

《计算机模拟》



第6讲 – 蒙特卡罗方法

胡贤良

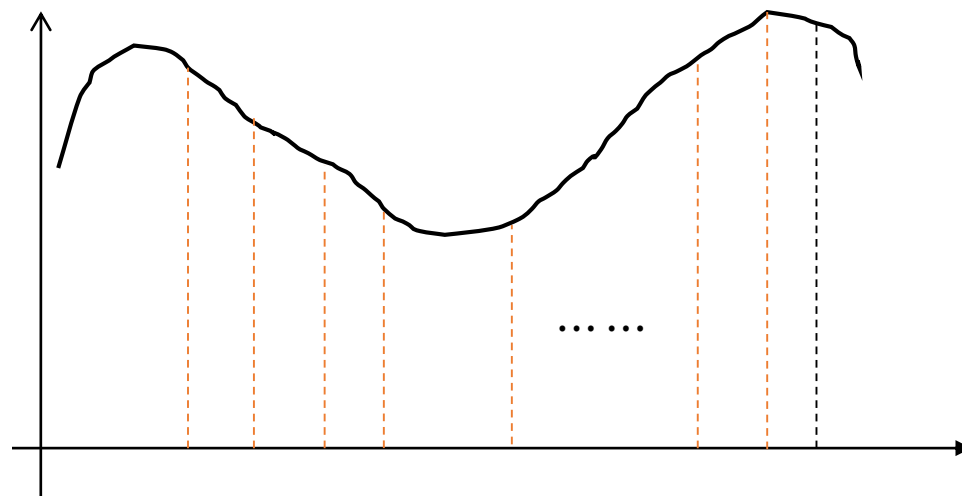
浙江大学数学科学学院

1.基本蒙特卡罗方法

以数值积分问题为例

问题：积分近似计算

- ① 传统数值方法
- ② 随机投点法
- ③ 样本平均(值)法



问题描述：已知函数 $f(x)$,及其定义域 $[a,b]$ 上的一个划分： $a = x_0 < x_1 < x_2 < \dots < x_n = b$,若记

$$T = \max\{x_{i+1} - x_i, \quad i = 0, 1, \dots, n-1\},$$

考察(上)和

$$S_1 = \sum_{i=0}^{n-1} (x_{i+1} - x_i) \sup\{f(x), x \in [x_i, x_{i+1}]\}$$

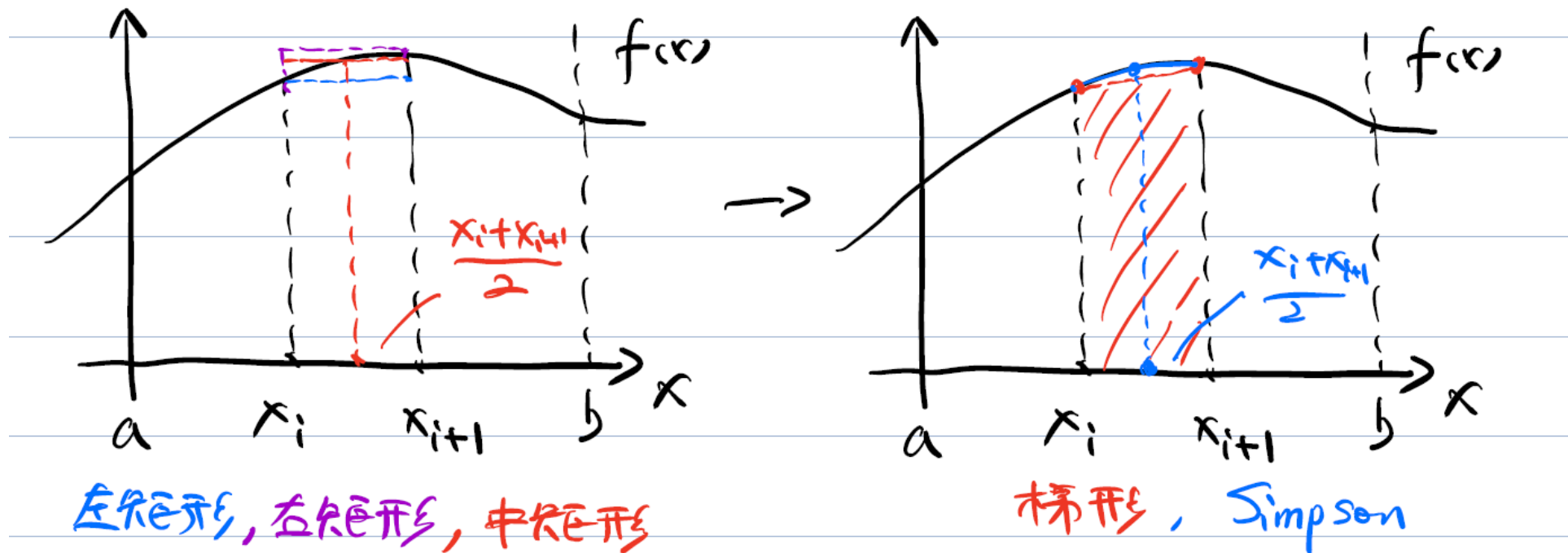
以及(下)和

$$S_2 = \sum_{i=0}^{n-1} (x_{i+1} - x_i) \inf\{f(x), x \in [x_i, x_{i+1}]\}$$

当 $T \rightarrow 0$ 时, 若 S_1, S_2 趋向于同一个极限 I , 且该极限与具体的划分无关, 则记Riemann积分

$$I = \int_a^b f(x) dx$$

1. 数值积分法 (传统方法)



而对于一个积分存在的 f , 我们知道

$$I = \lim_{T \rightarrow 0} \sum_{i=0}^{n-1} (x_{i+1} - x_i) \cdot f(\xi_i),$$

其中 $\xi_i \in [x_i, x_{i+1}]$ 可以是 $[x_i, x_{i+1}]$ 中任意一点.

比如取 $\xi = x_i$ 或 x_{i+1} , 上式就分别是复合左矩形或右矩形公式. 若 $\xi = \frac{x_i + x_{i+1}}{2}$, 则成为中矩形公式.

传统方法之殇：维数灾难

考虑计算 d 维(黎曼)积分问题

$$I = \int_D f(\mathbf{x}) d\mathbf{x} = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_d}^{b_d} f(x_1, x_2, \cdots, x_d) dx_1 dx_2 \cdots dx_d$$

其中, $0 \leq f(\mathbf{x}) \leq M, \forall \mathbf{x} \in D, D = \{a_1 \leq x_1 \leq b_1, a_2 \leq x_2 \leq b_2, \cdots, a_d \leq x_d \leq b_d\}$ 是一个 d 维积分区域。进一步地, 定义 $\mathbf{z} = (\mathbf{x}, y) = (x_1, x_2, \cdots, x_d, y)$ 为 $\Omega = D \times [0, M]$ 上 $d + 1$ 维均匀分布的随机向量, 且各分量相互独立。它们的联合概率密度函数

$$\varphi(\mathbf{x}, y) = \begin{cases} \frac{1}{M|D|}, & (\mathbf{x}, y) \in \Omega, \\ 0, & \text{otherwise.} \end{cases}$$

其中, $|D| = \prod_{i=1}^d (b_i - a_i)$.

维数灾难: 计算复杂度: $O(n^d)$

- 对于维数 d 不超过3的经典物理学问题, 尚可勉强工作。
- 对于现代物理学和数据分析来说, 维数 d 并不是一个小量!
- 亟需寻找关于维数增长不快的积分法!

2. 随机投点法

- 记被积函数的下方区域为：

$$S = \{(\mathbf{x}, y) | a_1 \leq x_1 \leq b_1, \dots, a_d \leq x_d \leq b_d, 0 \leq y \leq f(\mathbf{x})\}$$

- 随机投点 $z = (x, y)$ ，则该点落入区域 S 的概率

$$p = \frac{S}{|\Omega|} = \frac{I}{M|D|}.$$

只要估算出 p ，就可得出积分 I 的近似值！

- 生成 N 个相互独立的在 Ω 上均匀分布的随机向量

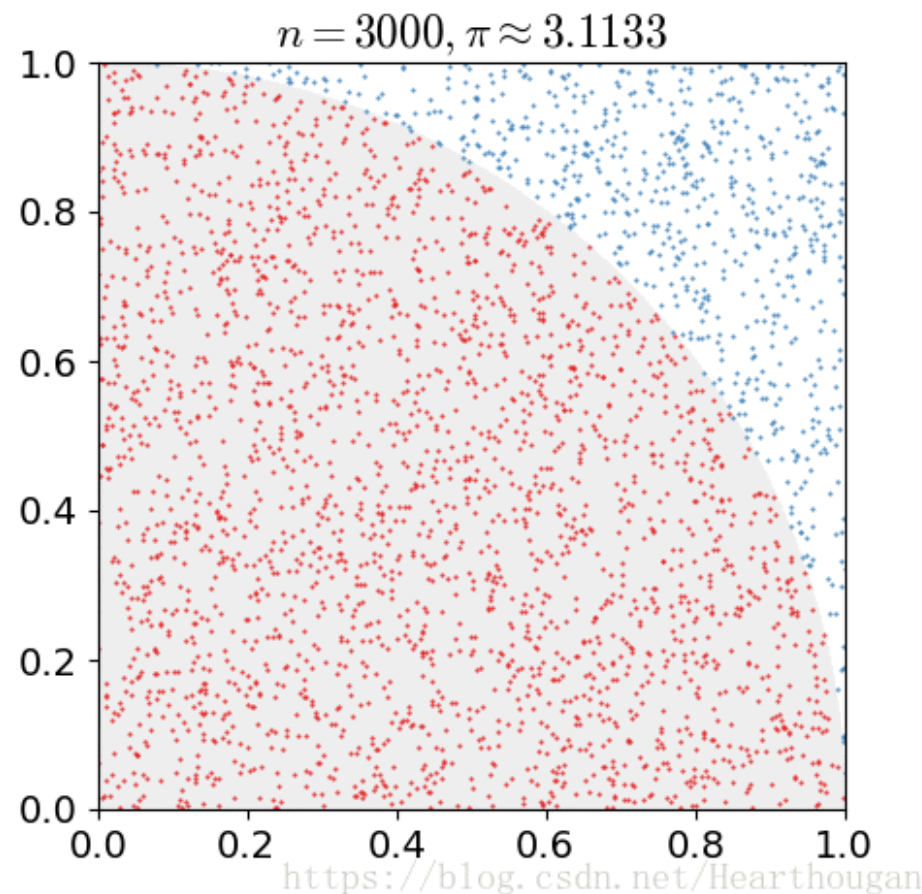
$$(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$$

- 用 n_H 表示满足 $y_i \leq f(x_i)$ 的落点数目，即落入 S 命中数。
此时，由强大数定律，可得

$$p_N = \frac{n_H}{N}.$$

- 进而得到 I 的近似值

$$I \approx I_N = \frac{n_H}{N} M|D|.$$

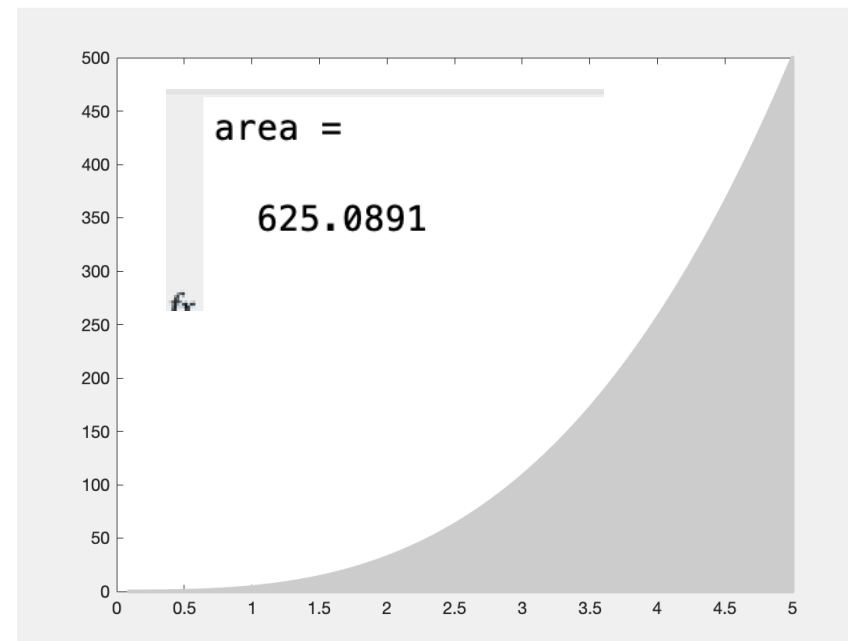


注：该方法并不需要被积函数是非负的，只要存在下界就可以，上述只是为了方便

算例1(ex1_4-integrate_2d.m)

```
ntrials=100000000;
x=5*rand(1,ntrials);
y=500*rand(1,ntrials);
nhits=sum(y<4*x.^3);
p=nhits/ntrials;
area=p*500*5
plot(x(y<4*x.^3),y(y<4*x.^3),'.','color',[0.8,0.8,0.8])
```

$$I = \int_0^5 4x^3 dx$$



计算 $\int_0^1 e^{-x^2/2} / \sqrt{2\pi} dx$,

精确值	10^3	10^4	10^5	10^6	10^7
0.3413447	0.342	0.344	0.34187	0.341539	0.341302

$\int_2^5 e^{-x^2/2} / \sqrt{2\pi} dx$,

真实值	10^3	10^4	10^5	10^6	10^7
0.02274985	0.02332792	0.02311736	0.02262659	0.02284152	0.02278524

算例2(ex1_4-integrate_3d.m)

- 求球体 $x^2 + y^2 + z^2 \leq 4$ 被圆柱面 $x^2 + y^2 = 2x$ 所截立体的体积。

分析：记D为 $y = \sqrt{2x - x^2} = \sqrt{1 - (x - 1)^2}$ 及x轴围成的闭区域：

$$V = 4 \int \int_D \sqrt{4 - x^2 - y^2} dx dy$$

利用重积分方法获得其准确值为9.6440.

$$\Omega = \{(x, y, z) | 0 \leq x \leq 2, 0 \leq y \leq 1, 0 \leq z \leq 2\}$$

```
function V=quad2(N)
for i =1:length(N)
    x=2*rand(N(i),1);
    y=rand(N(i),1);
    z=2*rand(N(i),1);
    %落到区域T内的点数
    nhit=sum((x.^2+y.^2+z.^2<=4)&((x-1).^2+y.^2<=1));
    V(i)=16*nhit/N(i);
end
```

```
>> V=quad2([100,1000,10000,10000,100000,1000000])
V =
    9.4400    9.6160    9.5360    9.7264    9.6528    9.6397
```


随机投点法 – 收敛速度

我们注意到， N 次随机落点击中区域 S 的数 n_H 服从二项分布 $n_H \sim B(N, p)$. 于是

$$E(n_H) = Np, \quad \text{var}(n_H) = Np(1 - p).$$

故

$$E(\hat{I}) = E(\hat{p}M|D|) = M|D|E(\hat{p}) = \frac{E(n_H)}{N}M|D| = pM|D| = I,$$

及

$$\text{var}(\hat{I}) = \frac{M^2|D|^2}{N^2} \text{var}(n_H) = \frac{M^2|D|^2}{N^2} Np(1 - p) = \frac{M^2|D|^2}{N} p(1 - p) = \frac{I}{N} (M|D| - I) \propto N^{-1}.$$

➤ 积分计算值的标准差

$$\sigma_{\hat{I}} = \sqrt{\text{var}(\hat{I})} = M|D| \sqrt{\frac{p(1 - p)}{N}}$$

- ▶ 从标准差看，算法的收敛速度很慢
- ▶ 积分的计算精度与积分的维度是无关的，这是最大的优点
- ▶ 维数 $d \geq 3$ 时能体现出算法优越性 (Simpson法的误差 $\propto N^{-\frac{1}{d}}$)

随机投点法 – 所需试验次数N的估算

- 令 $\delta = \frac{\epsilon}{\sigma_{\hat{I}}}$, 根据**中心极限定理**, 当N充分大时, 有 $\frac{|n_H - Np|}{\sqrt{Np(1-p)}} \sim N(0, 1)$

可知

$$P(|\hat{I} - I| < \epsilon) = P\left(\frac{|n_H - Np|}{\sqrt{NP(1-p)}} < \delta\right) = \Phi(\delta) - \Phi(-\delta) = 2\Phi(\delta) - 1.$$

其中, $\Phi(\cdot)$ 是标准正态分布函数。

- 如果上述误差概率要大于 $1 - \alpha$ (α 为任意正数/小概率值),

即 $\Phi(\delta) > 1 - \frac{\alpha}{2}$, 则 $\delta > Z_{\frac{\alpha}{2}}$ (即 $\frac{\epsilon}{\sigma_{\hat{I}}} > Z_{\frac{\alpha}{2}}$)

$$N \geq \frac{p(1-p)(M|D|z_{\alpha/2})^2}{\epsilon^2}$$

给出了为达到一定精度所必须的投点次数的估计数。例如 $\epsilon = 0.1$, 投点次数至少要多达 10^8 的数量级

3. 样本平均法(期望法)

将积分表示成某个随机变量 X 的数学期望

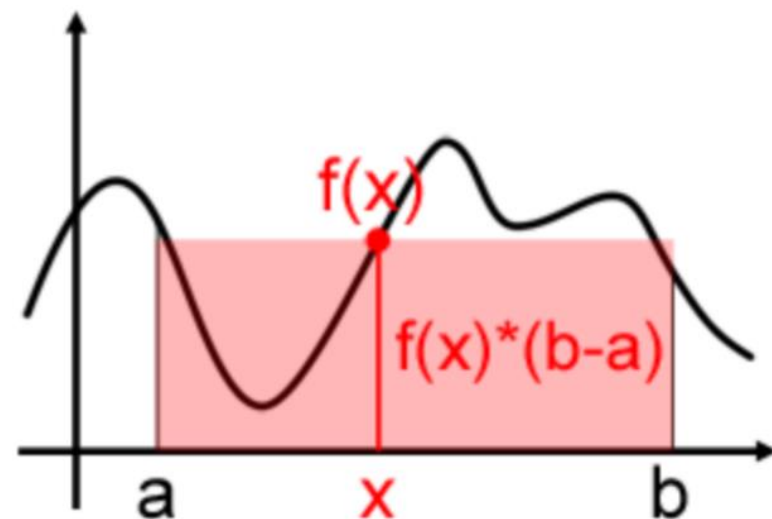
$$I = \int_a^b \frac{f(x)}{g_X(x)} g_X(x) dx = E\left(\frac{f}{g_X}\right)$$

为了计算简便，最简单的做法是取 X 为 (a,b) 上服从均匀分布的随机变量，即

$$g_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

于是，有

$$I = (b-a)\mathbf{E}(f(X))$$

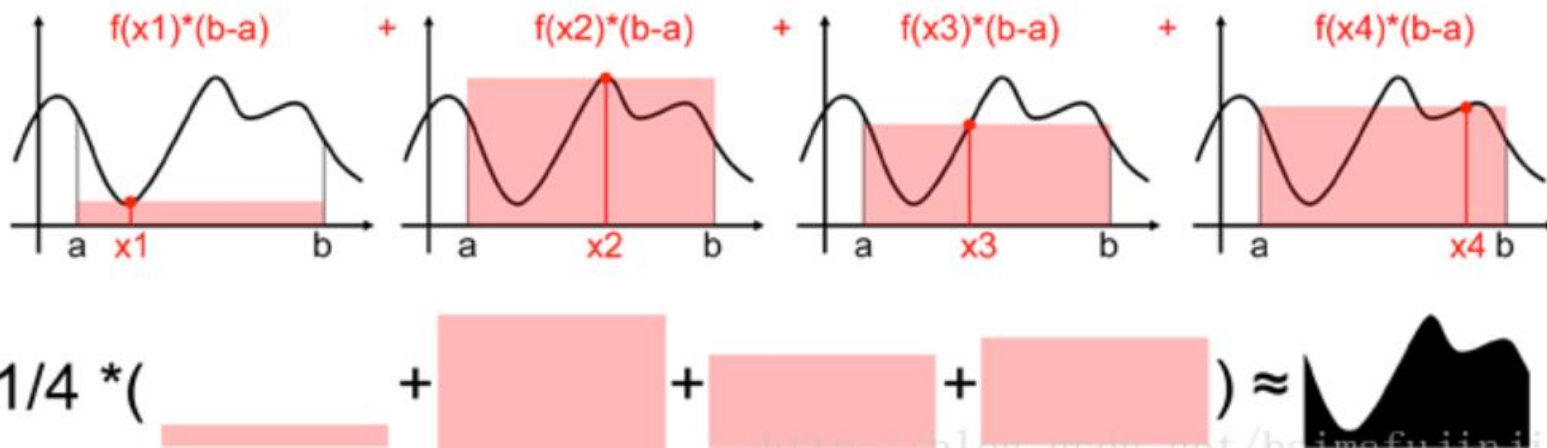


样本平均法 – 模拟方案

➤ 为了估计积分，我们只要产生N个在 (a,b) 上服从均匀分布的随机数 X_1, X_2, \dots, X_N ,并用样本均值

$$\hat{I} = (b - a) \frac{1}{N} \sum_{i=1}^N f(X_i)$$

来当积分的估计值。



▶ 容易将这个积分推广为高维情形：

从D中均匀产生N个点 $(x_1^{(i)}, x_2^{(i)}, \dots, x_d^{(i)})$, $i = 1, \dots, N$, 则该积分的近似值 $\hat{I} = \frac{|D|}{N} \sum_{i=1}^N f(x_1^{(i)}, x_2^{(i)}, \dots, x_d^{(i)})$

其中， $|D|$ 是区域D的体积 $|D| = \prod_{i=1}^d (b_i - a_i)$

算例(仅4行代码)

$$I = \int_0^5 4x^3 dx$$



$$I = \int_0^5 20x^3 \times \frac{1}{5} dx = \mathbf{E}(20X^3)$$



```
ntrials=100000000;  
x=unifrnd(0,5,1,ntrials);  
y=20*x.^3;  
s=mean(y)|
```



s =
624.9016

无偏性:

$$\begin{aligned} \mathbf{E}(\hat{I}) &= \mathbf{E}\left(\frac{(b-a)}{N} \sum_{i=1}^N f(X_i)\right) = \frac{(b-a)}{N} \sum_{i=1}^N \mathbf{E}(f(X_i)) \\ &= \frac{(b-a)}{N} \sum_{i=1}^N \int_a^b f(x) \frac{1}{(b-a)} dx = \int_a^b f(x) dx = I \end{aligned}$$

样本平均值法 – 精度/收敛速度

$$\begin{aligned}\sigma_{\hat{I}}^2 &= \text{var}\left(\frac{(b-a)}{N} \sum_{i=1}^N f(X_i)\right) = \frac{(b-a)^2}{N^2} \sum_{i=1}^N \text{var}(f(X_i)) \\ &= \frac{(b-a)^2}{N} \text{var}(f(X_i)) \\ &= \frac{(b-a)^2}{N} \int_a^b (f(x) - \mathbf{E}(f))^2 \frac{1}{(b-a)} dx \\ &= \frac{(b-a)}{N} \int_a^b \left(f(x) - \frac{I}{(b-a)}\right)^2 dx\end{aligned}$$

而 $\int_a^b \left(f(x) - \frac{I}{b-a}\right)^2 dx = \int_a^b f(x)^2 dx - \frac{I^2}{b-a}$, 故有

$$\sigma_{\hat{I}}^2 = \frac{1}{N} \left((b-a) \int_a^b f^2(x) dx - I^2 \right) = \frac{1}{N} (E(f^2) - I^2) \propto N^{-1}$$

这表明：样本平均法的精度/收敛速度与随机投点法是一样的！
进一步地，

$$M * I = \int_a^b M f(x) dx \geq \int_a^b f^2(x) dx$$

因此，平均值法得到的积分估计比随机投点法得到的更有效，即

$$\text{var}(\hat{I}) \leq \text{var}(I)$$

收敛性： 令 $Y = f(X)$ ，样本平均值法是以随机变量Y的简单抽样的算术平均作为近似值：

$$Y_N = \frac{1}{N} \sum_{i=1, \dots, N} Y_i$$

Y_1, Y_2, \dots, Y_N 独立同分布且期望有限，
根据强大数定律，可知

$$P(\lim_{N \rightarrow \infty} \bar{X}_N = E(X)) = 1$$

即当抽样数N充分大时，随机变量Y
简单抽样的算术平均以概率 1 收敛于
积分的真值！

样本平均值法- 误差估计

中心极限定理指出, 如果 Y_1, Y_2, \dots, Y_N 独立同分布且方差有限, 则

$$\lim_{N \rightarrow \infty} P\left(\frac{\sqrt{N}}{v} |\bar{Y}_N - \mathbf{E}Y| < \delta\right) = \frac{1}{\sqrt{2\pi}} \int_{-\sigma}^{\sigma} e^{-t^2/2} dt$$

当N充分大时, 就有如下的近似式

$$P(|\bar{Y}_N - \mathbf{E}Y| < \frac{z_{\alpha/2} v}{\sqrt{N}}) \approx \frac{2}{\sqrt{2\pi}} \int_0^{z_{\alpha/2}} e^{-t^2/2} dt = 1 - \alpha$$

其中, α 是显著水平($1-\alpha$ 为置信度), $\frac{z_{\alpha}}{2}$ 是标准正态分布在显著水平 $\frac{\alpha}{2}$ 的临界值, v 是Y的标准差.

► 上式表明, 如下不等式近似的以概率 $1-\alpha$ 成立, 且误差的收敛速度为 $O(N^{-1/2})$

$$|\bar{Y}_N - \mathbf{E}Y| < \frac{z_{\alpha/2} v}{\sqrt{N}}$$

► 注:

1.蒙特卡洛积分法的误差为概率意义上的误差!

2.误差中的标准差 v 是未知的, 需使用估计值 $\hat{v} = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y}_N)^2}$

2. 提高抽样效率

抽样(Sampling)

统计学定义: A sampling plan is just a method or procedure for specifying

how a sample will be taken from a population

- 简单(Simple)抽样: 直接抽取
- 系统(Systematic)采样: 平均分组, 简单灵活, 适用于大样本, 如人口普查
- 分层(Stratified)采样: 按影响较大的某种特征排序分组/层, 层间差异大。
- 整群(Cluster)采样:

按与主要研究指标无关特征分群, 互不重叠。群内差异大、群间差异小。

降低方差基本方法

一些通用性强、应用广泛的包括：

1. 分层抽样
2. 重要抽样
3. 控制变量
4. 对偶随机变量
5. 公共随机数
6. 条件期望
7. 样本分裂

从这些方法的原理上来看，主要运用了四类技巧：

- 只改变概率分布
- 只改变统计量
- 同时改变概率分布和统计量
- 只改变统计特性

1. 分层抽样

总体由差异明显的及部分组成时，按这些特征分组；只是改变概率分布，不改变统计量！

例[一维问题]：将长度 L 的区间分成 m 段/层， p_j 为第 j 层所占概率，随机变量 X_j 的概率分布为 $f(x_j)$ ，则

有统计量 $h_j(x)$ 取值 $h(x_j)$ ： $p_j = \int_{Z_j} f(x)dx$ ， $\sum_{j=1}^m p_j = 1$ 。这里第 j 层随机变量概率分布 $f(x_j) = p_j^{-1}f(x)$ 。

- 第 j 层统计量的期望

$$\mu_j = E(h_j) = \int_{x_{j,\min}}^{x_{j,\max}} h(x_j) f(x_j) dx_j$$

- 第 j 层分层抽样技巧统计量期望

$$\mu = E[h] = \sum_{j=1}^m p_j E[h(X_j)]$$

- 相应统计量估计值分别为

$$\hat{h}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} h(X_i^j), \quad \hat{h} = \sum_{j=1}^m p_j \hat{h}(X_i)$$

- 相应统计量估计值的方差分别为

$$\text{Var}[\hat{h}_j] = \frac{1}{p_j} \int_{x_{j,\min}}^{x_{j,\max}} (h(x_j) - E[h_j])^2 f(x_j) dx_j$$

$$\text{Var}[\hat{h}] = \sum_{j=1}^m p_j^2 \text{Var}[\hat{h}_j] / n_j$$

- 总样本数

$$n = \sum_{j=1}^m n_j$$

如何确定 n_j ?

(1) 比例分层抽样技巧

- 第 j 层样本数: $n_j = np_j, j = 1, \dots, m$
- 统计量估计值:

$$\hat{h} = \frac{1}{n} \sum_{j=1}^m \sum_{i=1}^{n_j} h(x_j^{(i)})$$

- 记整层抽样估计方差为

$$\begin{aligned} & \text{Var}[\hat{h}(X)] \\ &= \int h^2(x)f(x)dx - (E[h(X)])^2 \end{aligned}$$

- 可以证明:

$$\text{Var}[\hat{h}] \leq \text{Var}[\hat{h}(X)]$$

(2) 最优分层抽样技巧

- 令

$$\text{Var}[\hat{h}] = \sum_{j=1}^m p_j^2 \text{Var}[\hat{h}_j]/n_j := \sum_{j=1}^m p_j^2 \sigma_j^2 / n_j$$

取极小的 n_j , 即获得最小方差, 此时

$$n_j = np_j \sigma_j / \sum_{j=1}^m p_j \sigma_j, j = 1, 2, \dots, m$$

最优分层统计量估计值的方差为:

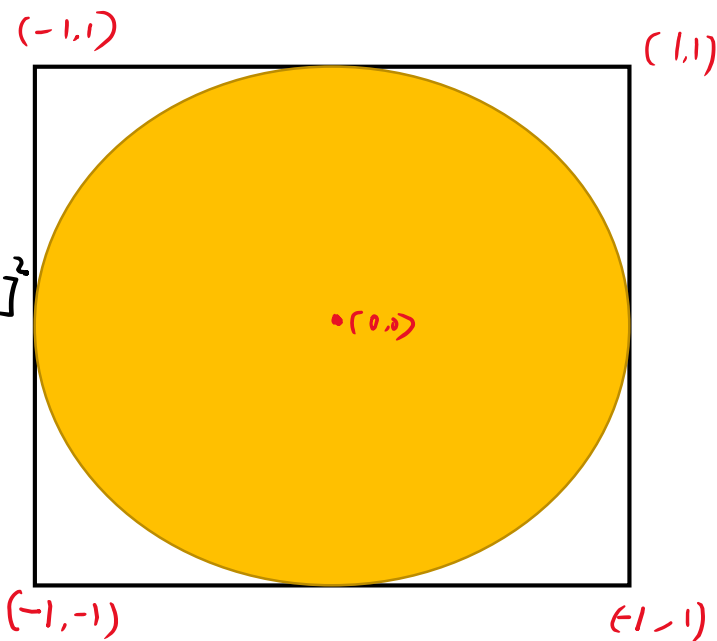
$$\text{Var}[h(n_1, n_2, \dots, n_m)] = \frac{1}{n} \left(\sum_{j=1}^m p_j \sigma_j \right)^2$$

例: 以圆周率随机投点为例

1. 直接模拟法: $X, Y \in U(-1, 1)$, $f(x) = f(y) = \frac{1}{2}$. $\begin{cases} X = 2U_1 - 1 \\ Y = 2U_2 - 1 \end{cases} \Rightarrow (U_1, U_2) \in [0, 1]^2$

$$\hat{\pi}_1 = 4\hat{p} = \frac{4}{n} \sum_{i=1}^n P(X_i, Y_i)$$

$$\text{其中, } P(X, Y) = \begin{cases} 0, & \text{当 } x^2 + y^2 > 1, \\ 1, & \text{当 } x^2 + y^2 \leq 1. \end{cases}$$



2. 分层抽样: $X_i \in [\frac{i-1}{n}, \frac{i}{n}]$. 均匀分布, 则 $f(x_i) = n$. 用逆变换法,

$$X_i = (U_i + i - 1) / n.$$

第 i 层随机变量 X_i 估计量可写成:

$$h(X_i) = h\left(\frac{U_i + i - 1}{n}\right),$$

$$\text{取 } h(x) = \sqrt{1 - x^2} \text{ (条件期望法)}$$

则分层抽样下, 圆周率估计值为:

$$\hat{\pi}_2 = 4\hat{h} = \frac{4}{n} \sum_{i=1}^n h(X_i) = \frac{4}{n} \sum_{i=1}^n h\left(\frac{U_i + i - 1}{n}\right)$$

注: U_i 为均匀分布随机数 on $[0, 1]$

模拟结果汇总:

次数	直接模拟 $\hat{\pi}_1$	分层抽样 $\hat{\pi}_2$
10^3	3.1760	3.145977796794
10^4	3.1404	3.1415929277487
10^7	3.1416	3.1415926535930
10^9	3.1416	3.1415926535896

2. 重要性抽样

顾名思义：只改变概率分布，起源于变量代换方法。记 X 为随机变量， $f(x)$ 是它的概率分布， $h(x)$ 表示统计量，则重要抽样可得到重要概率密度分布 $g(x)$ ，其权重为：

$$\omega(x) = \frac{f(x)}{g(x)} < 1$$

则模拟 n 次的统计量无偏估计为

$$\hat{h} = \frac{1}{n} \sum_{i=1}^n h(X_i) \omega(X_i)$$

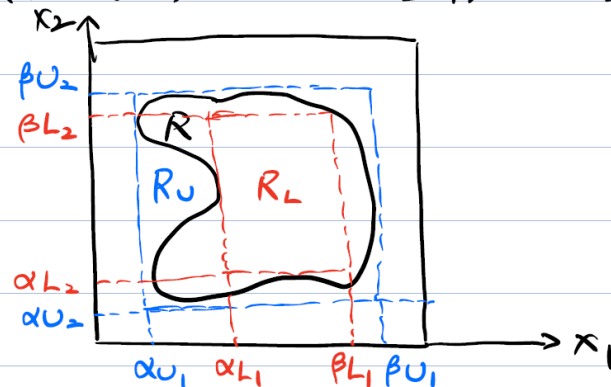
其数学期望也为 μ 。容易计算，其方差为

$$\text{Var}_g(h) = \int h^2(x) \omega^2(x) g(x) dx - \mu^2,$$

与直接模拟得到的统计量方差区别

$$\text{Var}_g(h) - \text{Var}_f(h) = \int h^2(x) (\omega(x) - 1) f(x) dx.$$

对于MC方法，提高效率和提高精度是等价的。如何用更少的采样点得到同样的估计是一个有意义的问题。以二维求面积为例，比如我们要求如下单位正方形内不规则图形 R 的面积



同时 R 有矩形上界 R_U 和矩形下界 R_L ，满足

$$R_L \subseteq R \subseteq R_U \subseteq [0,1]^2.$$

显然，我们应该在随机投点时，将点均匀地投到 $R_U \setminus R_L$ 区域中以提高投点效率，减小误差(var)。

重要性抽样法 – 基本思想

$$|\bar{Y}_N - \mathbf{E}Y| < \frac{z_{\alpha/2} \sigma}{\sqrt{N}}$$

事实上，上面已经指出，如果被积函数的方差较大，则采用上述方法会产生较大的误差。因此减小误差的有效方法是降低被积函数的方差，为此，我们再次考虑样本平均法：

$$I = \int_a^b f(x) dx = \int_a^b \frac{f(x)}{g(x)} g(x) dx = \mathbf{E}\left(\frac{f(X)}{g(X)}\right)$$

- 这里采用了简单的随机抽样法，这样，积分的计算精度直接与 f 的期望估算的精度相关，即由 f 的方差决定。
- 若是采用其他非均匀分布抽样的话，那么期望估算的误差就取决于 $\frac{f(X)}{g(X)}$ 的方差，这为我们减少误差提供了一条可行的路径。

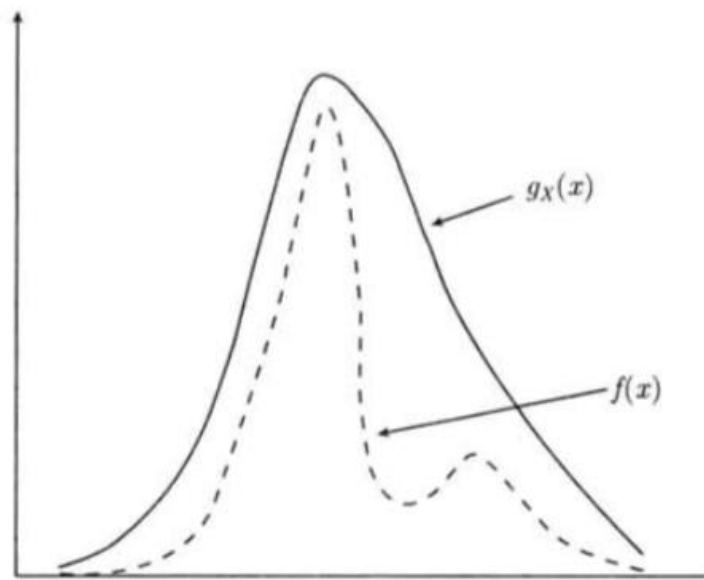
重要性抽样法 – 基本思想（续）

具体地，构造一个概率密度函数 $g_X(x)$,使其形状尽可能地与被积函数 $f(x)$ 接近，并且能方便地从概率密度函数为 $g_X(x)$ 的分布进行抽样（如可用逆变换法）

$$\text{令 } h(x) = \frac{f(x)}{g_X(x)}$$

将从概率密度函数为 $g_X(x)$ 的分布抽样出的随机数序列记为 X_1, X_2, \dots, X_N ,于是积分的近似值可用如下公式计算：

$$\mathbf{E}(h(x)) \approx \frac{1}{N} \sum_{i=1}^N h(X_i)$$



重要性抽样法 - 模拟步骤

(1)选择适合于 $f(x)$ 的概率密度函数 $g(x)$

(2)求 $g(x)$ 的分布函数 $G(x)$ 的反函数 $G^{-1}(x)$

(3)生成 N 个均匀分布的随机数 $U_1, U_2, \dots, U_N \sim U(1,0)$

(4)用逆变换法获得 N 个服从分布 $G(x)$ 的随机数

$$X_i = G^{-1}(U_i), i = 1, 2, \dots, N;$$

(5)计算平均值

$$\hat{I} = \frac{1}{N} \sum_{i=1}^N h(X_i)$$

算例4(ex4.m)

```
N = 1000000;  
f = @(x) exp(-x)./(1+x.^2);  
% 数值积分法  
quad0 = integral(f,0,1)  
% 样本平均法  
X = unifrnd(0,1,1,N);  
fg = f(X);  
quad1 = mean(fg)  
v1 = sqrt(var(fg))  
%重要性抽样法  
U = unifrnd(0,1,1,N);  
X = -log(1-U * (1 - exp(-1)));  
fg = f(X)./(exp(-X) / (1 - exp(-1)));  
quad2 = mean(fg)  
v2 = sqrt(var(fg))
```

$$g_X(x) = \frac{e^{-x}}{(1 - e^{-1})}, \quad 0 < x < 1$$

```
quad0 =  
    0.5248  
|  
quad1 =  
    0.5246  
  
v1 =  
    0.2450  
  
quad2 =  
    0.5248  
  
v2 =  
    0.0968  
,
```

重要性抽样法 – 无偏性、方差与减小误差

■无偏性

$$\begin{aligned}\mathbf{E}(\hat{I}) &= \mathbf{E}\left(\frac{1}{N} \sum_{i=1}^N h(X_i)\right) = \frac{1}{N} \sum_{i=1}^N \mathbf{E}(h(X_i)) \\ &= \frac{1}{N} \sum_{i=1}^N \int_a^b h(x) g_{\mathbf{X}}(x) dx = \int_a^b f(x) dx = I\end{aligned}$$

$$\mathcal{E}(\hat{I}) \propto N^{-1/2}$$

■积分近似值的方差

$$\begin{aligned}\text{var}(\hat{I}) &= \text{var}\left(\frac{1}{N} \sum_{i=1}^N h(X_i)\right) = \frac{1}{N^2} \sum_{i=1}^N \text{var}(h(X_i)) \\ &= \frac{1}{N} \text{var}(h(X_i)) = \frac{1}{N} \int_a^b [h(x) - E(h(X))]^2 g_X(x) dx \\ &= \frac{1}{N} \int_a^b (h(x) - I)^2 g_X(x) dx = \frac{1}{N} (E(h^2(x)) - I^2)\end{aligned}$$

$$\text{var}(\hat{I}) = \frac{1}{N} (E(h^2(x)) - I^2)$$

■只要满足条件 $E(h^2) < E(f^2)$ ，积分误差就会减小！

再谈Monte Carlo积分

要求 $f(x)$ 的积分

$$\int_a^b f(x) dx$$

而 $f(x)$ 的形式比较复杂积分不好求，考虑

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

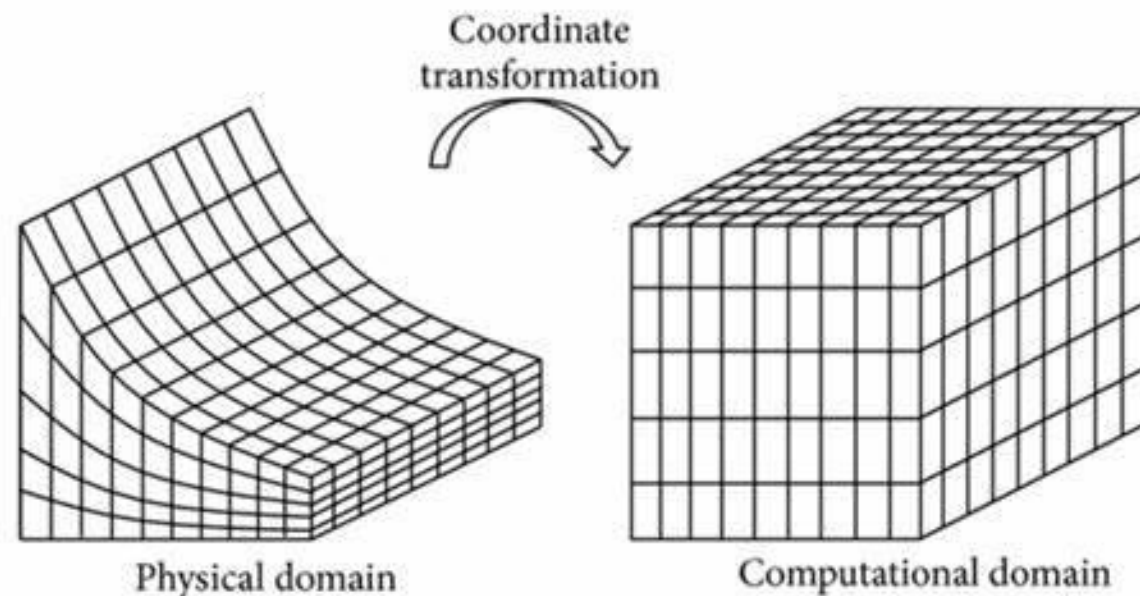
这样把 $g(x)$ 看做是 x 在区间内的概率分布，而把前面的分

数部分看做一个函数，

然后在 $g(x)$ 下抽取 n 个

样本，当 n 足够大时，

即可！



设 \mathcal{O} 为某一空间， n 为产生的总样本数， m 为链条达到平稳时的样本数，则MCMC方法的基本思路可概括为：

(1)构造Markov链。构造一条Markov链，使其收敛到平稳分布 $\pi(x)$ ；

(2)产生样本：由 \mathcal{O} 中的某一点 $x^{(0)}$ 出发，用(1)中的Markov链进行抽样模拟，产生点序列： $x^{(1)}, \dots, x^{(n)}$ ；

(3)蒙特卡洛积分：任一函数 $f(x)$ 的期望估计为： $E[f(t)] = \frac{1}{n-m} \sum_{t=m+1}^n f(x^{(t)})$ 。 [1]

Monte Carlo积分方法总结

优点:

- 受几何空间限制小
- 收敛速度与问题的维数无关
- 具有同时计算多个积分量的能力
- 误差容易确定（概率意义下）
- 程序简单

缺点

- 收敛速度慢
- 误差具有概率性

1. 随机投点法
2. 样本平均法
3. 重要性抽样法

**Calculate when you can,
Simulate when you can't!**

附录

3. 控制变量

控制变量的技巧 统计量

不改变概率分布，只改变统计量。做变换：

$$\underline{h^*(x)} = h(x) + \alpha(E[h_1] - h_1(x))$$

其中， α 为任意常数， $h_1(x)$ 为已知的控制变量。模拟n次：

$$\hat{h}^* = \frac{1}{n} \sum_{i=1}^n h^*(X_i) \quad \text{—— } h^* \text{ 的估计值}$$

1. 这里， $h(x)$ 与 $h_1(x)$ 可能正或负相关
2. 通过设定 α 最优值的正负来实现 h_1 对 h 的“控制”

3. 控制变量 - 统计量方差“缩减”

$$\text{Var}[h^*] = \text{Var}[h] + \alpha^2 \text{Var}[h_1] - 2\alpha \text{Cov}[h(X), h_1(X)]$$

α 的最优值取为: $\alpha^* = \text{Cov}[h(X), h_1(X)] / \text{Var}[h_1] = \text{Corr}[h(X), h_1(X)] \sqrt{\frac{\text{Var}[h]}{\text{Var}[h_1]}}$

此时, 有: $\text{Var}[h^*] = \text{Var}[h] - (\text{Cov}[h(X), h_1(X)])^2 / \text{Var}[h_1]$

$$= (1 - (\text{Corr}[h(X), h_1(X)])^2) \text{Var}[h]$$

最终, 可得 技巧统计量 与 直接模拟统计量 方差之比值为:

$$\frac{\text{Var}[h^*]}{\text{Var}[h]} = 1 - \frac{(\text{Cov}[h(X), h_1(X)])^2}{\text{Var}[h] \text{Var}[h_1]} = 1 - (\text{Corr}[h(X), h_1(X)])^2$$

控制变量法

与重要性抽样法相似的是控制变量法，它也需要找一个与被积函数 f 的形态相近似的可积函数 g ，然而在这里，我们是将两个函数相减，根据积分运算的线性性质

$$\int f(x)dx = \int (f(x) - g(x))dx + \int g(x)dx$$

- 选择 g 要求：积分易于计算！
- 上式右边第一项积分中 $(f-g)$ 方差比原来 f 的积分方差有显著减小
- 当 g_x 趋于0，被积函数 f/g_x 将可能趋于无穷，计算稳定性和精度都将下降
- 另一方面，该方法也省掉了从概率密度函数求分布函数这一计算

4.对偶随机变量

不改变概率分布，只改变统计量。对偶随机变量技巧的统计量：

$$h^*(x) = (1/2)(h_1(x) + h_2(x))$$

其中 $h_1(x), h_2(x)$ 负相关，称为对偶随机变量。抽样 n 次，统计量 h^* 的估计值为：

$$\hat{h} = (1/n) \sum_{i=1}^n h^*(X_i) = (1/2n) \sum_{i=1}^n (h_1(X_i) + h_2(X_i))$$

$h_1(x), h_2(x)$ 负相关越大，方差减小越大。

4.对偶随机变量-负相关统计量

考虑随机变量 $X \sim U(0,1)$, 统计量 h_1, h_2 是随机数 U 的函数:

$$h_1(x) = h_1(U), h_2(x) = h_2(1 - U)$$

对偶随机变量技巧的统计量:

$$h^*(X) = (1/2)(h_1(U) + h_2(1 - U))$$

其估计量为:

$$\hat{h}^* = (1/2n) \sum_{i=1}^n (h_1(U_i) + h_2(1 - U_i))$$

• 若进一步地, $X \sim N(\mu, \sigma^2)$, 则:

$$h_1(x) = h_1(X), h_2(x) = h_2(2\mu - X)$$

$$\hat{h}^* = (1/2n) \sum_{i=1}^n (h_1(X_i) + h_2(2\mu - X_i))$$

4.对偶随机变量-统计量方差

对偶随机变量技巧统计量方差为：

$$\text{Var}[h^*] = (1/4)(\text{Var}[h_1] + \text{Var}[h_2] + \text{Cov}[h_1(X), h_2(X)])$$

其中协方差为：

$$\text{Cov}[h_1(X), h_2(X)] = \text{Corr}[h_1(X), h_2(X)]\sqrt{\text{Var}[h_1]\text{Var}[h_2]}$$

若 $\text{Var}[h_1] = \text{Var}[h_2]$ ，则有

$$\text{Var}[h^*] = (1/2)(\text{Var}[h_1] + \text{Cov}[h_1(X), h_2(X)])$$

直接模拟方法的方差为 $(1/2)\text{Var}[h_1]$

若 h_1, h_2 负相关，则 $\text{Cov}[h_1(X), h_2(X)] < 0$ ，有

$$\text{Var}[h^*] < (1/2)\text{Var}[h_1]$$

5.公共随机数

不改变概率分布，只改变统计量。统计量改为：

$$h(X) = h_1(X) - h_2(X)$$

统计量的估计值为：

$$E[h] = E[h_1] - E[h_2]$$

两个统计量使用公共的随机数，即随机变量使用相同的样本值。

直接模拟方法统计量的方差为

$$\text{Var}[h_1] + \text{Var}[h_2]$$

公共随机数技巧统计量方差为

$$\text{Var}[h_1 - h_2] = \text{Var}[h_1] + \text{Var}[h_2] - 2\text{Cov}[h_1(X), h_2(X)]$$

若 h_1 ， h_2 正相关，即协方差为正值，则有

$$\text{Var}[h_1 - h_2] < \text{Var}[h_1] + \text{Var}[h_2]$$

6.条件期望

同时改变概率分布和统计量。条件方差公式：

$$\text{Var}[X] = \text{Var}[E[X|Z]] + E[\text{Var}[X|Z]]$$

由于方差总是非负的，有

$$\text{Var}[X] \geq \text{Var}[E[X|Z]]$$

有两个相关的随机变量 X, Z ，其概率分布分别为 $f(x), g(z)$ 。模拟 n 次，条件期望技巧算法如下：

- 1.由 $g(z)$ 抽样产生样本值 $Z_1, Z_2 \dots Z_n$
- 2.解析计算 $E[X|Z_i]$
- 3.随机变量 X 的估计值 $\hat{X} = (1/n) \sum_{i=1}^n E[X|Z_i]$

统计量：

统计量 $h(x)$ 是随机变量 X 的函数，所以也是随机变量，其条件方差为：

$$\text{Var}[h] = \text{Var}[E[h|Z]] + E[\text{Var}[h|Z]]$$

因此

$$\text{Var}[E[h|Z]] < \text{Var}[h]$$

其中 $\text{Var}[h]$ 是直接模拟方法统计量方差。

统计量 $h(x)$ 的估计值为

$$\bullet \hat{h} = (1/n) \sum_{i=1}^n E[h|Z_i = z_i]$$

7. 样本分裂

不改变概率分布，只改变统计量。把区域A分为重要区域S和不重要区域R。

当粒子向重要方向运动时，每通过一层，粒子分裂一次，变成 ν 个粒子，每个粒子的能量和方向保持不变，权重变为 $1/\nu$ 。第 k 个分裂粒子的贡献为 I_k ，记录 ν 个粒子的贡献。

当粒子向不重要方向运动的 h_i ，每通过一层进行轮盘赌，把 ν 个能量和方向相同的粒子压缩成一个粒子，权重为 ν ，以概率 $q = 1/\nu$ 继续跟踪这个粒子，记录一个粒子贡献为 I_1 ，以概率 $1 - q$ 结束游动历史。

第 i 次模拟第 j 次碰撞，分裂和轮盘赌技巧的贡献表示成两部分之和：

$$I_{ij} = \frac{1}{\nu} \sum_{k=2}^{\nu} I_k \eta(A \in S) + (I_1/q) \eta(A \in R, U \leq q).$$

模拟 n 次，每次模拟粒子碰撞 m 次，粒子贡献估计为：

$$\hat{I} = (1/n) \sum_{i=1}^n \sum_{j=0}^m I_{ij}$$

Homework 06

1. 课后作业 P221: 2, 5, 6
2. 请结合本讲内容和课本第十章:
 - ① 整理随机投点法和样本平均值法算法实现
 - ② 叙述两者的收敛性证明以及误差估计
3. 采用一种方差缩减技术提高模拟效率, 记录并比较所得结果。