2019302030053-胡哲-第一次作业

第三章习题

3-3

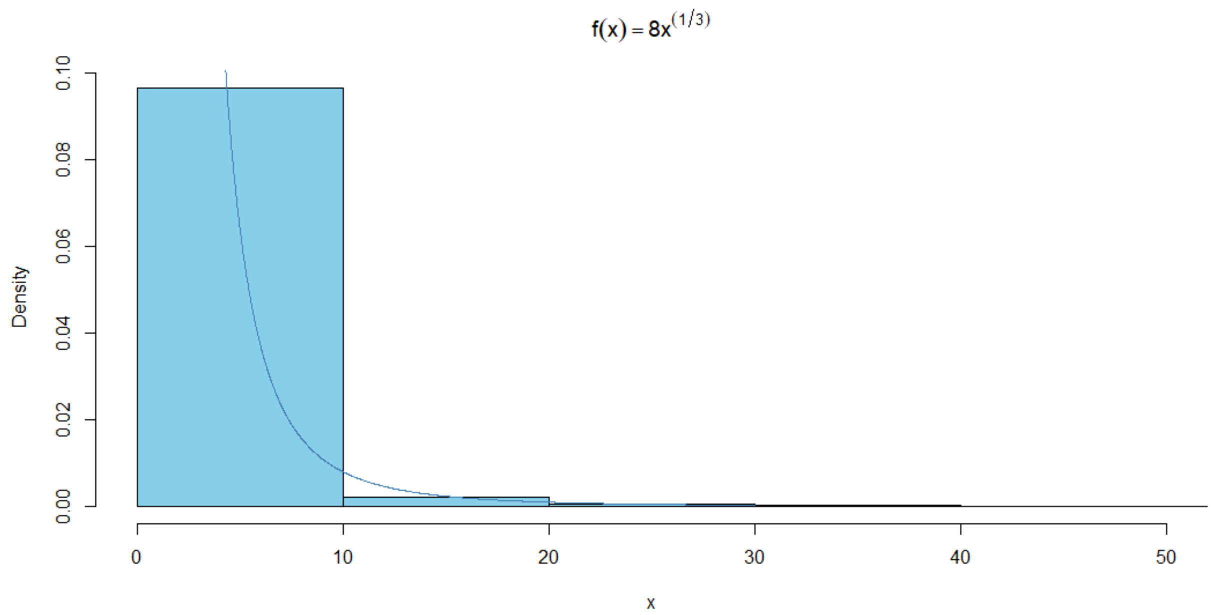
分析过程

由 题 意 $F(x) = 1 - (\frac{2}{x})^2, f(x) = 8x^{1/3}$， 利 用 逆 变 换 方 法 可 得 $F^{-1}(U) = \frac{2}{\sqrt{1-u}}$， $u \sim U(0,1)$。

代码

```
n <- 1000
u <- runif(n)
x <- 2/sqrt(1-u)
hist(x, prob = TRUE ,main = expression(f(x)==8*x^(1/3)),
     col = "skyblue",xlim = c(0,50)) # 样本密度直方图
y <- seq(2, 30, 0.01)
lines(y, 8/(y^3),col="steelblue") #密度曲线 f(x)
```

结果



结果分析

样本密度直方图和密度曲线大致吻合，说明生成效果较好。

3-4

分析过程

瑞利分布可以通过两个正态分布，生成以下内容摘自wiki：

Consider the two-dimensional vector $Y = (U, V)$ which has components that are bivariate normally distributed, centered at zero, and independent. Then U and V have density functions

$$f_U(x;\sigma) = f_V(x;\sigma) = \frac{e^{-x^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}.$$

Let X be the length of Y. That is, $X = \sqrt{U^2 + V^2}$. Then X has cumulative distribution function

$$F_X(x;\sigma) = \iint_{D_x} f_U(u;\sigma) f_V(v;\sigma)\, dA,$$

where D_x is the disk

$$D_x = \left\{(u,v) : \sqrt{u^2+v^2} \le x\right\}.$$

Writing the double integral in polar coordinates, it becomes

$$F_X(x;\sigma) = \frac{1}{2\pi\sigma^2}\int_0^{2\pi}\int_0^x re^{-r^2/(2\sigma^2)}\, dr\, d\theta = \frac{1}{\sigma^2}\int_0^x re^{-r^2/(2\sigma^2)}\, dr.$$

Finally, the probability density function for X is the derivative of its cumulative distribution function, which by the fundamental theorem of ca

$$f_X(x;\sigma) = \frac{d}{dx}F_X(x;\sigma) = \frac{x}{\sigma^2}e^{-x^2/(2\sigma^2)},$$

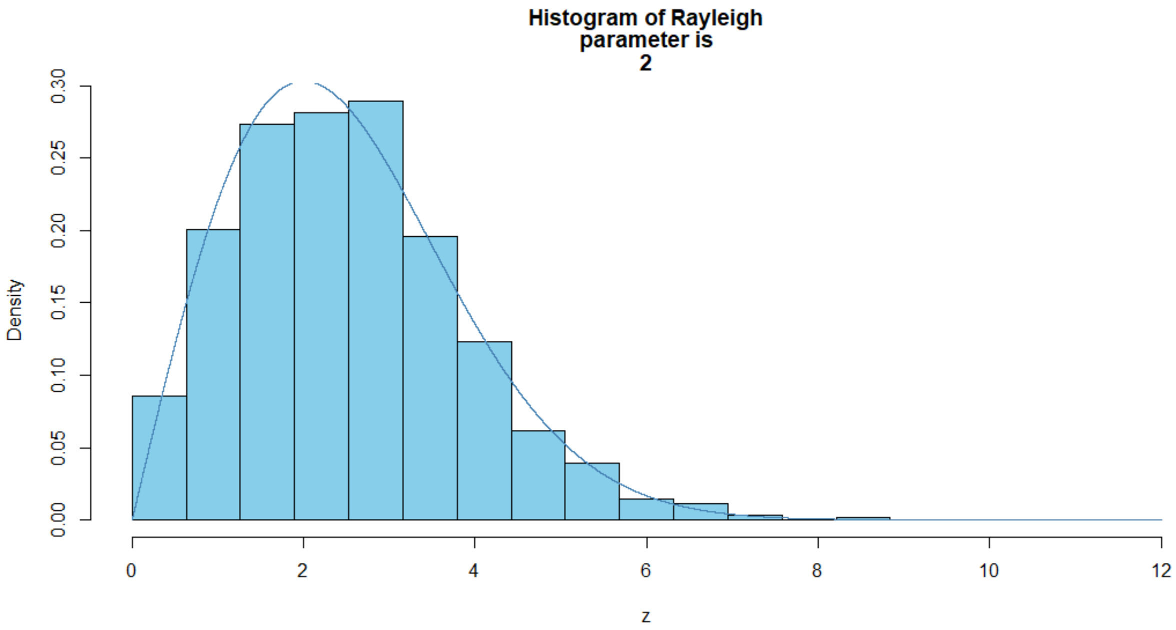
即令$Z = \sqrt{X^2 + Y^2}, X, Y \sim N(0, \sigma^2)$ 且独立，则 $Z \sim$ Rayleigh(σ) 分布.

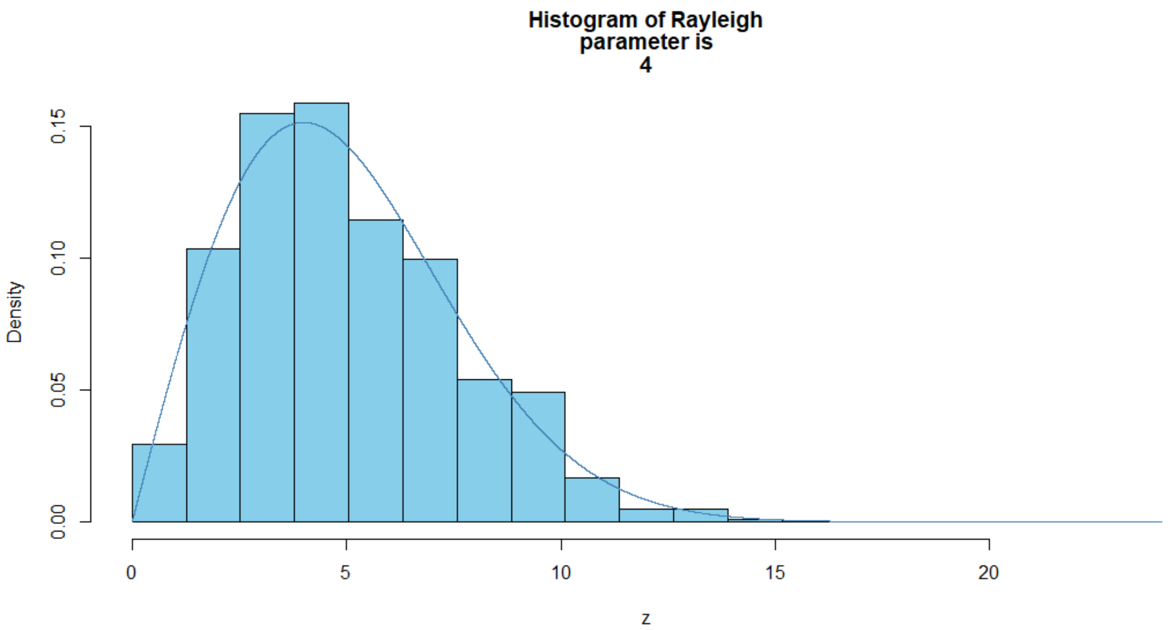
代码

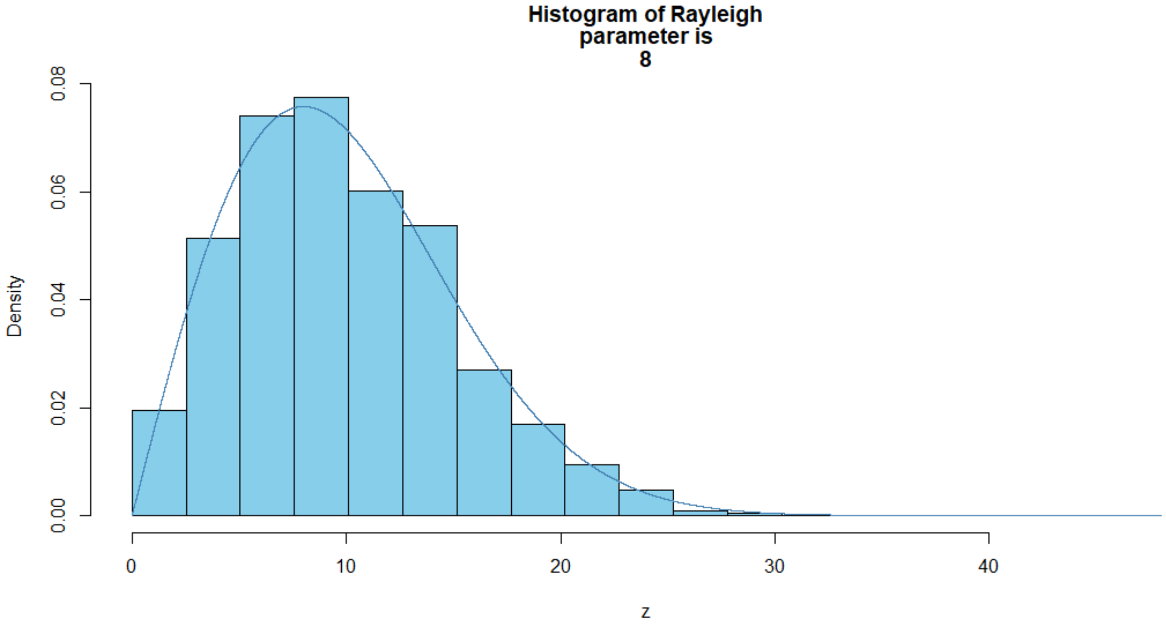
```
# 选取一系列sigma
sigma<-c(1,2,4,8,16,32)
for (i in 1:length(sigma)) {
  #设置种子以保证伪随机数的一致性
  set.seed(i)
  # 直方图的标题
  title<-c("Histogram of Rayleigh","parameter is",sigma[i])
  # 生成两个正态分布
  x<-rnorm(1000,0,sigma[i])
  y<-rnorm(1000,0,sigma[i])
  # 生成瑞利分布的随机数
  z<-sqrt(x^2+y^2)
  #绘制并检查
  hist(z,prob=TRUE,breaks = seq(0,6*sigma[i],length.out = 20)
       ,main = title,col = "skyblue")
  # 绘制Rayleigh 密度函数
  x1<-seq(0,6*sigma[i],length.out = 100000)
  y1<-(x1/sigma[i]^2)*exp(-(x1^2)/(2*sigma[i]^2))
  lines(x1,y1,col="steelblue")
}
```

结果

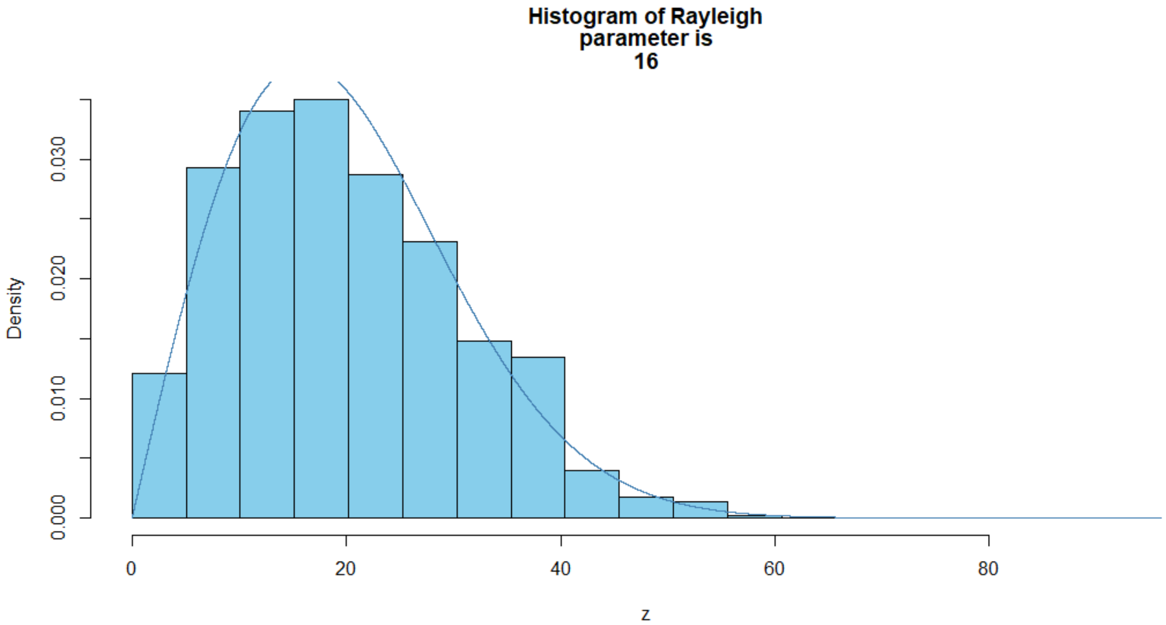
以下分别为不同取值σ对应直方图和密度曲线图



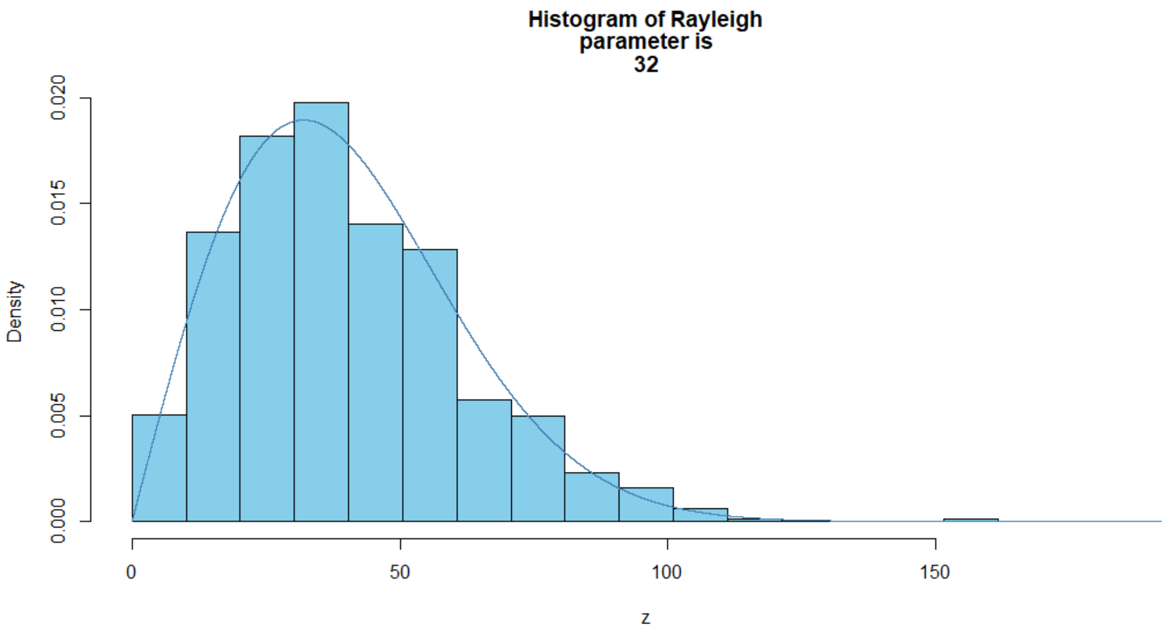




Histogram of Rayleigh
parameter is
8
Density
0.00
0.02
0.04
0.06
0.08
0
10
20
30
40
z



Histogram of Rayleigh
parameter is
16
Density
0.000
0.010
0.020
0.030
0
20
40
60
80
z



结果分析

样本密度直方图和密度曲线大致吻合，说明生成效果较好。

3-9

分析过程

令$y=\dfrac{1+x}{2}, f(y)=\dfrac{3}{4}(\dfrac{3}{4}-\dfrac{1}{4}y^2-\dfrac{1}{2}y)$,可见$Y\sim Beta(2,2)$,可以利用该变换方便的生成理论直方图和概率密度图像。

另一方面利用题目所给算法也可生成直方图和概率密度图像，二者可进行直观的比较。

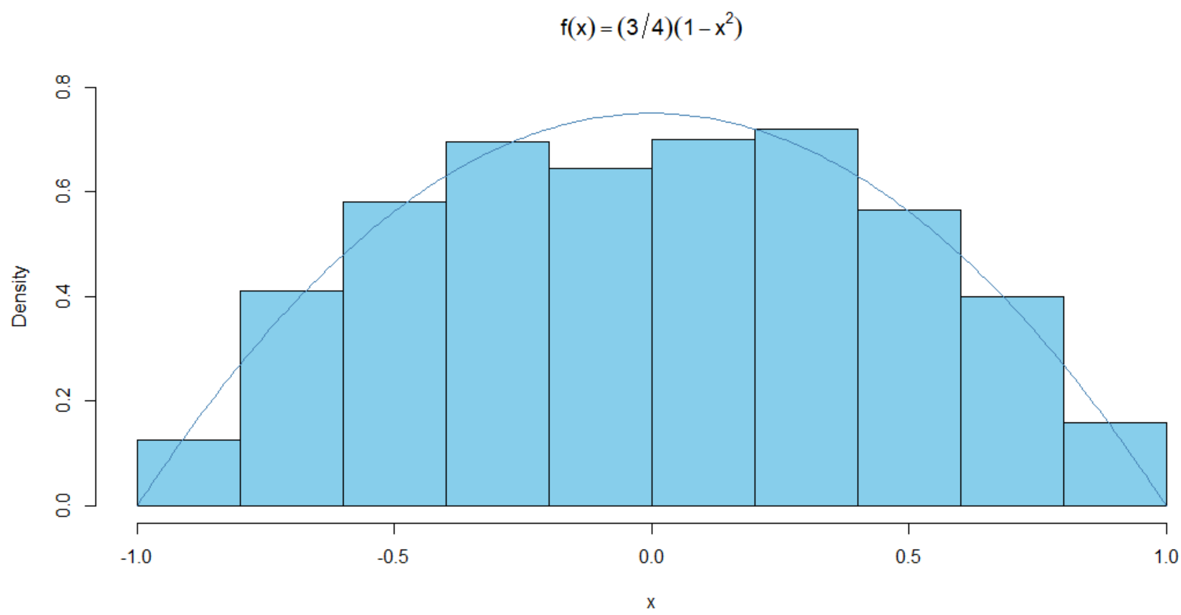
代码

```
#### 3.9 ####
n <- 1000 # 随机数个数
y <- rbeta(n,2,2) #令Y=(X+1)/2, Y~Be(2,2),由Y产生随机数
x <- 2*y-1 #将Y产生的随机数结果回代
hist(x,  prob = TRUE,main = expression(f(x)==(3/4)(1-x^2)),col = "skyblue") #
由此得到直方图
```

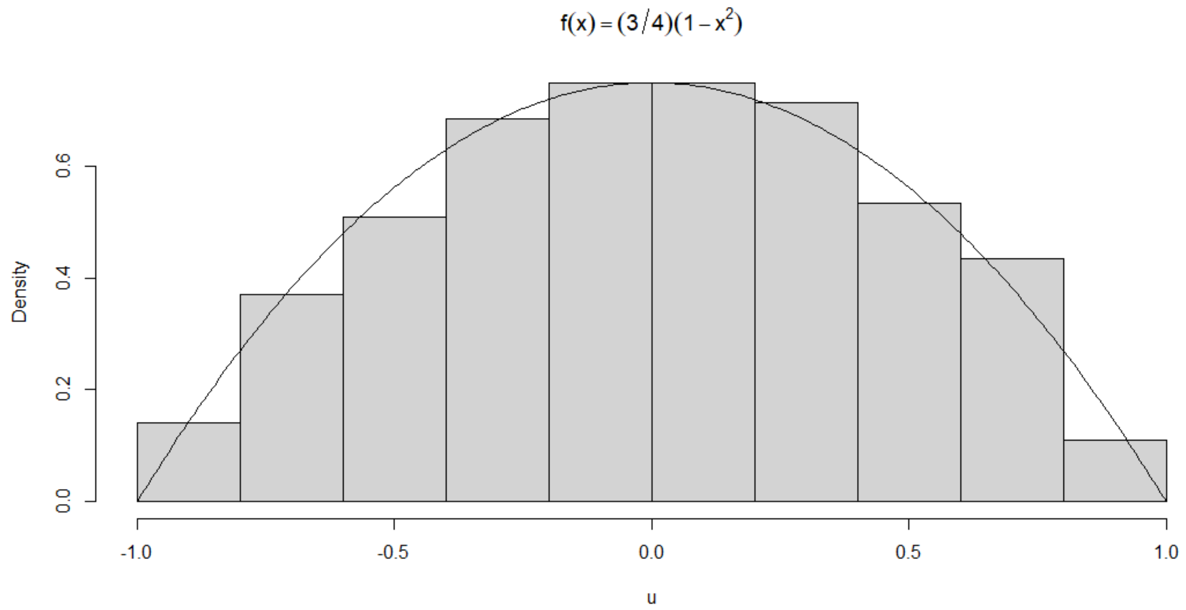
```
z <- seq(-1, 1, 0.01) #将结果进行拟合
lines(z, (3/4)*(1-z^2),col="steelblue")

n2 <- 1000  #通过题目给出的公式定义得到相同的结果图像并进行拟合
u <- vector(mode="numeric",length=1000)
for (i in 1:n2) {
  u1 <- runif(n2,-1,1)
  u2 <- runif(n2,-1,1)
  u3 <- runif(n2,-1,1)
  ifelse(abs(u3[i]) >= abs(u2[i]) && abs(u3[i]) >= abs(u1[i]),u[i] <- u2[i],
         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))
```

结果

理论图像:



算法生成图像:



结果分析

样本密度直方图和密度曲线大致吻合，而且与理论直方图图像较为接近，说明生成效果较好。

3-11

分析过程

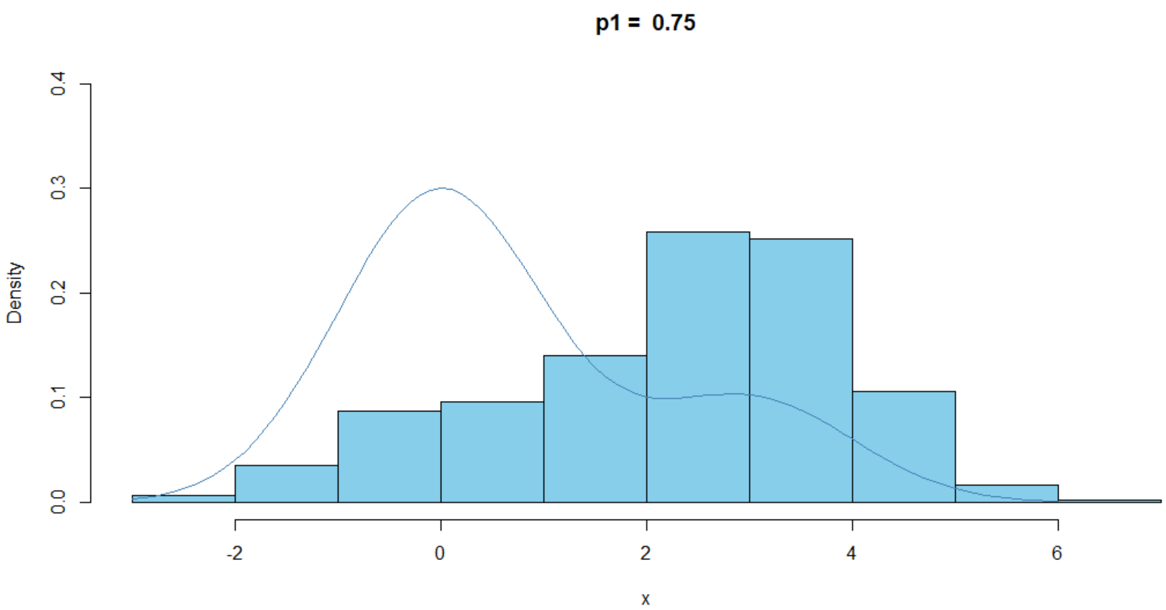
首先生成$p \sim Bio(0.75), X = pX_1 + (1-p)X_2$，$X_1 \sim N(0,1), X_2 \sim N(3,1)$,为了探究$p$对双峰的影响，可以写一个把$p$当成参数的函数。

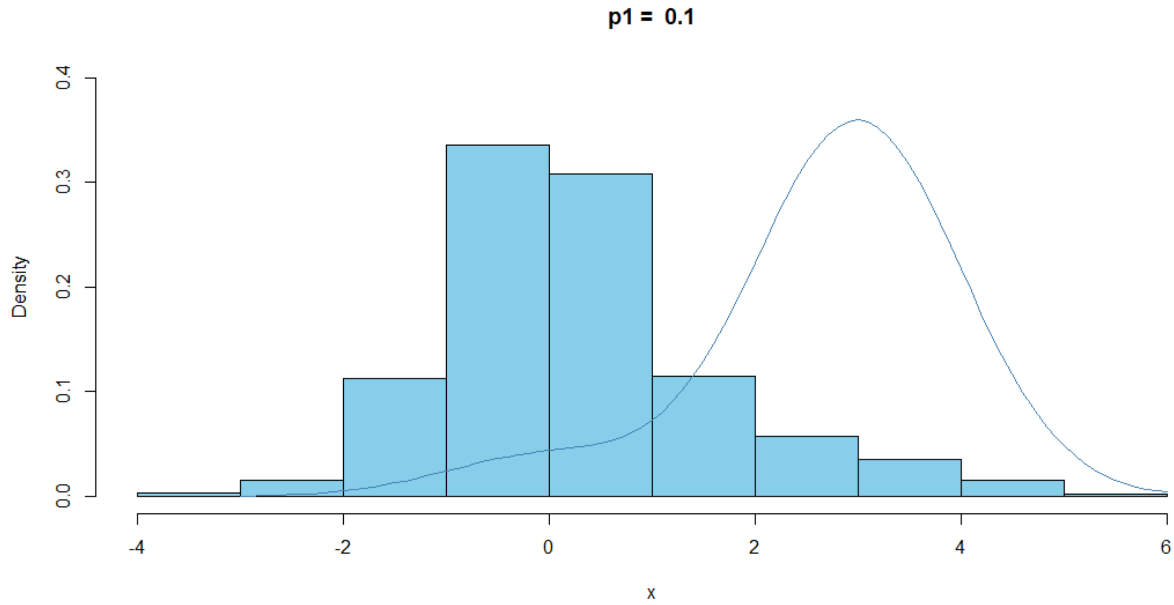
代码

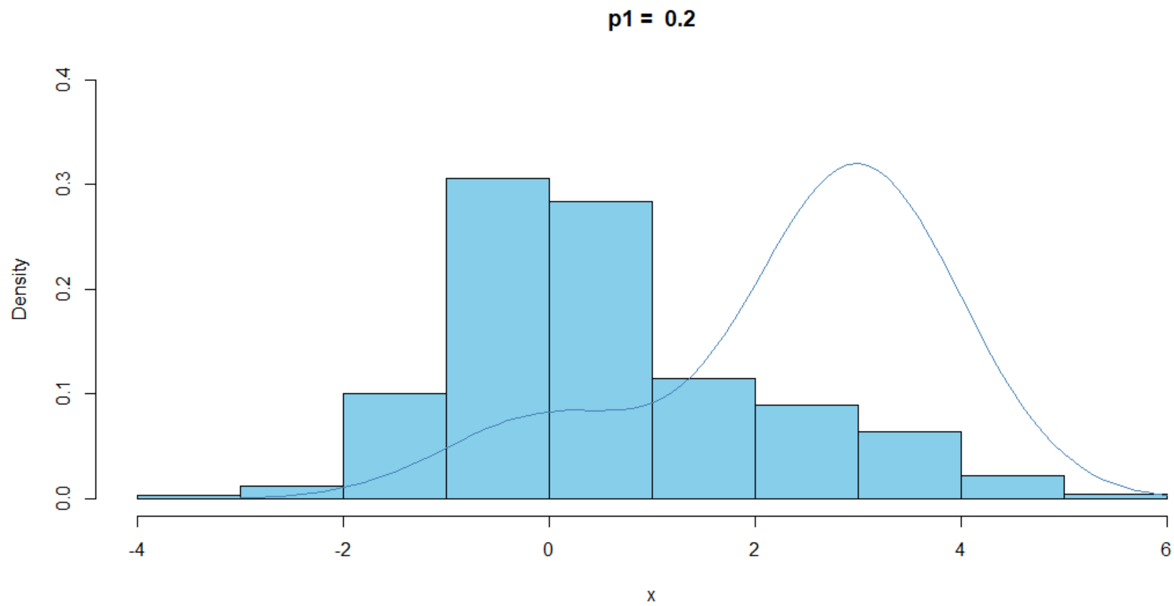
```
#### 3.11 ####
mixturehist<-function(p1){
set.seed(1012)
p<-rbinom(1000,1,prob = p1)
x1<-rnorm(1000)
x<-p*x1+(1-p)*x2 #Generate  samples
title<-paste('p1 = ',p1)
hist(x,probability = T,main = title,col = "skyblue")
```

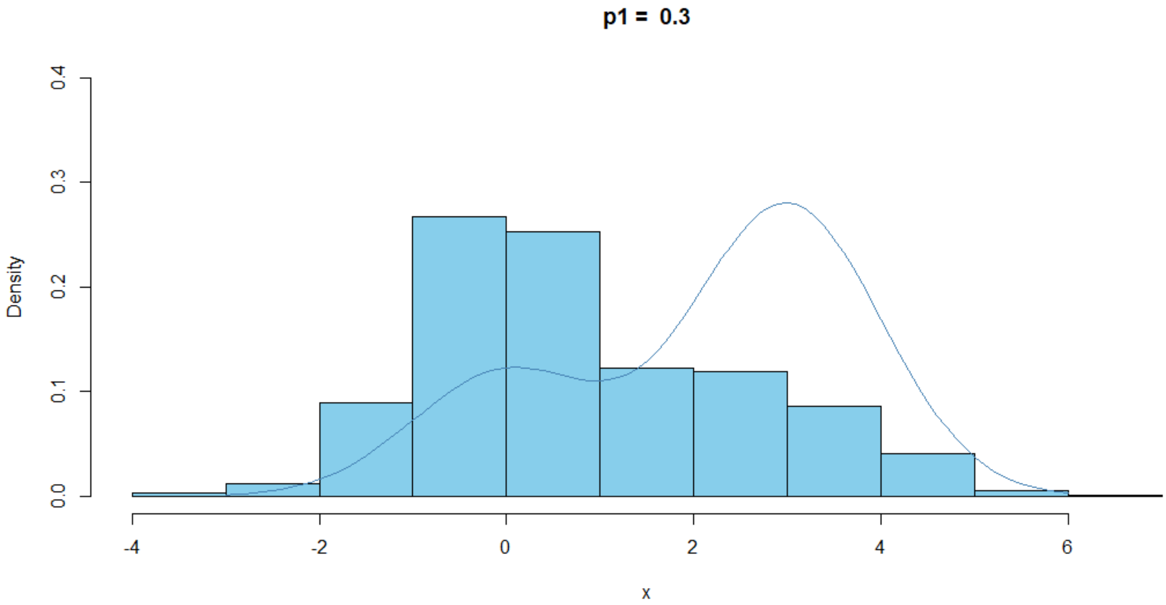
```
y<-seq(-3,6,0.1)
densityplot<-function(x){
  p1*dnorm(x)+(1-p1)*dnorm(x,3,1)
  }
lines(y,densityplot(y),col='steelblue')
}                        density(densityplot(y1 )
mixturehist(0.75)
for (p1 in seq(0.1,0.9,0.1)){
  mixturehist(p1)
}
```

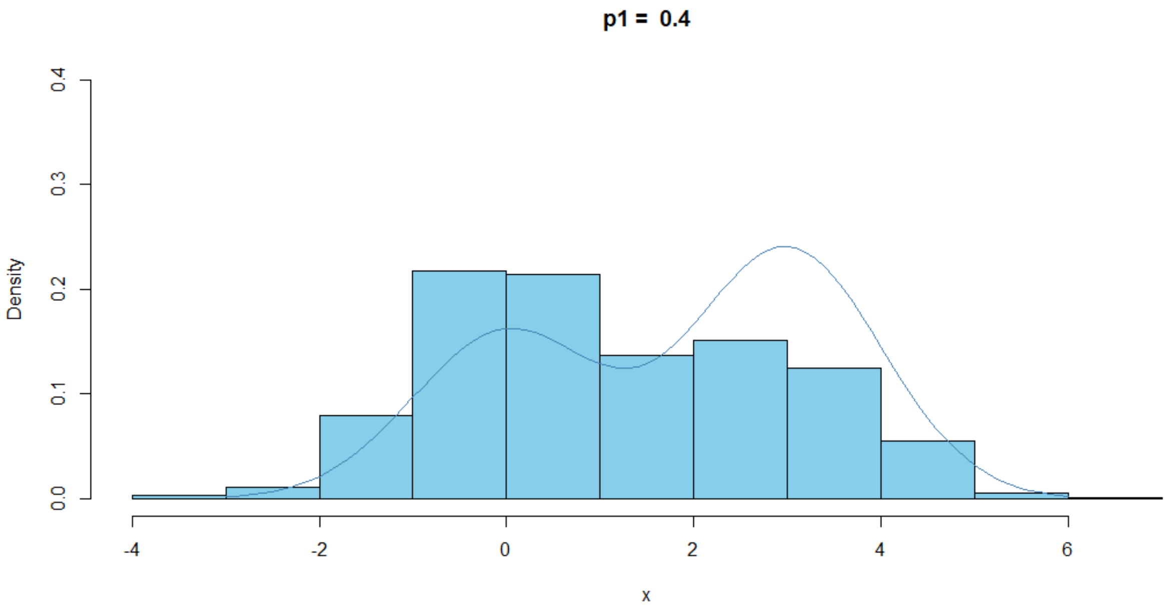
结果

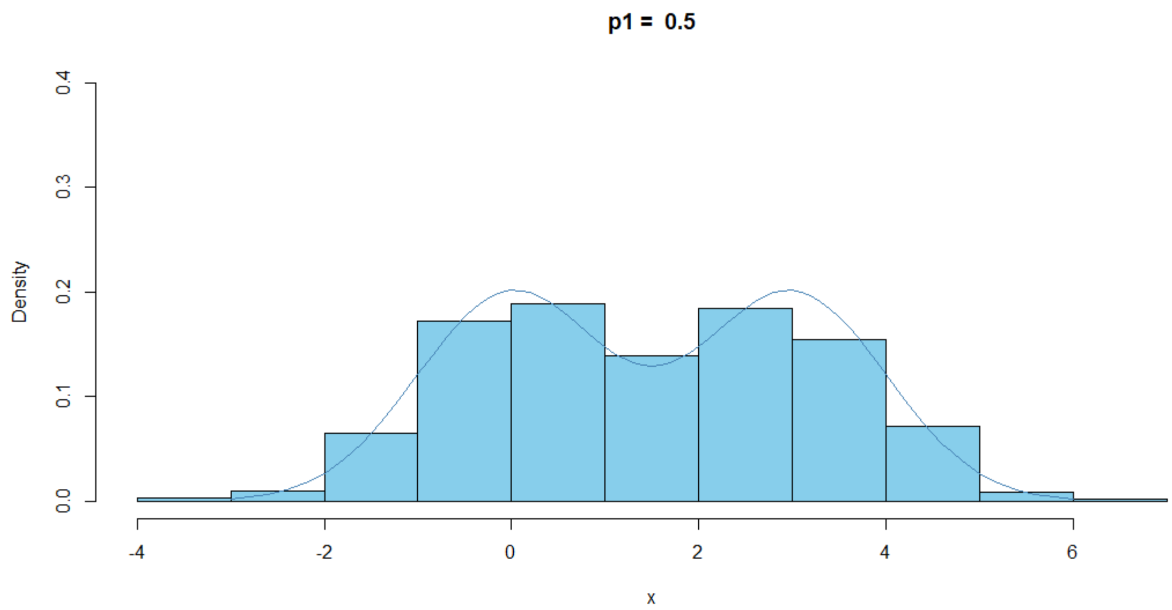


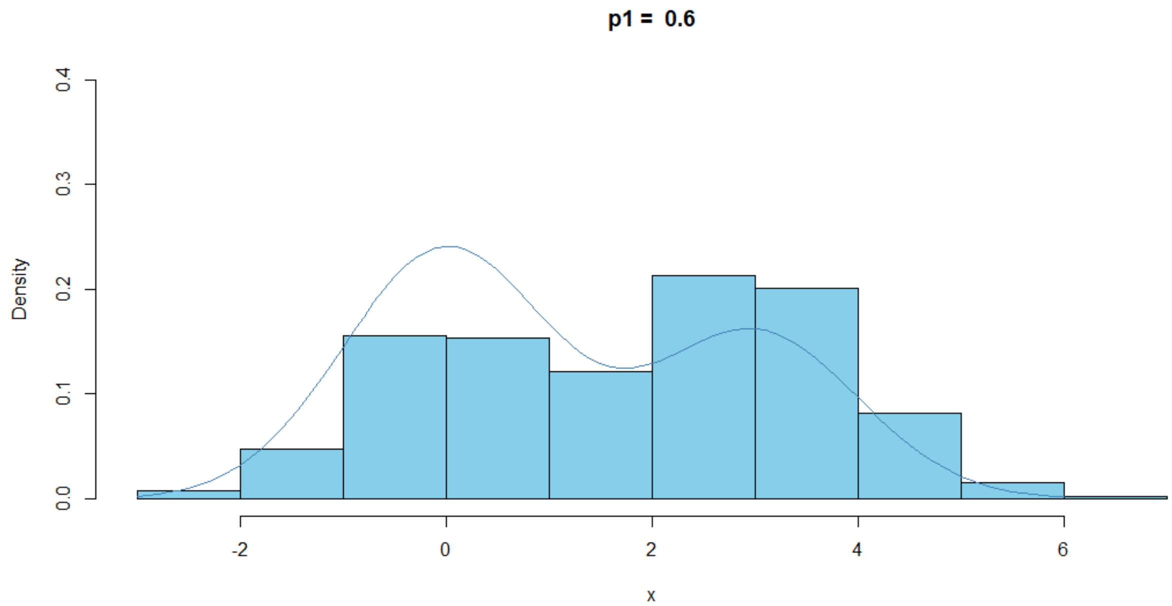

p1 = 0.1
Density
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
6
x

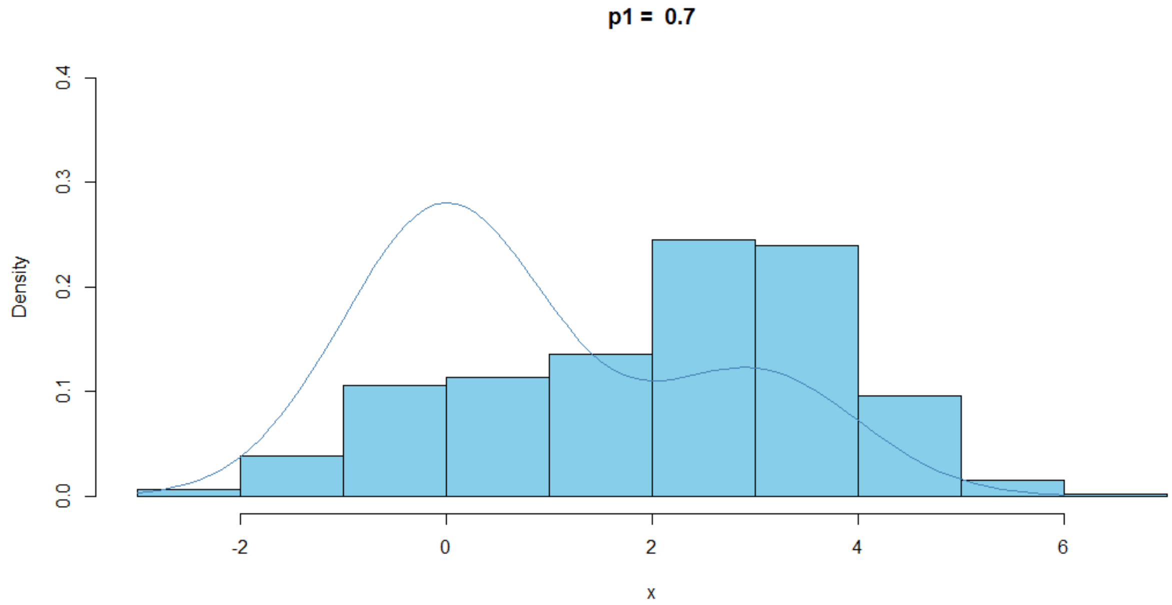

p1 = 0.2
Density
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
6
x


p1 = 0.3
Density
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
6
x

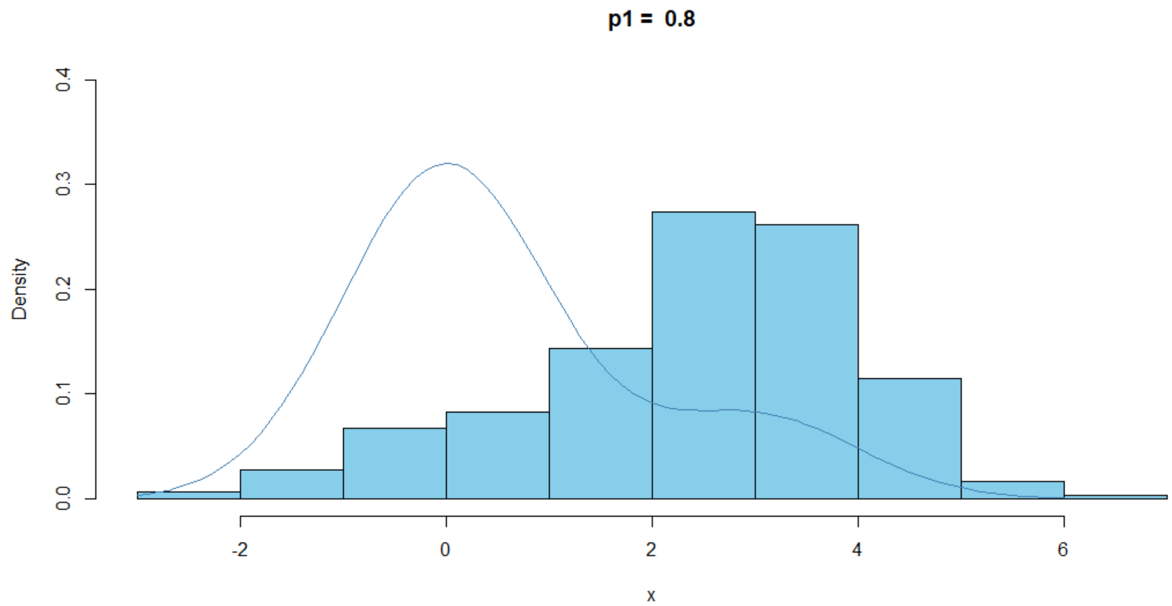

p1 = 0.4
Density
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
6
x


p1 = 0.5
Density
0.0
0.1
0.2
0.3
0.4
-4
-2
0
2
4
6
x

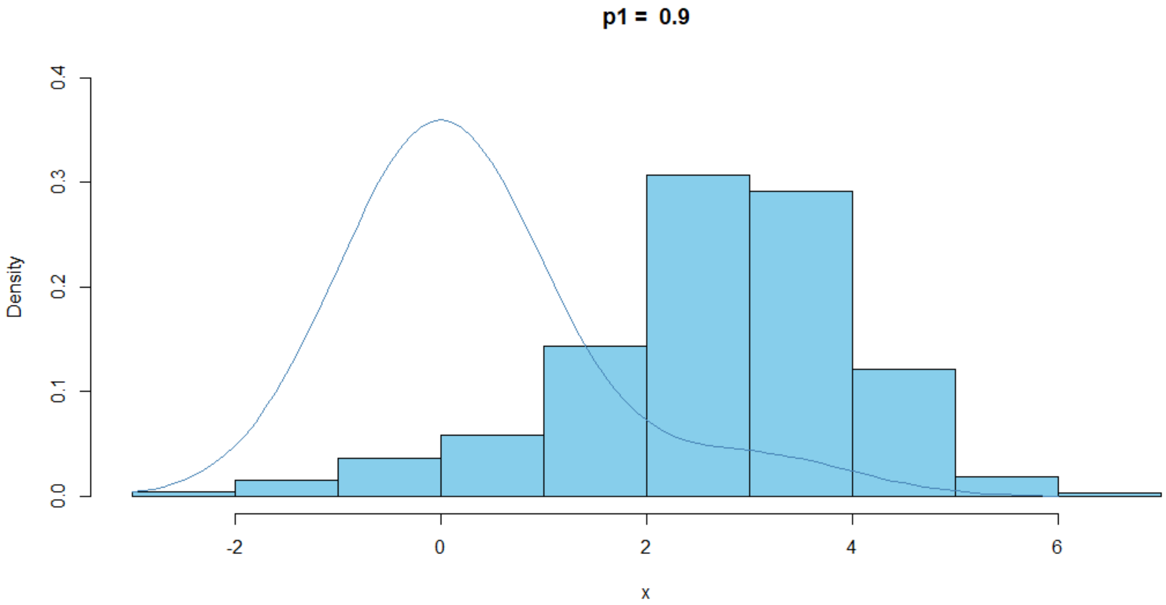

p1 = 0.6
Density
0.0
0.1
0.2
0.3
0.4
-2
0
2
4
6
x



p1 = 0.7
Density
0.0
0.1
0.2
0.3
0.4
-2
0
2
4
6
x



p1 = 0.8
Density
0.0
0.1
0.2
0.3
0.4
-2
0
2
4
6
x



结果分析

当$p \in (0.2, 0.8)$时会出现双峰现象

3-13

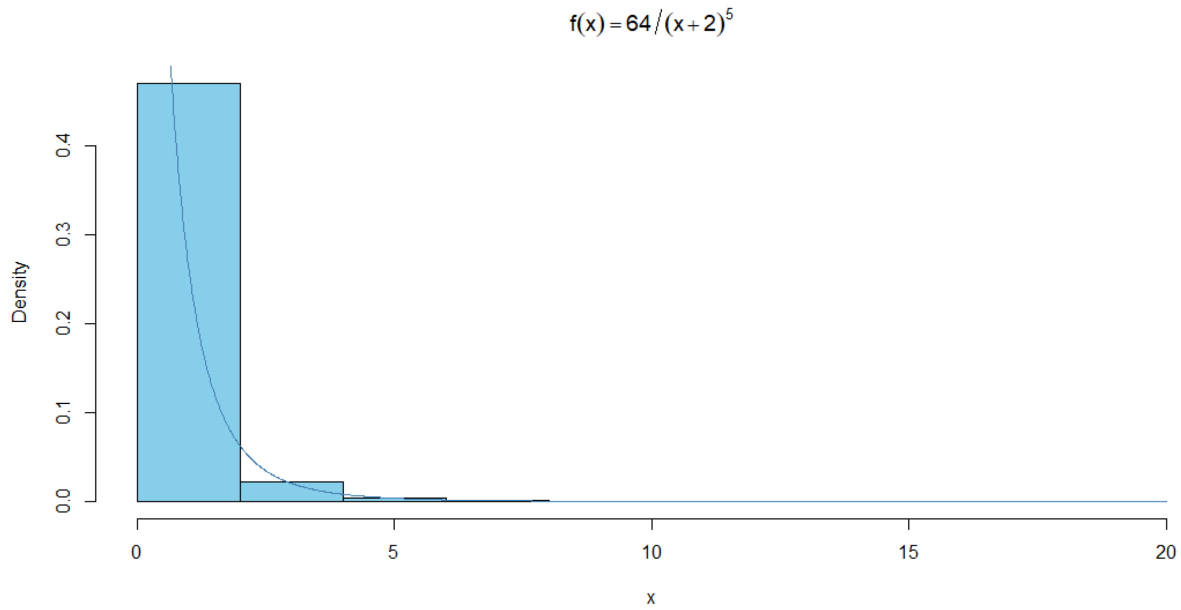
分析过程

利用逆变换方法可以进行模拟，由题意$r=4,\beta=2,F(y)=1-(\frac{2}{2+y})^4,F^{-1}(u)=(\frac{2}{1-u})^{1/4},f(y)=\frac{64}{(y+2)^5}$

代码

```
n <- 1000
u <- runif(n)
x <- (2/(1-u)^(1/4))-2
hist(x, prob = TRUE, main = bquote(f(x)==64/(x+2)^5),col = "skyblue") # 求导得密度函数
y <- seq(0, 20, 0.01) # 得到密度函数曲线
lines(y, 64/(y+2)^5,col="steelblue")
```

结果



结果分析

样本密度直方图和密度曲线大致吻合，说明生成效果较好。

3-14

分析过程

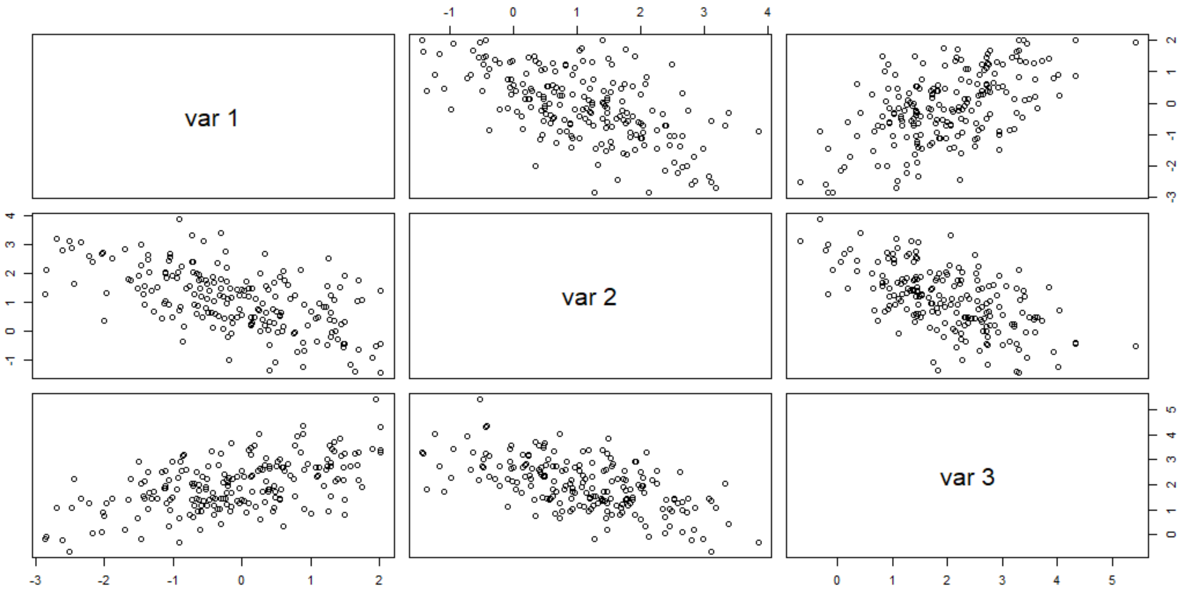
这题还需要分析吗

代码

```
mu=c(0,1,2)
Sigma=matrix(c(1,-.5,.5,-.5,1,-.5,.5,-.5,1),ncol = 3,byrow = F)
rmvn.Choleski <-
  function(n, mu, Sigma) {
    # generate n random vectors from MVN(mu, Sigma)
    # dimension is inferred from mu and Sigma
    d <- length(mu)
```

```
    Q <- chol(Sigma) # Choleski factorization of Sigma
    Z <- matrix(rnorm(n*d), nrow=n, ncol=d)
    X <- Z %*% Q + matrix(mu, n, d, byrow=TRUE)
    X
  }

X <- rmvn.Choleski(200, mu, Sigma)
# 绘图
pairs(X)
```

结果



结果分析

上图所示的样本数据散点图展示了多元正态分布的椭圆对称性，说明位置和相关性与相应二元正态分布的理论参数大致一致。

3-16

分析过程

按考试类型标准化，再计算协方差矩阵。

代码

```
#### 3.16 ###
library(bootstrap)# 没有就先下载
cov(scale(scor[,1:2]))
cov(scale(scor[,3:5]))
```

结果

```
# 闭卷                                    # 开卷
           mec       vec                          alg       ana       sta
mec 1.0000000 0.5534052                  alg 1.0000000 0.7108059 0.6647357
vec 0.5534052 1.0000000                  ana 0.7108059 1.0000000 0.6071743
                                         sta 0.6647357 0.6071743 1.0000000
```

3-20

分析过程

利用书上给的算法生成 $N(t) \sim Poss(\lambda), Then \quad Y_i \sim Gamma(shape, scale), X_i = \sum_{i=1}^{N(t)} Y_i$,之后再与理论值的期望和方差作比较

代码

```
 #### 3.20 ####
# shape: Gamma分布形状参数; scale:Gamma分布尺度参数
comp_poss <- function(lambda, shape, scale,size = 1000 ,t = 10) {
  # 到达间隔时间随速率λ呈指数分布。
  pp.exp = function (t0) {
    Tn = rexp(1000, lambda)
    Sn = cumsum(Tn)
    return(min(which(Sn > t0)) - 1)
  }

  # 生成服从泊松分布的N (t)
  ns = replicate(size, expr={ pp.exp(t)})
  # 生成题目描述的X(t)
  xs = sapply(ns, function (n) {
    ys = c(rgamma(n = n, shape = shape, scale = scale))
    sum(ys[1:n])
  })
  # 计算模拟值和理论值的差别
  # 样本
  mean.s = mean(xs)
  var.s = var(xs)

  # 理论
  mean.t = lambda * t * shape * scale
  var.t = (shape + 1) * shape * scale^2*lambda*t
  df = matrix(c(mean.s,mean.t,var.s,var.t),ncol = 4,
              dimnames =
list(c("value"),c("mean.s","mean.t","var.s","var.t")))
}
comp_poss(3,5,4)
```

结果

```
# 3 5 4
       mean.s mean.t    var.s var.t
value 595.898    600 14881.48 14400

# 1 2 3
        mean.s mean.t    var.s var.t
value 59.80842     60 511.6086   540

# 4 5 6
        mean.s mean.t    var.s var.t
value 1203.058   1200 42209.32 43200
```

结果分析

结果显示理论值和样本值的期望和方差都很接近，说明生成效果良好。