2．3 任意分布的伪随机变量的抽样

大多数的伪随机数变量并不满足[0，1]区间的均匀分布，而是具有各种不同形式的分布密度函数。

对一个具有分布密度函数 $f(x)$ 的伪随机变量的抽样是通过以下步骤来进行的：首先在[0，1]区间抽取均匀分布的伪随机数列，然后再从这个伪随机数列中抽取一个简单子样，使这个简单子样的分布满足分布密度函数 $f(x)$，并且各个伪随机数相互独立。实际上只要[0，1]区间上均匀分布的随机数具有好的独立性，则抽得的简单子样也一定具有和它同样好的独立性。

因此，对不均匀的伪随机变量抽样的关键问题是如何从均匀分布的伪随机变量样本中，抽取符合所要求的分布密度函数的简单子样。

迭加原则： 如果要产生分布密度函数为 $f(x)$ 的随机变量样本数列，我们可以把 $f(x)$ 变成分布概率密度函数 $h_i(x)$ 的和的形式，即：

$$f(x)=\sum_i h_i(x)$$

并按其中的分布密度函数 $h_i(x)$ 进行抽样作为 $f(x)$ 的抽样值，决定选择哪一个 $h_i(x)$ 进行抽样的原则是根据 $\int h_i(x)dx$ 的积分值作为权重随机地选择的。这就是蒙特卡洛方法的迭加原则。

在对复杂的分布密度函数的抽样时，伪随机变量抽样的迭加原则是十分有用的。

A． 离散型分布随机变量的直接抽样

如果离散型随机变量 x 以概率 p_i 取值 $x_i(i=1,2,...)$，则其分布函数为：

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

其中 p_i 应满足归一化条件：$\sum_i p_i=1$。该随机变量的直接抽样方法如下：首先选取在[0，1]区间上的均匀分布的随机数 ξ，然后判断满足如下不等式

$$F(x_{j-1})\le \xi < F(x_j)$$

的 j 值，与 j 对应的 x_j 就是所抽子样的一个抽样值，即 $\eta = x_j$。该子样具有分布函数 $F(x_j)$。

例：γ 光子与物质相互作用类型的抽样问题。

γ 光子与物质相互作用有三种类型：光电效应、康普顿效应和电子对效应。其中光电效应和电子对效应为光子吸收过程。设总截面为

$$\sigma_T=\sigma_e+\sigma_p+\sigma_s \quad .$$

1．选择均匀分布随机数 ξ，

2．若满足不等式 $\xi<\sigma_s/\sigma_T$，则发生康普顿散射；

3．若满足不等式 $\sigma_s/\sigma_T\le\xi<(\sigma_s+\sigma_e)/\sigma_T$，则发生光电效应；

4．若 $\xi\ge(\sigma_s+\sigma_e)/\sigma_T$，则产生电子对过程。

B．连续分布的随机变量抽样

一、直接抽样方法

直接抽样法又称为反函数法。设连续型随机变量η的分布密度函数为$f(x)$，在数学上它的分布函数应当为

$$F(x)=\int_{-\infty}^{x} f(x)dx \quad .$$

得到的$\eta=F^{-1}(\xi)$即为满足分布密度函数$f(x)$的一个抽样值。

证明：

该子样中$\eta \le x$的概率为：

$$p\{\eta \le x\}=p\{F^{-1}(\xi)\le x\}=p\{\xi \le F(x)\}=\int_{-\infty}^{0} 0\cdot dx+\int_{0}^{F(x)} 1\cdot dx=F(x) \quad .$$

优点是使用简单，应用范围较广。

缺点：在分布函数$F(x)$不能从分布密度函数$f(x)$解析求出时，或者求出的函数形式抽样太复杂的情况下，就不能采用这种方法。

例　对指数分布的直接抽样。

解　指数分布的问题可用于描述粒子运动的自由程，粒子衰变寿命或射线与物质作用长度等许多物理问题。它的分布密度函数为

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

它的分布函数为

$$F(x)=\int_{-\infty}^{x} f(t)dt=\int_{0}^{x} \lambda e^{-\lambda t}dt=1-e^{-\lambda x} \quad .$$

设ξ是[0,1]区间上的均匀分布的随机数，令$\xi=F(\eta)=1-e^{-\lambda\eta}$，解此方程得到

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

注意到$1-\xi$和ξ同样服从[0，1]区间的均匀分布，故有

$$\eta = -\frac{1}{\lambda}\ln\xi \quad .$$

例　对如下的分布密度函数抽样

$$f(x) = \left(\frac{\gamma-1}{x_0^{\gamma-1}}\right)x^{-\gamma}\text{，}\quad x_0 \le x, \gamma > 1\text{.}$$

解　（2.3.9）式的分布密度函数的对应分布函数为

$$F(x) = \int_{x_0}^{x} f(x)dx \Big/ \int_{x_0}^{+\infty} f(x)dx = 1-\left(\frac{x_0}{x}\right)^{\gamma-1} \quad .$$

在[0，1]区间上的随机抽取均匀分布的随机数ξ，令$\xi = F(\eta) = 1-\left(\frac{x_0}{x}\right)^{\gamma-1}$，解此方程，并考虑到到$1-\xi$和$\xi$都是[0，1]区间的均匀分布的伪随机数，得到

$$\eta = x_0\xi^{-1/(\gamma-1)} \quad .$$

二、　变换抽样法

基本思想： 将一个比较复杂的分布的抽样，变换为已经知道的、比较简单的分布的抽样。

例如,要对满足分布密度函数$f(x)$的随机变量η抽样。如果要对它进行直接抽样是比较困难的。

如果存在另一个随机变量δ，它的分布密度函数为$\phi(y)$，其抽样方法已经掌握，并且也比较简单．我们可以设法寻找一个适当的变换关系$x = g(y)$。如果$g(y)$的反函数存在，记为

$g^{-1}(x)=h(x)$，并且该反函数具有一阶连续导数。

根据概率论的知识，这时 x 满足的分布密度函数为 $\phi(h(x))\cdot|h'(x)|$。如果函数 $g(y)$ 选得合适，使得满足：

$$f(x)=\phi(h(x))\cdot|h'(x)|\,.$$

抽样步骤：首先对分布密度函数 $\phi(y)$ 抽样得到值 δ，然后通过变换 $\eta=g(\delta)$ 得到满足分布密度函数 $f(x)$ 的抽样值。

实际上，直接抽样法是 $\phi(y)$ 为在[0，1]区间上的均匀分布密度函数的特殊情况下，$g(y)=F^{-1}(y)$ 时的变换抽样。因而它是变换抽样的特殊情况。

二维情况下的变换抽样法与一维的情况完全是类似的。假如我们要对满足联合分布密度函数 $f(x,y)$ 的随机变量 η,δ 进行抽样。如果我们已经掌握了满足联合分布密度函数 $g(u,v)$ 的随机变量 η',δ' 的抽样方法，则可以寻找一个适当的变换

$$x=g_1(u,v),$$

$$y=g_2(u,v),$$

g_1,g_2 函数的反函数存在，记为

$$u=h_1(x,y)\ ,$$

$$v=h_2(x,y)\,.$$

该变换满足如下条件：

$$g(h_1(x,y),h_2(x,y))\cdot|J|=f(x,y)\,.$$

$|J|$ 表示函数变换的雅可比(Jacobi)行列式：

$$|J|=\begin{vmatrix}\dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y}\\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y}\end{vmatrix}.$$

这样就可以通过变换式，由满足分布密度函数 $g(u,v)$ 的抽样值 η',δ' 得到待求的满足分布密度函数 $f(x,y)$ 的抽样值 η,δ。

以上的处理要求变换函数 g_1 和 g_2 的反函数 h_1 和 h_2 具有一阶的连续非零导数。

变换抽样的缺点：对具体问题要找到所需要的变换关系式往往是比较困难的。

正态分布的抽样（变换抽样的具体应用）：

设随机变量 η 满足正态分布，它的分布密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}.$$

通常 $f(x)$ 记为 $N(\mu,\sigma^2)$，其中 μ 和 σ^2 分别是随机变量 η 的数学期望值和方差，即

$$E\{\eta\}=\mu,\qquad V\{\eta\}=\sigma^2.$$

当 $\mu=0,\sigma^2=1$ 时的分布称为标准正态分布，此时的分布密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}}\exp\{-x^2/2\}.$$

记为 $N(0,1)$。

通常我们只需考虑标准正态分布的抽样方法即可。因为假如随机变量 η 满足正态分布，随机变量 δ 满足标准正态分布，则 η 和 δ 之间满足关系式

$$\eta=\sigma\delta+\mu.$$

标准正态分布密度函数不能用一般函数解析积分求出分布函数 $F(x)$，因而不能直接应用从均匀分布的抽样值变换到标准正

态分布的抽样值。但是可以采用一个巧妙的办法将两个独立的均匀分布的随机变量u,v变换为标准正态分布的随机变量x,y。这就是做变换：

$$\left.\begin{aligned} x &= \sqrt{-2\ln u}\cos(2\pi v), \\ y &= \sqrt{-2\ln u}\sin(2\pi v). \end{aligned}\right\}$$

反解上式得到：

$$\left.\begin{aligned} u &= \exp\left\{-\frac{1}{2}\left(x^2+y^2\right)\right\} \equiv h_1(x,y) \\ v &= \frac{1}{2\pi}\tan^{-1}(y/x) \equiv h_2(x,y) \end{aligned}\right\}$$

按照概率理论，x 和 y 的联合分布密度函数为

$$f(x,y) = g(h_1(x,y),h_2(x,y))\cdot|J|\ .$$

由于u和v是独立的均匀分布的随机变量，它们的联合分布密度函数$g(u,v)=1$。可以证明：

$$f(x,y) = \frac{1}{2\pi}\exp\left\{-\frac{1}{2}\left(x^2+y^2\right)\right\}\ .$$

又因为$f(x,y)$可以写为：

$$f(x,y) = f(x)\cdot f(y)\ .$$

其中

$$f(x) = \frac{1}{\sqrt{2\pi}}\exp\left\{-x^2/2\right\}\ ,$$

$$f(y) = \frac{1}{\sqrt{2\pi}}\exp\left\{-y^2/2\right\}\ .$$

因此从上式中的任意一式给出的抽样值都满足标准正态分布。

上述正态分布的变换抽样法还可以做些改进，这就是所谓的 Maraglia 方法。其抽样过程：

（1） 产生[0，1]区间上的独立均匀分布随机数u和v。

（2） 计算$w=(2u-1)^2+(2v-1)^2$。

（3） 如果$w>1$，回到步骤（1）；否则，执行（4）。

（4） 计算$z=[-2\ln(w)/w]^{1/2}$，取$x=uz, y=vz$。

三、 舍选抽样法

舍选法是冯.诺曼(Von Neumann)为克服直接抽样和变换抽样方法的困难最早提出来的。

基本思想：按照给定的分布密度函数$f(x)$，对均匀分布的随机数序列$\{\xi_n\}$进行舍选。舍选的原则是在$f(x)$大的地方，保留较多的随机数ξ_i；在$f(x)$小的地方，保留较少的随机数ξ_i，使得到的子样中ξ_i的分布满足分布密度函数$f(x)$的要求。

这种方法对分布密度函数$f(x)$在抽样范围内有界，且其上界是容易得到的情况，是可以采用的。它使用起来十分灵活，计算也较简单，因而使用也比较广泛。

这种方法，对$f(x)$在抽样范围内函数值变化很大的时候，效率是很低的，因为大量的均匀分布抽样点被舍弃了。

1．第一类舍选法

设随机变量η在[a,b]上的分布密度函数为$f(x)$，$f(x)$的在区间[a,b]上的最大值存在，并等于

$$L=\max_{x\in[a,b]} f(x)=\frac{1}{\lambda}$$

显然这里$\lambda f(x)$在$x\in[a,b]$范围内的取值在[0，1]区间上。

采用舍选法的步骤为：

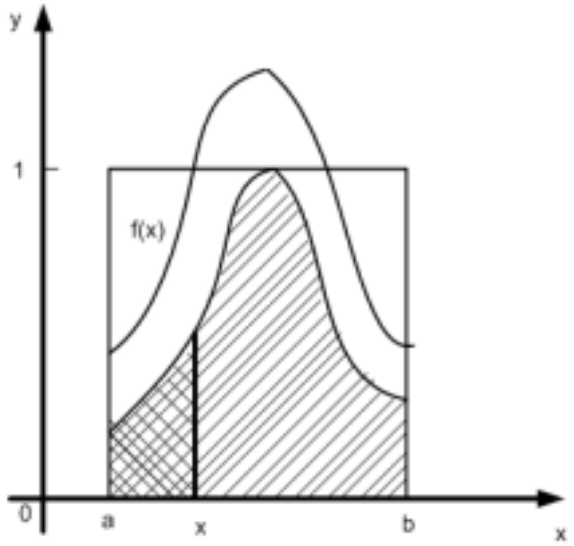
(a) 选用均匀的[0，1]区间的随机数ξ_1，构造出[a,b]区间上的均匀分布的随机数$\delta = a + (b-a)\xi_1$。

(b) 再选取独立的均匀分布于[0，1]区间上的随机数ξ_2，判断$\xi_2 \le \lambda f(\delta)$是否满足。如满足上面不等式，则执行（c）；如不满足，则返回到步骤（a)。

(c) 选取$\eta = \delta$作为一个抽样值。

重复上面三个步骤，就可以产生出随机数序列$\{\eta_n\}$，它满足分布密度函数$f(x)$。如图(2.3.1)所示，舍选抽样步骤（b）的判断不等式$\xi_2 \le \lambda f(\delta)$，是为了保证随机点$(\delta, \xi_2/\lambda)$落在$f(x)$曲线的下面。因为 x 取值在$[x, x+dx]$内的概率等于面积比

$$\frac{f(x)dx}{\int_a^b f(x)dx} = f(x)dx$$



上述抽样步骤得到的随机数数列是以分布密度函数$f(x)$分布的。由于随机点$(\delta, \xi_2/\lambda)$落在曲线$f(x)$以下才被接受，并且所有产生的点都落在面积$L(b-a)$的范围内。

采用该方法的抽样效率为

$$E=\frac{\int_a^b f(x)dx}{L(b-a)}=\frac{1}{L(b-a)} \quad .$$

显然我们希望效率能够越高越好。如果 L 很大（即 $f(x)$ 具有高峰），则此舍选抽样效率就不高。

例　对随机变量 η 抽样。它的分布密度函数为

$$f(x)=\begin{cases}2x, 0\le x\le 1,\\ 0, \text{其它}.\end{cases}$$

解　如果用直接抽样法，首先求出分布函数

$$F(x)=x^2 \quad .$$

抽取在[0，1]区间上的均匀分布的随机数 ξ。令

$$\xi=x^2 \quad .$$

则有

$$x=\sqrt{\xi} \quad .$$

x 为 η 的子样的一个个体。但是开方运算量较大，可改用舍选法来做。

$$L\equiv\max_{x\in[0,1]} f(x)=\max_{x\in[0,1]} 2x=2 \quad .$$

依照第一类舍选法步骤：

1. 依次产生独立的[0，1]区间上的均匀分布的随机数 ξ_1,ξ_2，
2. 判断 $\xi_2\le\frac{1}{L}f(\xi_1)=\xi_1$ 是否成立。
3. 若成立，则取 $x=\xi_1$；
4. 若上面不等式不成立，可以再产生一组 ξ_1,ξ_2 进行重复试验。

但实际上，因为 ξ_1,ξ_2 本来就是任意的，如果 $\xi_2\le\xi_1$ 不成立，必有 $\xi_1<\xi_2$。所以若 $\xi_2\le\xi_1$ 不成立，只要将 ξ_1 和 ξ_2 互换以下，这个不

等式就必定成立。所以可以取

$$x=\max(\xi_1,\xi_2)\quad .$$

一般高次幂的情况。设η满足分布密度函数

$$f(x)=\begin{cases} nx^{n-1}, x\in[0,1], n=1,2,... \\ 0, 其它. \end{cases}$$

用舍选法抽样，依次产生独立的[0，1]区间上的均匀分布的随机数$\xi_1,\xi_2...,\xi_n$，则取

$$x=\max(\xi_1,\xi_2,...,\xi_n)\quad .$$

2．第二类舍选法

假如$h(x)$和$f(x)$同是在$x\in[0,1]$区域上的分布密度函数，并且$f(x)$可以写为

$$f(x)=L\cdot\frac{f(x)}{Lh(x)}h(x)\equiv Lg(x)h(x)\quad .$$

其中 L 为常数，它要保证$|g(x)|\le 1$，即$L=\max\limits_{x\in[0,1]}\dfrac{f(x)}{h(x)}>1$。$g(x)$可视为另一个随机变量的分布密度函数。

抽样步骤：

（1） 在[0，1]区间上抽取均匀分布随机数ξ，并由$h(x)$分布密度函数抽样得到η_h。

（2） 判别$\xi\le g(\eta_h)$不等式是否成立。如果不成立，则返回到步骤（1）。

（3） 选取$\eta=\eta_h$作为服从分布密度函数$f(x)$的一个抽样值。

这种方法的抽样效率为 $E=1/L$。

例　采用第二类舍选抽样法来产生标准正态分布的随机抽样值。标准正态分布密度函数可以写为

$$f(x)=\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}\qquad(-\infty<x<+\infty)$$

解　由于相应的分布密度函数不存在反函数，故可以采用舍选法。令

$$L\equiv\sqrt{\frac{2e}{\pi}}\ ,$$

$$h(x)\equiv e^{-x},\qquad(0<x<+\infty)\ ,$$

$$g(x)\equiv\exp\{-(x-1)^2/2\}\ ,\qquad(0<x<+\infty)\ .$$

由于 $f(x)$ 是 x 的偶函数，因而可以在 $(0,+\infty)$ 区域上抽样后反射到 $(-\infty,0)$ 区间上的抽样值。这样我们可以只考虑 $(0,+\infty)$ 区域的抽样。此时在对 $f(x)=Lg(x)h(x)$ 的抽样中，

（1）对 $h(x)$ 的抽样可以用直接抽样法。由 $\eta_h=-\ln\xi_1$ 算出 η_h 的值，

（2）然后产生随机数 ξ_2，判别 $\xi_2\le g(\eta_h)$ 是否成立，也即判断不等式 $(\eta_h-1)^2\le-2\ln\xi_2$ 是否成立。

（3）如不成立，则舍弃，再重新由 $h(x)$ 直接抽样；

（4）如成立，则抽样值为 η_h。该抽样的效率为 $E=\sqrt{\frac{\pi}{2e}}$。

3.　第三类舍选法

如果分布密度函数可以表示成积分形式

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

其中 $g(x,y)$ 是二维随机向量 (x,y) 的联合分布密度函数，$h(x)$ 取

值在 y 的定义域上。常数 L 定义为

$$L = 1/\int_{-\infty}^{+\infty}\int_{-\infty}^{h(x)} g(x,y)dxdy > 1 \quad .$$

舍取抽样步骤：

（1） 由联合分布密度函数 $g(x,y)$ 抽取 (η_x, η_y) 随机向量值。

（2） 判别 $\eta_y \le h(\eta_x)$ 是否成立。若不成立，返回（1）。

（3） 取分布密度函数 $f(x)$ 的抽样值 $\eta = \eta_x$。

该方法的抽样效率为 $1/L$。

证明： 抽取的子样中 $\eta \le x$ 的概率等于在 $\eta_y \le h(\eta_x)$ 条件下，$\eta_x \le x$ 出现的概率。即

$$p\{\eta \le x\} = p\{\eta_x \le x \,|\, \eta_y \le h(\eta_x)\} = \frac{p\{\eta_x \le x, \eta_y \le h(\eta_x)\}}{p\{\eta_y \le h(\eta_x)\}}$$

$$= \frac{\int_{-\infty}^{x} dt_1 \int_{-\infty}^{h(t_1)} g(t_1,t_2)dt_2}{\int_{-\infty}^{+\infty} dt_1 \int_{-\infty}^{h(t_1)} g(t_1,t_2)dt_2} = \int_{-\infty}^{x} \left[L\int_{-\infty}^{h(t_1)} g(t_1,t_2)dt_2 \right] dt_1 \quad .$$

在此，我们应用了贝斯(Bayes)定理。

当 x ，y 相互独立时，则有 $g(x,y) = g_1(x)g_2(y)$。则

$$f(x) = Lg_1(x)\int_{-\infty}^{h(x)} g_2(y)dy \quad .$$

若进一步假定 $0 \le h(x) \le 1$，并且

$$g_2(y) = \begin{cases} 1, y \in [0,1] \\ 0, \text{其它} \end{cases}$$

则有 $f(x) = Lh(x)g_1(x)$，这正好属于第二类舍选法处理的分布密度函数类型。

例 各向同性方位角余弦的抽样。

解 此问题可以采用直接抽样法。由[0，1]区间上的均匀分布随机数 ξ 产生出 $[0,2\pi]$ 的均匀分布随机数 $\delta = 2\pi\xi$，方位角余弦的

抽样值为$\eta=\cos\delta$。但是由于余弦运算量较大，可以改用第三类舍选法。

方位角余弦的分布密度函数为

$$f(x)=\begin{cases}\dfrac{1}{\pi}\dfrac{1}{\sqrt{1-x^2}}, & |x|<1\\ 0, \text{其它}\end{cases}$$

取独立的在[0，1]区间上均匀分布的随机数ξ_1和ξ_2，定义

$$x=\frac{\xi_1^2-\xi_2^2}{\xi_1^2+\xi_2^2}\quad ,$$

$$y=\xi_1^2+\xi_2^2$$

反解公式所示方程得到

$$\xi_1=\sqrt{\frac{1}{2}y(1+x)}\equiv h_1(x,y)\quad ,$$

$$\xi_2=\sqrt{\frac{1}{2}y(1-x)}\equiv h_2(x,y)\quad .$$

现在我们来求出(x,y)所满足的联合分布密度函数。

$$g(x,y)=f_1(h_1(x,y),h_2(x,y))\cdot|J|$$

其中f_1为ξ_1，ξ_2的联合分布密度函数。由于ξ_1和ξ_2均为区间[0，1]上的独立均匀分布的随机数，因而$f_1(h_1(x,y),h_2(x,y))=1$。联合分布密度函数$g(x,y)$的计算结果为：

$$g(x,y)=\begin{cases}\dfrac{1}{4\sqrt{1-x^2}}, & \text{当}|x|<1,0<y<1,\\ 0, \text{其它}\end{cases}\quad .$$

可以得到

$$f(x)=\frac{4}{\pi}\int_{-\infty}^{1}g(x,y)dy\quad .$$

这相当于$L=\frac{4}{\pi}$，$h(x)=1$。

抽样步骤：

（1） 产生[0，1]区间上的均匀分布的独立随机数ξ_1和ξ_2，计算$x=\frac{\xi_1^2-\xi_2^2}{\xi_1^2+\xi_2^2}$和$y=\xi_1^2+\xi_2^2$。

（2） 判断$y\le h(x)=1$是否成立。如不成立返回（1）。

（3） 方位角余弦$\cos\phi$的抽样值$\eta=\frac{\xi_1^2-\xi_2^2}{\xi_1^2+\xi_2^2}$，$\sin\phi$的抽样值为$\eta'=\frac{2\xi_1\xi_2}{\xi_1^2+\xi_2^2}$。

这就同时求出$\sin\phi$的抽样值，但此时$\sin\phi$总是正的。这种方法的效率为$E=\frac{\pi}{4}\approx 0.785$。

改进后的抽样步骤：

（1） 产生[0，1]区域上的独立均匀分布的随机数ξ_1和ξ_2。令$x=\xi_1, y=2\xi_2-1$。

（2） 判断$x^2+y^2<1$是否成立。如果不等式不成立，则返回到(1)。

（3） 取$\cos\phi$的抽样值$\eta=\frac{x^2-y^2}{x^2+y^2}$，$\sin\phi$的抽样值为$\eta'=\frac{2xy}{x^2+y^2}$。

改进后的$\sin\phi$的抽样值就可以正可以负。

四、 复合抽样法

处理具有复合分布的随机变量的抽样。所谓复合分布是指随机变量 x 服从的分布与另一个随机变量 y 有关的分布。一般复合分布密度函数可以表示为

$$f(x)=\int_{-\infty}^{+\infty} g(x\mid y)h(y)dy \quad .$$

其中$g(x\mid y)$表示与参数 y 有关的 x 的条件分布密度函数，而$h(y)$

是 y 的分布密度函数。这时可以采取如下的方法来抽样：首先，由分布密度函数 $h(y)$ 抽取 y_h，然后由 $g(x|y_h)$ 抽取 x_g 的值：

$$\xi_f = x_{g(x|y_h)}.$$

上述抽样步骤的证明：

$$p(x \le \xi_f < x + dx) = p(x \le x_{g(x|y_h)} < x + dx)$$

$$= \int_{-\infty}^{+\infty} g(x|y)h(y)dydx = f(x)dx$$

所以 ξ_f 服从分布 $f(x)$。

1. 加分布抽样

作为复合抽样的特殊情况，在此首先介绍加分布抽样。数学上加分布的一般形式为

$$f(x) = \sum_n p_n h_n(x),$$

其中

$$0 < p_n < 1 \quad , \qquad \sum_n p_n = 1 \quad .$$

这即是意味作总体分布以概率 p_n 取分布 $h_n(x)$。

抽样的方法：

（1）取[0,1]区间上均匀分布随机数 ξ，解下面的不等式求得 n 。

$$\sum_{i=1}^{n-1} p_i < \xi \le \sum_{i=1}^{n} p_i \quad .$$

(2) 找到对应的 $h_n(x)$，并对其抽样，得到最后的抽样值 $\eta = \eta_{h_n}$。

这样的抽样步骤实际上是本节开始时介绍的迭加原则的应用。

例　球壳均匀分布的抽样．设球壳内外半径分别为 R_0 和 R_1，球壳内一点到球心距离为 r ，则 r 的分布密度函数为

$$f(r)=\frac{3r^2}{R_1^3-R_0^3},\qquad R_0\le r\le R_1 \quad .$$

解　用直接抽样法，取[0,1]区间上的均匀分布随机数 ξ，则 $\eta=\left[\left(R_1^3-R_0^3\right)\xi+R_0^3\right]^{1/3}$ 的取值就是以 $f(r)$ 分布的一个抽样值。

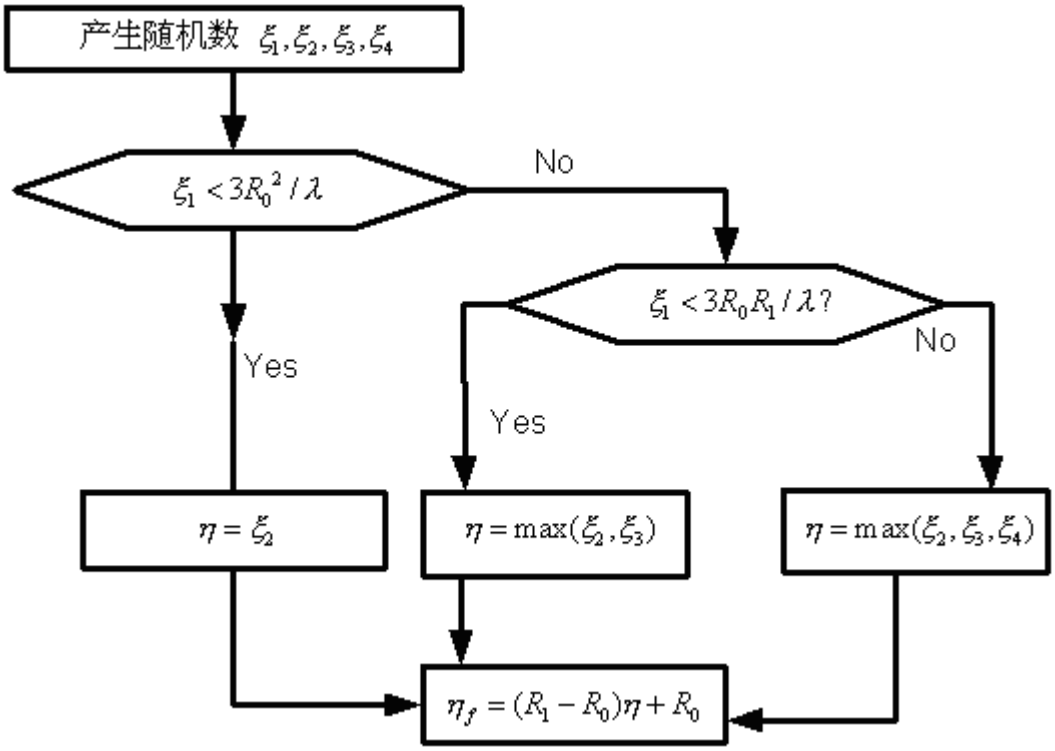
为了避免用运算量较大的开方运算，可以改用复合抽样。令

$$r=(R_1-R_0)x+R_0 \quad ,\qquad \lambda=R_1^2+R_1R_0+R_0^2 \quad .$$

则可以化为

$$f(x)=\frac{(R_1-R_0)^2}{\lambda}3x^2+\frac{3R_0(R_1-R_0)}{\lambda}2x+\frac{3R_0^2}{\lambda}\cdot 1 \quad .$$

抽样的程序框图：



2.　减分布抽样

此类抽样的分布密度函数为

$$f(x)=A_1g_1(x)-A_2g_2(x) \quad .$$

x 定义在区域[a,b]上，A_1和A_2为非负实数。令 m 为$g_2(x)/g_1(x)$的下界，即

$$m=\min_{x\in[a,b]}\frac{g_2(x)}{g_1(x)} \quad .$$

则

$$0<f(x)=g_1(x)\left[A_1-A_2\frac{g_2(x)}{g_1(x)}\right]\le g_1(x)(A_1-A_2m) \quad .$$

因为$A_1-A_2m>0$，所以

$$0<\frac{f(x)}{(A_1-A_2m)g_1(x)}\le 1$$

令

$$h_1(x)=\frac{f(x)}{(A_1-A_2m)g_1(x)}=\frac{A_1}{A_1-A_2m}-\frac{A_2}{A_1-A_2m}\frac{g_2(x)}{g_1(x)},$$

则$f(x)$可以写为：

$$f(x)=(A_1-A_2m)h_1(x)g_1(x)$$

我们可以知道$0<h_1(x)\le 1$．因而按第二类舍选法抽样即可。

抽样效率为：$E_1=\frac{1}{(A_1-A_2m)}$

类似上述方法，我们可以将$f(x)$写为

$$f(x)=\frac{A_1-A_2m}{m}h_2(x)g_2(x) \quad .$$

其中

$$h_2(x)=\frac{A_1m}{A_1-A_2m}\frac{g_1(x)}{g_2(x)}-\frac{A_2m}{A_1-A_2m} \quad , \quad 0<h_2(x)\le 1 \quad .$$

同样按第二类舍选抽样法，其效率为：$E_2=\frac{m}{(A_1-A_2m)}=mE_1$ ．

3． 乘加分布抽样

此类分布密度函数形式为

$f(x)=\sum_n H_n(x)g_n(x)$ ， $x\in[a,b]$

其中$H_n(x)\geq 0$。 为简单计，下面我们只考虑两项(n=2)的情况．对更多项(n>2) 情况的一般表示可以以此作推广。

设η的分布密度函数为：

$$f(x)=H_1(x)g_1(x)+H_2(x)g_2(x)$$

如果令

$p_1=\int_a^b H_1(x)g_1(x)dx$ ， $p_2=\int_a^b H_2(x)g_2(x)dx$ ．

则必有$p_1+p_2=1$。这样我们可以改写$f(x)$为：

$$f(x)=p_1\frac{H_1(x)}{p_1}g_1(x)+p_2\frac{H_2(x)}{p_2}g_2(x)=p_1g_1'(x)+p_2g_2'(x)$$ ．

上式所表示的分布密度函数形式就可以采用加分布抽样法。

我们也可以采用另一种方式，将公式改写为

$$f(x)=(M_1+M_2)\left\{\frac{M_1}{M_1+M_2}\frac{H_1(x)}{M_1}g_1(x)+\frac{M_2}{M_1+M_2}\frac{H_2(x)}{M_2}g_2(x)\right\}$$ ．

其中M_1和M_2分别是$H_1(x)$和$H_2(x)$在区域[a,b]上的上界。令

$p_1=\frac{M_1}{M_1+M_2}$ ， $p_2=\frac{M_2}{M_1+M_2}$ ．

$L_1=L_2=M_1+M_2$， $H_1(x)=M_1h_1(x)$ ， $H_2(x)=M_2h_2(x)$．

则

$$f(x)=p_1[L_1h_1(x)g_1(x)]+p_2[L_2h_2(x)g_2(x)]$$ ．

这样的分布密度函数形式就可以采用加分布抽样和第二类舍选法抽样。这种处理方法的效率不如前一种方法高，但省掉了

公式中的积分计算。

4. 乘减分布抽样

设分布密度函数 $f(x)$ 的形式为

$$f(x)=H_1(x)g_1(x)-H_2(x)g_2(x),\quad x\in[a,b]\ .$$

令

$$m=\min_{x\in[a,b]}\frac{H_2(x)g_2(x)}{H_1(x)g_1(x)}\ ,\qquad M=\max_{x\in[a,b]}H_1(x)\ ,$$

则有如下的关系:

$$0<f(x)=H_1(x)g_1(x)\left[1-\frac{H_2(x)g_2(x)}{H_1(x)g_1(x)}\right]\le H_1(x)g_1(x)(1-m)\le M_1(1-m)g_1(x)\ .$$

再令

$$h_1(x)=\frac{1}{M_1(1-m)}\left[H_1(x)-\frac{H_2(x)g_2(x)}{g_1(x)}\right]\ ,$$

则 $f(x)=M_1(1-m)h_1(x)g_1(x)$.

可知 $0<h_1(x)\le 1$，因而实际上抽样可以采用第二类舍选抽样法。采用如上类似的方法，不难也将分布密度函数 $f(x)$ 改写为

$$f(x)=M_2\frac{1-m}{m}h_2(x)g_2(x)\ .$$

其中 M_2 为 $H_2(x)$ 在[a,b]区间的上界．且

$$h_2(x)=\frac{m}{M_2(1-m)}\left[\frac{H_1(x)g_1(x)}{g_2(x)}-H_2(x)\right],$$

$h_2(x)$ 在[a,b]区间上满足 $0 < h_2(x) \le 1$.抽样方法与前面的抽样方法相同。