随机变量的生成

王宁宁

目录

1 简介

```
Nsim=10^4 #number of random numbers
x=runif(Nsim)
x1=x[-Nsim] #vectors to plot
x2=x[-1] #adjacent pairs
par(mfrow=c(1,3))
```

```r
hist(x)
plot(x1,x2)
acf(x)
```

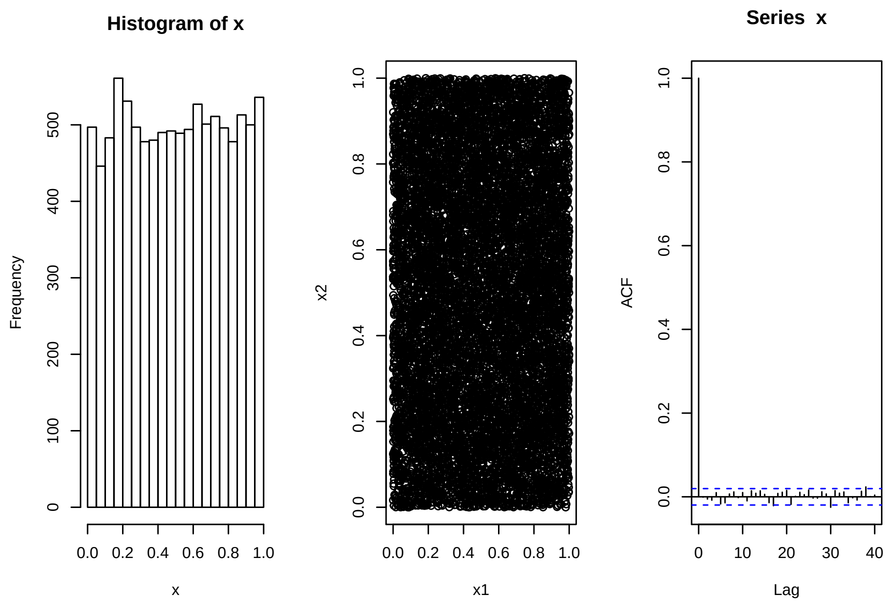

图 1:

随机数的生成都依赖均匀分布的随机数，但均匀分布的随机数所有的算法都是生成伪随机数，只要生产函数的种子确定了，随机数的序列就确定了。

```r
set.seed(1)
runif(5)
```

```
## [1] 0.2655087 0.3721239 0.5728534 0.9082078 0.2016819
```

```r
set.seed(1)
runif(5)
```

```
## [1] 0.2655087 0.3721239 0.5728534 0.9082078 0.2016819
```

```r
set.seed(2)
runif(5)
```

```
## [1] 0.1848823 0.7023740 0.5733263 0.1680519 0.9438393
```

2 反函数法

下面假定 $F(x)$ 严格单调：

2.1 原理

随机变量 $X \sim F(x)$，则 $F(X)$ 服从均匀分布 $U(0,1)$。
证明：

$$P(U \leq u) = P(F(X) \leq F(x)) = P(F^{-1}(F(X)) \leq F^{-1}(F(x))) = P(X \leq x)$$

2.2 推论

如果随机变量 $U \sim U(0,1)$，定义

$$F^{-}(u) = inf\{x; F(x) \geq U\}$$

则：$F^{-}(U) \sim F(x)$
根据上面的推论，如果某个分布可以求出反函数，那么只要有了均匀分布的随机数，就可以知道这个分布的随机数。

2.3 例 1

假定 $X \sim Exp(1)$，$F(x) = 1 - e^{-x}$，反函数为 $x = -log(1-u)$，所以：假定 $U \sim U(0,1)$ 则：$X = -logU \sim Exp(1)$

```r
set.seed(30)
Nsim=10^4
U=runif(Nsim)
X=-log(U) #transforms of uniforms
Y=rexp(Nsim) #exponentials from R
```

```
par(mfrow=c(1,2)) #plots
hist(X,freq=F,breaks=8,main="Exp from Uniform")
hist(Y,freq=F,breaks=8,main="Exp from R")
```

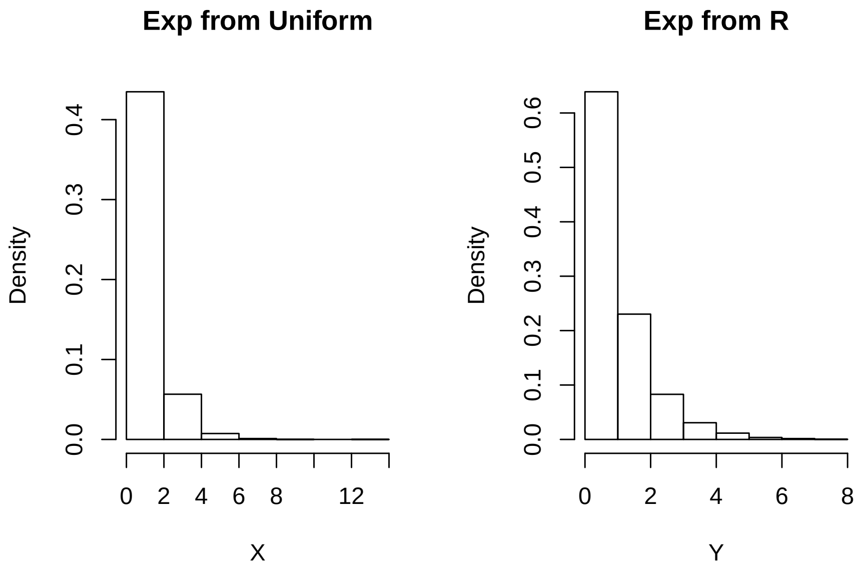

图 2:

2.4 练习：

假定 $X_i \sim Exp(1)$ 且相互独立，求下列分布：

$$Y = \sum_{j=1}^{n} X_j \tag{1}$$

$$Y = b\sum_{j=1}^{a} X_j \tag{2}$$

$$Y = \frac{\sum_{j=1}^{a} X_j}{\sum_{j=1}^{a+b} X_j} \tag{3}$$

3 正态分布的生成

Box-Muller 算法：
假定：$U_1, U_2 \sim U(0,1)$ 且相互独立，则：

$$X_1 = \sqrt{-2log(U_1)}cos(2\pi U_2)$$
$$X_2 = \sqrt{-2log(U_1)}sin(2\pi U_2)$$

相互独立且 $\sim N(0,1)$ 。

3.1 练习

使用 `system.time` 比较三种生成正态随机数的方式：

- 中心极限定理
- Box-Muller 方法
- 使用 `rnorm` 生成

想一下哪一种更有效率？

```
set.seed(20)
nsim=10000
u1=runif(nsim)
u2=runif(nsim)
X1=sqrt(-2*log(u1))*cos(2*pi*u2)
X2=sqrt(-2*log(u1))*sin(2*pi*u2)
U=array(0,dim=c(nsim,1))
for(i in 1:nsim)U[i]=sum(runif(12,-.5,.5))
par(mfrow=c(1,2))
hist(X1)
hist(U)
```

```
hist(rnorm(nsim))
```

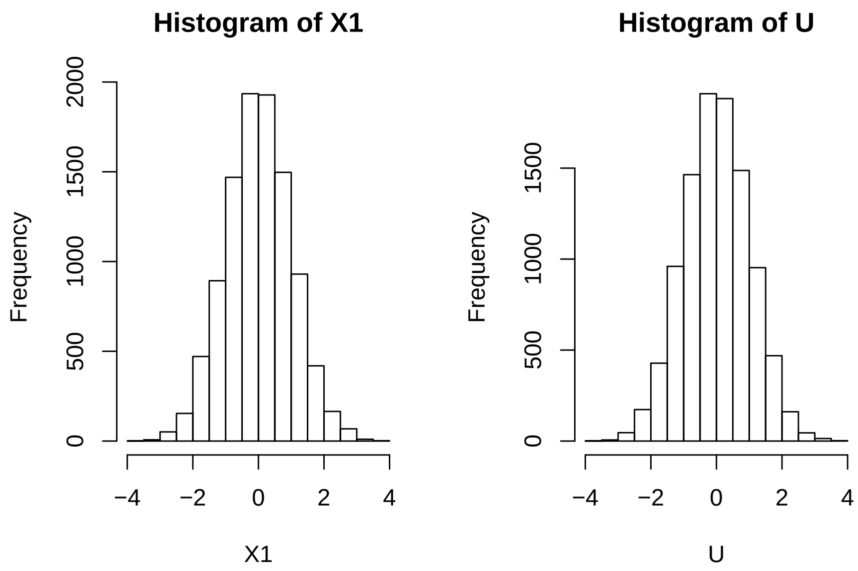
Histogram of X1
Frequency
0
500
1000
1500
2000
−4
−2
0
2
4
X1
Histogram of U
Frequency
0
500
1000
1500
−4
−2
0
2
4
U

图 3:

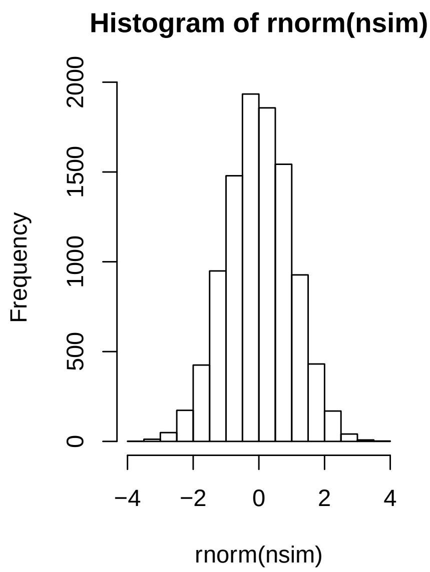
Histogram of rnorm(nsim)
Frequency
0
500
1000
1500
2000
−4
−2
0
2
4
rnorm(nsim)

图 4:

4 筛法

假设我们已经有了一种生成密度函数为 $g(x)$ 的随机变量的方法，现在利用这种方法生成密度函数为 $f(x)$ 的随机变量。

4.1 筛法原理

我们先生成来自 $g(x)$ 的随机变量 Y，然后以正比于 $\frac{F(Y)}{g(Y)}$ 的概率接收此值。

设：

$$\frac{f(y)}{g(y)} \leq M, \quad \forall y$$

4.2 筛法步骤

Algorithm 1 Accept–Reject Method

1. Generate $Y \sim g$, $U \sim \mathcal{U}_{[0,1]}$;
2. Accept $X = Y$ if $U \leq f(Y)/Mg(Y)$;
3. Return to 1 otherwise.

证明：

$$\begin{aligned}
P(Y \leq x | U \leq f(Y)/\{Mg(Y)\}) &= \frac{P(Y \leq x, U \leq f(Y)/\{Mg(Y)\})}{P(U \leq f(Y)/\{Mg(Y)\})} \\
&= \frac{\int_{-\infty}^{x} \int_{0}^{f(y)/\{Mg(y)\}} \mathrm{d}u \, g(y) \, \mathrm{d}y}{\int_{-\infty}^{\infty} \int_{0}^{f(y)/\{Mg(y)\}} \mathrm{d}u \, g(y) \, \mathrm{d}y} \\
&= \frac{\int_{-\infty}^{x} [f(y)/\{Mg(y)\}] \, g(y) \, \mathrm{d}y}{\int_{-\infty}^{\infty} [f(y)/\{Mg(y)\}] \, g(y) \, \mathrm{d}y} \\
&= \frac{\int_{-\infty}^{x} f(y) \, \mathrm{d}y}{\int_{-\infty}^{\infty} f(y) \, \mathrm{d}y} = P(X \leq x),
\end{aligned}$$

证毕。

4.3 例 2

使用参数为 1 的指数分布，利用筛法求下列密度函数（半对数分布）的随机数：

$$f(x)=\begin{cases}\frac{2}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) & \text{if } x\geq 0 \text{ and}\\ 0 & \text{otherwise.}\end{cases}$$

已知，参数为 λ 的指数分布的密度为：

$$g(x)=\begin{cases}\lambda\exp(-\lambda x) & \text{if } x\geq 0 \text{ and}\\ 0 & \text{otherwise.}\end{cases}$$

可得：

$$\frac{f(x)}{g(x)}=\frac{2}{\sqrt{2\pi}\lambda}\exp\left(-\frac{x^2}{2}+\lambda x\right).$$

对于所有的 $x\geq 0$：

$$c^*=\frac{2}{\sqrt{2\pi}\lambda}\exp\left(-\frac{\lambda^2}{2}+\lambda\cdot\lambda\right)=\sqrt{\frac{2}{\pi\lambda^2}}\exp\left(\lambda^2/2\right).$$

上述函数在 $\lambda=1$ 时达到最大：此时：

$$M=\sqrt{\frac{2}{\pi}}exp(\frac{1}{2})$$

设计 R 程序实现上述过程：

```
set.seed(20)

f <- function(x) {
  return((x>0) * 2 * dnorm(x,0,1))
}
g <- function(x) { return(dexp(x,1)) }
c <- sqrt(2 * exp(1) / pi)

rhalfnormal <- function(n) {
  res <- numeric(length=n)
  i <- 0
  while (i<n) {
```

```
    U <- runif(1, 0, 1)
    X <- rexp(1, 1)
    if (c * g(X) * U <= f(X)) {
      i <- i+1
      res[i] <- X;
    }
  }
  return(res)
}

X <- rhalfnormal(10000)
hist(X, breaks=50, prob=TRUE, ylim=c(0,1),
     main=NULL, col="gray80", border="gray20")
curve(f, min(X), max(X), n=500,
      ylim=c(0,1), ylab="f", add=TRUE)
```

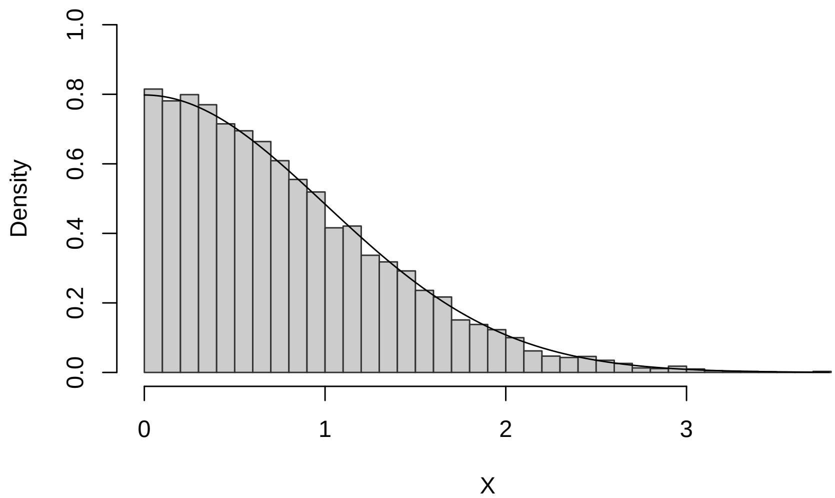

图 5:

5 混合分布

$$X = \begin{cases} X_1 & \alpha \\ X_2 & 1-\alpha \end{cases}$$

某个随机变量以 α 的概率来自随机变量 X_1，而以 $1-\alpha$ 的概率来自随机变量 X_2

算法：

1. 产生随机变量 X_1
2. 产生随机变量 X_2
3. 产生均匀分布随机数 U: 如果 $U \leq \alpha$，则令 $X = X_1$，否则令 $X = X_2$.