In [27]:
```r
par(family  = "Arial")
#install.packages("showtext")
library(showtext)
showtext_auto()
options(repr.plot.width=5, repr.plot.height=5)
```

1. CUSTOMERS' CHOICE OF BRANDS (*Total credits: 11p*)

   From a random sample of $n = 100$ customers, a store has noted for a certain product with three possible brands A, B, C that 38 customers bought brand A, 27 customers bought brand B, and 35 bought brand C. Let θ_i be the probability that a random customer, who buys the product, chooses brand i, where $i = A, B, C$.

   In problems (a), (b) and (c) you assume the prior $\theta_A \sim Beta\,(\alpha = 16, \beta = 24)$ and only consider that each customer chooses brand A or not brand A.

   (a) *Credits: 3p.* Compute the posterior probability that $\theta_A > 0.4$ and plot the posterior distribution of $1 - \theta_A$.

Postrior with prior beta and bernoulli

$$\text{Beta}(\alpha + s, \beta + f)$$

Postrior = beta(16 + 38 , 24+62) = beta(54,86)

In [2]:
```r
n = 100
A =38
B = 27
C = 35
beta = 24
alpha = 16
```

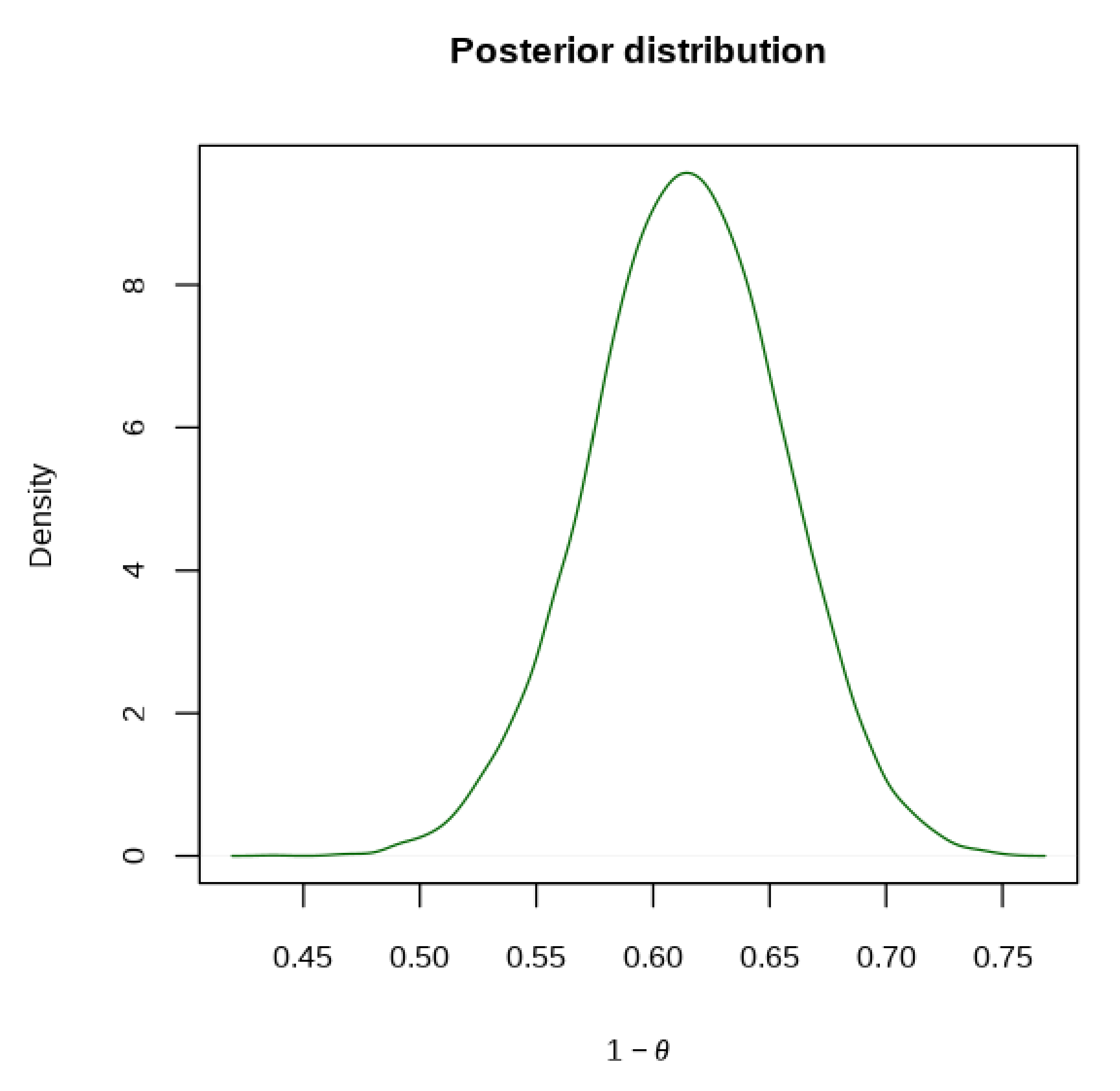
In [20]:
```r
print('theta > 0.4')
pbeta(q= 0.4, shape1 =54,shape2 = 86,lower.tail = FALSE)
```

```
[1] "theta > 0.4"
```

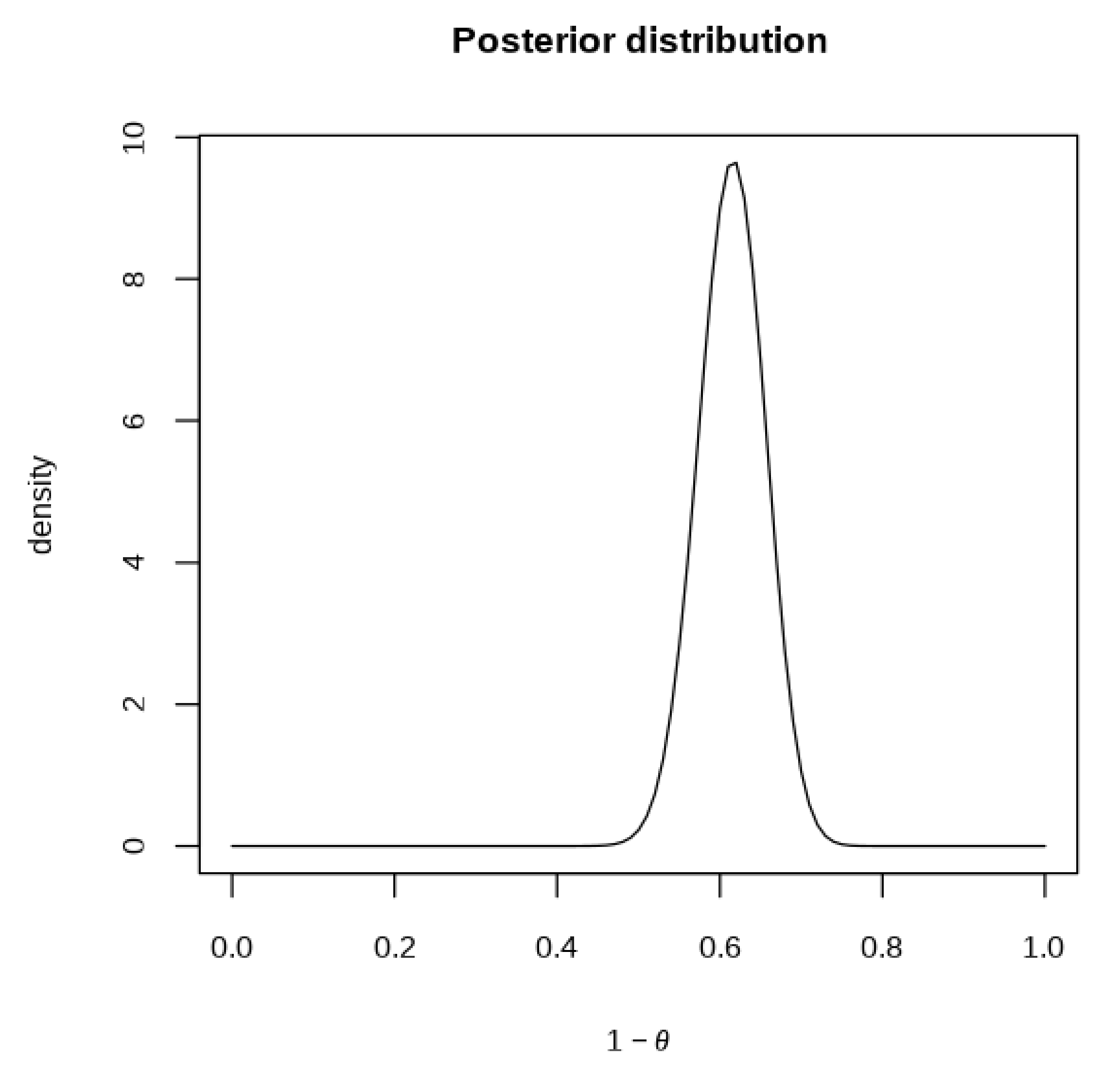
0.360038299240486

In [15]:
```r
# the posterior of 1- theta
post = rbeta(10000,shape1 = 54,shape2 = 86)
```

In [33]:
```r
plot(density(1 - post),main="Posterior distribution" ,xlab = expression(1 - theta),col = 'darkgreen')
```



In [37]:
```r
curve(dbeta(1-x,shape1 = 54,shape2 = 86),main='Posterior distribution',ylab = 'density',xlab = expres
sion(1-theta))
```



   (b) *Credits: 2p.* Compute a 95% equal tail credible interval for the ratio $\frac{1-\theta_A}{\theta_A}$ and interpret it.

In [49]:
```r
quantile((1-post)/post,c(0.025,0.975))
```

2.5%	1.13358380239789
97.5%	2.26134754673305

The ratio is the odds of not choosing brand A, i.e. it describes how many more times likely it is to not choose brand A compared to choosing brand A. The credible interval shows the values of the ratio with 95 % probability.

   (c) *Credits: 2p.* Compute the marginal likelihood for the model $x_1, \ldots, x_{100} | \theta_A \overset{iid}{\sim} Bernoulli\,(\theta_A)$.

The marginal likelihood:|

$$\frac{B(\alpha + s, \beta + f)}{B(\alpha, \beta)}$$

In [51]:
```r
beta(54,86)/(beta(16,24))
```

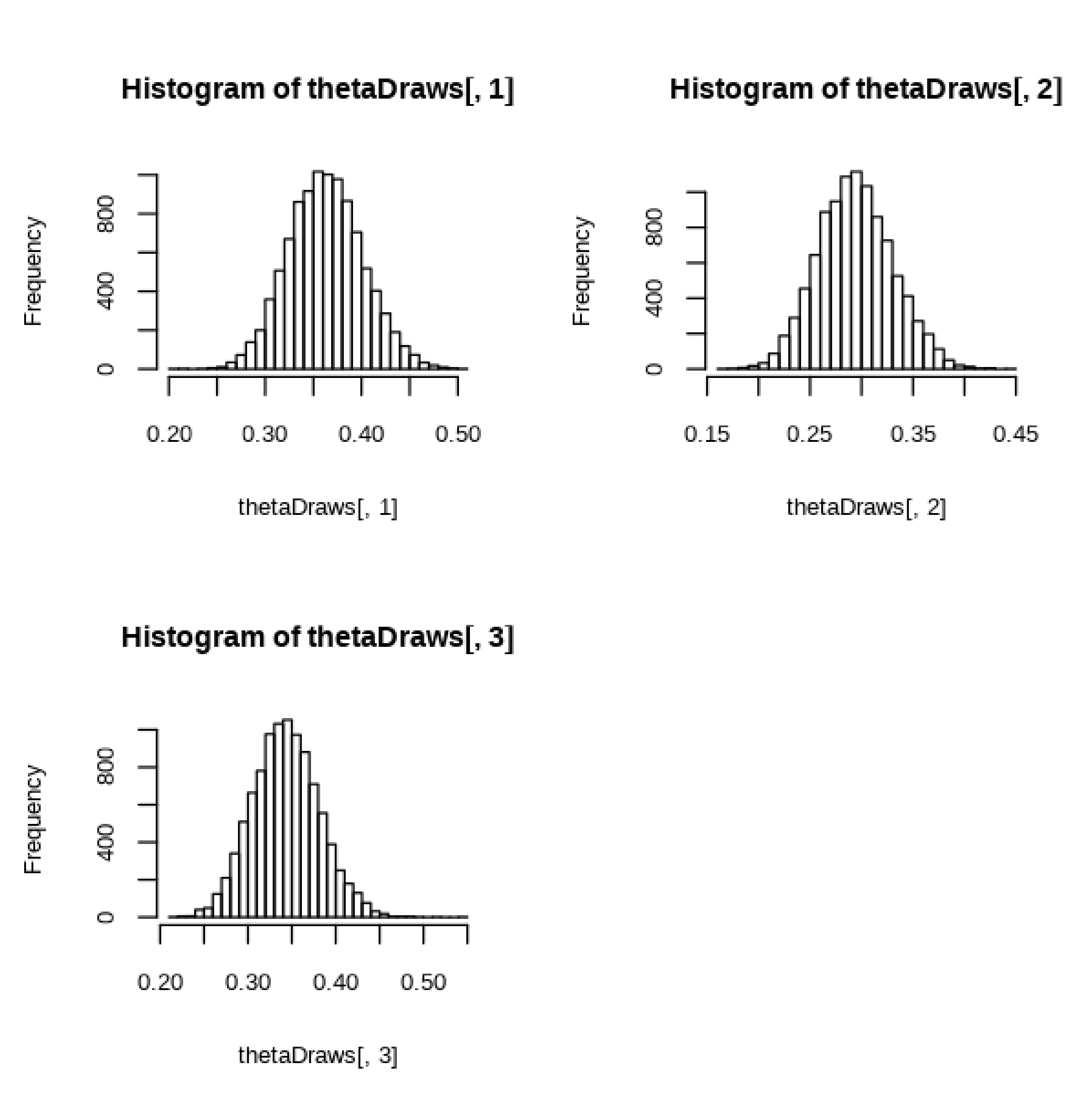
7.55677069331938e-30

   (d) *Credits: 4p.* Assume a Dirichlet prior for $(\theta_A, \theta_B, \theta_C)$ such that $E[\theta_A] = E[\theta_B] = E[\theta_C] = \frac{1}{3}$ and where the prior information is equivalent to a random sample of 60 customers. Compute the posterior probability that $\theta_A > \theta_C$.

In [53]:
```r
Dirichlet <- function(NDraws,y,alpha){
  K <- length(alpha)
  xDraws <- matrix(0,NDraws,K)
  thetaDraws <- matrix(0,NDraws,K) # Matrix where the posterior draws of theta are stored
  for (j in 1:K){
    xDraws[,j] <- rgamma(NDraws,shape=alpha[j]+y[j],rate=1)
  }
  for (ii in 1:NDraws){
    thetaDraws[ii,] <- xDraws[ii,]/sum(xDraws[ii,])
  }
  return(thetaDraws)
}

###########   Setting up data and prior   #################
y <- c(38,27,35) # Data of counts for each category
p <- y/sum(y)
alpha_const <- 1
alpha <- alpha_const*c(20,20,20) # Dirichlet prior hyperparameters
NDraws <- 10000 # Number of posterior draws

###########   Posterior sampling from Dirichlet   #################
thetaDraws <- Dirichlet(NDraws,y,alpha)

K <- length(y)
########### Summary statistics from the posterior sample #########
for (k in 1:K){
  mean(thetaDraws[,k])
  sqrt(var(thetaDraws[,k]))
}

sum(thetaDraws[,2]>thetaDraws[,1])/NDraws # p(theta2>theta3 | y)
# Posterior probability that Android has largest share, i.e. p(theta_2 > max(theta_1,theta_3,theta_4)
 | y)
Index_max <- matrix(0,NDraws,1)
for (ii in 1:NDraws){
Index_max[ii,1] <- which.max(thetaDraws[ii,])
}
mean(Index_max==2)

# Plot histograms of the posterior draws
plot.new() # Opens a new graphical window
par(mfrow = c(2,2)) # Splits the graphical window in four parts (2-by-2 structure)
hist(thetaDraws[,1],25) # Plots the histogram of theta[,1] in the upper left subgraph
hist(thetaDraws[,2],25)
hist(thetaDraws[,3],25)
#hist(thetaDraws[,4],25)
```

0.1429

0.0762



In [54]:
```r
Index_max <- matrix(0,NDraws,1)
for (ii in 1:NDraws){
Index_max[ii,1] <- which.max(thetaDraws[ii,])
}
mean(Index_max==1)
```

0.5731

In [55]:
```r
mean(thetaDraws[,1] > thetaDraws[,3])
```

0.6178

In [ ]: