蒙特卡罗方法(Monte Carlo simulation)

3.5 舍选抽样法

3.5 舍选抽样法（acceptance-rejection sampling)

直接抽样法的困难：

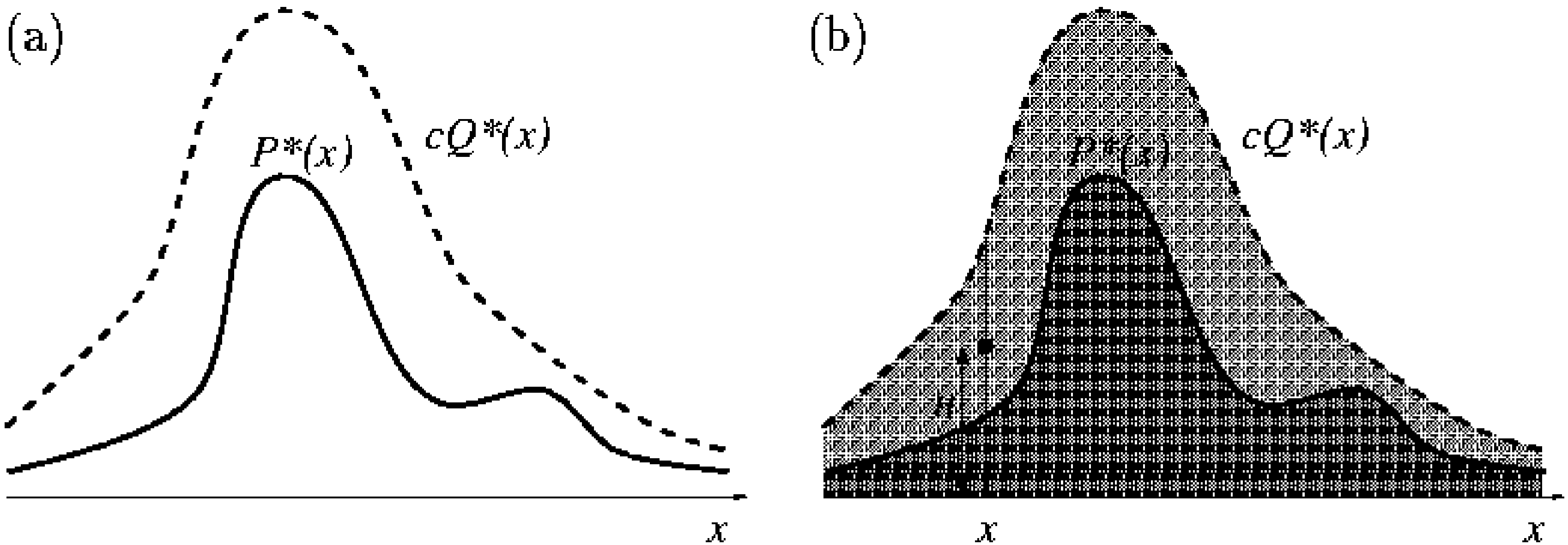
- 许多随机变量的累积分布函数无法用解析函数给出；
- 有些随机变量的累积分布函数的反函数不存在或难以求出；
- 即使反函数存在，但计算困难

- ➔舍选抽样法：抽取随机变量x的一个随机序列x_i,$i=1,2,\ldots$, 按一定的舍选规则从中选出一个子序列，使其满足给定的概率分布.
- 假定：
  - 随机变量x的值域为$[a,b]$;
  - X的概率密度函数：$f(x)=P^*(x)/Z$,（其中Z为归一化因子）🡐难以直接抽样
  - $Q(x)=Q^*(x)/Z_Q$ 是另外一个较为简单的函数(Z_Q为归一化因子）➔可用简单的方法进行抽样
  - 在x的整个取值范围内：$cQ^*(x) > P^*(x)$, 其中c为一常数
- 抽样方法：
  1. 产生两个随机数
     - 从$Q(x)$分布抽取x，$x\in[a,b]$
     - 由$[0, cQ^*(x)]$ 区间上的均匀分布抽取u，$u= cQ^*(x)\,\xi$，$\xi\in U[0,1]$
  2. 接收或舍弃取样值 x.
     - 如果 $u > P^*(x)$，舍弃，返回到1,重复上述过程；
     - 否则，接受

蒙特卡罗方法(Monte Carlo simulation)

3.5 舍选抽样法

- 几何解释

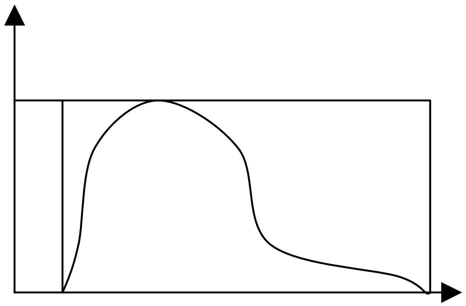


  – 在二维图上，随机选取位于曲线**cQ*(x)**下的点**[x,u]**；
  – 选取位于曲线**P*(x)**下的那些点，则这些点将服从概率密度为**P(x)**的分布
- 常数c的选取
  – 常数c应尽可能地小,因为抽样效率与c成反比
  – **c=max{P*(x)/Q*(x)}, x** $\in [a, b]$

蒙特卡罗方法(Monte Carlo simulation)

3.5 舍选抽样法

- **特例:**
  - 如果取**Q(x)=1,** *x∈[a,b]*,即均匀分布,则
    - X的抽样:x=(b-a)η+a, η *∈U[0,1]*
    - c可取为f(x)在[a,b]区间上的极大值



- 例1:标准正态分布的抽样,x∈[-a,a]
- $$p^*(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$$ 无法用直接抽样法,累积分布函数无解析表达式

$$Q^*(x)=\frac{1}{\pi}\frac{1}{1+x^2}$$

Breit-wigner or Cauchy分布

$$c=\max\{P^*(x)/Q^*(x)\}=\max\left\{\sqrt{\frac{\pi}{2}}(1+x^2)e^{-x^2/2}\right\}=\sqrt{\frac{2\pi}{e}}=1.52$$

$$Q(x)=Q^*(x)/\int_{-a}^{a}Q^*(x)dx=\frac{1}{2\arctan a}\frac{1}{1+x^2}$$

$$F_Q(x)=\int_{-a}^{x}Q(x')dx'=\frac{1}{2\arctan a}\left[\arctan x-\arctan(-a)\right]$$

蒙特卡罗方法(Monte Carlo simulation)

3.5 舍选抽样法

- 由Q(x)抽取x ← 直接抽样法

$$\eta \in U[0,1]$$
$$x = \tan\left(2\eta \arctan a + \arctan(-a)\right)$$

- 抽取u

$$Q = cQ^*(x) = \frac{1.52}{\pi}\frac{1}{1+x^2}$$
$$\xi \in U[0,1]$$
$$u = \xi Q$$

- 计算P*(x), 如果u<= P*(x), 接受x

```
float randac(unsigned long &i0, unsigned long &i1)
{
  const unsigned long a = 65539;
  const unsigned long b = 65539;
  unsigned long i2;
  unsigned long m = pow(2,31);
  i2 = (a * i1 + b * i0 ) % m;
  i0 = i1;   i1 = i2;
  return (float) i1/float(m);
}
```

蒙特卡罗方法(Monte Carlo simulation)

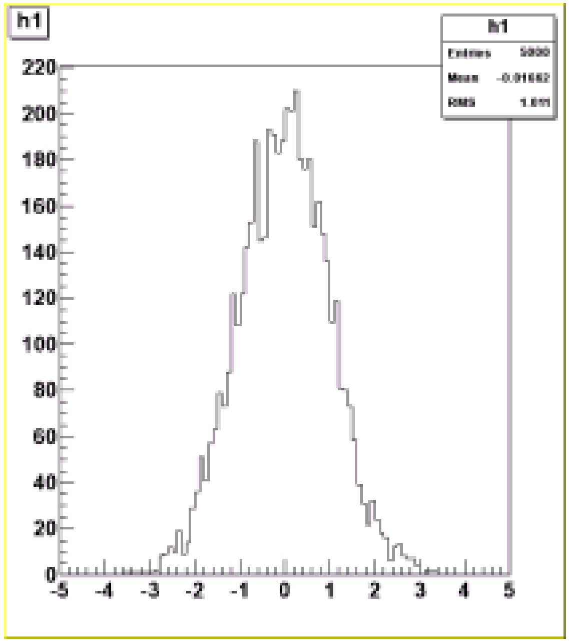
3.5 舍选抽样法

```
float gaussian_reject(double a)
{
  const float c = 1.52;
  static unsigned long i0 = 9, i1 = 11;
  while(true) {
    float eta = randac(i0,i1);
    float x = tan(eta * 2.0 * atan(a)+atan(-a));
    float q = c * 1/3.1415926*1.0/(1+x*x);
    float ksi = randac(i0,i1);
    float u = ksi*q;
    float p = 1/sqrt(2*3.1415926)*exp(-x*x/2.0);
    if(u <= p) break;
  }
  return x;
}
```

蒙特卡罗方法(Monte Carlo simulation)

3.5 舍选抽样法

```
void test()
{
  unsigned long i0 = 9, i1 = 11;
  c1 = new TCanvas("c1","Histogram Drawing Options",200,10,700,900);
  c1->Divide(1,2);
  TH1F * h1 = new TH1F("h1","h1",100,-5.0,5.0);
  for(int i=0; i < 5000; i++) {
    double x = gaussian_reject(5.0);
    h1->Fill(x);
  }
  c1->cd(2);h1->Draw();
}
```



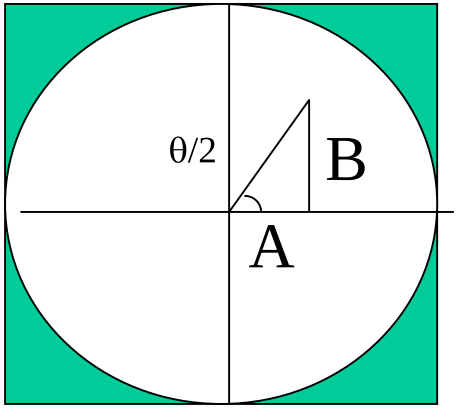
蒙特卡罗方法(Monte Carlo simulation)

3.5 舍选抽样法

- 例2：利用舍选法产生随机数C=cosθ, S=sinθ,其中θ为[0，2π]区间内均匀分布的随机数
- 方法1：先产生[0，2π]间均匀分布的随机数：θ= 2πr， r∈U[0,1],然后直接计算C和S ➔ 因需要计算三角函数，故此方法运算速度慢
- 方法2：利用舍选法可避免三角函数运算

$$C = \cos\theta = \cos^2\frac{\theta}{2} - \sin^2\frac{\theta}{2}$$

$$S = \sin\theta = 2\sin\frac{\theta}{2}\cos\frac{\theta}{2}$$



令A和B为单位圆内直角三角形的两个边，则有

$$\cos\frac{\theta}{2} = \frac{A}{\sqrt{A^2+B^2}}, \sin\frac{\theta}{2} = \frac{B}{\sqrt{A^2+B^2}}$$

$$\therefore C = \cos\theta = \frac{A^2-B^2}{A^2+B^2}, S = \sin\theta = \frac{2AB}{A^2+B^2}$$

因此，只要产生单位圆内的随机坐标A和B,就可用代数运算求出C和S,算法为

1. 产生两个[0,1]区间上均匀分布的随机数u_1和u_2;
2. 令$v_1=2u_1-1, v_2=u_2$,则$v_1\in U[-1,1], v_2\in U[0,1]$;
3. 计算$r^2=v_1^2+v_2^2$，如果$r^2>1$,转到1，重新产生；
4. 令$A=v_1, B=v_2$，计算C和S

蒙特卡罗方法(Monte Carlo simulation)

3.6 复合分布的抽样法

3.6 复合分布的抽样方法（composition method)

- **1961年由Marsaglia提出的方法**
- 设随机变量X的概率密度函数**f(x)**可写成一些**PDF**的线性叠加：

$$f(x)=\sum_k p_k f_k(x)$$

$$p_k \ge 0;\sum_k p_k =1; \qquad \int f_k(x)dx=1$$

- 抽样方法：
  - 利用离散型的随机变量的抽样方法抽取序号k;

$$r\in U[0,1]; \qquad \sum_{i=1}^{k-1} p_i \le r \le \sum_{i=1}^{k} p_i$$

  - 由$f_k(x)$抽取x
- 例：用复合法产生双指数分布随机数
  - 产生两个[0,1]区间均匀分布的随机数 r_1和r_2;
  - 如果r_1<=0.5, 按$f_1(x)$抽样；
  - 如果r_1>0.5, 按$f_2(x)$抽样；
  - $f_1(x)$和$f_2(x)$都可用直接抽样法

$$PDF \quad : f(x)=0.5e^{-|x|}$$

$$f(x)=0.5e^{x}I_{(-\infty,0)}(x)+0.5e^{-x}I_{[0,\infty)}(x)$$

$$I_{(-\infty,0)}(x)=\begin{cases}1 & -\infty<x<0\\ 0 & \text{其他}\end{cases}$$

$$I_{[0,\infty)}(x)=\begin{cases}1 & 0\le x<\infty\\ 0 & \text{其他}\end{cases}$$

$$\therefore\ p_1=0.5=p_2;\ f_1(x)=e^{x}(x<0);\ f_2(x)=e^{-x}(x\ge 0)$$

3.7 混合抽样法（mixed method)

混合法：$\begin{cases}\text{直接抽样法} \\ \text{舍选法}\end{cases}$ ➔亦称乘抽样法

适用的抽样场合：

- 概率密度函数难以积分➔无累积分布函数的解析表达式；
- 累积分布函数的反函数的解析表达式不存在；
- 概率密度函数存在尖峰（spiky);

➔直接抽样法不可用、舍选抽样法效率低

$$\int_a^b f(x)dx = C_f, \qquad \bar{f}(x) = \frac{1}{C_f} f(x)$$

假定：概率密度函数可写成下面的因子化形式

$$\boxed{p(x) = f(x)g(x)}$$

其中：

- f(x)包含了p(x)的峰值部分且可用直接抽样法进行抽样
- g(x)是一个相对变化平缓的函数，包含了p(x)函数的大部分的数学复杂性；

<u>抽样方法</u>：

1. 将f(x)归一化：$\int_a^b f(x)dx = C_f; \bar{f}(x) = {}^1\!/_{C_f}\, f(x)$
2. 令

$\bar{g}(x) = {}^1\!/_{M_g}\, g(x); \qquad \bar{g}(x) \le 1, x \in [a,b]$ Mg为g(x)在区间[a,b]上的极大值

蒙特卡罗方法(Monte Carlo simulation)

3.7 混合抽样法（mixed method)

3. 利用直接抽样法由$\bar{f}(x)$抽取$\boldsymbol{x}$;

4 . 抽取**[0,1]**区间均匀分布的随机数$\boldsymbol{r}$, 如果 $\bar{g}(x) \geq r$，则接受$\boldsymbol{x}$, 否则，返回到**3**重新抽样

<u>推广的形式</u>：

设概率密度函数可写成如下的形式：

$$p(x) = \sum_i f_i(x) g_i(x)$$

➔ 复合抽样法+混合抽样法==〉乘加抽样法

$$p(x) = \sum_i C_i \frac{f_i(x) g_i(x)}{C_i} = \sum_i C_i p_i'(x)$$

$$C_i = \int_a^b f_i(x) g_i(x) dx; \qquad p_i'(x) = \frac{f_i(x) g_i(x)}{C_i}$$

$$\int_a^b p(x) dx = \sum_i C_i = 1$$

抽样方法：

1. 采用符合抽样法，先确定**p(x)**的随机数应由哪一个 $p_i'(x)$ 抽取；

2. 在按$p_i'(x)$抽样时，采用上面的混合方法

蒙特卡罗方法(Monte Carlo simulation)

3.7 混合抽样法（mixed method)

例：**compton**散射

微分截面：

$$\Phi(E_0,E_1)=\frac{X_0 n\pi r_0^2 m_e c^2}{E_0^2}\left[\frac{1}{\varepsilon}+\varepsilon\right]\left[1-\frac{\varepsilon\sin^2\theta}{1+\varepsilon^2}\right]$$

$$\varepsilon=\frac{E_0}{E_1};\qquad E_1=E_0\frac{m_e c^2}{m_e c^2+E_0(1-\cos\theta)};\qquad E_1=E_{1\min}\to\varepsilon_0=\frac{m_e c^2}{m_e c^2+2E_0}$$

$$\varepsilon\in[\varepsilon_0,1]$$

如何抽取散射光子的能量？==〉乘加抽样法

$$\Phi(\varepsilon)\cong\left[\frac{1}{\varepsilon}+\varepsilon\right]\left[1-\frac{\varepsilon\sin^2\theta}{1+\varepsilon^2}\right]=f(\varepsilon)g(\varepsilon)=\left[\alpha_1 f_1(\varepsilon)+\alpha_2 f_2(\varepsilon)\right]g(\varepsilon)$$

$$\alpha_1=\ln\left(\frac{1}{\varepsilon_0}\right)=-\ln\varepsilon_0;\qquad f_1(\varepsilon)=\frac{1}{\alpha_1\varepsilon}\qquad F_1(\varepsilon)=\int_{\varepsilon_0}^{\varepsilon}f_1(\varepsilon')d\varepsilon'=\frac{1}{\alpha_1}\ln\frac{\varepsilon}{\varepsilon_0}$$

$$\alpha_2=\frac{(1-\varepsilon_0^2)}{2};\qquad f_2(\varepsilon)=\frac{\varepsilon}{\alpha_2}\qquad F_2(\varepsilon)=\int_{\varepsilon_0}^{\varepsilon}f_2(\varepsilon')d\varepsilon'=\frac{1}{\alpha_2}(\varepsilon^2-\varepsilon_0^2)$$

$$g(\varepsilon)=\left[1-\frac{\varepsilon}{1+\varepsilon^2}\sin^2\theta\right]\qquad\forall\varepsilon\in[\varepsilon_0,1]:0<g(\varepsilon)\le 1$$

抽样方法：r_1,r_2,r_3是三个在[0,1]区间上均匀分布的随机数

1. 确定由哪一个f函数来抽取ε：如果$r_1<\alpha_1/(\alpha_1+\alpha_2)$，选择$f_1(\varepsilon)$, 否则选$f_2(\varepsilon)$；
2. 根据f_1或f_2抽取ε：直接抽样法

蒙特卡罗方法(Monte Carlo simulation)

3.7 混合抽样法（mixed method)

$$f_1 : F_1(\varepsilon) = r_2 \rightarrow \varepsilon = \varepsilon_0 e^{\alpha_1 r_2}$$
$$f_2 : F_1(\varepsilon) = r_2 \rightarrow \varepsilon = \varepsilon_0^2 - \alpha_2 r_2$$

3. 计算sin²θ:

$$\sin^2\theta = t(2-t) \qquad t \equiv (1-\cos\theta) = \frac{m_e c^2}{E_0}\frac{1-\varepsilon}{\varepsilon}$$

4.计算g(ε), 如果 $g(\varepsilon) \geq r_3$ 接受ε， 否则返回到第一步

3.8 近似抽样法（列表法)

近似抽样法：

用近似的分布函数取代欲抽取的概率密度函数，一般是采用列表的形式将连续型的概率分布变成分离型的概率分布。

使用的场合：

– 概率密度函数的形式非常复杂，在模拟过程中进行计算需花费相当多的CPU时间;

– 概率密度函数无解析形式，只能用数值或曲线的形式表示。

基本方法:

设概率密度函数：f(x), x∈[a,b]

1. 将区间[a,b]分成n个子区间，分点为

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b$$

2. 分点对应的函数值为:

$$f_i = f(x_i), \qquad i = 1,2,\cdots n$$

蒙特卡罗方法(Monte Carlo simulation)

3.8 近似抽样法（列表法)

3. 对于每一个小区间，利用一简单的函数$f_a(x)$来近似地表示原概率分布函数，并使$f_a(x)$在该区间内的积分与$f(x)$在该区间内的积分相等，即俩者的概率相等。
4. 利用$f(x)$计算每一个子区间的概率值 $p_i=\int_{x_{i-1}}^{x_i} f(x)dx$ 并计算$f_a(x)$的累积分布函数在x_i处的值 $F_a(x_i)=\sum_{j=1}^{i} p_j, i=1,2,\cdots,n$,并将$F_a(x_i)$和x_i存入数据表中;
5. 抽样方法:
   - 随机选择子区间:选取[0,1]区间内均匀分布的随机数r,找出满足$F_a(x_{i-1})\le r\le F_a(x_i)$的子区间$[x_{i-1},x_i]$;
   - 根据$f_a(x)$的特点,确定欲抽取的随机数ξ；

几点说明：

1. 分点的选取:
   - $f_0,f_1,\ldots,f_n$应能充分反映$f(x)$的变化状况,即,在 $f(x)$ 变化迅速的区域分点密一点,变化缓慢的区域分点稀一点.
2. $f_a(x)$的选取:
   - 阶梯近似:
   - 线性近似
   - 二次曲线近似
   - 样条函数近似

蒙特卡罗方法(Monte Carlo simulation)

3.8 近似抽样法（列表法)

阶梯近似

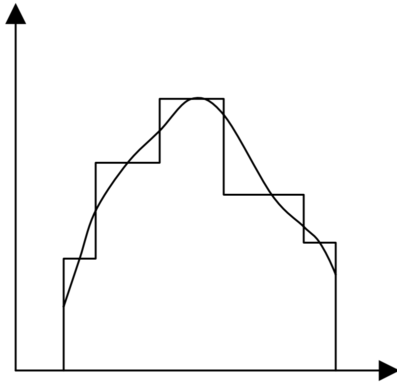
将$f_a(x)$ 取为阶梯函数，在每一个子区间中$f_a(x)$都是均匀分布

$$f_a(x)=\frac{\int_{x_{i-1}}^{x_i} f(x)\,dx}{x_i-x_{i-1}}=\frac{p_i}{x_i-x_{i-1}},\qquad i=1,2,\cdots,n$$

$$\int_{x_{i-1}}^{x_i} f_a(x)\,dx=\int_{x_{i-1}}^{x_i}\frac{p_i}{x_i-x_{i-1}}dx=p_i$$

$$\sum_{i=1}^{n} p_i=\sum_{i=1}^{n}\int_{x_{i-1}}^{x_i} f(x)\,dx=\int_{x_0}^{x_n} f(x)\,dx=\int_a^b f(x)\,dx=1$$

$$F_a(x_i)=\sum_{j=1}^{i} p_i\qquad i=1,2,\cdots,n$$



抽样方法：

1. 选取[*0*，*1*]区间均匀分布的随机数r_i;
2. 找出满足下式的分点x_{j-1}和x_j: $F_a(x_{j-1})<r_i\le F_a(x_j)$
3. 欲抽取的随机数为：

$$\xi_i=x_{j-1}+(x_j-x_{j-1})\frac{r_i-F_a(x_{j-1})}{F_a(x_j)-F_a(x_{j-1})}$$

蒙特卡罗方法(Monte Carlo simulation)

3.8 近似抽样法（列表法)

- 线性近似

利用一系列的折线来近似原分布$f(x)$，即将$f_a(x)$取为

$$f_a(x) = C\left\{ f_{i-1} + \frac{x - x_{i-1}}{x_i - x_{i-1}}(f_i - f_{i-1}) \right\} \qquad x \in [x_{i-1}, x_i], i = 1,2,\cdots,n$$

其中C为归一化因子，使得每一子区间内原分布和近似分布的积分概率相等

$$\int_{x_{i-1}}^{x_i} f_a(x)dx = \int_{x_{i-1}}^{x_i} f(x)dx, \qquad i = 1,2,\cdots,n$$

抽样方法：

抽取r1,r2

↓

找出满足下式要求的x_{j-1},x_j

$$F_a(x_{j-1}) < r_1 \le F_a(x_j)$$

↓

$$r_1 - F_a(x_{j-1}) \le \left\{ (f_j - f_{j-1})r_2 + f_{j-1} \right\} \frac{x_j - x_{j-1}}{2}$$

$\le$：

$$\xi_i = x_{j-1} + (x_j - x_{j-1})r_2$$

$>$：

$$\xi_i = x_{j-1} + (x_j - x_{j-1})(1 - r_2)$$

蒙特卡罗方法(Monte Carlo simulation)

3.9 多维分布的抽样

- 将一维分布的抽样方法推广到多维分布
- 条件密度法：利用条件概率密度将多维模拟转化为一维模拟问题