第二章 由已知分布的随机抽样

1. 随机抽样及其特点
2. 直接抽样方法
3. 挑选抽样方法
4. 复合抽样方法
5. 复合挑选抽样方法
6. 替换抽样方法
7. 随机抽样的一般方法
8. 多维分布的随机抽样

第二章 由已知分布的随机抽样

本章叙述由已知分布抽样的各主要方法，并给出在粒子输运问题中经常用到的具体实例。

1. 随机抽样及其特点

由已知分布的随机抽样指的是由己知分布的总体中抽取简单子样。**随机数序列**是由单位均匀分布的总体中抽取的简单子样，属于一种特殊的由已知分布的随机抽样问题。本章所叙述的由任意已知分布中抽取简单子样，是在假设随机数为已知量的前提下，使用严格的数学方法产生的。

为方便起见，用 X_F 表示由己知分布 $F(x)$ 中产生的简单子样的个体。对于连续型分布，常用分布密度函数 $f(x)$ 表示总体的己知分布，用 X_f 表示由己知分布密度函数 $f(x)$ 产生的简单子样的个体。另外，在抽样过程中用到的伪随机数均称随机数。

伪随机数

在各种随机变量中，［０，１］上均匀分布的随机数是实现由已知分布抽样的基本量，产生随机数的方法也与其它随机变量的抽样方法不同。

由于随机数在蒙特卡罗方法中占有极其重要的位置，我们用专门的符号ξ表示。

在连续型随机变量的分布中，最简单且最基本的分布是单位均匀分布。由该分布抽取的简单子样称随机数序列，其中每一个体称为随机数。

单位均匀分布也称为［0，1］上的均匀分布，其分布密度函数为：

$$f(x)=\begin{cases}1, & 0\le x\le 1\\ 0, & \text{其他}\end{cases}$$

分布函数为：

$$F(x)=\begin{cases}0, & x<0\\ x, & 0\le x\le 1\\ 1, & x>1\end{cases}$$

在计算机上通常是用数学方法产生随机数，即用如下递推公式：

$$\xi_{n+k} = T(\xi_n, \xi_{n+1}, \cdots, \xi_{n+k-1}), \quad n = 1, 2, \cdots$$

产生随机数序列。

经常使用的是 $k=1$ 的情况，其递推公式为：

$$\xi_{n+k} = T(\xi_n)$$

对于给定的初始值 ξ_1，确定 ξ_{n+1}，n = 1，2...

例如：乘同余方法

$$x_{i+1} = a \cdot x_i, \qquad (\mathrm{mod}\, M)$$

$$\xi_{i+1} = \frac{x_{i+1}}{M}, \qquad i = 1, 2, \cdots$$

为了便于在计算机上使用，通常取M $=2^s$ ，其中 s 为计算机中二进制数的最大可能有效位数。

可以取 $x_1=$ 奇数， $a=5^{2k+1}$ ，其中 k 为使 5^{2k+1} 在计算机上所能容纳的最大整数，即 a 为计算机上所能容纳的 5 的最大奇次幂。一般地， $s=32$ 时， $a=5^{13}$ ； $s=48$ ， $a=5^{15}$ 等。

乘同余方法是使用的最多、最广的方法，在计算机上被广泛地使用。

伪随机数序列具有周期性，有容量的限制。例如，上述乘同余法产生的伪随机数序列的最大容量 $\lambda(M)=2^{s-2}$ 。

2. 直接抽样方法

对于任意给定的分布函数 $F(x)$ ，直接抽样方法如下：

$$X_n = \inf_{F(t) \geq \xi_n} t, \quad n = 1, 2, \cdots, N$$

其中， ξ_1 ， ξ_2 ，… ， ξ_N 为随机数序列。为方便起见，将上式简化为：

$$X_F = \inf_{F(t) \geq \xi} t$$

若不加特殊说明，今后将总用这种类似的简化形式表示， ξ 总表示随机数。

➢ 证明

下面证明用前面介绍的方法所确定的随机变量序列 X_1 ，X_2 ，…，X_N 具有相同分布 $F(x)$ 。

$$F_{X_n}(x) = P(X_n < x) = P(\inf_{F(t)\geq\xi_n} t < x)$$

$$= P(\xi_n < F(x)) = F(x)$$

对于任意的 n 成立，因此随机变量序列 X_1 ，X_2 ，…，X_N 具有相同分布 $F(x)$ 。另外，由于随机数序列 ξ_1 ，ξ_2 ，…，ξ_N 是相互独立的，因此，由它所确定的 X_1 ，X_2 ，…，X_N 也是相互独立的。

1）离散型分布的直接抽样方法

对于任意离散型分布： $F(x)=\sum_{x_i<x} P_i$

其中 x_1 ， x_2 ，…为离散型分布函数的跳跃点， P_1 ， P_2 ，…为相应的概率，根据前述直接抽样法，有离散型分布的直接抽样方法如下：

$$X_F=x_I,\quad \text{当}\sum_{\mathrm{i}=1}^{\mathrm{I}-1}\mathrm{P_i}<\xi\le\sum_{\mathrm{i}=1}^{\mathrm{I}}\mathrm{P_i}$$

$$P(X_F=x_I)=P\left(\sum_{\mathrm{i}=1}^{\mathrm{I}-1}\mathrm{P_i}<\xi\le\sum_{\mathrm{i}=1}^{\mathrm{I}}\mathrm{P_i}\right)=P_I$$

$$\left(P(a<\xi\le b)=\int_a^b dx=b-a\right)$$

例 1. 二项分布的抽样

二项分布为离散型分布，其概率函数为：

$$P(x=n)=P_n=C_N^n P^n (1-P)^{N-n}$$

其中， P 为概率。对该分布的直接抽样方法如下：

$$X_F=n, \quad 当\sum_{i=0}^{n-1} P_i < \xi \le \sum_{i=0}^{n} P_i$$

例 2. 泊松 (Possion) 分布的抽样

泊松 (Possion) 分布为离散型分布，其概率函数为：

$$P(x=n)=P_n=e^{-\lambda}\frac{\lambda^n}{n!}$$

其中，λ>0 。对该分布的直接抽样方法如下：

$$X_F=n, \qquad 当\sum_{i=0}^{n-1}\frac{\lambda^i}{i!}<\xi\cdot e^{\lambda}\le\sum_{i=0}^{n}\frac{\lambda^i}{i!}$$

例 3. 掷骰子点数的抽样

掷骰子点数 $X=n$ 的概率为：

$$P(X=n)=\frac{1}{6}$$

选取随机数 ξ，如

$$\frac{n-1}{6}<\xi\leq\frac{n}{6}$$

则

$$X_F=n$$

在等概率的情况下，可使用如下更简单的方法：

$$X_F=[6\cdot\xi]+1$$

其中 [] 表示取整数。

例 4. 碰撞核种类的确定

中子或光子在介质中发生碰撞时，如介质是由多种元素组成，需要确定碰撞核的种类。假定介质中每种核的宏观总截面分别为 Σ_1 ，Σ_2 ，…，Σ_n，则中子或光子与每种核碰撞的概率分别为：

$$P_i = \frac{\Sigma_i}{\Sigma_t} \quad i = 1,2,\cdots,n$$

其中 $\Sigma_t = \Sigma_1 + \Sigma_2 + \ldots + \Sigma_n$。碰撞核种类的确定方法为：产生一个随机数 ξ，如果

$$\sum_{i=1}^{I-1} P_i < \xi \leq \sum_{i=1}^{I} P_i$$

则中子或光子与第 I 种核发生碰撞。

例 5. 中子与核的反应类型的确定

假设中子与核的反应类型有如下几种：弹性散射，非弹性散射，裂变，吸收，相应的反应截面分别为 Σ_{el}，Σ_{in}，Σ_f，Σ_a。则发生每一种反应类型的概率依次为：

$$P_{el}=\frac{\Sigma_{el}}{\Sigma_t} \qquad P_{in}=\frac{\Sigma_{in}}{\Sigma_t} \qquad P_f=\frac{\Sigma_f}{\Sigma_t} \qquad P_a=\frac{\Sigma_a}{\Sigma_t}$$

其中反应总截面 $\Sigma_t = \Sigma_{el} + \Sigma_{in} + \Sigma_f + \Sigma_a$。

反应类型的确定方法为：产生一个随机数 ξ

$$\xi \le P_{el} \xrightarrow{\le} \text{弹性散射}$$

$$\downarrow >$$

$$\xi \le P_{el} + P_{in} \xrightarrow{\le} \text{非弹性散射}$$

$$\downarrow >$$

$$\xi \le P_{el} + P_{in} + P_f \xrightarrow{\le} \text{裂变}$$

$$\downarrow >$$

吸收

2) 连续型分布的直接抽样方法

对于连续型分布，如果分布函数 $F(x)$ 的反函数 $F^{-1}(x)$ 存在，则直接抽样方法是：

$$X_F = F^{-1}(\xi)$$

$$F_{X_F}(x) = P(X_F < x) = P(F^{-1}(\xi) < x)$$
$$= P(\xi < F(x)) = F(x)$$

例 6. 在［a，b］上均匀分布的抽样

在［a，b］上均匀分布的分布函数为：

$$F(x)=\begin{cases}0 & \text{当} x<a \\ \dfrac{x-a}{b-a} & \text{当} a\le x\le b \\ 1 & \text{当} x>b\end{cases}$$

则

$$X_F=a+(b-a)\cdot\xi$$

例 7. β 分布

β 分布为连续型分布，作为它的一个特例是：

$$f(x) = 2x, \qquad 0 \le x \le 1$$

其分布函数为：

$$F(x) = \int_{\infty}^{x} f(t)dt = \int_{0}^{x} 2tdt = x^2, \qquad 0 \le x \le 1$$

则

$$X_F = \sqrt{\xi}$$

例 8. 指数分布

指数分布为连续型分布，其一般形式如下：

$$f(x) = a \cdot e^{-ax}, \qquad x \geq 0$$

其分布函数为：

$$F(x) = \int_{-\infty}^{x} f(t)dt = \int_{0}^{x} a \cdot e^{-at}dt = 1 - e^{-ax}, \qquad x \geq 0$$

则

$$X_F = -\frac{1}{a}\ln(1-\xi)$$

因为 1 － ξ 也是随机数，可将上式简化为

$$X_F = -\frac{1}{a}\ln\xi$$

连续性分布函数的直接抽样方法对于分布函数的反函数存在且容易实现的情况，使用起来是很方便的。但是对于以下几种情况，直接抽样法是不合适的。

1) 分布函数无法用解析形式给出，因而其反函数也无法给出。
2) 分布函数可以给出其解析形式，但是反函数给不出来。
3) 分布函数即使能够给出反函数，但运算量很大。

下面叙述的挑选抽样方法是克服这些困难的比较好的方法。

3. 挑选抽样方法

为了实现从已知分布密度函数 $f(x)$ 抽样，选取与 $f(x)$ 取值范围相同的分布密度函数 $h(x)$ ，如果

$$M = \sup_{-\infty < x < \infty} \frac{f(x)}{h(x)} < \infty$$

则挑选抽样方法为：

$$\xi \le \frac{f(X_h)}{M \cdot h(X_h)} \quad >$$

$$\downarrow \le$$

$$X_f = X_h$$

即从 $h(x)$ 中抽样 x_h，以 $\frac{f(x_h)}{M\cdot h(x_h)}$ 的概率接受它。下面证明 x_f 服从分布密度函数 $f(x)$。

证明：对于任意 x

$$P(x \le X_f < x+dx) = P\left(x \le X_h < x+dx \middle/ \xi \le \frac{f(X_h)}{M\cdot h(X_h)} \right)$$

$$= \frac{P\left(x \le X_h < x+dx, \xi \le \frac{f(X_h)}{M\cdot h(X_h)} \right)}{P\left(\xi \le \frac{f(X_h)}{M\cdot h(X_h)} \right)}$$

$$= \frac{\int_{x}^{x+dx} \int_{0}^{\frac{f(X_h)}{M \cdot h(X_h)}} h(X_h) dX_h \cdot d\xi}{\int_{-\infty}^{\infty} \int_{0}^{\frac{f(X_h)}{M \cdot h(X_h)}} h(X_h) dX_h \cdot d\xi}$$

$$= \frac{\int_{x}^{x+dx} \frac{f(X_h)}{M \cdot h(X_h)} \cdot h(X_h) dX_h}{\int_{-\infty}^{\infty} \frac{f(X_h)}{M \cdot h(X_h)} \cdot h(X_h) dX_h}$$

$$= \frac{\int_{x}^{x+dx} f(X_h) dX_h}{\int_{-\infty}^{\infty} f(X_h) dX_h} = f(x) dx$$

使用挑选抽样方法时，要注意以下两点：选取 $h(x)$ 时要使得 $h(x)$ 容易抽样且 M 的值要尽量小。因为 M 小能提高抽样效率。抽样效率是指在挑选抽样方法中进行挑选时被选中的概率。按此定义，该方法的抽样效率 E 为：

$$E = P\left(\xi \le \frac{f(X_h)}{M \cdot h(X_h)}\right)$$

$$= \int_{\infty}^{\infty} \frac{f(X_h)}{M \cdot h(X_h)} \cdot h(X_h) dX_h = \frac{1}{M}$$

所以，M 越小，抽样效率越高。

当 $f(x)$ 在［0，1］上定义时，取 $h(x)=1$，$X_h=\xi$，$M=\sup\limits_{0\le x\le 1} f(x)$

此时挑选抽样方法为

$$\xi' \le \frac{f(\xi)}{M} \quad >$$

$$\downarrow \ \le$$

$$X_f=\xi$$

例 9. 圆内均匀分布抽样

令圆半径为 R_0，点到圆心的距离为 r，则 r 的分布密度函数为

$$f(r)=\begin{cases}\dfrac{2r}{R_0^2} & 当 0\le r\le R_0\\ 0 & 其它\end{cases}$$

分布函数为

$$F(r)=\frac{r^2}{R_0^2}$$

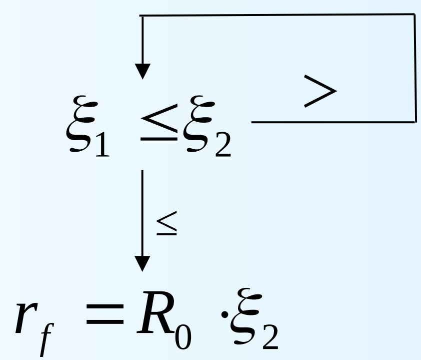
容易知道，该分布的直接抽样方法是

$$r_f=R_0\cdot\sqrt{\xi}$$

由于开方运算在计算机上很费时间，该方法不是好方法。下面使用挑选抽样方法：取

$$h(r)=\frac{1}{R_0}，\quad \frac{f(r)}{h(r)}=\frac{2r}{R_0}，\quad M=2，\quad r_h=R_0\cdot\xi$$

则抽样框图为



显然，没有必要舍弃 $\xi_1 > \xi_2$ 的情况，此时，只需取 $r_f = R_0 \cdot \xi_1$ 就可以了，亦即

$$r_f = R_0 \cdot \max(\xi_1, \xi_2)$$

另一方面，也可证明 $\sqrt{\xi}$ 与 $\max(\xi_1, \xi_2)$ 具有相同的分布 $F(r) = r^2$。

4. 复合抽样方法

在实际问题中，经常有这样的随机变量，它服从的分布与一个参数有关，而该参数也是一个服从确定分布的随机变量，称这样的随机变量服从复合分布。例如，分布密度函数

$$f(x)=\sum_{n=1}^{\infty} P_n \cdot f_n(x)$$

是一个复合分布。其中 $P_n \geq 0$ ，$n=1$ ，2 ，...，$\sum_{n=1}^{\infty} P_n = 1$

$f_n(x)$ 为与参数 n 有关的分布密度函数， $n=1$ ，2 ，...，

参数 n 服从如下分布 $F(y)=\sum_{n<y} P_n$

复合分布的一般形式为：

$$f(x) = \int f_2(x/y)dF_1(y)$$

其中 $f_2(x/y)$ 表示与参数 y 有关的条件分布密度函数，$F_1(y)$ 表示分布函数。

复合分布的抽样方法为：首先由分布函数 $F_1(y)$ 或分布密度函数 $f_1(y)$ 中抽样 Y_{F_1} 或 Y_{f_1}，然后再由分布密度函数 $f_2(x/Y_{F_1})$ 中抽样确定 $X_{f_2(x/Y_F)}$

$$X_f = X_{f_2(x/Y_{F_1})}$$

证明：

$$p(x \le X_f < x + dx) = p(x \le X_{f_2(x/Y_{F_1})} < x + dx)$$
$$= \int f_2(x/Y)dxdF_1(Y) = f(x)dx$$

所以，X_f所服从的分布为 $f(x)$。

例 10. 指数函数分布的抽样

指数函数分布的一般形式为：

$$E_n(x)=\begin{cases} n\int^{\infty}\dfrac{e^{-xy}}{y^n}dy & 当 x\geq 0 \\ 0 & 其它 \end{cases}$$

引入如下两个分布密度函数：

$$f_1(y)=\begin{cases} n\cdot y^{-n-1} & 当 y\geq 1 \\ 0 & 其它 \end{cases}$$

$$f_2(x/y)=\begin{cases} y\cdot e^{-xy} & 当 x\geq 0 \\ 0 & 其它 \end{cases}$$

则

$$E_n(x) = \int_1^{\infty} f_2(x/y) f_1(y) dy$$

使用复合抽样方法，首先从 $f_1(y)$ 中抽取 y

$$Y_{f_1} = \frac{1}{\sqrt[n]{\xi}} = \frac{1}{\max(\xi_1, \xi_2, \cdots, \xi_n)}$$

再由 $f_2(x/Y_{F_1})$ 中抽取 x

$$X_f = \frac{-\ln \xi_{n+1}}{Y_{f_1}}$$

$$= -\max(\xi_1, \xi_2, \cdots, \xi_n) \cdot \ln \xi_{n+1}$$

5. 复合挑选抽样方法

考虑另一种形式的复合分布如下：

$$f(x)=\int H(x,y)\,f_2(x/y)\,dF_1(y)$$

其中 $0\le H(x,y)\le M$ ，$f_2(x/y)$ 表示与参数 y 有关的条件分布密度函数，$F_1(y)$ 表示分布函数。抽样方法如下：

$$\xi \le \frac{H(X_{f_2(x/Y_{F_1})},\,Y_{F_1})}{M} \quad >$$

$$\downarrow \le$$

$$X_f = X_{f_2(x/Y_{F_1})}$$

证明：

$$P(x \le X_f < x+dx) = P\left(x \le X_{f_2} < x+dx \middle/ \xi \le \frac{H(X_{f_2}, Y_{F_1})}{M} \right)$$

$$= \frac{P\left(x \le X_{f_2} < x+dx, \xi \le \frac{H(X_{f_2}, Y_{F_1})}{M} \right)}{P\left(\xi \le \frac{H(X_{f_2}, Y_{F_1})}{M} \right)}$$

$$= \frac{\int_x^{x+dx} \int_{\infty}^{\infty} \int_0^{\frac{H(x,y)}{M}} f_2(x/y)dxdF_1(y)d\xi}{\int_{\infty}^{\infty} \int_{\infty}^{\infty} \int_0^{\frac{H(x,y)}{M}} f_2(x/y)dxdF_1(y)d\xi}$$

$$= \frac{\int_x^{x+dx} \int_{\infty}^{\infty} \frac{H(x,y)}{M} f_2(x/y)dxdF_1(y)}{\int_{\infty}^{\infty} \int_{\infty}^{\infty} \frac{H(x,y)}{M} f_2(x/y)dxdF_1(y)}$$

$$= \int_{\infty}^{\infty} H(x,y) f_2(x/y)dF_1(y)dx = f(x)dx$$

抽样效率为： $E=1/M$

6. 替换抽样方法

为了实现某个复杂的随机变量 y 的抽样，将其表示成若干个简单的随机变量 x_1 ，x_2 ，… ，x_n 的函数

$$y = g(x_1, x_2, \cdots, x_n)$$

得到 x_1 ，x_2 ，… ，x_n 的抽样后，即可确定 y 的抽样，这种方法叫作替换法抽样。即

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

例 11. 散射方位角余弦分布的抽样

散射方位角 φ 在［0,2π］上均匀分布，则其正弦和余弦 $\sin\varphi$ 和 $\cos\varphi$ 服从如下分布：

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

直接抽样方法为：

$$\sin\varphi=\sin 2\pi\xi$$

$$\cos\varphi=\cos 2\pi\xi$$

令 $\varphi=2\theta$，则 θ 在［$0,\pi$］上均匀分布，作变换

$$x=\rho\cos\theta$$
$$y=\rho\sin\theta$$

其中 $0\le\rho\le1$ ，$0\le\theta\le\pi$，则

$$\cos\theta=\frac{x}{\sqrt{x^2+y^2}}$$

$$\sin\theta=\frac{y}{\sqrt{x^2+y^2}}$$

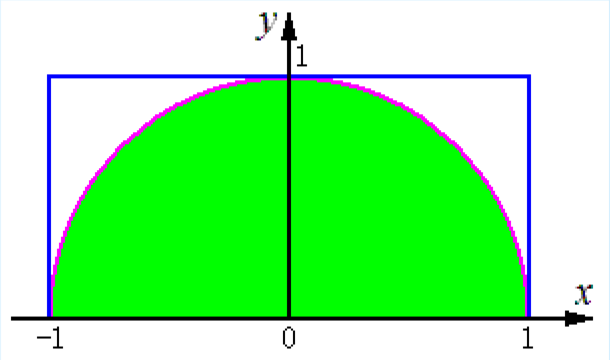
(x,y) 表示上半个单位圆内的点。如果 (x,y) 在上半个单位圆内均匀分布，则 θ 在［$0,\pi$］上均匀分布，由于

$$\cos\varphi=\cos2\theta=\cos^2\theta-\sin^2\theta=\frac{x^2-y^2}{x^2+y^2}$$

$$\sin\varphi=\sin2\theta=2\sin\theta\cos\theta=\frac{2xy}{x^2+y^2}$$

因此抽样 sinφ 和 cosφ 的问题就变成在上半个单位圆内均匀抽样 (x,y) 的问题。

为获得上半个单位圆内的均匀点，采用挑选法，在上半个单位圆的外切矩形内均匀投点（如图）。



$$x = \eta_1 \quad y = \xi_2$$

舍弃圆外的点，余下的就是所要求的点。

抽样方法为：

$$\eta_1^2 + \xi_2^2 \le 1 \quad >$$

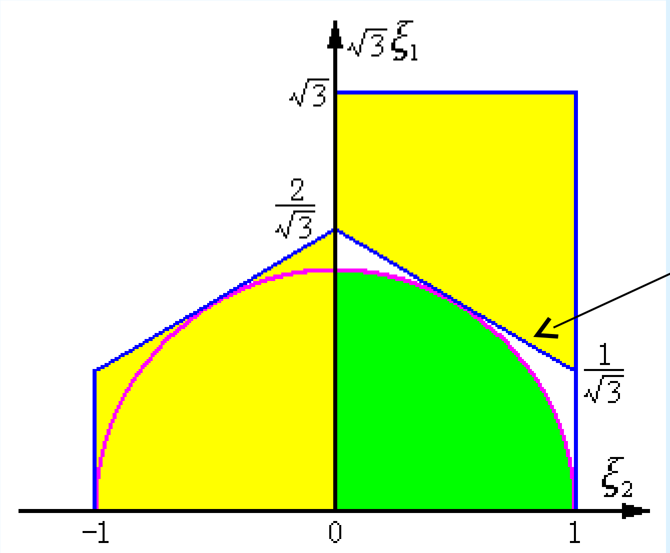
$$\downarrow \le$$

$$\cos\varphi = \frac{\eta_1^2 - \xi_2^2}{\eta_1^2 + \xi_2^2}, \sin\varphi = \frac{2\eta_1\xi_2}{\eta_1^2 + \xi_2^2}$$

抽样效率

E=π/4≈0.785

为实现散射方位角余弦分布抽样，最重要的是在上半个单位圆内产生均匀分布点。下面这种方法，首先在单位圆的半个外切正六边形内产生均匀分布点，如图所示。



$$\sqrt{3}\xi_1 = \frac{2-\xi_2}{\sqrt{3}}$$

$$3\xi_1 = 2-\xi_2$$

于是便有了抽样效率更高的抽样方法：

$$3\xi_1 \le 2-\xi_2$$

$$\downarrow \; >$$

$$\xi_1 = 1-\xi_1,\ \xi_2 = \xi_2 - 1$$

$$\downarrow$$

$$3\xi_1^2+\xi_2^2 \le 1$$

$$\downarrow \; \le$$

$$\cos\varphi = \frac{3\xi_1^2-\xi_2^2}{3\xi_1^2+\xi_2^2},\ \sin\varphi = \frac{2\sqrt{3}\xi_1\xi_2}{3\xi_1^2+\xi_2^2}$$

抽样效率

$$E = \pi/2\sqrt{3} \approx 0.906$$

例 12. 正态分布的抽样

标准正态分布密度函数为：

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

引入一个与标准正态随机变量 X 独立同分布的随机变量 Y，则（X,Y）的联合分布密度为：

$$f(x,y)=\frac{1}{2\pi}e^{-(x^2+y^2)/2}$$

作变换

$$x=\rho\cos\varphi$$
$$y=\rho\sin\varphi$$

则（ρ,φ）的联合分布密度函数为：

$$f(\rho,\varphi)=\frac{\rho}{2\pi}e^{-\rho^2/2}$$

由此可知，ρ 与 φ 相互独立，其分布密度函数分别为

$$f_1(\rho)=\rho\cdot e^{-\rho^2/2}$$

$$f_2(\varphi)=\frac{1}{2\pi}$$

分别抽取 ρ ，φ ：

$$\rho=\sqrt{-2\ln\xi_1}$$

$$\varphi=2\pi\xi_2$$

从而得到一对服从标准正态分布的随机变量 X 和 Y：

$$X_f = \sqrt{-2\ln\xi_1}\cdot\cos(2\pi\xi_2)$$

$$Y_f = \sqrt{-2\ln\xi_1}\cdot\sin(2\pi\xi_2)$$

对于一般的正态分布密度函数 $N(\mu,\sigma^2)$ 的抽样，其抽样结果为：

$$\tilde{X}_f = \mu + \sigma\cdot X_f$$

$$\tilde{Y}_f = \mu + \sigma\cdot Y_f$$

例 13. β 分布的抽样

β 分布密度函数的一般形式为：

$$f(x)=\frac{n!}{(k-1)!(n-k)!}x^{k-1}(1-x)^{n-k} \qquad 0\le x\le 1$$

其中 n ，k 为整数。为了实现 β 分布的抽样，将其看作一组简单的相互独立随机变量的函数，通过这些简单随机变量的抽样，实现 β 分布的抽样。设 x_1 ，x_2 ，… ，x_n 为一组相互独立、具有相同分布 $F(x)$ 的随机变量，ζ_k 为 x_1 ，x_2 ，… ，x_n 按大小顺序排列后的第 k 个，记为：

$$\zeta_k = R_k(x_1,x_2,\cdots,x_n)$$

则 ζ_k 的分布函数为：

$$F_{\zeta_k}(x) = \sum_{i=k}^{n} C_n^i \cdot \left[F(x)\right]^i \cdot \left[1 - F(x)\right]^{n-i}$$

当 $F(x)=x$ 时，

$$F_{\zeta_k}(x) = \sum_{i=k}^{n} C_n^i \cdot x^i \cdot (1 - x)^{n-i}$$

不难验证，ζ_k 的分布密度函数为 β 分布。因此，β 分布的抽样可用如下方法实现：

选取 n 个随机数，按大小顺序排列后取第 k 个，即

$$X_f = R_k(\xi_1, \xi_2, \cdots, \xi_n)$$

7. 随机抽样的一般方法

1) 加抽样方法
2) 减抽样方法
3) 乘抽样方法
4) 乘加抽样方法
5) 乘减抽样方法
6) 对称抽样方法
7) 积分抽样方法

1）加抽样方法

加抽样方法是对如下加分布给出的一种抽样方法：

$$f(x)=\sum_{n=1}^{\infty}P_n\cdot f_n(x)$$

$$\sum_{n=1}^{\infty}P_n=1$$

其中 $P_n\geq 0$，且 $f_n(x)$ 为与参数 n 有关的分布密度函数，$n=1$，2，…。

由复合分布抽样方法可知，加分布的抽样方法为：首先抽样确定 n'，然后由 $f_{n'}(x)$ 中抽样 x，即：

$$X_f=X_{f_{n'}},\quad 当\sum_{n=1}^{n'-1}P_n<\xi\leq\sum_{n=1}^{n'}P_n$$

例 14. 多项式分布抽样

多项式分布密度函数的一般形式为：

$$f(x)=\sum_{i=0}^{\infty}a_i x^i$$

将 $f(x)$ 改写成如下形式：

$$f(x)=\sum_{i=0}^{\infty}\frac{a_i}{i+1}\cdot(i+1)x^i=\sum_{i=0}^{\infty}P_i\cdot f_i(x)$$

则该分布的抽样方法为：

$$X_f=\max(\xi_1,\cdots,\xi_{n+1}),\quad 当\sum_{i=0}^{n-1}\mathrm{P_i}<\xi\le\sum_{i=0}^{n}\mathrm{P_i}$$

例 15. 球壳内均匀分布抽样

设球壳内半径为 R_0，外半径为 R_1，点到球心的距离为 r，则 r 的分布密度函数为

$$f(r)=\begin{cases}\dfrac{3r^2}{R_1^3-R_0^3} & \text{当} R_0\le r\le R_1 \\ 0 & \text{其它}\end{cases}$$

分布函数为

$$F(r)=\frac{r^3-R_0^3}{R_1^3-R_0^3}$$

该分布的直接抽样方法是

$$r_f=\left[(R_1^3-R_0^3)\xi+R_0^3\right]^{1/3}$$

为避免开立方根运算，作变换：

$$r = (R_1 - R_0) \cdot x + R_0$$

则 $x \in [0,1]$，其分布密度函数为：

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

其中

$$\lambda = R_0^2 + R_0 R_1 + R_1^2$$

则 x 及 r 的抽样方法为：

$$\xi_1 \le \frac{3R_0^2}{\lambda}$$

$\le$：$X_f = \xi_2$

$>$：

$$\xi_1 \le \frac{3R_0R_1}{\lambda}$$

$\le$：$X_f = \max(\xi_2, \xi_3)$

$>$：$X_f = \max(\xi_2, \xi_3, \xi_4)$

$$r_f = (R_1 - R_0)\cdot X_f + R_0$$

2) 减抽样方法

减抽样方法是对如下形式的分布密度所给出的一种抽样方法：

$$f(x) = A_1 f_1(x) - A_2 f_2(x)$$

其中 A_1 、 A_2 为非负实数， $f_1(x)$ 、 $f_2(x)$ 均为分布密度函数。

减抽样方法分为以下两种形式：

以上两种形式的抽样方法，究竟选择哪种好，要看 $f_1(x)$ 、 $f_2(x)$ 哪一个容易抽样，如相差不多，选用第一种方法抽样效率高。

（1）将 $f(x)$ 表示为

$$f(x)=f_1(x)\left[A_1 - A_2\frac{f_2(x)}{f_1(x)}\right]$$

令 m 表示 $f_2(x)$ ／ $f_1(x)$ 的下界，使用挑选法，从 $f_1(x)$ 中抽取 X_{f_1}

$$\xi \le \frac{A_1}{A_1 - mA_2} - \frac{A_2}{A_1 - mA_2}\frac{f_2(X_{f_1})}{f_1(X_{f_1})} \quad >$$

$$\downarrow \le$$

$$X_f = X_{f_1}$$

抽样效率为：

$$E=\frac{1}{A_1 - mA_2}$$

（2）将 $f(x)$ 表示为

$$f(x) = f_2(x)\left[A_1 \frac{f_1(x)}{f_2(x)} - A_2\right]$$

使用挑选法，从 $f_2(x)$ 中抽取 X_{f_2}

$$\xi \le \frac{mA_1}{A_1 - mA_2}\,\frac{f_1(X_{f_2})}{f_2(X_{f_2})} - \frac{mA_2}{A_1 - mA_2} \quad >$$

$$\downarrow \le$$

$$X_f = X_{f_2}$$

抽样效率为：

$$E' = \frac{m}{A_1 - mA_2} = mE$$

例 16. β 分布抽样

β 分布的一个特例：

$$f(x) = 2(1 - x), \quad 0 \le x \le 1$$

取 $A_1 = 2$ ，$A_2 = 1$ ，$f_1(x) = 1$ ，$f_2(x) = 2x$ ，此时 $m = 0$ ，则根据第一种形式的减抽样方法，有

$$\xi_1 \le 1 - \xi_2 \quad (>) \qquad \xrightarrow{\le} \qquad X_f = \xi_2$$

或

$$\xi_2 \le 1 - \xi_1 \quad (>) \qquad \xrightarrow{\le} \qquad X_f = \xi_2$$

由于 $1-\xi_1$ 可用 ξ_1 代替，该抽样方法可简化为：

$$\begin{tikzcd} \xi_2 \le \xi_1 \arrow[loop right, ">"] \arrow[d, "\le"] \\ X_f = \xi_2 \end{tikzcd}$$

对于 $\xi_2 > \xi_1$ 的情况，可取 $X_f = \xi_1$ ，因此

$$X_f = \min(\xi_1, \xi_2)$$

与 β 分布的推论相同。

3) 乘抽样方法

如下形式的分布称为乘分布：

$$f(x)=H(x)f_1(x)$$

其中 $H(x)$ 为非负函数， $f_1(x)$ 为任意分布密度函数。

令 M 为 $H(x)$ 的上界，乘抽样方法如下：

$$\xi \le \frac{H(X_{f_1})}{M} \quad >$$

$\le$

$$X_f = X_{f_1}$$

抽样效率为：$E=\dfrac{1}{M}$

例 17. 倒数分布抽样

倒数分布密度函数为：

$$f(x)=\frac{1}{\ln a}\cdot\frac{1}{x},\quad 1\le x\le a$$

其直接抽样方法为：

$$X_f=a^{\xi}=e^{\ln a\,\xi}$$

下面采用乘抽样方法，考虑如下分布族：

$$f_i(x)=\frac{1}{i(a^{1/i}-1)}\frac{x^{1/i}}{x},\quad 1\le x\le a$$

其中 $i=1$，2，…，该分布的直接抽样方法为：

$$X_{f_i}=[(a^{1/i}-1)\xi+1]^i$$

利用这一分布族，将倒数分布 $f(x)$ 表示成：

$$f(x)=H(x)f_i(x)$$

其中，$H(x)=\dfrac{i(a^{1/i}-1)}{\ln a\cdot x^{1/i}}$， $M=\dfrac{i(a^{1/i}-1)}{\ln a}$， $\dfrac{H(x)}{M}=\dfrac{1}{x^{1/i}}$，

乘法分布的抽样方法如下：

$$\begin{tikzcd}
\xi_1[(a^{1/i}-1)\xi_2+1]\le 1 \arrow[loop right, ">"] \arrow[d, "\le"] \\
X_f=[(a^{1/i}-1)\xi_2+1]^i
\end{tikzcd}$$

该分布的抽样效率为： $E=\dfrac{\ln a}{i(a^{1/i}-1)}$

例 18. 麦克斯韦 (Maxwell) 分布抽样

麦克斯韦分布密度函数的一般形式为：

$$f(x)=\frac{2\beta^{3/2}}{\sqrt{\pi}}\sqrt{x}\cdot e^{-\beta\cdot x},\ x\geq 0$$

使用乘抽样方法，令

$$f_1(x)=\frac{2\beta}{3}e^{-\frac{2}{3}\beta\cdot x},\quad x\geq 0$$

该分布的直接抽样方法为：

$$X_{f_1}=-\frac{3}{2\beta}\ln\xi_2$$

此时

$$H(x)=\frac{3\beta^{1/2}}{\sqrt{\pi}}\sqrt{x}\cdot e^{-\frac{1}{3}\beta\cdot x},\quad x\ge 0$$

$$M=\sqrt{\frac{27}{2\pi\cdot e}}$$

则麦克斯韦分布的抽样方法为：

$$\begin{tikzcd} \xi_1^2 \le -e\xi_2\cdot\ln\xi_2 \arrow[loop right, ">"] \arrow[d, "\le"] \\ X_f=-\frac{3}{2\beta}\ln\xi_2 \end{tikzcd}$$

该分布的抽样效率为：

$$E=\sqrt{\frac{2\pi\cdot e}{27}}\approx 0.795$$

4) 乘加抽样方法

在实际问题中，经常会遇到如下形式的分布：

$$f(x)=\sum_{n=1}^{\infty}H_n(x)f_n(x)$$

其中 $H_n(x)$ 为非负函数， $f_n(x)$ 为任意分布密度函数， n=1 ，2 ，…。不失一般性，只考虑 n=2 的情况：

$$f(x)=H_1(x)f_1(x)+H_2(x)f_2(x)$$

将 $f(x)$ 改写成如下的加分布形式：

$$f(x)=P_1\frac{H_1(x)}{P_1}f_1(x)+P_2\frac{H_2(x)}{P_2}f_2(x)$$
$$=P_1f_1^*(x)+P_2f_2^*(x)$$

其中

$$P_1 = \int H_1(x) f_1(x) dx$$
$$P_2 = \int H_2(x) f_2(x) dx$$
$$f_1^*(x) = \frac{H_1(x)}{P_1} f_1(x)$$
$$f_2^*(x) = \frac{H_2(x)}{P_2} f_2(x)$$

乘加抽样方法为：

$$\xi_1 \le P_1$$

$$\xi_1 \le P_1 \xrightarrow{\le} \xi_2 \le \frac{H_1(X_{f_1})}{M_1} \qquad\qquad \xi_1 \le P_1 \xrightarrow{>} \xi_2 \le \frac{H_2(X_{f_2})}{M_2}$$

$$\xi_2 \le \frac{H_1(X_{f_1})}{M_1} \xrightarrow{>} \xi_2 \le \frac{H_1(X_{f_1})}{M_1} \qquad\qquad \xi_2 \le \frac{H_2(X_{f_2})}{M_2} \xrightarrow{>} \xi_2 \le \frac{H_2(X_{f_2})}{M_2}$$

$$\xi_2 \le \frac{H_1(X_{f_1})}{M_1} \xrightarrow{\le} X_f = X_{f_1} \qquad\qquad \xi_2 \le \frac{H_2(X_{f_2})}{M_2} \xrightarrow{\le} X_f = X_{f_2}$$

该方法的抽样效率为：

$$E_1 = P_1 \cdot \frac{P_1}{M_1} + P_2 \cdot \frac{P_2}{M_2} = \frac{P_1^2}{M_1} + \frac{P_2^2}{M_2}$$

这种方法需要知道 P_1 的值 ($P_2=1-P_1$)，这对有些分布是很困难的。下面的方法可以不用计算 P_1：

对于任意小于 1 的正数 P_1，令 $P_2=1-P_1$；

$$H(x,y)=\begin{cases}\dfrac{H_1(x)}{P_1}, & 1<y\le 2\\[2ex] \dfrac{H_2(x)}{P_2} & y>2\end{cases}$$

则采用复合挑选抽样方法，有：

$$M=\max\left(\frac{M_1}{P_1},\frac{M_2}{P_2}\right)$$

$$f_2(x/y)=\begin{cases}f_1(x), & 1<y\le 2\\ f_2(x), & y>2\end{cases}$$

$$E=\min\left(\frac{P_1}{M_1},\frac{P_2}{M_2}\right)$$

$$F_1(y)=\sum_{n<y}P_n$$

当取

$$P_1=\frac{M_1}{M_1+M_2}\qquad P_2=\frac{M_2}{M_1+M_2}$$

时，抽样效率最高

$$E_2=\frac{1}{M_1+M_2}$$

这时，乘加抽样方法为：

$\le$ $\quad$ $>$

$$\xi_1\le\frac{M_1}{M_1+M_2}$$

$>$ $\quad$ $>$

$$\xi_2\le\frac{H_1(X_{f_1})}{M_1}\qquad\qquad \xi_2\le\frac{H_2(X_{f_2})}{M_2}$$

$\le$ $\quad$ $\le$

$$X_f=X_{f_1}\qquad\qquad X_f=X_{f_2}$$

由于

$$
\begin{aligned}
E_1 - E_2 &= \frac{P_1^2}{M_1} + \frac{P_2^2}{M_2} - \frac{1}{M_1 + M_2} \\
&= \frac{M_2(M_1 + M_2)P_1^2 + M_1(M_1 + M_2)P_2^2 - M_1M_2}{M_1M_2(M_1 + M_2)} \\
&= \frac{M_2^2P_1^2 + M_1^2P_2^2 + M_1M_2(P_1^2 + P_2^2) - M_1M_2}{M_1M_2(M_1 + M_2)} \\
&= \frac{M_2^2P_1^2 + M_1^2P_2^2 - 2M_1M_2P_1P_2}{M_1M_2(M_1 + M_2)} \\
&= \frac{(M_2P_1 - M_1P_2)^2}{M_1M_2(M_1 + M_2)} \geq 0
\end{aligned}
$$

可知第一种方法比第二种方法的抽样效率高。

例 19. 光子散射后能量分布的抽样

令光子散射前后的能量分别为 α 和 α'（以 m_0c^2 为单位，m_0 为电子静止质量，c 为光速），$x=\alpha/\alpha'$，则 x 的分布密度函数为：

$$f(x/\alpha)=\frac{1}{K(\alpha)}\left[\left(\frac{\alpha+1-x}{\alpha\cdot x}\right)^2+\frac{1}{x}-\frac{1}{x^2}+\frac{1}{x^3}\right],\quad 1\le x\le 1+2\alpha$$

该分布即为光子散射能量分布，它是由著名的 *Klin – Nishina* 公式确定的。其中 $K(\alpha)$ 为归一因子：

$$K(\alpha)=\left[1-\frac{2(\alpha+1)}{\alpha^2}\right]\ln(1+2\alpha)+\frac{1}{2}+\frac{4}{\alpha}-\frac{1}{2(1+2\alpha)^2}$$

把光子散射能量分布改写成如下形式:

$$f(x/\alpha)=\frac{1}{K(\alpha)}\left\{\left[\left(\frac{\alpha+1-x}{\alpha}\right)^2+1\right]\frac{1}{x^2}+\frac{(x-1)^2}{x^3}\right\}$$

在［1, 1+2α］上定义如下函数：

$$f_1(x/\alpha)=\frac{1+2\alpha}{2\alpha}\cdot\frac{1}{x^2}$$

$$f_2(x/\alpha)=\frac{1}{2\alpha}$$

$$H_1(x/\alpha)=\frac{2\alpha}{K(\alpha)(1+2\alpha)}\left[\left(\frac{\alpha+1-x}{\alpha}\right)^2+1\right]$$

$$H_2(x/\alpha)=\frac{2\alpha}{K(\alpha)}\frac{(x-1)^2}{x^3}$$

则有

$$f(x/\alpha) = H_1(x/\alpha) \cdot f_1(x/\alpha) + H_2(x/\alpha) \cdot f_2(x/\alpha)$$

使用乘加抽样方法：

$$X_{f_1} = \frac{1+2\alpha}{1+2\alpha\xi_2}$$

$$X_{f_2} = 1+2\alpha\xi_2$$

$$H_1(x/\alpha) \le = \frac{4\alpha}{K(\alpha)(1+2\alpha)} = M_1$$

$$H_2(x/\alpha) \le \frac{8\alpha}{27 \cdot K(\alpha)} = M_2$$

光子散射能量分布的抽样方法为：

$$\xi_1 \le \frac{27}{4\alpha + 29}$$

≤

$$X_{f_1} = \frac{1+2\alpha}{1+2\alpha\xi_2}$$

$$\xi_3 \le \frac{1}{2}\left[\left(\frac{\alpha+1-X_{f_1}}{\alpha}\right)^2 + 1\right]$$

≤

$$X_f = X_{f_1}$$

>

$$X_{f_2} = 1 + 2\alpha\xi_2$$

$$\xi_3 \le \frac{27}{4}\frac{(X_{f_2}-1)^2}{X_{f_2}^3}$$

≤

$$X_f = X_{f_2}$$

该方法的抽样效率为：

$$E = \frac{1}{M_1 + M_2} = \frac{27(1+2\alpha)K(\alpha)}{4\alpha(4\alpha+29)}$$

5) 乘减抽样方法

乘减分布的形式为：

$$f(x) = H_1(x) f_1(x) - H_2(x) f_2(x)$$

其中 $H_1(x)$ 、 $H_2(x)$ 为非负函数， $f_1(x)$ 、 $f_2(x)$ 为任意分布密度函数。

与减抽样方法类似，乘减分布的抽样方法也分为两种。

（1）将 $f(x)$ 表示为

$$f(x)=f_1(x)H_1(x)\left[1-\frac{H_2(x)f_2(x)}{H_1(x)f_1(x)}\right]$$

令 $H_1(x)$ 的上界为 M_1，$\dfrac{H_2(x)f_2(x)}{H_1(x)f_1(x)}$ 的下界为 m，使用乘抽样方法得到如下乘减抽样方法：

$$\xi \le \frac{1}{M_1(1-m)}\left[H_1(X_{f_1})-\frac{H_2(X_{f_1})f_2(X_{f_1})}{f_1(X_{f_1})}\right] \quad >$$

$$\downarrow \le$$

$$X_f = X_{f_1}$$

（2）将 $f(x)$ 表示为

$$f(x) = f_2(x)H_2(x)\left[\frac{H_1(x)f_1(x)}{H_2(x)f_2(x)} - 1\right]$$

令 $H_2(x)$ 的上界为 M_2，使用乘抽样方法，得到另一种乘减抽样方法：

$$\xi \le \frac{m}{M_2(1-m)}\left[\frac{H_1(X_{f_2})f_1(X_{f_2})}{f_2(X_{f_2})} - H_2(X_{f_2})\right] \xrightarrow{>}$$

$$\downarrow \le$$

$$X_f = X_{f_2}$$

例 20. 裂变中子谱分布抽样

裂变中子谱分布的一般形式为：

$$f(E) = C \cdot e^{-E/A} \cdot \mathrm{sh}\sqrt{BE}, \quad E_{\min} \le E \le E_{\max}$$

其中 A ，B ，C ，$E_{\min}$ ，$E_{\max}$ 均为与元素有关的量。令

$$f_1(E) = f_2(E) = \frac{\lambda}{\gamma} e^{-E/\gamma}, \quad E_{\min} \le E \le E_{\max}$$

其中 λ 为归一因子，γ 为任意参数。

相应的 $H_1(E)$ ，$H_2(E)$ 为：

$$H_1(E)=\frac{C\gamma}{2\lambda}\exp\left\{-\left(\frac{1}{A}-\frac{1}{\gamma}\right)E+\sqrt{BE}\right\}$$

$$H_2(E)=\frac{C\gamma}{2\lambda}\exp\left\{-\left(\frac{1}{A}-\frac{1}{\gamma}\right)E-\sqrt{BE}\right\}$$

于是裂变中子谱分布可以表示成乘减分布形式：

$$f(E)=H_1(E)f_1(E)-H_2(E)f_2(E)$$

容易确定 $H_1(E)$ 的上界为：

$$M_1=\frac{C\gamma}{2\lambda}\exp\left\{\frac{1}{4}\frac{B}{1/A-1/\gamma}\right\}$$

为提高抽样效率，应取 γ 使得 M_1 达到最小，此时

$$\gamma=A\left[1+\frac{AB}{8}\left(\sqrt{1+\frac{16}{AB}}+1\right)\right]$$

取 $m=0$，令 $\alpha=\frac{1}{A}-\frac{1}{\gamma}$, $\beta=\frac{B}{4\alpha}$

则裂变中子谱分布的抽样方法为：

$$E_{f_1}=-\gamma\ln\left[\exp\left(-\frac{E_{\min}}{\gamma}\right)-\frac{\xi_1}{\lambda}\right]$$

$$\downarrow$$

$$(\ln\xi_2+\alpha E_{f_1}+\beta)^2\le BE_{f_1} \quad \xrightarrow{>}$$

$$\downarrow \le$$

$$E_f=E_{f_1}$$

抽样效率

$$E=\frac{1}{M_1}=\frac{2\lambda}{C\gamma}e^{-\beta}$$

6) 对称抽样方法

对称分布的一般形式为：

$$f(x) = f_1(x) + H(x)$$

其中 $f_1(x)$ 为任意分布密度函数，满足偶函数对称条件，$H(x)$ 为任意奇函数，即对任意 x 满足：

$$f_1(x) = f_1(-x)$$
$$H(x) = -H(-x)$$

对称分布的抽样方法如下：取 $\eta=2\xi - 1$

$$\eta \le \frac{H(X_{f_1})}{f_1(X_{f_1})}$$

$\le$：$X_f = X_{f_1}$　　　　$>$：$X_f = -X_{f_1}$

证明：

因为 $\eta=2\xi-1$，$\eta\le x$ 相当于 $\xi \le \frac{1}{2}\left[1+x\right]$，因此

$$P(X_f < x) = P\left(X_f < x, \xi \le \frac{1}{2}\left[1+\frac{H(X_{f_1})}{f_1(X_{f_1})}\right]\right) + P\left(X_f < x, \xi > \frac{1}{2}\left[1+\frac{H(X_{f_1})}{f_1(X_{f_1})}\right]\right)$$

$$= P\left(X_{f_1} < x, \xi \le \frac{1}{2}\left[1+\frac{H(X_{f_1})}{f_1(X_{f_1})}\right]\right) + P\left(- X_{f_1} < x, \xi > \frac{1}{2}\left[1+\frac{H(X_{f_1})}{f_1(X_{f_1})}\right]\right)$$

$$= \int_{\infty}^{x} \frac{1}{2}\left[1+\frac{H(X_{f_1})}{f_1(X_{f_1})}\right] \cdot f_1(X_{f_1}) dX_{f_1} + \int_{x}^{\infty} \frac{1}{2}\left[1-\frac{H(X_{f_1})}{f_1(X_{f_1})}\right] \cdot f_1(X_{f_1}) dX_{f_1}$$

$$= \int_{\infty}^{x} \frac{1}{2}\left[f_1(X_{f_1}) + H(X_{f_1})\right] \cdot dX_{f_1} + \int_{x}^{\infty} \frac{1}{2}\left[f_1(X_{f_1}) - H(X_{f_1})\right] \cdot dX_{f_1}$$

$$= \int_{\infty}^{x} \frac{1}{2}\left[f_1(X_{f_1}) + H(X_{f_1})\right] \cdot dX_{f_1} + \int_{\infty}^{x} \frac{1}{2}\left[f_1(X_{f_1}) + H(X_{f_1})\right] \cdot dX_{f_1}$$

$$= \int_{\infty}^{x} \left[f_1(X_{f_1}) + H(X_{f_1})\right] \cdot dX_{f_1} = \int_{\infty}^{x} f(X_{f_1}) \cdot dX_{f_1}$$

例 21. 质心系各向同性散射角余弦分布抽样

在质心系各向同性散射的假设下，为得到实验室系散射角余弦，需首先抽样确定质心条散射角余弦：

$$f(\mu_C)=\frac{1}{2}, \quad -1 \le \mu_C \le 1$$

$$\mu_C \equiv \cos\theta_C = \eta = 2\xi - 1$$

再利用下面转换公式 :

$$\mu_L \equiv \cos\theta_L = \frac{1+A\mu_C}{\sqrt{1+A^2+2A\mu_C}}$$

得到实验室系散射角余弦 μ_L。其中 A 为碰撞核质量， θ_C、 θ_L 分别为质心系和实验室系散射角。

为避免开方运算，可以使用对称分布抽样。

根据转换公式可得：

$$\mu_C = \frac{1}{A}\left[\mu_L\sqrt{A^2 - 1 + \mu_L^2} + \mu_L^2 - 1\right]$$

依照质心系散射各向同性的假定，可得到实验室系散射角余弦 μ_L 的分布如下：

$$f(\mu_L) = \frac{1}{2A}\left[\frac{A^2 - 1 + 2\mu_L^2}{\sqrt{A^2 - 1 + \mu_L^2}} + 2\mu_L\right], \quad -1 \le \mu_L \le 1$$

该密度函数中的第一项为偶函数，第二项为奇函数，因而是对称分布。其中

$$f_1(\mu_L) = \frac{1}{2A}\frac{A^2 - 1 + 2\mu_L^2}{\sqrt{A^2 - 1 + \mu_L^2}}, \quad -1 \le \mu_L \le 1$$

从 $f_1(\mu_L)$ 的抽样可使用挑选法

$$\xi_2 \le \frac{A^2-1+2\eta_1^2}{\sqrt{A^2-1+\eta_1^2}} \Bigg/ \frac{A^2+1}{A} \qquad >$$

$$\Big\downarrow \le$$

$$\mu_L = \eta_1$$

然后再以

$$\eta \le 2\mu_L \Bigg/ \frac{A^2-1+2\mu_L^2}{\sqrt{A^2-1+\mu_L^2}}$$

的概率决定接受或取负值。

上述公式涉及开方运算，需要进一步简化。

注意以下事实：对于任意 $0\le a\le 1$

$$P(\xi_2 \le a) = P(\eta_2^2 \le a^2) = a$$

令

$$a = \frac{A^2 - 1 + 2\eta_1^2}{\sqrt{A^2 - 1 + \eta_1^2}} \bigg/ \frac{A^2 + 1}{A}$$

则上述挑选抽样中的挑选条件简化为：

$$\left(\frac{A^2 + 1}{A}\eta_2\right)^2 \le \frac{(A^2 - 1 + 2\eta_1^2)^2}{A^2 - 1 + \eta_1^2}$$

另一方面，在$\eta_2^2 \le a^2$即$-1 \le \eta_2/a \le 1$的条件下，η_2/a 在［－1, 1］上均匀分布，故可令 $\eta = \eta_2/a$ ，则最终决定取正负值的条件简化为：

$$\frac{A^2 + 1}{A}\eta_2 \le 2\eta_1$$

于是，得到质心系各向同性散射角余弦分布的抽样方法为：

$$\left(\frac{A^2+1}{A}\eta_2\right)^2 \le \frac{(A^2-1+2\eta_1^2)^2}{A^2-1+\eta_1^2} \qquad >$$

$$\Big\downarrow \le$$

$$\le \qquad \frac{A^2+1}{A}\eta_2 \le 2\eta_1 \qquad >$$

$$\mu_L = \eta_1 \qquad\qquad \mu_L = -\eta_1$$

7）积分抽样方法

如下形式的分布密度函数

$$f(x)=\frac{\int_{\infty}^{H(x)} f_0(x,y)dy}{\int_{\infty}^{\infty}\int_{\infty}^{H(x)} f_0(x,y)dxdy}$$

称为积分分布密度函数，其中 $f_0(x，y)$ 为任意二维分布密度函数， $H(x)$ 为任意函数。该分布密度函数的抽样方法为：

$$Y_{f_0} \le H(X_{f_0}) \quad >$$

$$\downarrow \le$$

$$X_f = X_{f_0}$$

证明：对于任意 x

$$P(x \le X_f < x + dx) = P\left(x \le X_{f_0} < x + dx \big/ Y_{f_0} \le H(X_{f_0})\right)$$

$$= \frac{P\left(x \le X_{f_0} < x + dx, Y_{f_0} \le H(X_{f_0})\right)}{P\left(Y_{f_0} \le H(X_{f_0})\right)}$$

$$= \frac{\int_{\infty}^{H(x)} f_0(x,y)dxdy}{\int_{\infty}^{\infty}\int_{\infty}^{H(x)} f_0(x,y)dxdy} = f(x)dx$$

例 22. 各向同性散射方向的抽样

为了确定各向同性散射方向 $\boldsymbol{\Omega} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$ ，根据公式：

$$u = \sin\theta\cos\varphi$$

$$v = \sin\theta\sin\varphi$$

$$w = \cos\theta$$

对于各向同性散射， $\cos\theta$ 在［－ 1, 1］上均匀分布， φ 在［0, 2π］上均匀分布。由于

$$\sin\theta = \sqrt{1 - \cos^2\theta} = \sqrt{1 - w^2}$$

直接抽样需要计算三角函数和开方。

定义两个随机变量：

$$x = \frac{\xi_1^2 - A^2\eta_2^2 - A^2\eta_3^2}{\xi_1^2 + A^2\eta_2^2 + A^2\eta_3^2}$$

$$y = \xi_1^2 + A^2\eta_2^2 + A^2\eta_3^2$$

可以证明，当 $\eta_2^2 + \eta_3^2 \le 1$ 时，随机变量 x 和 y 服从如下分布：

$$f_0(x, y) = \frac{1}{2\sqrt{2}A^2}\sqrt{\frac{y}{1+x}}$$

定义区域为：

$$-1 \le x \le 1, \quad 0 \le y \le \min\left(\frac{2A^2}{1-x}, \quad \frac{2}{1+x}\right)$$

令 $r^2 = \eta_2^2 + \eta_3^2 \le 1$

则 ***r*** 为单位圆内均匀分布的随机变量， $f(r) = 2r$

$$x = \frac{\xi_1^2 - A^2\eta_2^2 - A^2\eta_3^2}{\xi_1^2 + A^2\eta_2^2 + A^2\eta_3^2} = \frac{\xi_1^2 - A^2 r^2}{\xi_1^2 + A^2 r^2}$$

$$y = \xi_1^2 + A^2\eta_2^2 + A^2\eta_3^2 = \xi_1^2 + A^2 r^2$$

$$\xi_1 = \sqrt{\frac{y(1+x)}{2}}$$

$$r = \sqrt{\frac{y(1-x)}{2A^2}}$$

则

$$f_0(x,y)=f(\xi_1)f(r)\cdot\left\|\begin{matrix}\frac{\partial\xi_1}{\partial x} & \frac{\partial\xi_1}{\partial y}\\ \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y}\end{matrix}\right\|$$

$$=2\sqrt{\frac{y(1-x)}{2A^2}}\cdot\left\|\begin{matrix}\sqrt{\frac{y}{8(1+x)}} & \sqrt{\frac{(1+x)}{8y}}\\ -\sqrt{\frac{y}{8A^2(1-x)}} & \sqrt{\frac{(1-x)}{8A^2y}}\end{matrix}\right\|=\frac{1}{2\sqrt{2}A^2}\sqrt{\frac{y}{1+x}}$$

而

$$y(1+x)=2\xi_1^2\le 2,\quad y(1-x)=2A^2r^2\le 2A^2$$

$$-1\le x\le 1,\quad 0\le y\le\min\left(\frac{2A^2}{1-x},\quad\frac{2}{1+x}\right)$$

令

$$H(x)=(\tfrac{1+x}{2})^{1/3}$$

则

$$f(x)=\frac{\int_{\infty}^{(\frac{1+x}{2})^{1/3}} f_0(x,y)dy}{\int_{\infty}^{\infty}\int_{\infty}^{(\frac{1+x}{2})^{1/3}} f_0(x,y)dxdy}$$

$$=\frac{\int_{\infty}^{(\frac{1+x}{2})^{1/3}} \sqrt{\frac{y}{1+x}}dy}{\int_{1}^{1}\int_{\infty}^{(\frac{1+x}{2})^{1/3}} \sqrt{\frac{y}{1+x}}dxdy}=\frac{1}{2}$$

所以，x 与 $w=\cos\theta$ 同分布。同时，$f(x)$ 又符合积分分布的特征，故可以用积分抽样方法抽取 x。

抽样过程：$y \le H(x)$ 时，$\left(\frac{1+x}{2}\right)^{1/3}$ 接受 x。

即

$$y^3 \le \frac{1+x}{2} \Rightarrow y^4 \le \frac{y(1+x)}{2} = \xi_1^2 \Rightarrow y^2 \le \xi_1$$

而

$$r^2 = \frac{y(1-x)}{2A^2} \le \frac{H(x)(1-x)}{2A^2} = \frac{\left(\frac{1+x}{2}\right)^{1/3}(1-x)}{2A^2} \le \left.\frac{3}{4\cdot\sqrt[3]{4}\cdot A^2}\right|_{x=-\frac{1}{2}}$$

令 $A = \sqrt{3}/\sqrt[3]{16}$，则属于上述积分限内的 y 一定满足

条件 $\eta_2^2 + \eta_3^2 = r^2 \le 1$。

另外，由于点 (η_2,η_3) 在园内均匀分布，令

$$\cos\varphi=\frac{\eta_2}{\sqrt{\eta_2^2+\eta_3^2}}=\frac{\eta_2}{r}$$

$$\sin\varphi=\frac{\eta_3}{\sqrt{\eta_2^2+\eta_3^2}}=\frac{\eta_3}{r}$$

则 φ 在 ［0, 2π］ 上均匀分布。

而

$$\sin\theta = \sqrt{1 - x^2} = \sqrt{1 - \left(\frac{\xi_1^2 - A^2 r^2}{\xi_1^2 + A^2 r^2}\right)^2} = \frac{2A\xi_1 r}{\xi_1^2 + A^2 r^2}$$

所以

$$u = \sin\theta\cos\varphi = \frac{2A\xi_1 r}{\xi_1^2 + A^2 r^2}\cdot\frac{\eta_2}{r} = \frac{2A\xi_1\eta_2}{\xi_1^2 + A^2\eta_2^2 + A^2\eta_3^2}$$

$$v = \sin\theta\sin\varphi = \frac{2A\xi_1 r}{\xi_1^2 + A^2 r^2}\cdot\frac{\eta_3}{r} = \frac{2A\xi_1\eta_3}{\xi_1^2 + A^2\eta_2^2 + A^2\eta_3^2}$$

$$w = \cos\theta = \frac{\xi_1^2 - A^2\eta_2^2 - A^2\eta_3^2}{\xi_1^2 + A^2\eta_2^2 + A^2\eta_3^2}$$

各向同性散射方向的抽样方法为：

$$(\xi_1^2 + A^2\eta_2^2 + A^2\eta_3^2)^2 \le \xi_1$$

$>$

$\le$

$$u = \sin\theta\cos\varphi = \frac{2A\xi_1\eta_2}{\xi_1^2 + A^2\eta_2^2 + A^2\eta_3^2}$$

$$v = \sin\theta\sin\varphi = \frac{2A\xi_1\eta_3}{\xi_1^2 + A^2\eta_2^2 + A^2\eta_3^2}$$

$$w = \cos\theta = \frac{\xi_1^2 - A^2\eta_2^2 - A^2\eta_3^2}{\xi_1^2 + A^2\eta_2^2 + A^2\eta_3^2}$$

抽样效率为： $E = \pi / 12A^2 \approx 0.555$

8. 多维分布的随机抽样

为方便起见，这里仅讨论二维分布的情况，对于更高维数的分布，可用类似的方法处理。

对于任意二维分布密度函数，总可以用其边缘分布密度函数和条件分布密度函数的乘积表示：

$$f(x,y)=f_1(x)\cdot f_2(y/x)$$

其中 $f_1(x)$ ，$f_2(y|x)$ 分别为分布 $f(x,y)$ 的边缘分布密度函数和条件分布密度函数，即

$$f_1(x)=\int_{\infty}^{\infty} f(x,y)dy$$

$$f_2(y/x)=\frac{f(x,y)}{f_1(x)}=\frac{f(x,y)}{\int_{\infty}^{\infty} f(x,y)dy}$$

二维分布密度函数的抽样方法是：

首先由 $f_1(x)$ 中抽取 X_{f_1}，再由 $f_2(y|X_{f_1})$ 中抽样确定 Y_{f_2}。

对于多维分布密度函数，也可直接采用类似于一维分布密度函数的抽样方法。例如，对如下形式的二维分布密度函数：

$$f(x,y)=H(x,y)\,f_1(x,y)$$

其中 $H(x,y)$ 为非负函数，$f_1(x,y)$ 为任意二维分布密度函数。设 M 为 $H(x,y)$ 的上界，则有二维分布的乘抽样方法如下：

$$\xi \le \frac{H(X_{f_1},Y_{f_1})}{M} \quad >$$

$$\downarrow \le$$

$$X_f = X_{f_1},\quad Y_f = Y_{f_1}$$

例 25. 下面二维分布密度函数的抽样

$$f(x,y)=\frac{e^{-xy}}{x} \quad x\geq 1, y\geq 0$$

将 $f(x,y)$ 写为

$$f(x,y)=f_1(x)\cdot f_2(y/x)$$

其中

$$f_1(x)=\frac{1}{x^2}$$

$$f_2(y/x)=x\cdot e^{-xy}$$

用直接抽样方法分别从 $f_1(x)$ 和 $f_2(y|X_{f_1})$ 中抽样，得到

$$X_{f_1}=1/\xi_1$$

$$Y_{f_2}=-\ln\xi_2/X_{f_1}=-\xi_1\ln\xi_2$$