统计计算第一次作业

Code

苏锦华 2017201620

题目要求

问题1、给定$\lambda = 0.5$, 用直接抽样法产生5000个具有如下密度函数的随机数。并且画出该组随机数的直方图，和相应的密度函数曲线。问该直方图是否和密度函数很接近。

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0(\lambda > 0); \\ 0, & x \leq 0. \end{cases}$$

问题2、 产生5000个具有如下分布列的随机数。并且验证取不同值的概率和估计概率是否相近。

x_i	1	2	3	4	5	6
p_i	0.05	0.05	0.1	0.1	0.6	0.1

问题3、 设X_1, X_2为独立随机变量，密度函数分别是$f_1(x)$和$f_2(x)$. 令$\xi = g(X_1, X_2)$，则ξ的密度函数是什么？

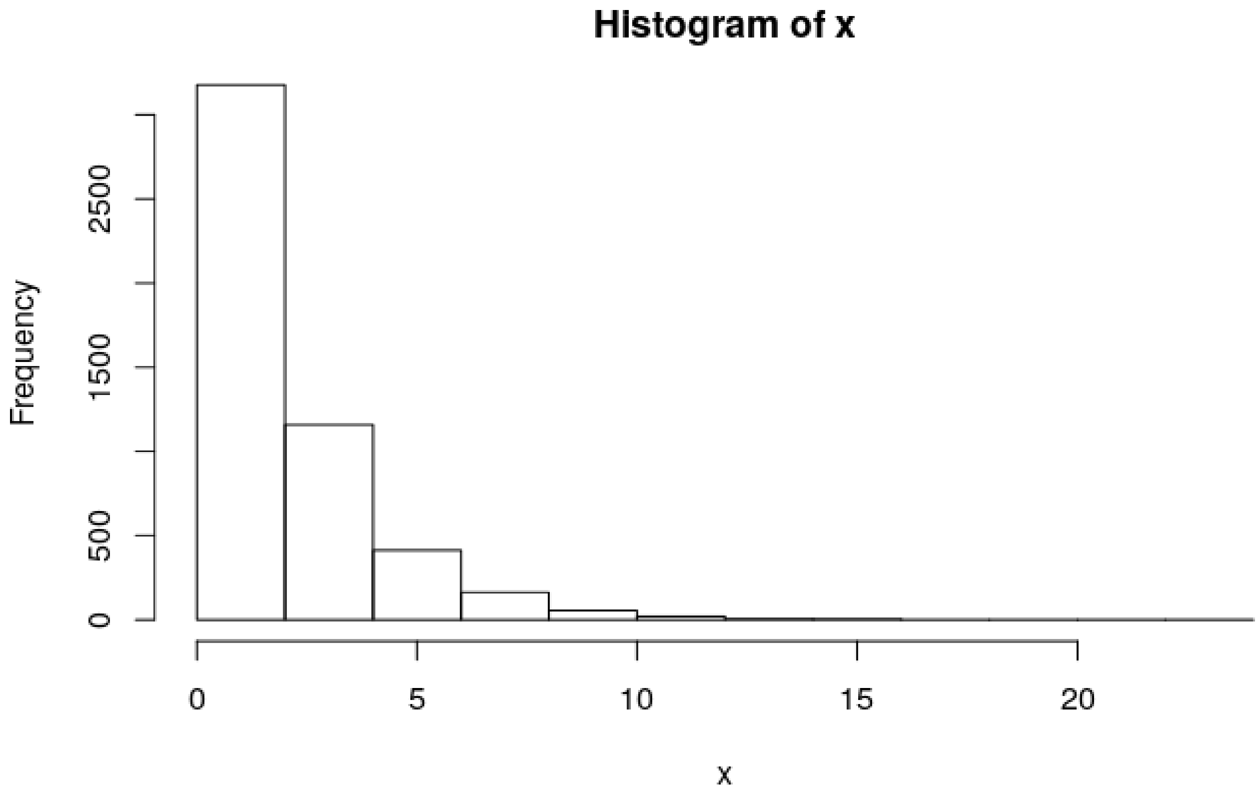
问题4、 找两个可以用舍选抽样获得随机数的例子，一个在有限区间密度函数是非零的，一个在无限区间密度函数是非零的。并且说明如何对这两个例子用舍选法进行抽样。

问题1

根据直接抽样法公式$R = F(\xi) = \int_0^{\xi} \lambda e^{-\lambda x} dx = 1 - e^{-\lambda \xi}$,经变换得$\xi = -\frac{1}{\lambda}\ln(1 - R)$

Hide

```
r <- runif(5000,min=0,max=1)
x <- -2 * log(1-r)
hist(x)
```



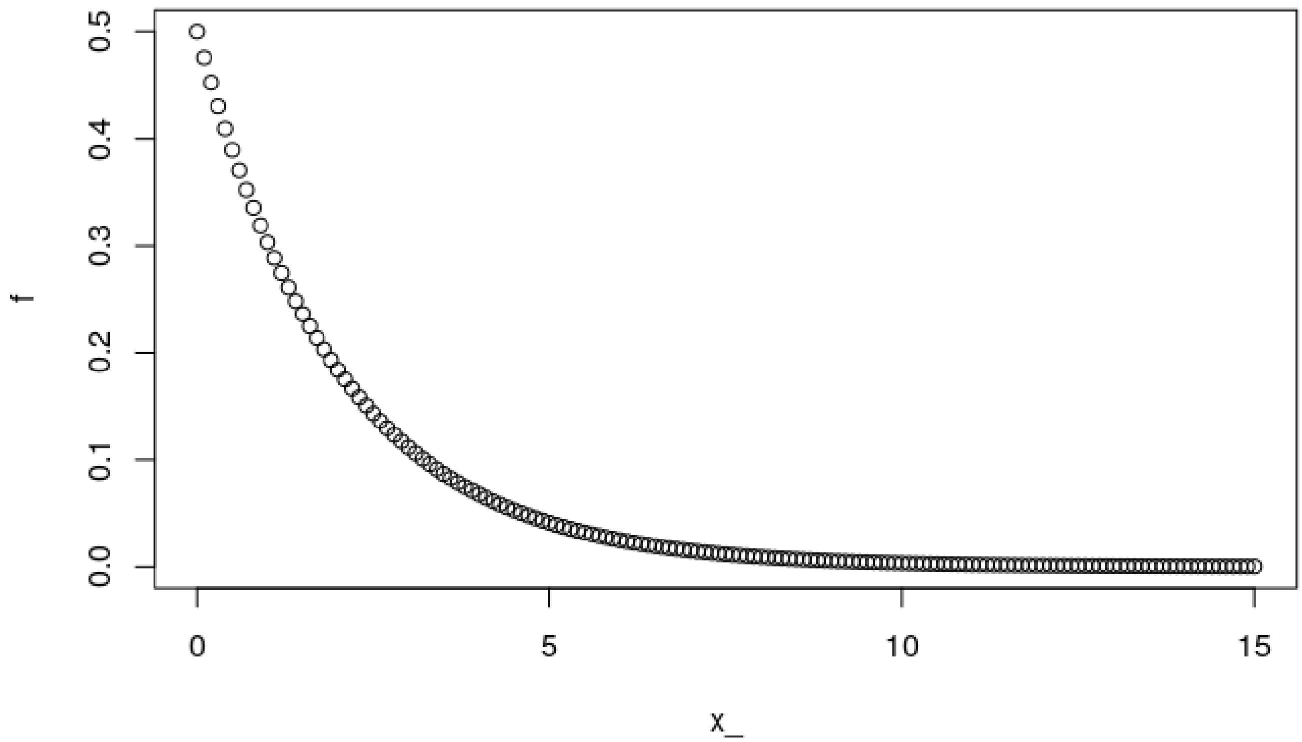
Hide

```
x_ <- seq(0,15,0.1)
```

再画出密度曲线进行对比,发现形状很接近。

Hide

```
f <- 0.5 * exp(-0.5 * x_)
plot(x_,f)
```

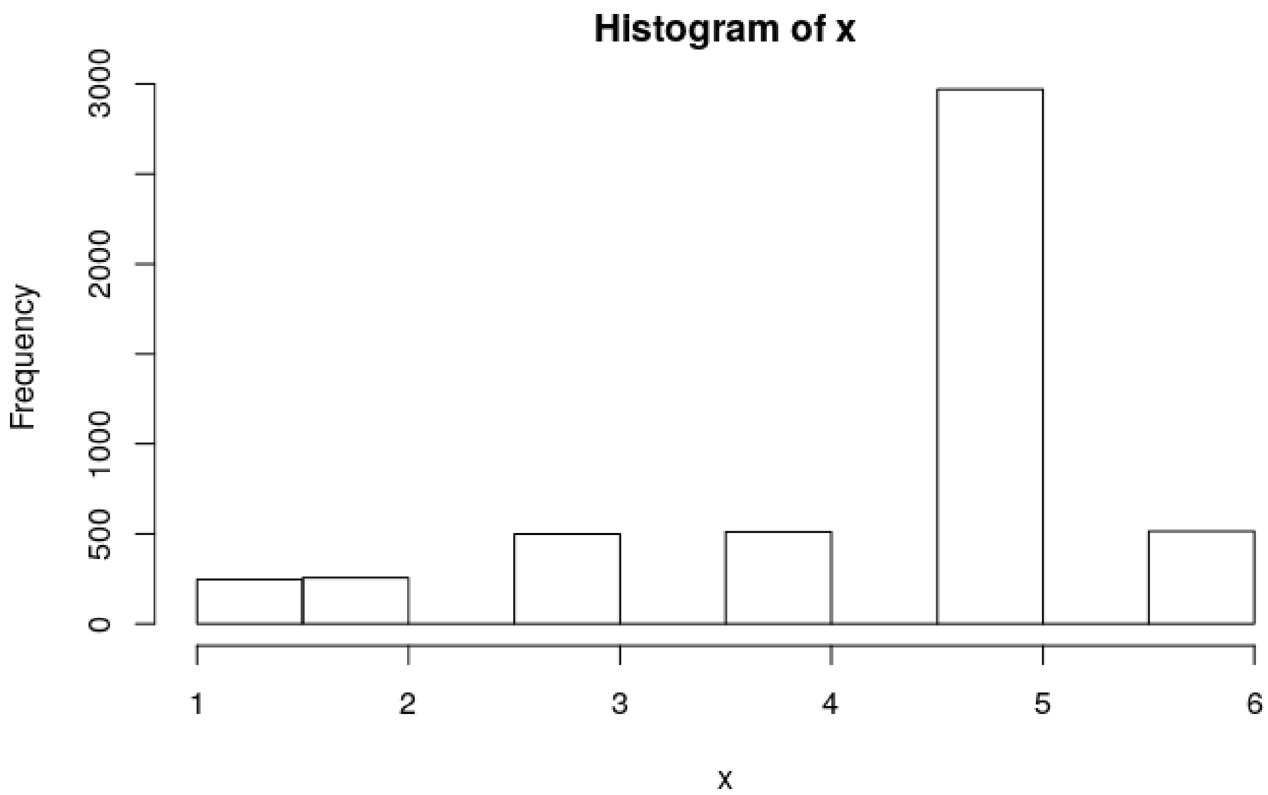


问题2

根据直接抽样法，抽取0-1的均匀随机数r，以x的累积概率值作为分段点，取r落入概率区间右端点所对应的x取值，作为生成结果。

Hide

```
r <- runif(5000)
x <- rep(1,times=5000)
for (i in 1:5000) {
  tmp <- r[i]
  if(tmp > 0.05 & tmp <= 0.1){
    x[i] <- 2
  } else{
     if(tmp > 0.1 & tmp <= 0.2){
       x[i] <- 3
     }else{
       if(tmp > 0.2 & tmp <= 0.3){
         x[i] <- 4
       }
       else{
         if(tmp > 0.3 & tmp <= 0.9){
           x[i] <- 5
         } else {
           if(tmp > 0.9){
             x[i] <- 6
           }
         }
       }
     }
  }
}
hist(x)
```



直方图显示结果符合离散概率分布，取频率作为概率估计值，发现误差最大为0.006,未经统计检验，直观感觉符合要求。

Hide

```
p <- rep(0,6)
for(i in 1:6) {
  p[i] <- length(which(x==i)) / 5000
  print(cat("P(x=",i ,")=",p[i],sep=""))
}
```

```
P(x=1)=0.0496NULL
P(x=2)=0.0514NULL
P(x=3)=0.0998NULL
P(x=4)=0.1022NULL
P(x=5)=0.594NULL
P(x=6)=0.103NULL
```

问题3

受到2维变换抽样公式启发，构造辅助变量$\eta = X_1$。求f1与f2的唯一反变换，记为$X_1 = h_1(\xi, \eta), X_2 = h_2(\xi, \eta)$,求$(X_1, X_2)$关于$(\xi, \eta)$的Jacobi行列式。

由变换抽样公式可得$p(\xi, \eta) = f(h_1(\xi, \eta), h_2(\xi, \eta))|J| = f_1(h_1(\xi, \eta))f_2(h_2(\xi, \eta))|J|$,两边再乘以$d\eta$进行积分，最终得到独立的密度函数。

问题4

使用PPT中的例子：

有限区间:

$\xi \to \beta(a, b)$ - 先求p(x)最大值D: $D = \frac{1}{B(a.b)}(\frac{a-1}{a+b-2})^{a-1}(\frac{b-1}{a+b-2})^{b-1}, x = \frac{a-1}{a+b-2}$,令h(x)=p(x)/D - 生成U(0,1)分布的X - 生成与X独立的同样U(0,1)分布的R - 如果R<=H(X)，令Z=X，并输出密度为p(x)的随机数Z;否则从第二步开始重新抽样

无限区间:正半轴半正态分布

- 注意到半正态分布可以分解出一个指数分布$p(x) = \sqrt{\frac{2}{\pi}}e^{-\frac{(x-1)^2}{2}}e^{-x} = C_0h(x)f(x)$
- 生成参数为1的指数分布的X
- 生成与X独立的同样U(0,1)分布的R
- 如果R<=H(X)，令Z=X，并输出密度为p(x)的随机数Z;否则从第二步开始重新抽样