The first Homework of Statistical Computing

Lin Zikai

2023 年 9 月 22 日

目录

1 Exercise 3.5

1.1 R code

Listing 1: R code of Exercise 3.5

```
Inv_trans <- function(n){
  rs = runif(n);
  for(i in 1:n){
    temp = rs[i]
    if (temp <= 0.1){
      rs[i] = 0;
    }else if(temp <= 0.3){
      rs[i] = 1;
    }else if(temp <= 0.5){
      rs[i] = 2;
    }else if(temp <= 0.7){
```

```
      rs[i] = 3;
    }else{
      rs[i] = 4;
    }
  }
  rs
}

RVs = Inv_trans(1000);
mySize = 400;
theoretical = c(.1, .2, .2, .2, .3)
T = matrix(rep(0,5*5),5,5)
for (i in 1:5){
  mySample = sample(RVs,size=mySize, replace=F)
  empirical = table(mySample)/mySize
  T[i,] = empirical
}
T = rbind(theoretical,T)
rownames(T) = c('theoretical prob','empirical prob 1',
                'empirical prob 2','empirical prob 3',
                'empirical prob 4','empirical prob 5')
colnames(T) = c(0,1,2,3,4)
print(T)
write.csv(T,file='3.5_result.csv')
```

1.2 experiment result

I have repeated the experiment for 5 times. Compared with the theoretical probability, the five relative frequency tables are as follows:

表 1: experiment results of Exercise 3.5

	0	1	2	3	4
theoretical prob	0.1000	0.2000	0.2000	0.2000	0.3000
empirical prob 1	0.1100	0.2275	0.1875	0.2025	0.2725
empirical prob 2	0.1200	0.2125	0.1900	0.1825	0.2950
empirical prob 3	0.1375	0.1925	0.1650	0.2125	0.2925
empirical prob 4	0.1200	0.1950	0.1825	0.2025	0.3000
empirical prob 5	0.1075	0.2275	0.2250	0.1725	0.2675

2 Exercise 3.6

Suppose Y is a random variable generated from density function $g(y)$, and U from Uniform$(0,1)$. According to the procedure of the Acceptance-Rejection method, we have

$$P(\text{accepted}|Y=y) = P(U < \frac{f(y)}{cg(y)}|Y=y) = \frac{f(y)}{cg(y)}$$

$$\begin{aligned} &P(\text{Y is accepted for one iteration}) \\ &= \int_{-\infty}^{+\infty} P(\text{accepted}|Y=y)g(y)\,\mathrm{d}y \end{aligned}$$

$$
\begin{aligned}
&= \int_{-\infty}^{+\infty} \frac{f(y)}{cg(y)} g(y)\,\mathrm{d}y \\
&= \frac{1}{c}\int_{-\infty}^{+\infty} f(y) \\
&= \frac{1}{c}
\end{aligned}
$$

And finally we have

$$
\begin{aligned}
f_{Y|accepted}(y) &= \frac{f(Y=y \text{ and } \text{accpeted})}{P(\text{Y is accepted for one iteration})} \\
&= \frac{P(\text{accepted}|Y=y)g(y)}{P(\text{Y is accepted for one iteration})} \\
&= \frac{\frac{f(y)}{cg(y)}g(y)}{\frac{1}{c}} \\
&= f(y)
\end{aligned}
$$

3 Exercise 3.9

3.1 R code

Listing 2: R code of Exercise 3.9

```
fe <- function(n){
  U1 = -1 + 2*runif(n)
  U2 = -1 + 2*runif(n)
  U3 = -1 + 2*runif(n)
  for(i in 1:n){
    if (abs(U3[i])>=abs(U2[i]) && abs(U3[i])>=abs(U1[i])){
      U3[i] = U2[i]
    }
  }
  U3
}

hist(fe(10000),col='orange',breaks=20,freq = F,xlab = 'x',ylab = 'density estimate',main='
    histogram density estimate by 10000 samples ')
xx = seq(-1,1,length = 1000)
yy = 0.75*(1-xx^2)
lines(xx,yy,col='blue',lwd=2)
```

3.2 histogram density estimate

The histogram (orange bars) density estimate by 10000 samples is as follows:

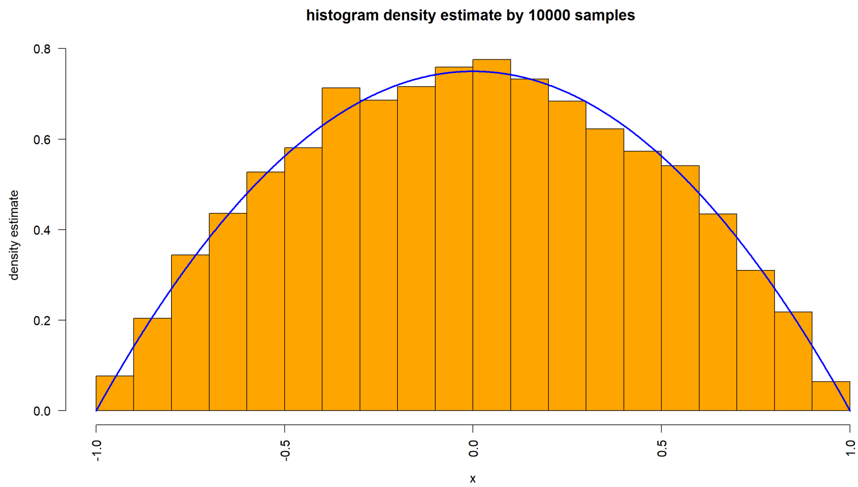

图 1: histogram density estimate of f_e

The blue line is f_e density function, which is consistent with the histogram.

4 Exercise 3.10

Assume that the variates generated by the algorithm in Exercise 3.9 is from a density $f(x)$. From the Law of Total probability, we have

$$
\begin{aligned}
f(x) &= f_{U_3}(x \Big| |U_3| \text{ is not maximum}) P(|U_3| \text{ is not maximum}) + f_{U_2}(x \Big| |U_3| \text{ is maxium}) P(|U_3| \text{ is maximum}) \\
&= f_{U_3}(x) P(|U_3| \text{ is not maximum} \Big| U_3 = x) + f_{U_2}(x) P(|U_3| \text{ is maximum} \Big| U_2 = x) \\
&= \frac{1}{2}[1 - P(|U_3| \text{ is maximum} \Big| U_3 = x)] + \frac{1}{2} P(|U_3| \text{ is maximum} \Big| U_2 = x) \\
&= \frac{1}{2}[1 - P(|U_1| < |x|, |U_2| < |x| \Big| U_3 = x)] + \frac{1}{2} P(|U_3| > |U_1| \Big| U_2 = x, |U_3| > |x|) P(|U_3| > |x| \Big| U_2 = x) \\
&= \frac{1}{2}(1 - x^2) + \frac{1}{2} \frac{\int_{|x|}^{1} t \, \mathrm{d}t}{(1 - |x|)} (1 - |x|) \\
&= \frac{3}{4}(1 - x^2)
\end{aligned}
$$

Therefore, the algorithm given in Exercise 3.9 generates variates from the density f_e.

5 Exercise 3.12 and 3.13

5.1 R code

Listing 3: R code of Exercise 3.12 and 3.13

```
Exp_Gamma_mixture <- function(r,beta,n){
  U1 = rgamma(n,r,beta)
  U2 = rexp(n,U1)
}

r = 4
```

```
beta = 2
results = Exp_Gamma_mixture(r,beta,1000)
hist(results,col='orange',probability = T,breaks=30,xlab='x',ylab = 'density estimate',
    main='density histogram of the sample from Exponential-Gamma mixture and Pareto density
        curve')
xx = seq(min(results),max(results),length=1000)
yy = beta^r * r * (beta+xx)^(-r-1)
lines(xx,yy,col='blue',lwd=2)
```

5.2 density histogram

The empirical and theoretical (Pareto) distributions are as follows:

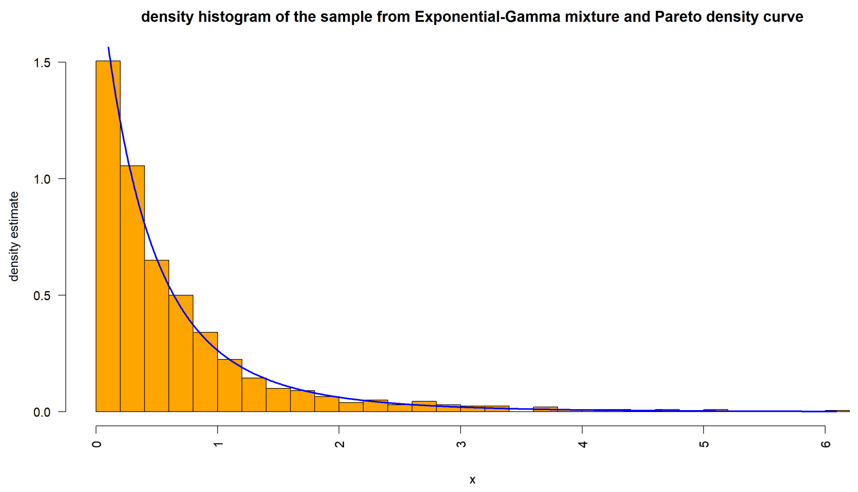

图 2: Histogram of sample from Exp-Gamma mixture and the Pareto density Curve

We can observe that the density histogram of the sample from Exponential-Gamma mixture is consistent with the Pareto density curve.

6 Exercise 3.14

6.1 R code

Listing 4: R code of Exercise 3.14

```
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
  }

Sigma = c(1,-.5,.5,-.5,1,-.5,.5,-.5,1)
```

```
Sigma = matrix(Sigma,nrow=3)
miu = c(0,1,2)
samples = rmvn.Choleski(200,miu,Sigma)

X1 = samples[,1]
X2 = samples[,2]
X3 = samples[,3]
pairs(samples)
apply(samples,MARGIN = 2,FUN = mean)
cor(X1,X2)
cor(X1,X3)
cor(X2,X3)
```

6.2 scatter plots for each pair of variables

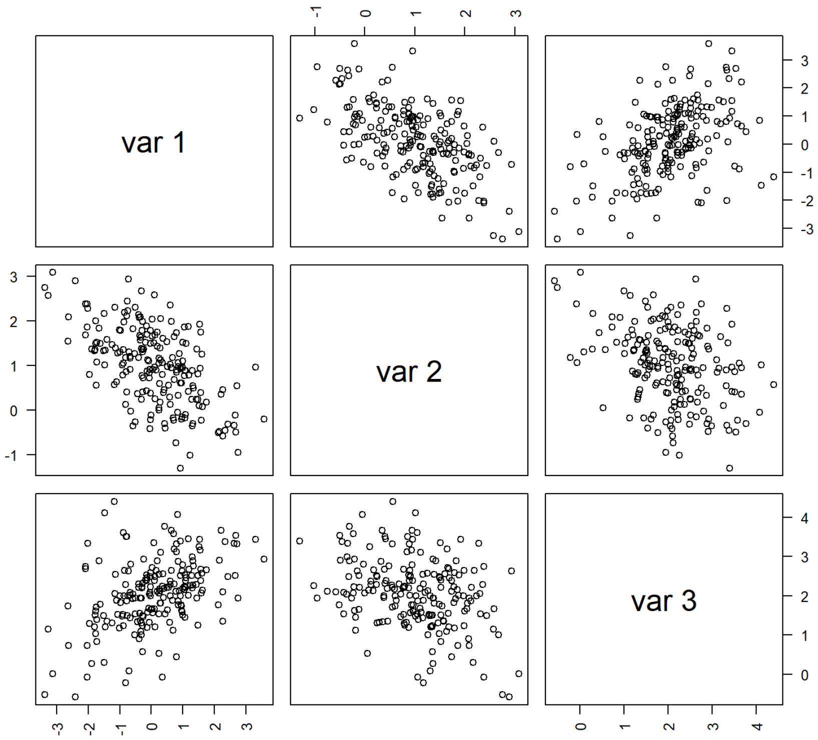

图 3: scatter plots for each pair of variables

From fig3, we can see that the scatter plot of var1 and var2 (Plot-12 for short) is centered at approximately (0,1), scatter plot of var1 and var3 (Plot-13 for short) is centered at approximately (0,2), and scatter plot of var2 and var3 (Plot-23 for short) is centered at (1,2), which correspond to the parameters $\mu = (0, 1, 2)$. On the other hand, Plot-12 and Plot-23 are distributed around the main diagonal, while Plot-13 is distributed around the secondary diagonal, which are consistent with the parameters $\rho_{12} = \rho_{23} = -0.5$, $\rho_{13} = 0.5$.