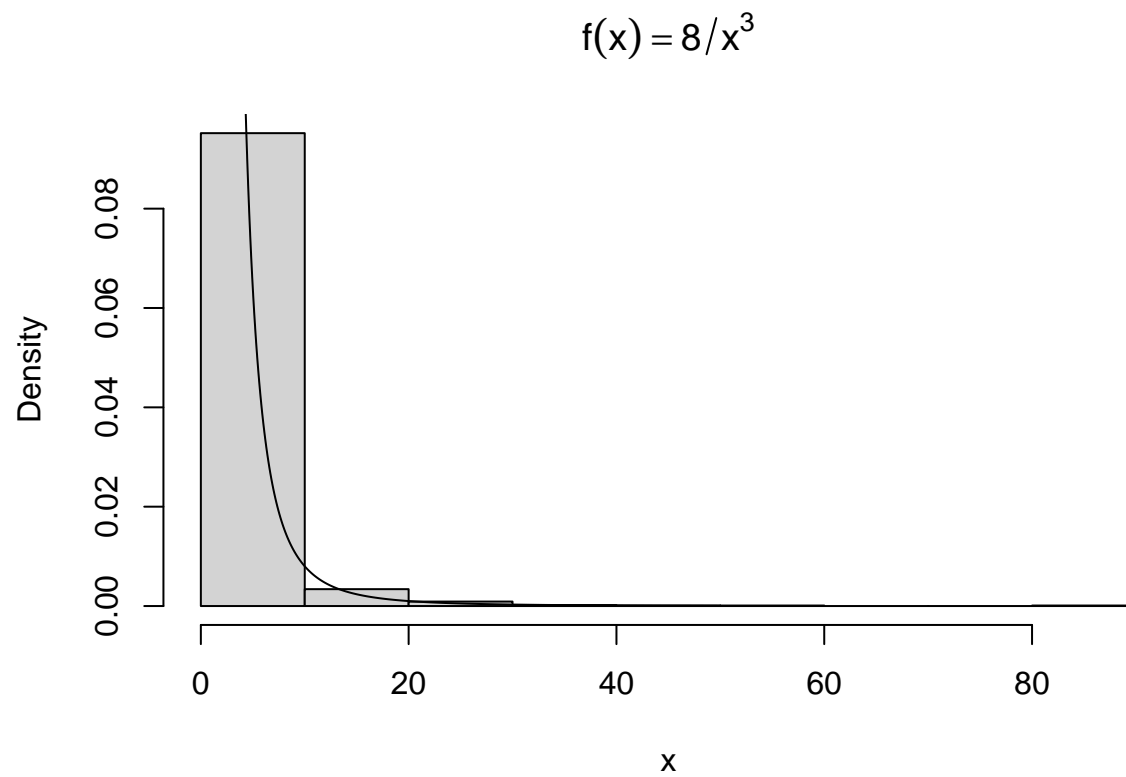# HW2

Getong Zhong

2023-02-06
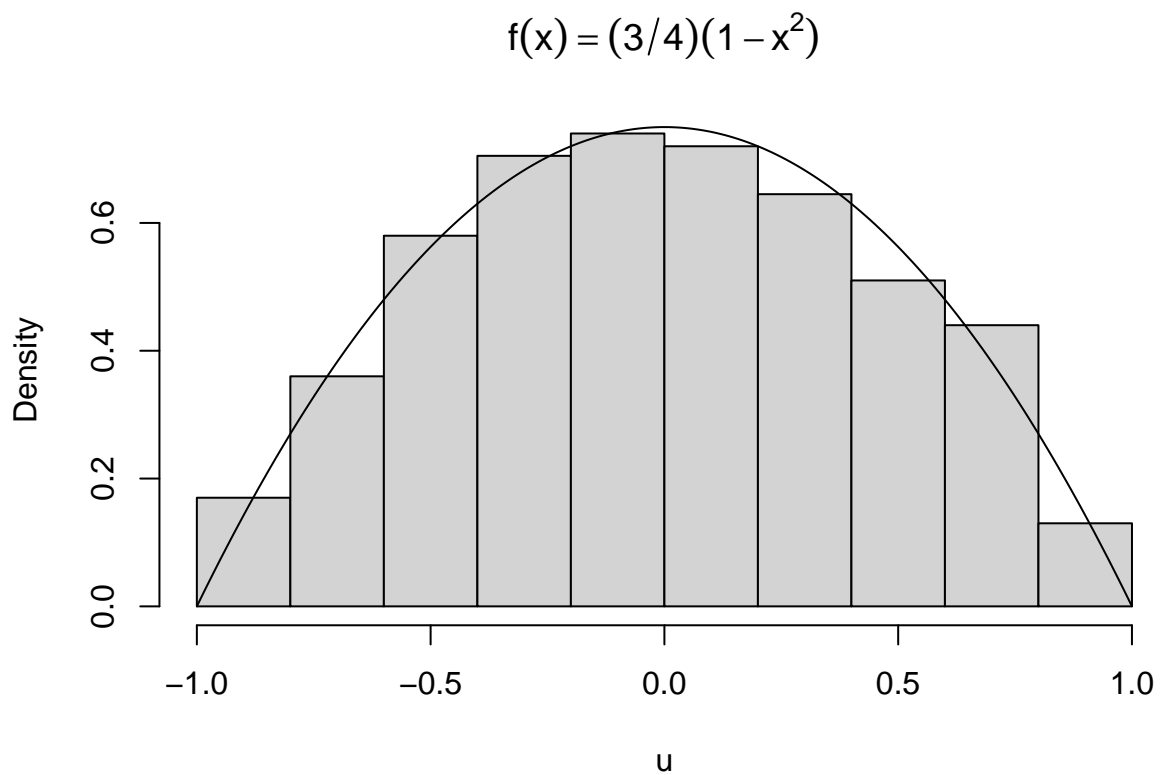
3.3

```r
set.seed(1111)
n <- 1000
u <- runif(n)
x <- sqrt(4/(1-u))
hist(x, prob = TRUE, main = expression(f(x)==8/x^3))
y <- seq(0, 50, 0.01)
lines(y, 8/(y^3))
```

$$f(x) = 8/x^3$$

3.9

```r
set.seed(1111)
n <- 1000
u <- 0
for (i in 1:n) {
  u1 <- runif(n,-1,1)
  u2 <- runif(n,-1,1)
  u3 <- runif(n,-1,1)
  if(abs(u3[i]) >= abs(u2[i]) && abs(u3[i]) >= abs(u1[i])){
    u[i] <- u2[i]}
  else
   { u[i] <- u3[i]}
}
hist(u, prob = TRUE, main = expression(f(x)==(3/4)(1-x^2)))
z <- seq(-1, 1, 0.01)
lines(z, (3/4)*(1-z^2))
```


3.11

$p = 0.75$
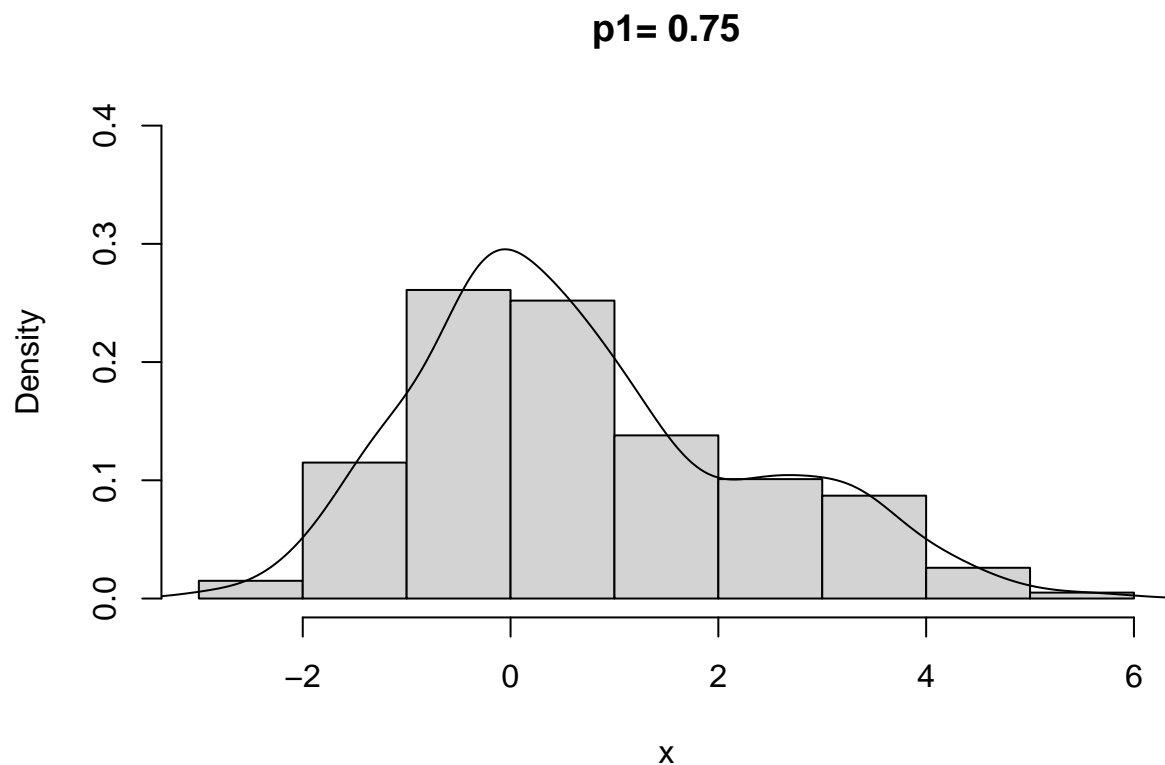
```
set.seed(1111)
n <- 1000
x1 <- rnorm(1000,0,1)
x2 <- rnorm(1000,3,1)

p1 <- 0.75
p2 <- 1-p1
u <- runif(n)
k <- as.integer(u>p2)
x <- k*x1+(1-k)*x2

hist(x, prob = TRUE, ylim = c(0,.4), main = "p1= 0.75")
lines(density(x))
```
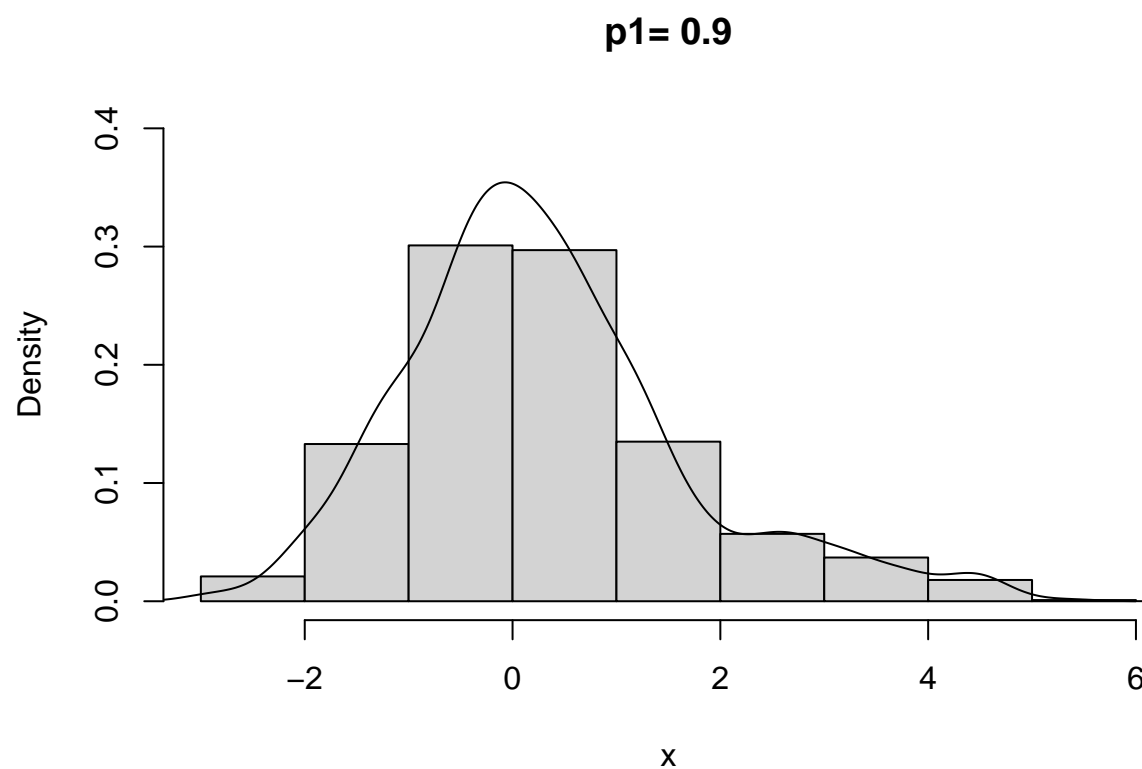
p1= 0.75

p = 0.9

```
set.seed(1111)
p1 <- 0.9
p2 <- 1-p1
u <- runif(n)
k <- as.integer(u>p2)
x <- k*x1 + (1-k)*x2

hist(x, prob=  TRUE, ylim = c(0,.4), main = "p1= 0.9")
lines(density(x))
```

p1= 0.9

p = 0.6

```
set.seed(1111)
p1 <- 0.6
p2 <- 1-p1
u <- runif(n)
k <- as.integer(u>p2)
x <- k*x1 + (1-k)*x2

hist(x, prob=  TRUE, ylim = c(0,.4), main = "p1= 0.6")
lines(density(x))
```
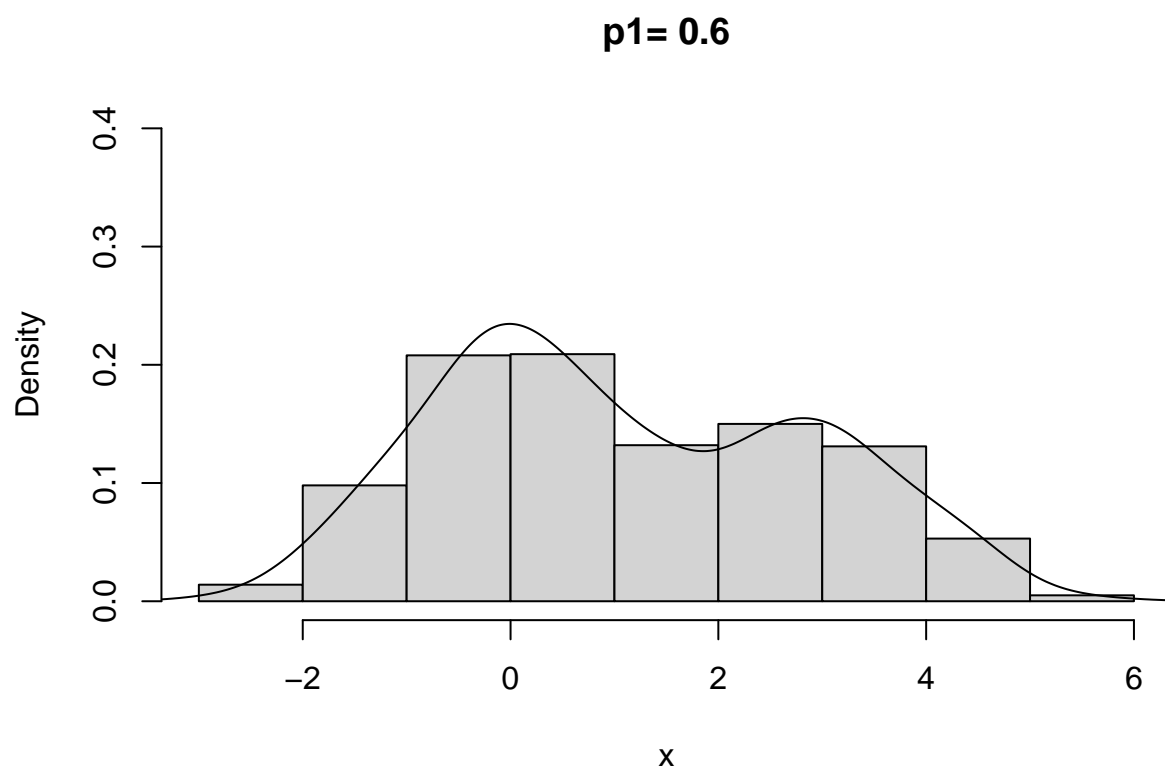
p1= 0.6

p = 0.5

```
set.seed(1111)
p1 <- 0.5
p2 <- 1-p1
u <- runif(n)
k <- as.integer(u>p2)
x <- k*x1 + (1-k)*x2

hist(x, prob = TRUE, ylim = c(0,.4), main = "p1 = 0.5")
lines(density(x))
```
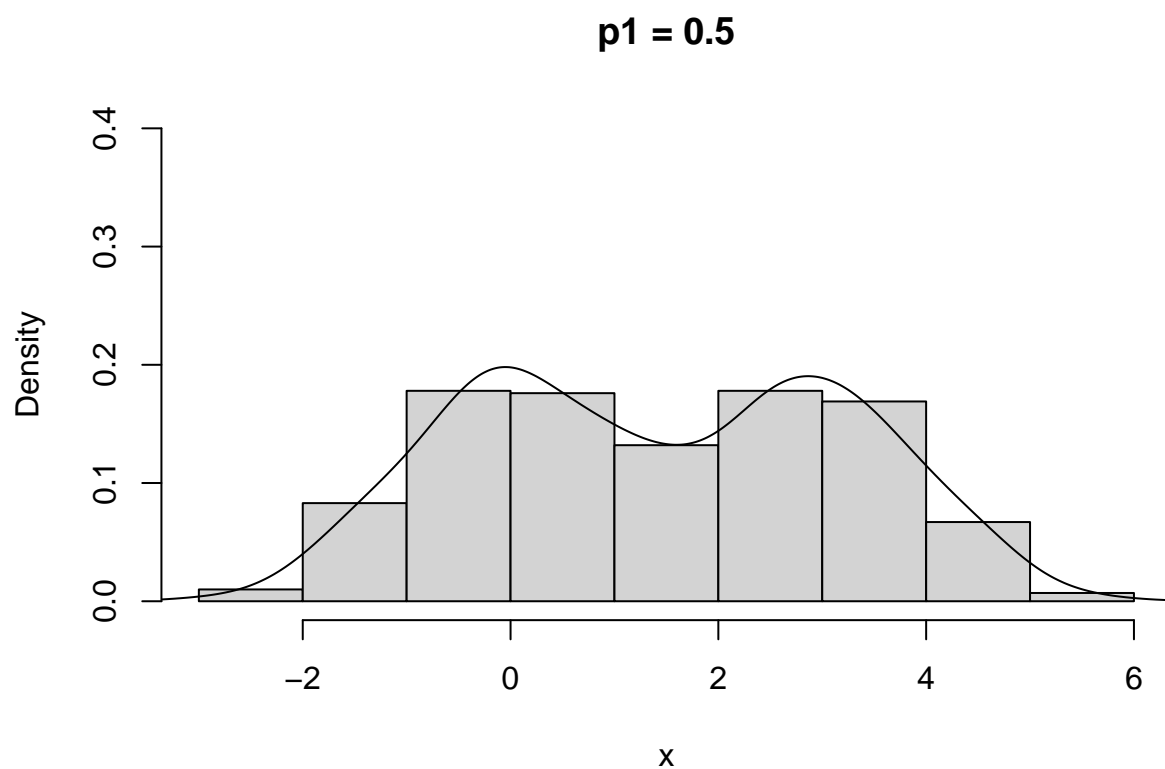
p1 = 0.5
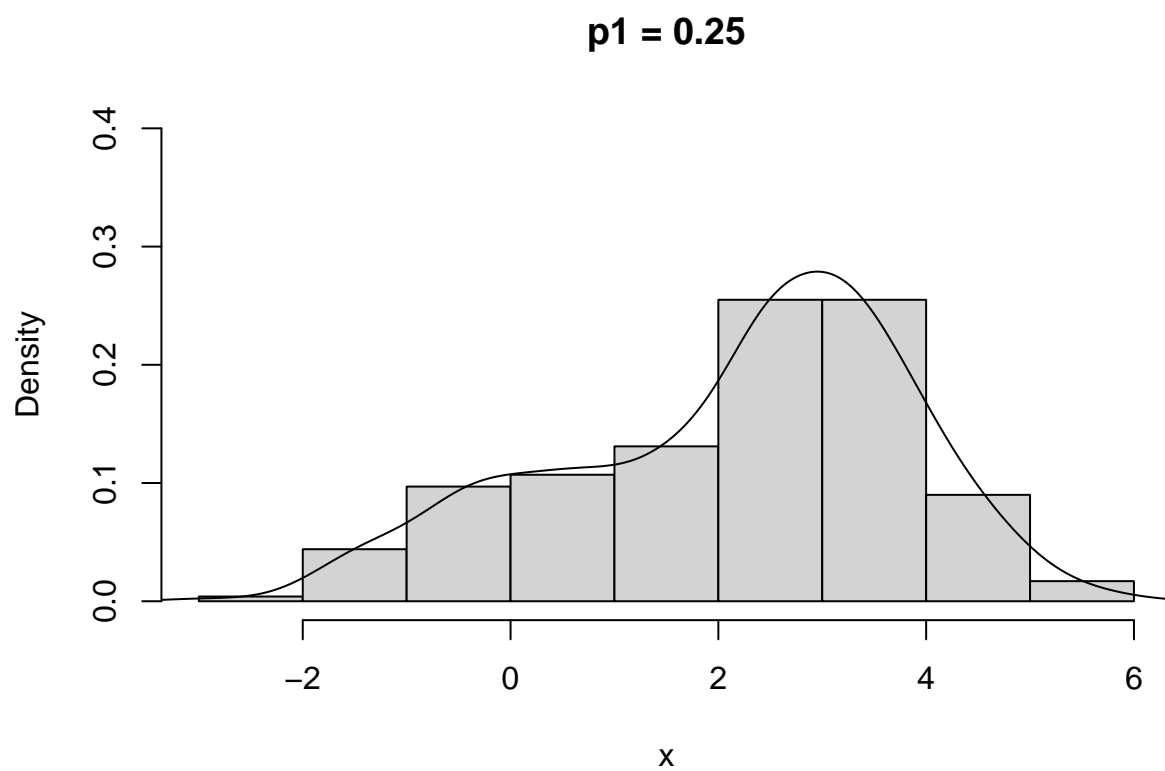

p = 0.25

```
set.seed(1111)
p1 <- 0.25
p2 <- 1-p1
u <- runif(n)
k <- as.integer(u>p2)
x <- k*x1 + (1-k)*x2
hist(x, prob = TRUE, ylim = c(0,.4), main = "p1 = 0.25")
lines(density(x))
```

p1 = 0.25

p = 0.1

```
set.seed(1111)
p1 <- 0.1
p2 <- 1-p1
u <- runif(n)
k <- as.integer(u>p2)
x <- k*x1 + (1-k)*x2

hist(x, prob = TRUE, ylim = c(0,.4), main = "p1 = 0.1")
lines(density(x))
```
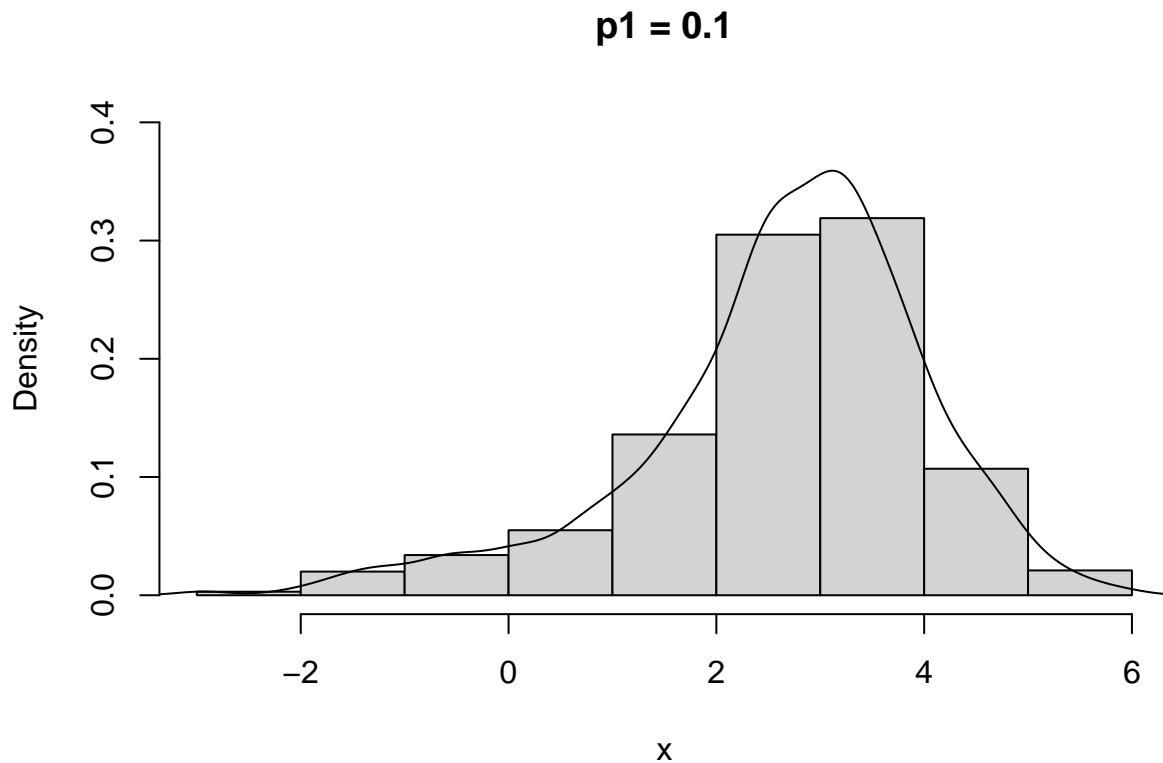
p1 = 0.1

From the graphs of we can conjecture that the values of p1 that produce bimodel mixtures is around 0.1 to 0.9.

3.14

```r
mvn_gen <- function(n, mu, sigma, factorization = "Cholesky") {
  d <- length(mu)
  Z <- matrix(rnorm(n*d), nrow = n, ncol = d)
  if(factorization == "Cholesky") { Q <- chol(sigma) } else {
    if(factorization == "Spectral") {
      ev <- eigen(sigma)
      lambda <- ev$values
      P <- ev$vectors
      Q <- P %*% diag(sqrt(lambda)) %*% t(P) } else {
        stop("Arg factorization must be 'Cholesky' or 'Spectral'.")
      } }
  mu <- matrix(mu, nrow = d, ncol = 1)
  J = matrix(1, nrow = n, ncol = 1)
  X <- Z %*% Q + J %*% t(mu)
  return(data.frame(X))
}

sig <- matrix(c(1, -0.5, 0.5, -0.5, 1, -0.5, 0.5, -0.5, 1), 3, 3, byrow = TRUE)
set.seed(1111)
```
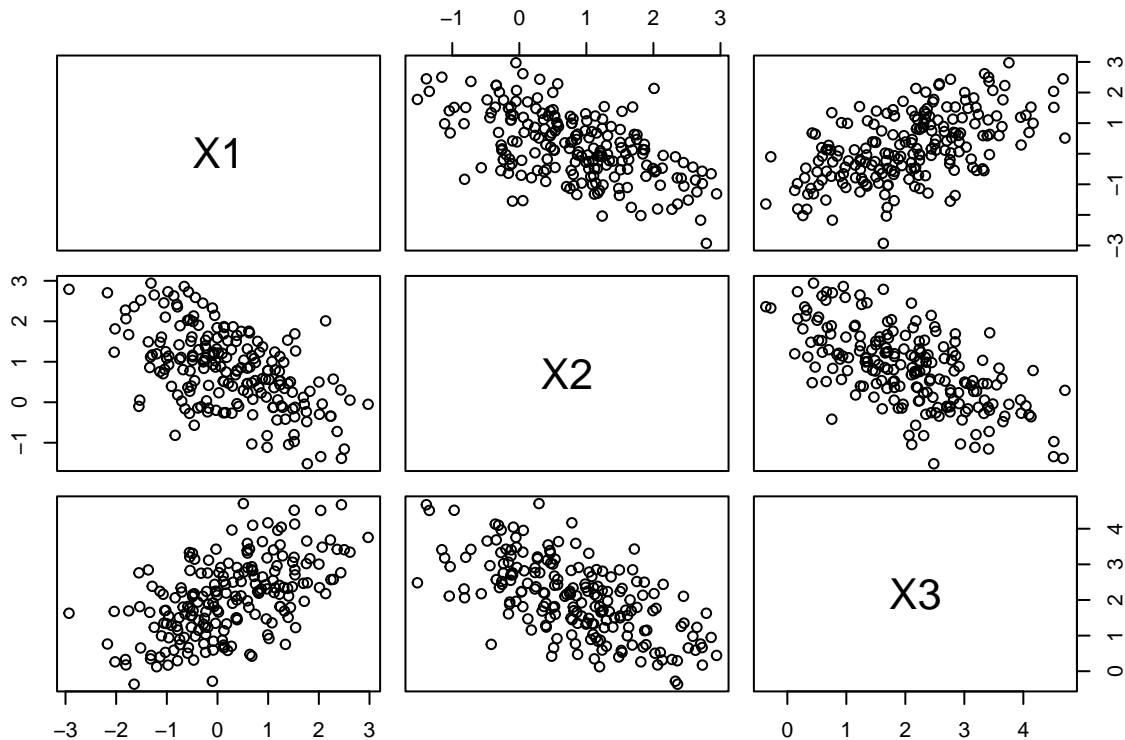
```
mv1 <- mvn_gen(n = 200, mu = c(0, 1, 2), sigma = sig, factorization = "Cholesky")
cov2cor(sig)
```

```
##      [,1] [,2] [,3]
## [1,]  1.0 -0.5  0.5
## [2,] -0.5  1.0 -0.5
## [3,]  0.5 -0.5  1.0
```

```
pairs(mv1)
```


As we take a look at the correlation plot between variables, we can see that between x1 and x2, x2 and x3, there are clear and similar negative pattern about -0.5 and between x1 and x3 there is a clear positive pattern about 0.5 so they all approximately agree with the theoretical parameters of the corresponding bivariate normal distribution.