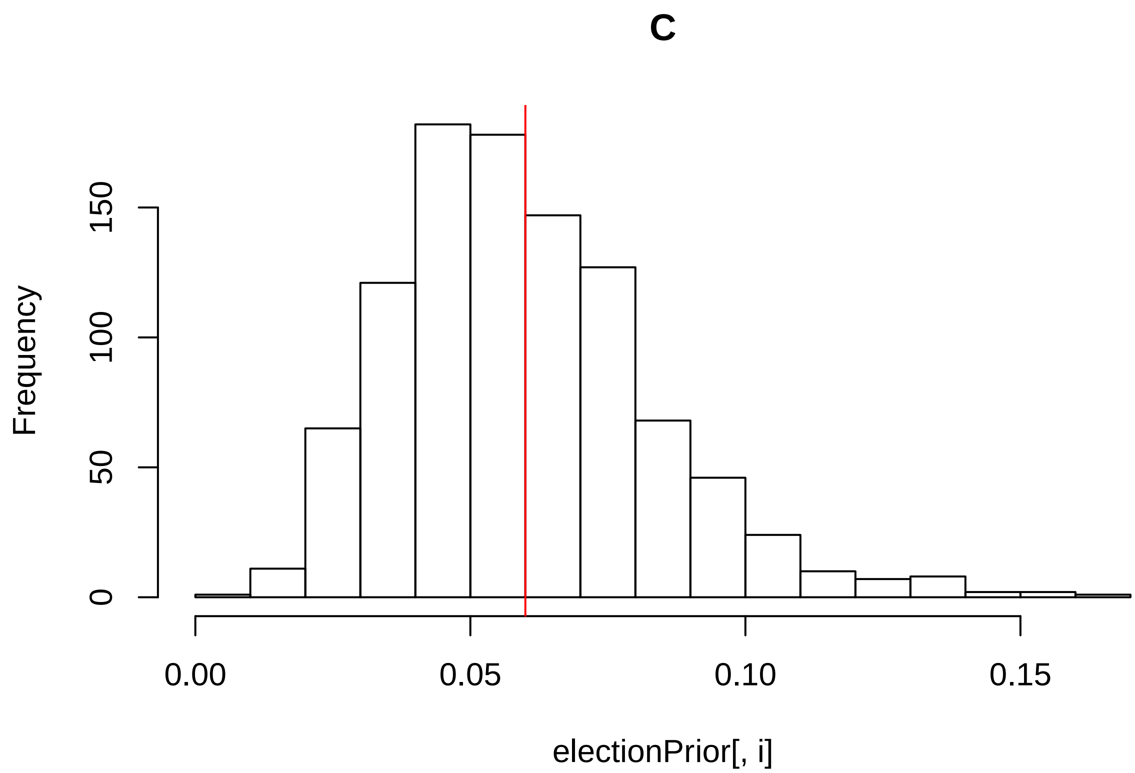
```r
library(gtools)
set.seed(4711)

Parties <- c("S", "M", "SD", "MP", "C", "V", "FP", "KD", "FI")
n <- 1000
alpha <- c(31, 23, 13, 7, 6, 6, 6, 5, 3)

electionPrior <- rdirichlet(n = n, alpha = alpha)

for (i in 1:9){
    party <- Parties[i]
    hist(electionPrior[,i], main = party)
    abline(v = alpha[i] / sum(alpha), col = "red")
}
```

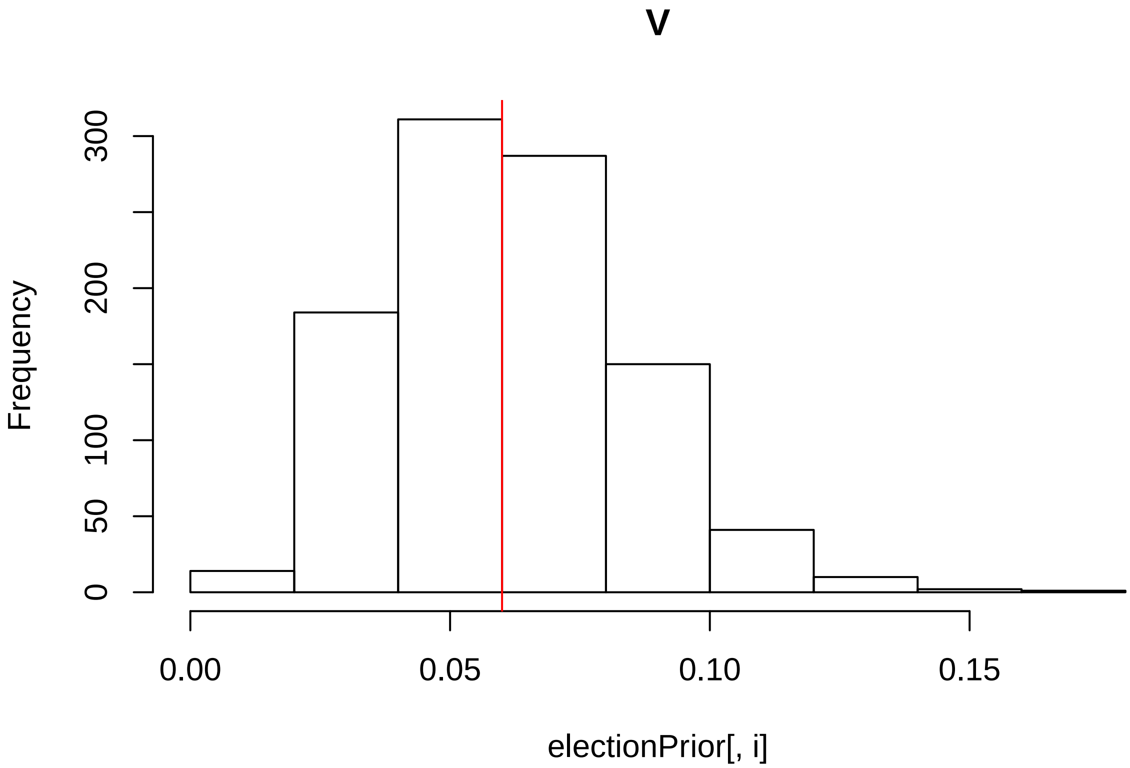


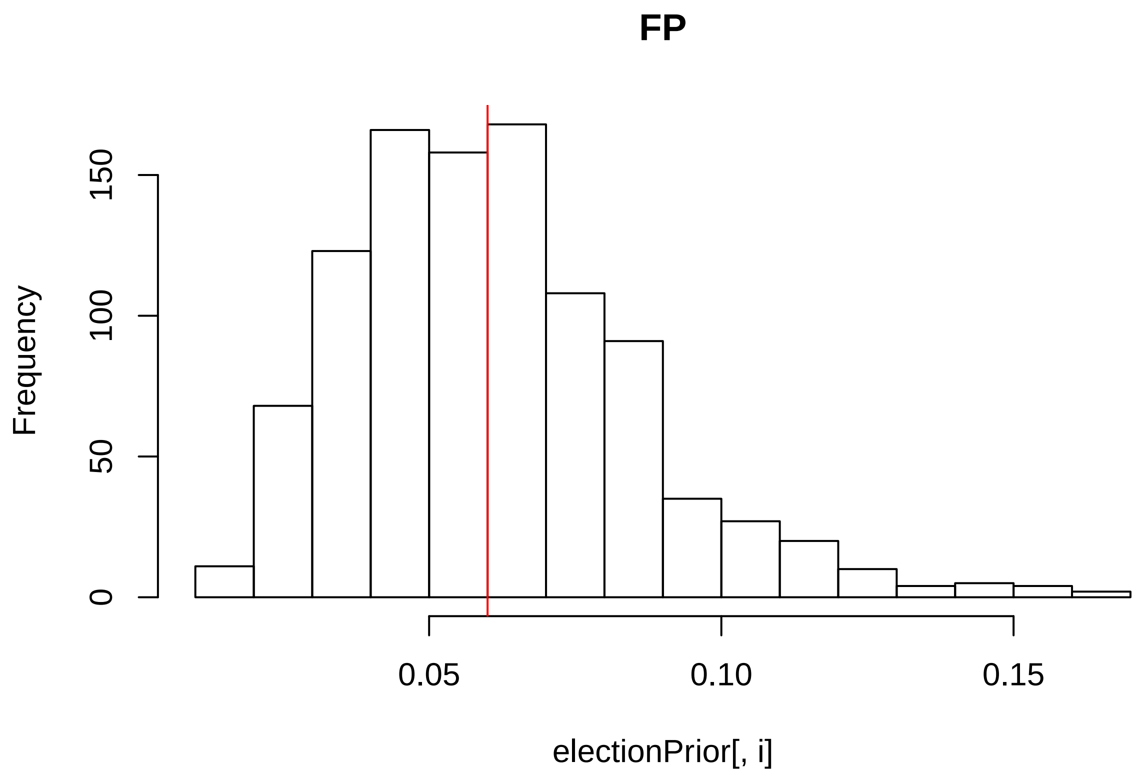

M
Frequency
0
50
100
150
0.15
0.20
0.25
0.30
0.35
0.40
electionPrior[, i]

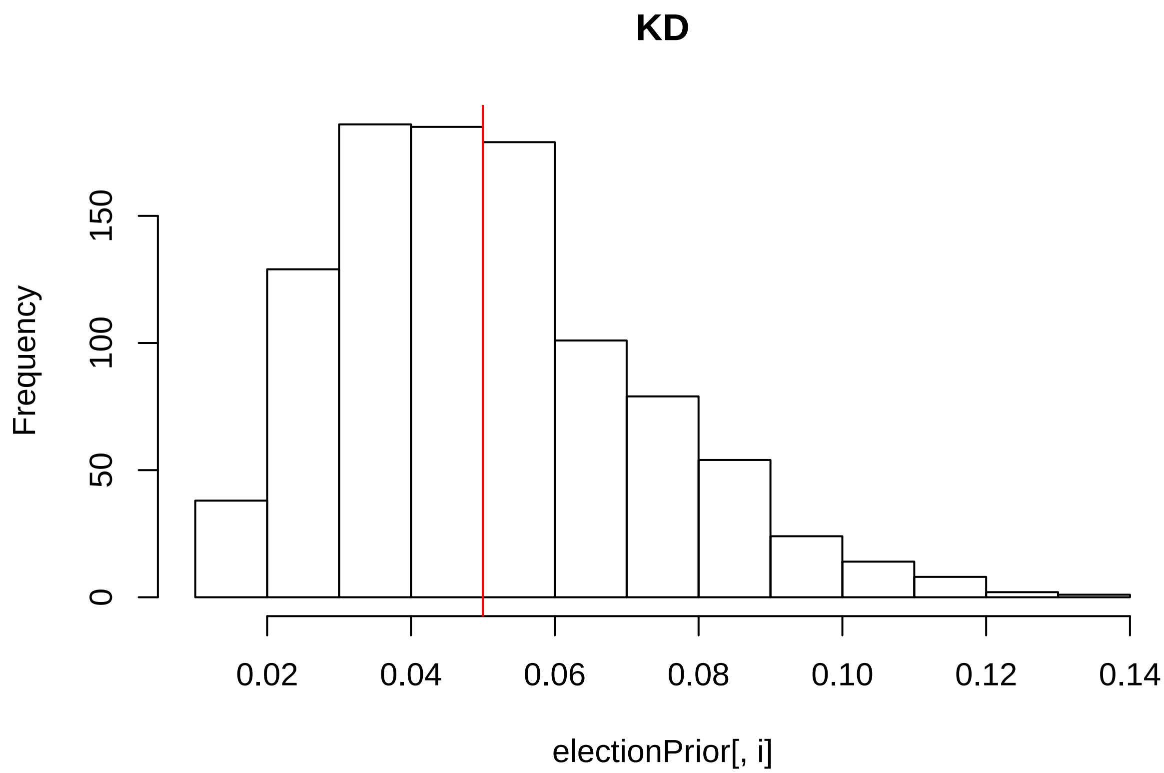

SD
Frequency
0
50
100
150
200
250
0.05
0.10
0.15
0.20
0.25
electionPrior[, i]

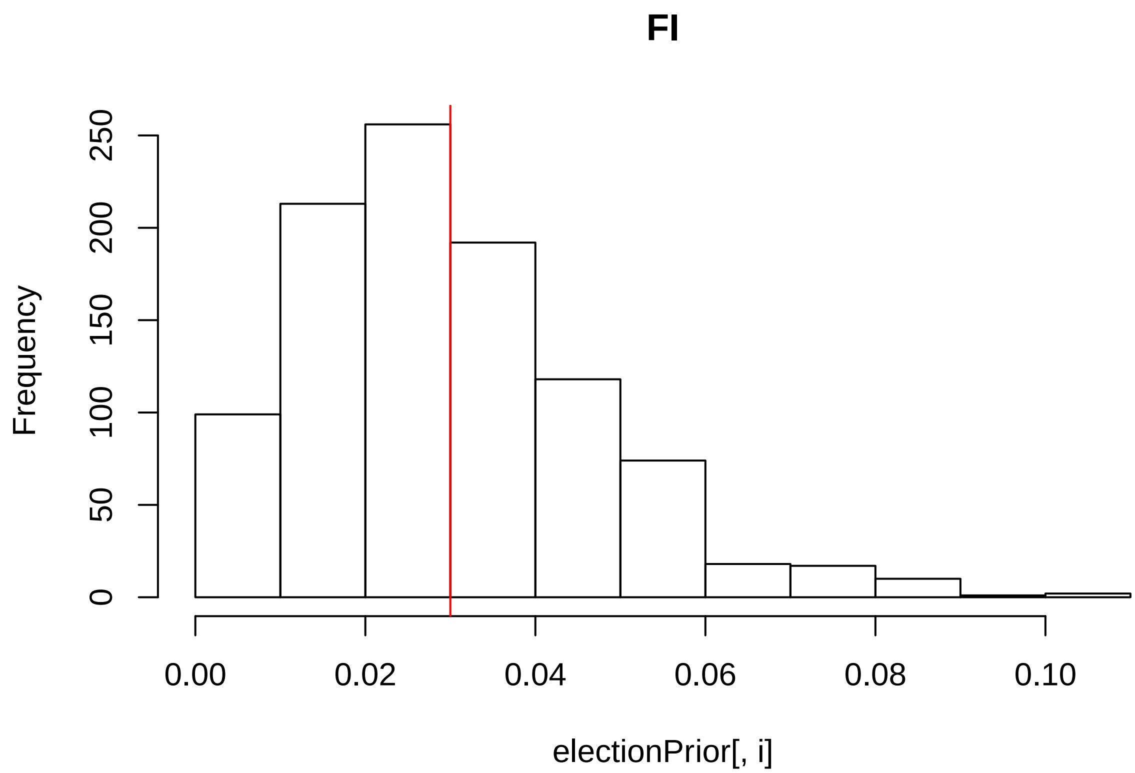

MP
Frequency
0
50
100
200
300
0.00
0.05
0.10
0.15
electionPrior[, i]


C
Frequency
0
50
100
150
0.00
0.05
0.10
0.15
electionPrior[, i]


V
Frequency
0
50
100
200
300
0.00
0.05
0.10
0.15
electionPrior[, i]


FP
Frequency
0
50
100
150
0.05
0.10
0.15
electionPrior[, i]


KD
Frequency
0
50
100
150
0.02
0.04
0.06
0.08
0.10
0.12
0.14
electionPrior[, i]



(2)

Den undersökning vi valde var en undersökning från Skop som publicerades 2018-09-06 och där de frågade 2928 personer.

> row 125: 2018-sep Skop 17.3 6.2 7.9 6.9 23.1 10.6 5.7 17.7 0.8 1 2928 2018-09-06 2018-08-30 2018-09-06 FALSE Skop

Parti	%	antal (200 pers)
S	23.1	46
M	17.3	35
SD	17.7	35
MP	5.7	11
C	7.9	16
V	10.6	21
L	6.2	12
KD	6.9	14
FI	0.8	2

```r
pollResults <- c(23.1, 17.3, 17.7, 5.7, 7.9, 10.6, 6.2, 6.9, 0.8)
for (proportion in pollResults){
    print(round((proportion/100)*200, digits = 0))
}
```

```
## [1] 46
```

```
## [1] 35
## [1] 35
## [1] 11
## [1] 16
## [1] 21
## [1] 12
## [1] 14
## [1] 2
```

(3)

aposteriorifördelningen för andelenper riksdagsparti

```
model <- c(46, 35, 35, 11, 16, 21, 12, 14, 2)
n <- 10000
barrier <- 0.04
Parties <- c("S", "M", "SD", "MP", "C", "V", "FP", "KD", "FI")
results <- rdirichlet(n = n, alpha = alpha + model)
normalizedResults <- matrix(nrow = n, ncol = 9)

for (i in 1:n){
  for (j in 1:9){
      if (results[i, j] > barrier) {
          normalizedResults[i, j] = results[i, j]
      } else {
          normalizedResults[i, j] = 0
      }
  }
}
for (i in 1:n){
    sum_ <- sum(normalizedResults[i, ])
    for (j in 1:9){
        normalizedResults[i, j] <- normalizedResults[i, j] / sum_
    }
}
```

(a) Vad är sannolikheten att de rödgröna är större än alliansen ?

```
S <- 1
M <- 2
MP <- 4
SD <- 3
C <- 5
V <- 6
FP <- 7
KD <- 8

RedGreenLargerAmount <- 0
for (i in 1:n){
    AlliedSum <- normalizedResults[i, M] + normalizedResults[i, FP] +
                 normalizedResults[i, C] + normalizedResults[i, KD]
    redGreenSum <- normalizedResults[i, S] + normalizedResults[i, MP] + normalizedResults[i, V]
    RedGreenLargerAmount <- RedGreenLargerAmount + (redGreenSum > AlliedSum)
}
print(RedGreenLargerAmount / n)
```

```
## [1] 0.6209
```

(b) Vad ar sannolikheten att Sverigedemokraterna (SD) är störe än Moderaterna (M)?

```r
SDLargerAmount <- 0
for (i in 1:n){
    SDLargerAmount <- SDLargerAmount + (normalizedResults[i, SD] > normalizedResults[i, M])
}
print(SDLargerAmount / n)
```

```
## [1] 0.1662
```

(c) Vad är sannolikheten att Kristidemokraterna inte skulle komma in i Riksdagen, dvs de far mindre än 4 % ?

```r
KDOutAmount <- 0
for (i in 1:n){
    KDOutAmount <- KDOutAmount + (normalizedResults[i, KD] == 0)
}
print(KDOutAmount / n)
```

```
## [1] 0.0278
```

(d) Vad ar sannolikheten att Miljöpartiet (MP) skulle åka ur Riksdagen ?

```r
MPOutAmount <- 0
for (i in 1:n){
    MPOutAmount <- MPOutAmount + (normalizedResults[i, MP] == 0)
}
print(MPOutAmount / n)
```

```
## [1] 0.0499
```

(e) Skapa ett sannolikhetsintervall (95 %) för Socialdemokraterna

På grund av två delat så kommer alpha att delas på 2.

$$\frac{\alpha}{2} = \frac{1 - 0.95}{2} = \frac{0.05}{2} = 0.025$$

```r
lower <- quantile(prob=0.025, x = normalizedResults[,S])
higher <- quantile(prob=0.975, x = normalizedResults[,S])

print(lower)
```

```
##      2.5%
## 0.2178666
```

```r
print(higher)
```

```
##     97.5%
## 0.3239201
```