随机数的产生

许王莉

中国人民大学　统计学院

非均匀随机数的产生

(1). 产生非均匀随机数的一般方法

(2). 常用连续分布的抽样法

(3). 常用离散分布的抽样法

产生非均匀随机数的一般方法

- 直接抽样法(反函数法)
- 变换抽样法
- 值序抽样法
- 舍选抽样法
- 复合抽样法(合成法)
- 近似抽样法

定义

随机数定义: 设随机变量$\eta \sim F(x)$, 则称随机变量η 的随机抽样序列$\{\eta_i\}$为分布$F(x)$ 的随机数;

定理

定理1：设$F(x)$是连续且严格单调上升的分布函数，它的反函数存在，且记为$F^{-1}(x)$，即$F(F^{-1}(x))=x$.

(1):若随机变量ξ的分布函数为$F(x)$, 则$F(\xi)\sim U(0,1)$;

(2):若随机变量$R\sim U(0,1)$, 则$F^{-1}(R)$的分布函数为$F(x)$.

(1)证明：设随机变量$F(\xi)$的分布函数为$F_1(u)$,

当$U \in [0,1]$,

$$F_1(u) = P(F(\xi) \leq u) = P(\xi \leq F^{-1}(u)) = F(F^{-1}(u)) = u;$$

当$u < 0$,

$$F_1(u) = 0; u > 1, F_1(u) = 1;$$

(2)证明：设随机变量$F^{-1}(R)$的分布函数为$F_2(x)$, 则

$$F_2(x) = P(F^{-1}(R) \leq x) = P(R \leq F(x)) = F_R(F(x)) = F(x);$$

因为$R \sim U(0,1)$, 对任意$F(x) \in [0,1]$，有

$$F_R(F(x)) = F(x)$$

.

推论

已知$\xi \sim G(x)$, 设$F(x)$是一个分布函数，且反函数存在，则$\eta = F^{-1}(G(\xi)) \sim F(x)$.

证明：根据定理1(1), 如果$\xi \sim G(x)$, 则

$$R = G(\xi) \sim U(0,1);$$

根据定理1(2),

$$\eta = F^{-1}(R) = F^{-1}(G(\xi)) \sim F(x);$$

直接抽样法(反函数法)

连续分布的直接抽样法

根据以前的定理知，如果随机变量$R \sim U(0,1)$，$F^{-1}(\cdot)$是分布函数$F(x)$的反函数，则

$$\xi = F^{-1}(R) \sim F(x), \tag{1}$$

如果分布$F(x)$有概率密度函数$f(x)$, 根据定理知

$$R = F(\xi) = \int_{-\infty}^{\xi} f(x)dx. \tag{2}$$

利用上述公式，由均匀随机数$\{r_i\}$直接产生$F(x)$分布随机数的方法称为直接抽样法或发你函数法，上述公式称为直接抽样公式。

例

例

产生$[a,b]$上均匀分布的随机数。

解　已知$[a,b]$上均匀分布的随机变量ξ的密度函数为

$$f(x)=\begin{cases}\frac{1}{b-a}, & x\in[a,b];\\ 0, & \text{其他}.\end{cases}$$

根据公式(2)，$R=\int_a^{\xi}\frac{1}{b-a}dx=\frac{\xi-a}{b-a}$。即得抽样公式

$$\xi=(b-a)R+a.$$

为了得到$U(a,b)$随机数，先产生均匀随机数r_i，

令$\xi_i=(b-a)r_i+a$为$[a,b]$区间的均匀分布随机数。

直接抽样法(反函数法)

例

产生密度函数为$f(x)$的随机数，其中

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

解：由(2)式得

$$R=\int_0^{\xi}\lambda e^{-\lambda x}dx=1-e^{-\lambda\xi},$$

由此解出$\xi=-\frac{1}{\lambda}\ln(1-R)$，由于$R$与$1-R$均为$[0,1]$上均匀分布随机变量，抽样公式常取为

$$\xi=-\frac{1}{\lambda}\ln(R).$$

直接抽样法(反函数法)

离散分布的直接抽样法

设ξ的概率分布为$P(\xi = x_i) = p_i (i = 1, 2, \cdots)$，不妨设$x_1 < x_2 < x_3 < \cdots$，$\xi$的分布函数为

$$F(x) = P(\xi \le x) = \sum_{x_i \le x} p_i.$$

$F(x)$仅在至多可列个点$x_1, x_2, \cdots$上有正的跳跃值。产生离散分布$F(x)$随机数的直接抽样法如下：

(1). 产生$R \sim U(0, 1)$;

(2). 取$\xi = \begin{cases} x_i, & 若 F(x_{i-1}) < R \le F(x_i)(i = 2, 3, \cdots); \\ x_1, & R \le F(x_1). \end{cases}$

则$\xi \sim F(x)$。

证明：显然ξ的可能取值为$x_1, x_2, \cdots$，且

$$P(\xi = x_1) = P(R \le F(x_1)) = F(x_1) = p1;$$

$$P(\xi = x_i) = P\{F(x_{i-1}) < R \le F(x_i)\} = F(x_i) - F(x_{i-1}) = p_i \quad (i = 2, 3$$

所以$\xi \sim F(x)$.

直接抽样法(反函数法)

例

已知离散随机变量ξ的概率分布表为

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

试用框图描述用直接抽样方法产生ξ随机数的过程。

直接抽样法(反函数法)

例

已知离散随机变量ξ的概率分布为

$$P(\xi = k) = \frac{1}{n} \quad (k = 1, 2, \cdots, n),$$

试用直接抽样方法产生ξ随机数（也称为离散均匀分布）。

直接抽样法(反函数法)

解：随机变量ξ的分布函数为

$$F(k) = P(\xi \le k) = \frac{k}{n},$$

由离散分布的直接抽样法，

$$F(k-1) < R \le F(k) \Leftrightarrow \frac{k-1}{n} < R \le \frac{k}{n} \Leftrightarrow k-1 < nR \le k.$$

故产生ξ随机数的直接抽样方法为：

(1). 产生$r \sim U(0,1)$;

(2). 令$\xi = [nr] + 1$ （$[\cdot]$表示取整数部分）;

(3). 重复(1)-(2)，产生的$\{\xi_i\}$即为ξ随机数序列。

直接抽样法(反函数法)

直接抽样法是一个常用的方法，对连续型分布或离散分布都有效。对于连续随机变量，应用直接抽样法，首先必须求得其分布函数的反函数$F^{-1}(x)$，有些分布函数的反函数不能用初等函数表出，如正态分布和Gamma分布等，抽样公式不能精确表出。

直接抽样法是一种特殊的变换抽样

一维变换抽样公式

定理

设随机变量X具有密度函数$f(x)$，$Y=g(X)$是随机变量X的函数，又设$x=g^{-1}(y)=h(y)$存在且有一阶连续倒数。则$Y=g(X)$的密度函数为

$$p(y)=f[h(y)]\cdot|h'(y)|.$$

例

用变换抽样法求$[a,b]$区间上的均匀随机数。

由概率论知，若$R \sim U(0,1)$，则$(b-a)R+a \sim U(a,b)$。变换抽样公式

$$\xi = g(R) = (b-a)R + a. \tag{3}$$

只需产生均匀随机数，带入变换公式(3)就可得$[a,b]$区间上的均匀随机数。

二维变换抽样公式

定理

设随机向量(X,Y)具有二维联合密度$f(x,y)$，令

$$\begin{cases} u = g_1(x,y), \\ v = g_2(x,y). \end{cases}$$

设g_1, g_2的反变换存在唯一，记为$\begin{cases} x = h_1(u,v), \\ y = h_2(u,v). \end{cases}$，并设$h_1, h_2$的一阶偏导数存在；函数变换的Jacobi行列式

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial vu} \end{vmatrix} \neq 0.$$

则随机变量U,V的二维联合密度为

例

用二维变换抽样法产生标准正态随机数。

解：设r_1, r_2为相互独立的均匀分布随机数，令

$$\begin{cases} u = \sqrt{-2\ln r_1}\cos 2\pi r_2, \\ v = \sqrt{-2\ln r_1}\sin 2\pi r_2, \end{cases} \tag{4}$$

则u, v相互独立且为$N(0,1)$随机数。

事实上，由公式(4)，可解出

$$\begin{cases} r_1 = \exp\{-\frac{1}{2}(u^2+v^2)\}, \\ r_2 = \frac{1}{2\pi}\arctan\frac{v}{u}, \end{cases}$$

变换Jacobi行列式为

$$J = \begin{vmatrix} \frac{\partial r_1}{\partial u} & \frac{\partial r_1}{\partial v} \\ \frac{\partial r_2}{\partial u} & \frac{\partial r_2}{\partial vu} \end{vmatrix} = \begin{vmatrix} -ur_1 & -vr_1 \\ -\frac{1}{2\pi}\frac{v}{u^2+v^2} & \frac{1}{2\pi}\frac{u}{u^2+v^2} \end{vmatrix}$$

$$= -\frac{r_1}{2\pi}\frac{u^2}{u^2+v^2} - \frac{r_1}{2\pi}\frac{v^2}{u^2+v^2}$$

$$= -\frac{1}{2\pi}\exp\{-\frac{1}{2}(u^2+v^2)\}.$$

由定理，(4)式确定的随机变量U, V的二维联合密度为

$$p(u,v) = f[h_1(u,v), h_2(u,v)]|J| = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}.$$

显然U, V均服从$N(0,1)$分布，且相互独立。

n维变换抽样公式

设$(X_1, X_2, \cdots, X_n)'$为n维随机向量，令$\xi = g(X_1, X_2, \cdots, X_n)$，若$\xi \sim F(x)$，则通过$n$维变换公式

$$\xi = g(X_1, X_2, \cdots, X_n)$$

给出由$X_1, X_2, \cdots, X_n$的随机数产生分布为$F(x)$的随机数，常见的n维变换函数有：

$$\xi_1 = X_1 + X_2 + \cdots + X_n,$$

$$\xi_2 = X_1^2 + X_2^2 + \cdots + X_n^2,$$

$$\xi_3 = \max(X_1, X_2, \cdots, X_n),$$

$$\xi_4 = \min(X_1, X_2, \cdots, X_n), \cdots.$$

例

用n维变换抽样方法产生χ^2分布随机数。

解　可以证明，若$X_1, X_2, \cdots, X_n$独立同$N(0,1)$分布，令$\xi = \sum_{i=1}^n X_i^2$，则$\xi \sim \chi^2(n)$。利用n维变换公式

$$\xi \sim \sum_{i=1}^{n} X_i^2$$

产生n个标准正态随机数后即可得到$\chi^2(n)$分布随机数。

值序抽样法

将样本$X_1, X_2, \cdots, X_n$按取值由小到大进行重排得到值序统计量(或称次序统计量)，记为$X_{(1)}, X_{(2)}, \cdots, X_{(n)}$。所谓值序抽样法就是利用值序统计量产生随机数的方法。

定理

设$X_1, X_2, \cdots, X_n$独立同分布函数$F(x)$ (密度函数为$f(x)$)，次序统计量为$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$，则$(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$的联合密度函数为

$$g(y_1, \cdots, y_n) = \begin{cases} n!f(y_1)\cdots f(y_n), & -\infty < y_1 < \cdots < y_n < \infty; \\ 0, & \text{其他}. \end{cases}$$

而$X_{(l)}$的密度函数和分布函数分别为

$$f_{nl}(x) = \frac{n!}{(l-1)!(n-l)!}[F(x)]^{l-1}[1-F(x)]^{(n-l)}f(x)$$

$$F_{nl}(x) = \frac{n!}{(l-1)!(n-l)!}\int_0^{F(x)} t^{l-1}(1-t)^{n-l}dt$$

值序抽样法

特别当$X_1, X_2, \cdots, X_n$独立同$U(0,1)$分布时，值序统计量$X_{(l)}$的密度函数为

$$f_{nl}(x) = \begin{cases} \frac{n!}{(l-1)!(n-l)!} x^{l-1}(1-x)^{(n-l)}, & 0 < x < 1; \\ 0, & \text{其他}. \end{cases}$$

设$R_1, R_2, \cdots, R_n$独立同$U(0,1)$分布，根据上式有

$$R_{(n)} = \max(R_1, \cdots, R_n) \sim \beta(n, 1),$$
$$R_{(1)} = \min(R_1, \cdots, R_n) \sim \beta(1, n),$$

一般地$R_{(l)} \sim \beta(l, n+1-l)(l = 1, 2, \cdots, n-1, n)$。适当选择$n, l$，就可以得到整参数的Beta分布随机数。

由值序抽样法定理，利用均匀分布的值序统计量可产生Beta分布随机数。

例

试用值序抽样法产生*Beta*分布随机数。

解　Beta分布的密度函数为

$$f(x)=\begin{cases}\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}, & 0<x<1;\\ 0, & \text{其他}.\end{cases}$$

我们有$E(X)=\frac{a}{a+b}$，$Var(X)=\frac{ab}{(a+b)^2(a+b+1)}$。

例如为产生$\beta(5,10)$随机数，先产生$n=14$个均匀随机数$r_1,r_2,\cdots,r_{14}$，按从小到大的次序重新排列，则$r_{(5)}$为$\beta(5,10)$随机数。

一般地产生$\beta(a,b)$随机数的步骤（a,b为整数）如下

(1). 产生$n=a+b-1$个均匀随机数$r_1,r_2,\cdots,r_n$;

(2). 排序$r_{(1)}\le r_{(2)}\le\cdots\le r_{(n)}$;

(3). 令$\xi=r_{(a)}$，并输出服从$\beta(a,b)$分布的随机数ξ。

舍选抽样法

上述几种抽样法，其实都可统称为"直接法"，因它们都是由已知分布的随机数(如均匀随机数)按某个变换公式直接得到$F(x)$（或密度$f(x)$）随机数的抽样方法。这里介绍的舍选抽样法是一种"非直接法"，它对已知的随机数，先通过某个检验条件决定取舍，才能得到$F(x)$随机数。当以上几种方法不适用或效率不高时，它是很有用的。

定理

设$f(x), g(x)$为分布密度函数，$h(\cdot)$为给定函数（不必是密度函数）。如果按下述方法进行舍选抽样；

(1). 生成$X \sim f(x)$；

(2). 生成$Y \sim g(x)$，且X与Y独立；

(3). 如果随机数X, Y满足：$Y \le h(X)$，令$Z = X$；否则转到*(1)*，则Z的分布密度为

$$p(z) = \frac{f(z)G(h(z))}{\int_{-\infty}^{+\infty} f(y)G(h(y))dy},$$

其中$G(y) = \int_{-\infty}^{y} g(x)dx$。

证明：由抽样过程知

$$
\begin{aligned}
P(Z \leq z) &= P(X \leq z | Y \leq h(X)) \\
&= \frac{P(X \leq z, Y \leq h(X))}{P(Y \leq h(X))} = \frac{\int_{-\infty}^{z} \int_{-\infty}^{h(z)} f(x)g(y)dydx}{\int_{-\infty}^{+\infty} \int_{-\infty}^{h(z)} f(x)g(y)dydx} \\
&= \frac{\int_{-\infty}^{z} f(x)G(h(x))dx}{\int_{-\infty}^{+\infty} f(x)G(h(x))dx} \\
&= \int_{-\infty}^{z} \frac{f(x)G(h(x))}{\int_{-\infty}^{+\infty} f(x)G(h(x))dx} dx,
\end{aligned}
$$

故Z的密度函数为$p(z) = \frac{f(x)G(h(x))}{\int_{-\infty}^{+\infty} f(x)G(h(x))dx}$。

由上述定理易见，舍选法产生的随机数，其密度函数的形式为$Cf(z)G(h(z))$。一般地设$(X, Y) \sim g(x, y)$，则舍选法产生的随机数，其密度函数的形式为$C\int_{-\infty}^{h(z)} g(z, y)dy$.

设随机变量Z的分布密度$p(z)$有上界函数$M(x)$，

$$p(z) \leq M(z) \quad (\text{对一切z}),$$

且$C = \int_{-\infty}^{\infty} M(x)dx < \infty$，令$f(z) = M(x)/c$；取$g(y)$为均匀分布密度。则用舍选抽样方法产生$Z$随机数的抽样过程为

(1). 生成$X \sim f(x)$;

(2). 生成$R \sim g(y)$（即产生$U(0,1)$随机数R），且X与R独立；

(3). 如果$R \leq p(X)/M(x)$，令$Z = X$；否则转到(1)，则Z的分布密度为$Z \sim p(z)$。

证明：由定理，取$h(x)=p(x)/M(x)$，它满足$0\le h(x)\le 1$，故均匀分布的分布函数$G(h(x))=h(x)$。以上步骤产生的随机数Z的密度函数为

$$\frac{f(z)G(h(z))}{\int_{-\infty}^{+\infty} f(x)G(h(x))dx}=\frac{\frac{M(z)}{C}\frac{p(z)}{M(z)}}{\int_{-\infty}^{+\infty}\frac{p(x)}{C}dx}=p(z).$$

舍选抽样法I

上述算法连续在(1)-(3)步中循环，直到第(1)、(2)步产生的一对数(X,R)满足$R \le p(X)/M(x)$，从而得$Z = X$为止。上界函数$M(\cdot)$的选取除满足$p(x) \le M(x)$（对一切x）外，希望由此得到密度$f(x) = M(x)/C$的随机数容易生成；故对有限区间$[a,b]$上的密度函数$p(x)$，常取$M(x)$恒为常数，$f(x)$就是$[a,b]$上均匀分布的密度函数。

例

试产生$\beta(a,b)$的随机数ξ.

解： $\beta(a,b)$的密度函数为

$$p(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{(b-1)} \quad (0 < x < 1).$$

当$x = \frac{a-1}{a+b-2}$时，$p(x)$取最大值D

$$D = \frac{1}{B(a,b)} \left(\frac{a-1}{a+b-2}\right)^{a-1} \left(\frac{b-1}{a+b-2}\right)^{b-1}.$$

由舍选抽样I的抽样过程如下：

(1). 生成$X \sim U(0,1)$;

(2). 生成$R \sim U(0,1)$，且与X独立;

(3). 如果$R \leq p(x)/D = h(X)$时，令$Z = X$，并输出密度为$p(x)$的随机数Z；否则转到(1)重新抽样。

例

试产生密度函数$p(x)$的随机数ξ，其中ξ的取值在有限区间$[a,b]$上，且$\sup_{x\in[a,b]} p(x) = f_0 < \infty$。

解：　取$M(x) = \begin{cases} f_0, & x \in [a,b]; \\ 0, & \text{其他}. \end{cases}$ 则

$$\begin{aligned} f(x) &= \frac{M(x)}{\int_a^b f_0 dx} = \begin{cases} \frac{1}{b-a}, & x \in [a,b]; \\ 0, & \text{其他}. \end{cases} \\ h(x) &= p(x)/M(x). \end{aligned}$$

抽样过程为:

(1). 产生$[a, b]$上均匀随机数X;

(2). 产生均匀随机数R，且与X独立;

(3). 若$R \le p(X)/f_0$，令$Z = X$；否则重新产生均匀随机数（转(1)），直至一对随机数(X, R)满足$R \le p(X)/f_0$时为止。

例

设$Z \sim p(z), z \in (0,2)$，已知

$$0.3 \le p(z) \le \frac{z+1}{2}, \quad z \in (0,2)$$

试用舍选抽样方法I产生密度为$p(z)$的随机数Z.

解：取上界函数$M(x) = \frac{x+1}{2}$, 则$\int_0^2 M(x)dx = 2$。

令$f(x) = \frac{M(x)}{2} = \frac{x+1}{4}$，则改进的舍选抽样I的抽样过程如下：

(1). 生成$R_1 \sim U(0,1)$，则$X = \sqrt{1+8R_1} - 1 \sim f(x)$;

(2). 生成$R_2 \sim U(0,1)$，且与R_1独立；

(3). 如果$R_2 \le 0.3/M(X)$时，令$Z = X$，否则转到(4);

(4). 如果$R_2 \le p(X)/M(X) = h(X)$时，令$Z = X$，否则转到(1)。

则$Z \sim p(x)$。

在一般的舍选抽样I中判断不等式$R_2 \leq p(X)/M(X)$时，必须计算$p(X)$和$M(X)$，通常$p(X)$比较复杂，为减少计算量，当$p(x)$存在某个简单的下界函数$m(x)$时，可以采用类似上面的例子来改进舍选抽样方法。

近似线性密度的抽样法

已知密度函数$g(x)$在$(s, s+h)$范围取值，且近似线性，即可以用两条平行线作为$g(x)$的界

$$a - \frac{b(x-s)}{h} \leq g(x) \leq b - \frac{b(x-s)}{h},$$

产生$g(x)$随机数的算法如下：

(1). 产生$R_1, R_2 \sim U(0,1)$；令$U = \min(R_1, R_2)$，

$V = \max(R_1, R_2)$；

(2). 如果$V \leq a/b$，转(4)；

(3). 如果$V > U + \frac{1}{b} g(s = hU)$，转(1)；

(4). 令$X = s + hU$，则X为所求密度函数$g(x)$的随机数。

设随机变量Z的密度函数$p(z)$可表示为

$$p(z) = Lh(z)f(z),$$

其中$L > 1$, $0 \le h(z) \le 1$, $f(z)$为随机变量X的密度，则Z的抽样过程可表示为

(1). 产生$X \sim f(x)$;

(2). 产生$R \sim U(0,1)$，且与X独立;

(3). 如果(X,R)满足条件$R \le h(X)$，令$Z = X$；否则转到(1)，则$Z \sim p(x)$。

证明：根据上述定理，Z的密度为

$$p(z) = \frac{f(z)G(h(z))}{\int_{-\infty}^{+\infty} f(x)G(h(x))dx} = \frac{f(z)h(z)}{\int_{-\infty}^{+\infty} f(x)h(x)dx} = Lf(x)h(x),$$

其中$L = (\int_{-\infty}^{+\infty} f(x)h(x)dx)^{-1}$。

例

试用舍选法产生半正态分布的随机数Z，Z的密度函数为

$$p(z)=\begin{cases}\sqrt{2/\pi}e^{-\frac{x^2}{2}}, & x\geq 0;\\ 0, & x<0.\end{cases}$$

解：因$p(x)=\sqrt{\frac{2}{\pi}}e^{-\frac{x^2}{2}}=\sqrt{\frac{2}{\pi}}e^{-\frac{(x-1)^2}{2}}e^{-x}=Lh(x)f(x)$, 其中$f(x)=e^{-x}(x>0)$为指数分布的密度函数，$h(x)=e^{-\frac{(x-1)^2}{2}}$的值域$\in[0,1]$，$L=\sqrt{2/\pi}>1$，则舍选抽样过程可用下图描述。注意：检验条件

$$R\leq h(X)=e^{-\frac{(X-1)^2}{2}}\Leftrightarrow -\ln R\geq\frac{(X-1)^2}{2}\Leftrightarrow\frac{(X-1)^2}{2}\leq Y\quad (Y=-\ln R)$$

舍选抽样法II

产生图??

舍选抽样法III

设随机变量Z的密度函数$p(z)$可表为

$$p(x) = L\int_{-\infty}^{h(z)} g(z,y)dy,$$

其中$g(x,y)$为随机向量(X,Y)的联合密度函数；$h(x)$在Y的定义域上取值；L为规格化常量，则随机变量Z的抽样过程可用如下所示：

(1). 产生随机向量$(X,Y) \sim g(x,y)$

(2). 如果$Y \le h(X)$，令$Z = X$；输出Z，否则转(1)。

证明：Z的分布函数

$$
\begin{aligned}
P(Z \leq z) &= P(X \leq z | Y \leq h(X)) \\
&= \frac{P(X \leq z, Y \leq h(X))}{P(Y \leq h(X))} \\
&= \frac{\int_{-\infty}^{z} \int_{-\infty}^{h(z)} g(x,y)dydx}{\int_{-\infty}^{+\infty} \int_{-\infty}^{h(z)} g(x,y)dydx} \\
&= \int_{-\infty}^{z} [L \int_{-\infty}^{h(z)} g(x,y)dy]dx.
\end{aligned}
$$

故Z的密度函数为

$$p(z) = L \int_{-\infty}^{h(z)} g(z,y)dy, \text{其中，} L = (\int_{-\infty}^{+\infty} \int_{-\infty}^{h(z)} g(x,y)dydx)^{-1}.$$

例

试用舍选抽样法产生密度为$p(x)$的随机数，其中

$$p(x) = \begin{cases} \frac{1}{\pi\sqrt{1-x^2}}, & |x| < 1; \\ 0, & \text{其他}. \end{cases}$$

解：根据表3-3知，若$R \sim U(0,1)$，则$\xi = \sin 2\pi R$（或$\xi = \cos 2\pi R \sim p(x)$），可以用变换抽样法产生$p(x)$随机数，但变换抽样法必须计算三角函数值，运算速度慢。以下舍选法，可以得到效率更高的抽样法。

设$R_1, R_2 \sim U(0,1)$且独立，令

$$\begin{cases} X = \frac{R_1^2 - R_2^2}{R_1^2 + R_2^2}, \\ Y = R_1^2 + R_2^2, \end{cases}$$

则(X, Y)的联合密度函数为

$$g(x,y) = \begin{cases} \frac{1}{4\sqrt{1-x^2}}, & |x| < 1, 0 < y < \frac{1}{1+|x|}, \\ 0, & \text{其他}. \end{cases}$$

这时$p(x) = L\int_{-\infty}^{1} g(x,y)dy = \frac{L}{4\sqrt{1-x^2}}(|x| < 1)$，其中$L = 4/\pi$，根据舍选抽样II，可得$p(x)$的抽样过程如下

舍选抽样法III

在此例中，$h(x) = 1$，当$R_1, R_2 \sim U(0,1)$时，
有$U = R_1 \sim U(0,1)$，$V = 2R_2 - 1 \sim U(-1,1)$，

(1). 产生均匀随机数R_1, R_2；

(2). 令$U = R_1, V = 2R_2 - 1$

(1). 如果$U^2 + V^2 \leq 1$，计算$Z = U^2 - V^2$，$Y = U^2 + V^2$。
令$X = Z/Y$，输出X。否则转到(1)。

复合抽样法(合成法)

复合抽样法是1961年Marsaglia提出的，当我们要抽取的分布函数$F(x)$可以表成几个分布函数$F_1(x), F_2(x), \cdots$的线性组合，且$F_j(x)$随机数容易得到时，我们不是直接产生$F(x)$随机数，而是采用复合抽样法由$F_j(x)$的随机数来产生$F(x)$随机数。假设对所有x，随机变量ξ的分布函数$F(x)$ 可写成

$$F(x) = \sum_j p_j F_j(x). \tag{5}$$

复合抽样法(合成法)

若ξ是连续型随机变量，密度函数$f(x)$可写为

$$f(x)=\sum_j p_j f_j(x), \tag{6}$$

其中$p_j\geq 0, \sum_j p_j=1$，每个$F_j(x)$（或$f_j(x)$）都是分布函数（或密度函数），公式(5)和(6) 就是复合抽样公式，它们给出的抽样方法如下：

(1). 产生一个正的随机整数J，使得：

$$P\{J=j\}=p_j, \quad (j=1,2,\cdots).$$

(2). 产生分布为$F_J(x)$（或$f_J(x)$）的随机数，即为ξ的随机数。

复合抽样法(合成法)

重复(1)-(2)步骤，即可产生ξ的随机数序列，可以证明由(1)-(2)两步得到的随机数$\xi \sim F(x)$。事实上，

$$
\begin{aligned}
P\{\xi \leq x\} &= P\{(\xi \leq x) \cap \sum_j (J = j)\} \\
&= \sum_j P\{(\xi \leq x) | J =)\} P\{J = j\} \\
&= \sum_j F_j(x) p_j = F(x).
\end{aligned}
$$

复合抽样法(合成法)

例

设$0 < a < 1$时梯形分布的密度函数为

$$f(x) = \begin{cases} a + 2(1-a)x, & x \in [0,1] \\ 0, & \text{其他}. \end{cases}$$

试用复合抽样法产生密度为$f(x)$的随机数。

解：首先把$f(x)$下的面积分成S_1和S_2两部分，长方形的面积$S_1 = a$，三角形的面积$S_2 = 1 - a(S_1 + S_2 = 1)$。

设$f_1(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & \text{其他}. \end{cases}$ 为$[0,1]$区间上均匀分布的密度函数；

$f_2(x) = \begin{cases} 2x, & x \in [0,1] \\ 0, & \text{其他}. \end{cases}$ 为三角形分布密度（或$\beta(2,1)$分布）。

复合抽样法(合成法)

则复合抽样公式为

$$f(x) = af_1(x) + (1-a)f_2(x).$$

三角形分布的随机数可用变换抽样法（或值序抽样法）产生。如U, V独立且$\sim U(0,1)$，则$\max(U,V) \sim f_3(x)$。图?? 的是复合抽样法的框图。

以上介绍的几类抽样法从理论上讲都是精确的，即除了用伪随机数代替随机数形成的误差外，不含系统误差。当分布$F(x)$比较复杂，以上介绍的方法难以实现外，还可以用近似方法。这里介绍的近似抽样法除含有上述误差外，选取的方法中还含有系统性误差。当然在实际应用中，这些误差与模拟计算的误差相比较可忽略不计。

近似抽样法(利用中心极限定理)

(1) 利用中心极限定理

例

试用近似抽样法产生$N(0,1)$随机数。

解：根据中心极限定理，产生均匀随机数$r_1, r_2, \cdots, r_n$，记$\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r_i$，则$E(\bar{r}) = \frac{1}{2}$，$Var(\bar{r}) = \frac{1}{2n}$，$N(0,1)$随机数的近似抽样公式为

$$u = \sqrt{12n}(\bar{r} - \frac{1}{2}) \tag{7}$$

近似抽样法(利用中心极限定理)

在实际应用中，常取$n = 6$或$n = 12$。特别在$n = 12$时，

$$u = \sum_{i=1}^{12} r_i - 6.$$

利用以上公式，产生12个均匀随机数，即可得到一个标准正态随机数。若注意到r_i和$1 - r_i$同为均匀随机数，则产生$N(0,1)$随机数的近似公式可化为

$$u = \sum_{i=1}^{6} (r_i - r_{2i-1}).$$

Butler抽样法(1970)是复合抽样法和近似抽样法的综合。Butler抽样法的思想是在密度函数$f(x)$的分解公式中让权系数$p_j = \frac{1}{m}(j = 1, 2, \cdots, m)$；而$f_j(x)$用最简单的线性函数近似。

近似抽样法(Butler抽样法)

做法如下（设总体分布函数为$F(x)$）；

(1). 确定分点$x_j(j=0,1,\cdots,m)$，使

$$\int_{x_{j-1}}^{x_j} f(x)dx = \frac{1}{m} \quad (j=1,2,\cdots,m).$$

(2). 对密度函数$f(x)$进行分解

$$f(x) = \sum_{i=1}^{m} p_i f_i(x),$$

其中$p_i = \frac{1}{m}(i=1,2,\cdots,m)$,

$$f_i(x) = \begin{cases} mf(x), & x \in (x_{i-1}, x_i] \\ 0, & \text{其他}. \end{cases} \quad (i=1,2,\cdots,m).$$

(3). 在小区间$(x_{i-1}, x_i]$上用直线作为曲线$f_i(x)$的近似，然后对线性函数构成的分布函数f_i^*下的面积分解为两部分（见图3-13），按符合抽样法可得到

$$f_i(x) \approx \sum_{i=1}^{m} d_i f_{i1}(x) + (1 - d_i) f_{i2}(x),$$

其中，$d_i = \frac{\text{三角形面积}}{\text{梯形面积}} = \frac{|f(x_i) - f(x_{i-1})|}{f(x_i) + f(x_{i-1})}$,

$$f_{i1} = \begin{cases} \frac{1}{(x_i - x_{i-1})^2}(x - x_{i-1}), if \quad f(x_i) > f(x_{i-1}), \\ \frac{-2}{(x_i - x_{i-1})^2}(x - x_i), if \quad f(x_i) < f(x_{i-1}) \end{cases}$$

$$f_{i2} = \frac{1}{(x_i - x_{i-1})}, \quad x \in (x_{i-1}, x_i].$$

近似抽样法(Butler抽样法)

综合之，$f(x)$有近似复合抽样公式：

$$f(x) \approx \sum_{i=1}^{m} \frac{1}{m}[d_i f_{i1}(x) + (1 - d_i) f_{i2}(x)].$$

近似抽样法(Butler抽样法)

Butler抽样法的具体步骤如下：

首先把$(-\infty, +\infty)$用分点$-\infty = x_0, x_1, \cdots, x_n = +\infty$分成$m$个小区间，使

$$\int_{x_{i-1}}^{x_i} f(x)dx = \frac{1}{m} \quad (i = 1, 2, \cdots, m).$$

在实际应用中，取$x_0 = a$，使$F(a) \approx 0$；取$x_m = b$，使$F(b) \approx 1$。

(1). 产生$r \sim U(0,1)$，令$i = [mr + 1]$;

(2). 产生$r-1, r_2 \sim U(0,1)$，当$r_i \le d_i$时，令

$$Z = \begin{cases} x_{i-1} + (x_i - x_{i-1})\sqrt{r_i}, & \text{if} \quad f(x_i) > f(x_{i-1}) \\ x_i - (x_i - x_{i-1})\sqrt{1 - r_i}, & \text{if} \quad f(x_i) < f(x_{i-1}) \end{cases}$$

否则，令$Z = x_{i-1} + (x_i - x_{i-1})r_i$，则由此得到的$Z$为近似服从分布$f(x)$的随机数。

设 $F(x)$ 为一般分布的分布函数,首先把 $(-\infty,+\infty)$ 用分点 $-\infty=a_0,a_1,a_2,\cdots,a_{m-1},a_m=\infty$ 分为 m 个小区间,在第 k 段 $(a_{k-1},a_k]$ 上由 $F(x)$ 定义分布函数 $F_k(x)$,即令

$$F_k(x)=\begin{cases}\frac{F(x)-F(a_{k-1})}{p_k}, & x\in(a_{k-1},a_k],\\ 0, & x\le a_{k-1},\\ 1, & x>a_k,\end{cases}$$

其中 $p_k=F(a_k)-F(a_{k-1})$.显然 $F_k(x)$ 是取值在 $(a_{k-1},a_k]$ 上的分布函数,而且 $F(x)$ 可分解为

$$F(x)=\sum_{i=1}^{m}p_iF_i(x) \tag{4.18}$$

近似抽样法(一般分布近似抽样)

在每一小段上用线性函数作为分布函数的近似,则

$$F_k(x) \approx \frac{x - a_{k-1}}{a_k - a_{k-1}}, \quad x \in (a_{k-1}, a_k].$$

故 $F(x)$ 的近似抽样法的具体步骤如下:

首先把 $(-\infty, +\infty)$ 用分

点 $-\infty = a_0, a_1, a_2, \cdots, a_{m-1}, a_m = \infty$ 分为 m 个小区间,记

$$p_j = F(a_j) - F(a_{j-1}) \quad (j = 1, 2, \cdots, m).$$

在实际应用中取 $a_0 = a$ 使 $F(a) \approx 0$;取 $a_m = b$ 使 $F(b) \approx 1$.

①用直接抽样法产生离散分布随机数J:

$$J \sim P\{J = k\} = p_k \quad (k = 1, 2, \cdots, m);$$

②产生 $r_J \sim U(0,1)$,令 $X = a_{J-1} + (a_J - a_{J-1})r_1$,则 X 为近似地服从分布函数 $F(x)$ 的随机数.

近似抽样法(经验分布)

以上介绍的抽样法都是产生给定分布随机数的抽样法.当对实际过程进行模拟计算时,调查得来的数据 $x_1, x_2, \cdots, x_n$ 来自哪类总体分布 $F(x)$ 是未知的.但由观测数据可求出经验分布函数 $F_n(x)$,且 $F_n(x) \approx F(x)$.这种直接由经验分布函数或观测数据出发,产生总体 $F(x)$ 随机数的方法,称为经验分布函数法.

近似抽样法(经验分布)

(I)已知原始观测数据

设已知观测数据 $x_1, x_2, \cdots, x_n$ 来自某总体,其分布函数为 $F(x)$.将 $x_1, x_2, \cdots, x_n$ 排序: $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$. n 个点将 $[x_{(1)}, x_{(n)}]$ 分为 $n-1$ 个小区间,假定数据落入每个小区间的概率为 $\frac{1}{n-1}$,且在每个小区间是均匀分布的.具体抽样法如下:

①产生 $r \sim U(0,1)$,记 $p = (n-1)r$,令 $I = [p] + 1$;

②令 $X = x_{(I)} + (p - [p])(x_{(I+1)} - x_{(I)})$,则 X 为近似地服从分布函数 $F(x)$ 的随机数.

(II)已知分组观测频数

设已知 n 个观测数据在 m 个连续的小区间 $[a_0,a_1),[a_1,a_2),\cdots,[a_{m-1},a_m]$ 内的观测频数为 $n_1,n_2,\cdots,n_m(n_1+n_2+\cdots+n_m=n)$.利用这些观测频数可以给出经验分布函数 $F_n(x)$,那么产生总体分布随机数的抽样法如下：

①产生 $r\sim U(0,1)$,若 $F_n(a_{k-1})<r\leq F_n(a_k)$,则取 $J=k$;

②令 $X=a_{k-1}+\frac{r-F_n(a_{k-1})}{F_n(a_k)-F_n(a_{k-1})}(a_k-a_{k-1})$, 则 X 为近似地服从总体分布 $F(x)$ 的随机数.

二常用连续分布的抽样法

1.正态分布

2.指数分布

3.Gamma分布

4.Beta分布

5.卡方n分布,F(m,n)分布与t(n)分布

6.其他连续分布

正态分布(一)

(1)基于中心极限定理的近似抽样法

(见例4.16)设 $r_1, \cdots, r_n$ 为均匀随机数,近似抽样公式为

$$U = \sum_{i=1}^{6}(r_{2i} - r_{2i-1}) \dot{\sim} N(0,1).$$

(2)Box和Muller(1958)提出的变换抽样法

(见例4.6)设 r_1, r_2 为均匀随机数，变换抽样公式为

$$\begin{cases} U_1 = \sqrt{-2\ln r_1}\cos 2\pi r_2, \\ U_2 = \sqrt{-2\ln r_1}\sin 2\pi r_2. \end{cases}$$

由两个独立的均匀随机数,利用变换公式可得两个独立的 $N(0,1)$ 随机数.此算法是较常用的抽样法,但计算量较大(因需调用标准函数计算 $\sin x, \cos x$).

(3)修正变换抽样法

变换抽样法因必须计算 $\sin 2\pi R$ 和 $\cos 2\pi R$,计算量大.修正变换抽样法利用舍选抽样法产生 $\sin 2\pi R$ 和 $\cos 2\pi R$ 随机数.

设 $R \sim U(0,1)$,令 $\begin{cases} \xi = \sin 2\pi R \triangleq \sin 2\alpha, \\ \eta = \cos 2\pi R \triangleq \cos 2\alpha. \end{cases}$ 显然 $\alpha = \pi R \sim U(0,\pi)$,对单位圆内的随机点 $P(X,Y)$,它与相应的 α 有以下关系(见图3-14):

$$\sin\alpha = \frac{Y}{\sqrt{X^2+Y^2}}, \quad \cos\alpha = \frac{X}{\sqrt{X^2+Y^2}}$$

故

$$\xi = sin2\alpha = 2sin\alpha cos\alpha$$
$$= \frac{2XY}{X^2+Y^2},$$
$$\eta = cos2\alpha = cos^2 - sin^2\alpha$$
$$= \frac{X^2-Y^2}{X^2+Y^2}$$

产生 ξ,η 随机数可以通过单位圆内的随机点得到.综合之,产生 $N(0,1)$ 随机数的修正变换抽样法的步骤为:

①产生相互独立的均匀随机数 $r_1.r_2, r_3$;

②计算 $u_1 = 2r_2 - 1, u_2 = r_3$,则 $u_1 \sim U(-1, 1)$;

③如果 $\omega = u_1^2 + u_2^2 > 1$,则转到① 重新抽样;否则令

$$\begin{cases} X_1 = \sqrt{-2lnr_1}\frac{u_1^2-u_2^2}{\omega}, \\ X_2 = \sqrt{-2lnr_1}\frac{2u_1u_2}{\omega}, \end{cases}$$

则 $(X_1, X_2)' \sim N_2(0, I_2)$ (即产生的 X_1 和 X_2 为相互独立的标准正态分布随机数).

(4)"极坐标"抽样法

这是Marsaglia(1962)给出的对变换抽样法的改进方法,它取消了三角函数的运算,效率可比变换抽样法提高.极坐标法的具体步骤如下:

①产生 $r_1, r_2 \sim U(0,1)$;令 $V_1 = 2r_1 - 1, V_2 = 2r_2 - 1$;

②如果 $W = V_1^2 + V_2^2 > 1$,转到①;否则令 $Y = \sqrt{\frac{-2lnW}{W}}$,则 $X_1 = V_1 Y, X_2 = V_2 Y$ 为相互独立的 $N(0,1)$ 随机数.

事实上,

因 $(V_1,V_2)\sim f(v_1,v_2)=\begin{cases}1/4, & -1\le v_1,v_2\le 1,\\ 0, & \text{其他},\end{cases}$ 且 $P\{W=V_1^2+V_2^2\le 1\}=\frac{\pi}{4}$.故 (X_1,X_2) 的联合分布函数为

$$
\begin{aligned}
&P\{X_1\le x_1, X_2\le x_2|V_1^2+V_2^2\le 1\}\\
&=P\{(X_1,X_2)\in D\}/\frac{\pi}{4}\ (D=\{(X_1,X_2)|X_1\le x_1,X_2\le x_2,W\le 1\})\\
&=\frac{4}{\pi}P\{(V_1,V_2)\in D^*\}\quad (D^*\text{是}D\text{在}OV_1V_2\text{平面上对应的区域})\\
&=\frac{4}{\pi}\iint\limits_D \frac{1}{4}\left\|\frac{\partial(v_1,v_2)}{\partial(x_1,x_2)}\right\|dx_1dx_2=\frac{1}{\pi}\iint\limits_D\frac{\omega}{2}dx_1dx_2\\
&=\frac{1}{2\pi}\iint\limits_D e^{-\frac{1}{2}(x_1^2+x_2^2)}dx_1dx_2.
\end{aligned}
$$

(5)Hasting有理逼近方法(近似直接抽样法)

由定理1.1知,若 $R \sim U(0,1)$,则 $\Phi^{-1}(R) \sim N(0,1)$.用反函数法产生 $N(0,1)$ 随机数的困难在于 Φ^{-1} 不能用初等函数表示.但用有理数可以逼近 $\Phi^{-1}(x)$ .Hasting给出的有理逼近方法是最好的一种.具体方法如下:

①产生 $r \sim U(0,1)$;

②计算 $y = \sqrt{-2ln\alpha}$,其中 $\alpha = \begin{cases} r, & 当 r \leq 0.5, \\ 1-r, & 当 r > 0.5; \end{cases}$

③令 $X = sign(r - \frac{1}{2})(y - \frac{c_0+c_1y+c_2y^2}{1+d_1y+d_2y^2+d_3y^3})$, 其中

$$c_0 = 2.515517, c_1 = 0.802853, c_2 = 0.010328,$$

$$d_1 = 1.432788, d_2 = 0.189269, d_3 = 0.001308,$$

则 X 是 $N(0,1)$ 随机数.

正态分布(六)

(6)Kahn密度逼近法

Kahn对半正态密度函数给出了一个渐进函数,即

$$\sqrt{\frac{2}{\pi}}e^{-\frac{x^2}{2}} \approx \sqrt{\frac{2}{\pi}}\frac{4e^{-ax}}{(1+e^{-ax})^2} \triangleq g(x),$$

适当选取 a 使 $g(x)$ 为密度函数.用直接抽样法对密度为 $g(x)$ 的分布进行抽样,得抽样公式为:

$$u = \sqrt{\frac{\pi}{8}} ln\frac{1+r}{1-r} \quad (r \sim U(0,1)).$$

u 近似为半正态分布;然后随机地确定符号,即得正态随机数.

(7)复合舍选抽样法

这是Marsaglia和Bray(1964年)提出复合舍选抽样法.将正态密度函数 $\varphi(x)$ 分解成个密度函数的概率和:

$$\varphi(x) = p_1 f_1(x) + p_2 f_2(x) + p_3 f_3(x) + p_4 f_4(x).$$

复合抽样的具体抽样步骤如下:

①以概率 $p_1 = 0.8638$ 产生均匀随机数 r_1, r_2, r_3 ,则

$$X = 2(r_1 + r_2 + r_3 - 1.5) \quad (X \sim f_1(x)).$$

②以概率 $p_2 = 0.1107$ 产生均匀随机数 r_1, r_2 ,则

$$X = 1.5(r_1 + r_2 - 1) \quad (X \sim f_2(x)).$$

③以概率 $p_3 = 0.0228002039$ 产生均匀随机数 r_1, r_2 , 令 $U = 6r_1 - 3, V = 0.358r_2$,当 $V < g(U)$ 时,则 $X = U \quad (X \sim f_3(x))$,

其中

$$g(u)=\begin{cases} 17.49731196e^{-\frac{u^2}{2}} \\ \quad +2.15787544(|u|-1.5) \\ \quad +4.73570326(u^2-3), & 当 |u|<1, \\ 17.49731196e^{-\frac{u^2}{2}} \\ \quad +2.15787544(|u|-1.5) \\ \quad -2.36785163(3-|u|)^2, & 当 1\le |u|<1.5, \\ 17.49731196e^{-\frac{u^2}{2}} \\ \quad -2.36785163(3-|u|)^2, & 当 1.5\le |u|<3, \\ 0, & 当 |u|\ge 3. \end{cases}$$

正态分布(七)

④以概率 $p_4 = 0.0026997961$ 产生随机数 r_i ,令 $v_i = 2r_i - 1(i = 1, 2, \cdots)$,直到 $\omega = v_i^2 + v_{i+1}^2 \le 1$ 时计算

$$S = v_1(\frac{9 - 2ln\omega}{\omega})^{\frac{1}{2}}, T = v_2(\frac{9 - 2ln\omega}{\omega})^{\frac{1}{2}}.$$

若 $|S| \le 3$ 且 $|T| \le 3$ 时,则重新抽样;否则令

$$X = \begin{cases} S, & \text{当}|S| > 3\text{时}, \\ T, & \text{当}|S| \le 3\text{且}|T| > 3\text{时}, \end{cases}$$

则 $X \dot{\sim} N(0, 1)$.

指数分布(一)

(1)变换抽样法(直接抽样法)

设 $R \sim U(0,1)$,产生 $e(\lambda)$ 随机数的变换抽样公式为

$$X = -\frac{1}{\lambda} lnR.$$

利用以上抽样公式由 R 得到 X ,算法很简单,但在计算机上计算自然对数是比较费时间的.为了提高抽样速度,下面给出两个产生 $e(1)(\lambda = 1)$ 随机数的算法.

指数分布(二)

(2)Von Neumann的抽样方法

(3)Marsaglia的抽样方法

Marsaglia通过复杂的计算推导,给出了 $\lambda = 1$ 的指数分布随机数 X 的抽样公式:

$$X = M + min(r_1, r_2, \cdots, r_N),$$

其中 M, N 是离散随机变量,且

$$P\{M = m\} = \frac{e - 1}{e^{m+1}} \quad (m = 0, 1, 2, \cdots),$$

$$P\{N = n\} = \frac{1}{n!(e - 1)} \quad (n = 1, 2, \cdots).$$

Gamma分布

Gamma分布

Matlab中:random

$X \sim Beta(a,b)$ 分布

Matlab中:random

$E(X) = \frac{a}{a+b}$;

$Var(X) = \frac{ab}{(a+b)^2(a+b+1)}$

卡方n分布,F(m,n)分布与t(n)分布

Matlab中:random

(1)三角形分布的直接抽样法

三角形分布(triang(a,b,m))的概率密度函数(见图3-17)为

$$f(x)=\begin{cases}\frac{2(x-a)}{(b-a)(m-a)}, & a<x\le m,\\ \frac{2(b-x)}{(b-a)(b-m)}, & m<x\le b,\\ 0, & \text{其他}.\end{cases}$$

其他连续分布(一)

当 $a=0, b=1$ 时的三角形分布称为标准三角形分布.若 $X_1 \sim triang(0,1,c)(0 \le c \le 1)$,令 $X = a+(b-a)X_1$,则 $X \sim triang(a,b,m)$(其中 $m=a+(b-a)c$).我们以下只讨论 $triang(0,1,c)$ 随机数的产生方法.

$triang(0,1,c)$ 的分布函数为

$$F(x)=\begin{cases}0, & x<0,\\ \frac{x^2}{c}, & 0\le x<c,\\ 1-\frac{(1-x)^2}{(1-c)}, & c\le x<1,\\ 1, & x\ge 1.\end{cases}$$

其反函数为$F^{-1}(r)=\begin{cases}(cr)^{\frac{1}{2}}, & 0\le r<c,\\ 1-\sqrt{(1-c)(1-r)}, & c\le r\le 1.\end{cases}$

其他连续分布(一)

故 $triang(0,1,c)$ 随机数的反函数抽样法为:

①产生 $r \sim U(0,1)$;

②若 $r \le c$,令 $X = (cr^{\frac{1}{2}})$;否则令 $X = 1 - \sqrt{(1-c)(1-r)}$,

那么由此产生的 X 为 $triang(0,1,c)$ 随机数.

(2)威布尔(Weibull)分布的直接抽样法

Weibull分布(W(m,a))的密度函数为

$$f(x) = \frac{m}{a}x^{m-1}e^{-\frac{x^m}{a}} \quad (x > 0; m > 0, a > 0),$$

分布函数为 $F(x) = 1 - e^{-\frac{x^m}{a}}(x > 0)$;其反函数 $F^{-1}(r) = [-aln(1-r)]^{\frac{1}{m}}$.故威布尔分布的直接抽样法如下:

①生成 $r \sim U(0,1)$;

②令 $X = [-alnr]^{\frac{1}{m}}$,并输出服从 $W(m,a)$ 分布的随机数 X .

(3)对数正态分布的变换抽样法

利用对数正态分布与正态分布的关系,具体算法如下:

①产生 $U \sim N(0,1)$;

②计算 $Y = \sigma U + \mu$;

③令 $X = e^{Y}$,并输出服从对数正态分布(记为 $LN(\mu,\sigma^2)$) 的随机数 X .

注意:参数 μ,σ^2 并不是对数正态随机变量 X 的均值和方差.经计算可知 $E(X) = e^{\mu+\frac{\sigma^2}{2}} \triangleq \mu.$, $Var(X) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1) \triangleq \sigma^2.$.

(4)柯西分布

柯西分布(C(0,1))的密度函数为

$$f(x)=\frac{1}{\pi(1+x^2)},\quad x\in(-\infty,\infty).$$

方法一　柯西分布的直接抽样法.

柯西分布的分布函数为 $F(x)=\frac{1}{\pi}arctgx+\frac{1}{2}$,其反函数 $F^{-1}(r)=tg[\pi r-\pi/2]$.故柯西分布的直接抽样法如下:

①生成 $r\sim U(0,1)$;

②令 $X=tg[\pi r-\pi/2]$,并输出服从 $C(0,1)$ 分布的随机数 X .

方法二　柯西分布的变换抽样法.

利用柯西分布与正态分布的关系有以下抽样法:

①生成 $X_1,X_2\sim N(0,1)$;

②令 $X=X_1/X_2$ 并输出服从 $C(0,1)$ 分布的随机数 X .

三常用连续分布的抽样法

1. 二项分布B(n,p)
2. 泊松分布
3. 几何分布与负二项分布

二项分布B(n,p)

Matlab中:random

泊松分布

泊松分布

Matlab中:random

几何分布与负二项分布

几何分布与负二项分布

Matlab中:random

一.随机向量的一般抽样方法

二.多维正态随机向量的抽样法

(1)条件分布法

设随机向量 $X=(X_1,\cdots,X_n)'$ 的联合密度函数为 $f(x_1,x_2,\cdots,x_n)$,则有

$$f(x_1,x_2,\cdots,x_n)=f_1(x_1)f_2(x_2|x_1)\cdots f_n(x_n|x_1,\cdots,x_{n-1}),$$

其中 $f_2(x_2|x_1),\cdots,f_n(x_n|x_1,\cdots,x_{n-1})$ 均为条件密度.如 $n=3$ 时,

$$f_2(x_2|x_1)=\int_{-\infty}^{\infty}\frac{f(x_1,x_2,x_3)}{f_1(x_1)}dx_3,$$

$$f_3(x_3|x_1,x_2)=\frac{f(x_1,x_2,x_3)}{f_1(x_1)\cdot f_2(x_2|x_1)}.$$

随机向量的一般抽样方法

随机向量 X 的抽样方法如下:

①产生 $x_1 \sim f_1(x_1)$;

②以 x_1 为已知参数,产生 $x_2 \sim f_2(x_2|x_1)$;

③以 x_1, x_2 为已知参数,产生 $x_3 \sim f_3(x_3|x_1, x_2)$;

④一般地以 $x_1, x_2, \cdots, x_{n-1}$ 为已知参数,产生 $x_n \sim f_n(x_1, \cdots, x_{n-1})$;

⑤令 $X = (x_1, x_2, \cdots, x_n)'$,并输出联合密度为 $f(x_1, x_2, \cdots, x_n)$ 的随机向量 X.

(2)舍选法

随机向量的舍选法是随机变量舍选法的直接推广.下面讨论最简单的情况.

设随机向量 X 在平行多面体 $a_i \le x_i \le b_i (i = 1, \cdots, n)$ 上具有密度函数 $f(x_1, x_2, \cdots, x_n)$,且上界:$f_0 = \sup\limits_{a_i \le x_i \le b_i} f(x_1, x_2, \cdots, x_n)$ 取有限值.产生 $n+1$ 个均匀随机数 $r_0, r_1, \cdots, r_n \sim U(0,1)$;若

$$f_0 r_0 \le f[(b_1 - a_1) r_1 + a_1, \cdots, (b_n - a_n) r_n + a_n]$$

成立,则有

$$X = ((b_1 - a_1) r_1 + a_1, \cdots, (b_n - a_n) r_n + a_n)' \sim f(x_1, x_2, \cdots, x_n).$$

多维正态随机向量的抽样法