蒙特卡罗（Monte carlo）方法

【实验目的】

- 了解 **Monte carlo** 随机摸拟方法的具体应用
- 学习使用 **MATLAB** 软件中有关随机数生成函数

【实验准备与内容】

1．Monte carlo 方法

蒙特卡罗(**Monte Carlo**)方法，又称计算机随机模拟方法，是一种基于"随机数"的计算方法。这一方法源于美国在第二次世界大战研制原子弹的"曼哈顿计划"。该计划的主持人之一数学家冯诺伊曼用驰名世界的赌城-摩纳哥的 **Monte Carlo** 来命名这种方法。

Monte Carlo 方法的基本思想很早以前就被人们所发现和利用。早在 **17** 世纪，人们就知道用事件发生的"频率"来决定事件的"概率"。**1777** 年，蒲丰（**Buffon**）提出著名的 **Buffon** 投针试验永来近

似计算圆周率 **π**。本世纪 **40** 年代电子计算机的出现，特别是近年来高速电子计算机的出现，使得人们在计算机上利用数学方法大量、快速地模拟这样的试验成为可能。目前这一方法已经广泛地运用到数学、物理、管理、生物遗传、社会科学等领域，并显示出特殊的优越性。

2．伪随机数

实际应用中的随机数通常都是某些数学公式计算而产生的伪随机数，这样的伪随机数从数学意义上讲不是严格的随机数，但是，只要伪随机数能够通过随机数的一系列统计检验，我们就可以将它当作真随机数而放心使用。这样我们就可以方便、经济、重复地产生随机数。理论上要求伪随机数产生器具备以下特征：良好的统计分布特性，高效率的伪随机数产生，伪随机数产生的循环周期长，伪随机数可以重复产生等。到目前为止，已经提出了各种分布的伪随机数产生方法，在这里，我们不准备从数学上介绍随机数的产生原理，而只给出伪随机数的 **MATLAB** 产生函数。

- 生成各类分布的随机数统一生成函数

random('name',A1,A2,A3,m,n)——生成以 **A1,A2,A3** 为参数分布为 **name** 的 $m \times n$ 阶随机数矩阵，其中，‘**name**’为包含特定分布名称的字符串(如表*.*)，**A1,A2,A3** 是分布参数矩阵，根据分布不同，

各参数的含义也不相同，且其中一些参数也不是必须的。

● 针对不同分布的随机数产生函数

MATLAB 还提供了不同分布的随机数产生函数，这些函数的具体使用格式及参数含义较 **random** 函数更明确（如表*.*）。

表*.*

特定分布字符串‘**name**’	分布名称	随机数产生函数	函数格式
'beta‘	β 分布	**Betarnd**	**Betarnd(A,B,m,n)**
'bino‘	二项分布	**Binornd**	**Binornd(N,P,m,n)**
'chi2'	χ^2 分布	**chi2rnd**	**chi2rnd(V,m,n)**
'exp‘	指数分	**Exprnd**	**Exprnd(MU,m,n)**

	布		
'f'	**F** 分布	**frnd**	**Frnd(V1,V2,m,n)**
'gam'	γ分布	**Gamrnd**	**Gamrnd(A,B,m,n)**
'geo'	几何分布	**Geornd**	**Geornd(P,m,n)**
'hyge'	超几何分布	**Hygernd**	**Hygernd(M,K,N,mm,nn)**
'logn'	对数正态分布	**Lognrnd**	**Lognrnd(Mu,Sigma,m,n)**
'nbin'	负二项分布	**Nbinrnd**	**Nbinrnd((R,P,m,n)**
'ncf‘	非中心 **F** 分布	**Ncfrnd**	**Ncfrnd(Nu1,Nu2,Delta,m,n)**

'nct‘	非中心 **t** 分布	**Nctrnd**	**Nctrnd(V,Delta,m,n)**
'ncx2'	非中心 χ^2 分布	**ncx2rnd**	**ncx2rnd(V,Delta,m,n)**
'norm'	正态分布	**Normrnd**	**Normrnd(Mu,Sigma,m,n)**
'poiss'	泊松分布	**Poissrnd**	**Poissrnd(Lambda,m,n)**
'rayl'	**Rayleigh** 分布	**raylrnd**	**Raylrnd(B,m,n)**
't'	**t** 分布	**Trnd**	**Trnd(V,m,n)**
'unif'	连续均匀分布	**unifrnd**	**Unifrnd(A,B,m,n)**

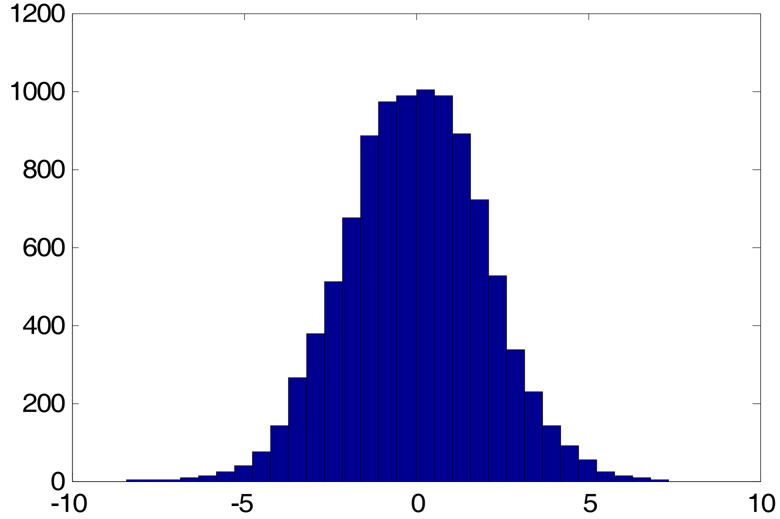
'unid'	离散均匀分布	**Unidrnd**	**Unidrnd(N,m,n)**
'weib‘	**Weibull** 分布	**weibrnd**	**Weibrnd(A,B,m,n)**

为了说明函数的使用方法，我们举两个例子。

例：用 **random** 函数生成均值为 **0**，标准差为 **2** 的正态分布随机数。

rn=random('norm',0,2,10000,1); ％生成正态分布随机数

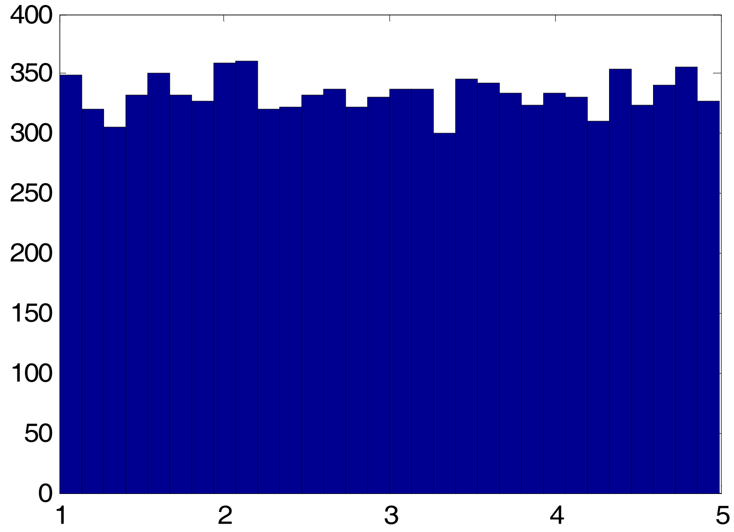
hist(rn，30) ％绘出随机数的直方图



例：生成区间[1,5]的服从均匀分布的随机数

rn=unifrnd(1,5,10000,1); ％生成均匀分布随机数

hist(rn，30) ％绘出随机数的直方图

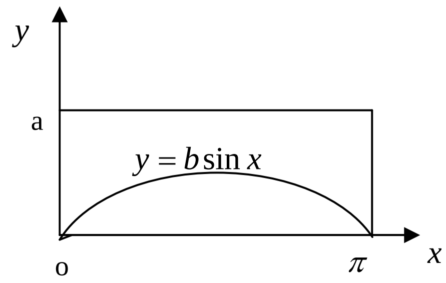


3. Buffon 投针试验

在平滑的桌面上画一组相距 **2a** 的平行线束，向此桌面上投一枚长为 **2b** 的细针，为避免针与两平行线同时相交的复杂情况，假定$a>b>0$。现在，我们来求针与平行线相交的概率。

设 **M** 为针的中点，**y** 为 **M** 与最近平行线的距离，**x** 为与针与平行线的交角（$0\le x\le \pi$），于是，针与平行线相交的充要条件是$0\le y\le b\sin x$,故相交的概率为图中曲线$y=b\sin x$与x轴所夹图形的面积占图

中长方形面积的百分比。



即

$$p = \frac{1}{\pi a}\int_0^{\pi} b\sin x dx = \frac{2b}{\pi a}$$

从而

$$\pi = \frac{2b}{pa}$$

用 **N** 表示投针的次数，**n** 表示其中针与平行线相交的次数，由贝努里（**Bernoulli**）定理知，当 **N** 充分大时，频率接近于概率，即 $p \approx \frac{n}{N}$，从而 $\pi \approx \frac{2bN}{na}$。这就是 **Buffon** 用随机实验近似求圆周率 π 的基本公式。

我们不妨用产生随机数的办法来模拟投针实验：对每一次投针实验，针与平行线的交角x（$0 \le x \le \pi$）和针的中点 **M** 与最近平行线的距离y（$0 \le y \le a$）是一随机变量，其分别服从$[0,\pi]$和$[0,a]$上的均匀分布。因此，**N** 次投针实验可以通过以上产生随机数的函数生成，然后根据针与直线相交的充要条件$0 \le y \le b\sin x$，即可判断针与直线是否相交，从而可以计算出 **n** 的值，达到近似计算圆周率π的目的。具体程序如下

```
function [pai,number]=buffon(a,b,N)
% 2a，2b 分别为平行线间距和针长，N 为投针次数
x=unifrnd(0,pi,N,1);
y=unifrnd(0,a,N,1);
number=0; %  相交计数器
for i=1:N
if y(i)<=b*sin(x(i))
number=number+1;
```

```
end
end
pai=2*b*N/(a*number)
```

其计算结果如下：

a	b	N	相交次数	π的估计值
45	36	10000	5139	3.1134
45	36	100000	50992	3.1377
45	36	500000	254722	3.1407

4．积分计算

定积分的计算是 **Monte Carlo** 方法引入计算数学的开端，在实际问题中，许多需要计算多重积分的复杂问题，用 **Monte Carlo** 方法一般都能够很有效地予以解决，尽管 **Monte Carlo** 方法计算结果的精度不很高，但它能很快地提供出一个低精度的模拟结果也是很有价值的。而且，在多重积分中，由于 **Monte Carlo** 方法的计算误差与积分重数无关，因此它比常用的均匀网格求积公式要优越。

4．1 随机投点法

（1） 设计算的定积分为 $I=\int_a^b f(x)dx$，其中 a,b 为有限数，被积函数 $f(x)$ 是连续随机变量 ς 的概率密度函数，因此 $f(x)$ 满足如下条件：

$$f(x)\text{ 非负，且 }\int_{-\infty}^{+\infty} f(x)dx=1 \qquad (1)$$

显然 I 是一个概率积分，其积分值等于概率 $P(a\le\varsigma<b)$。

下面按给定分布 $f(x)$ 随机投点的办法，给出如下 **Monte Carlo** 近似求积算法：

step1： 产生服从给定分布的随机变量值 $x_i, i=1,2,\ldots,N$；

step2: 检查x_i是否落入积分区间。如果条件$a \le x_i < b$满足，则记录x_i落入积分区间一次。

假设在 **N** 次实验以后，x_i落入积分区间的总次数为 **n**，那么用

$$\bar{I} = \frac{n}{N}$$

作为概率积分的近似值，即

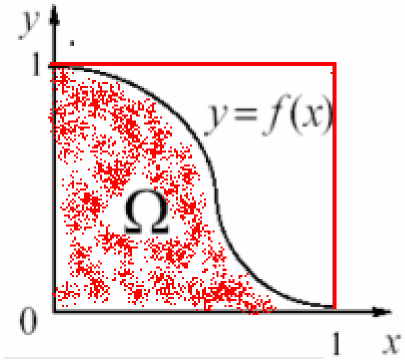
$$I \approx \frac{n}{N}$$

（**2**）如果要计算的定积分为$I = \int_a^b f(x)dx$，其中a,b为有限数，但被积函数$f(x)$不是某随机变量的的概率密度函数。在这种情况下，当然我们可以对积分作变换，将积分变换成满足（**1**）的条件。但在变换以后，需要产生新的分布随机变量，因此常会遇到很多困难和比较复杂的计算。

上述求积方法需要产生给定分布的随机变量，它适合于解决特殊类型的概率积分。如果只用随机数完成随机投点，那么，下面的方法可以解决较为广泛的一类积分问题。

（3） 设 $f(x)$ 是[0,1]上的连续函数，且 $0 \le g(x) \le 1$，需计算积分 $I = \int_a^b f(x)dx$，如下图所示阴影部分的面积。在图中单位正方形内均匀地投点 (ς,η)，则该随机点落入曲线 $y = f(x)$ 下面的概率为

$$P\{y \le f(x)\} = \int_0^1 f(x)dx = I$$



因此，给出如下 **Monte Carlo** 近似求积算法：

step1：产生两组[0，1]区间内的随机数 $x_i, y_i, i = 1,2,\ldots,N$，并把 (x_i, y_i) 作为随机点 (ς,η) 的可取坐标；

step2：检查（x_i， y_i）是否落入 Ω 内，如果条件 $y_i \le f(x_i)$ 满足，则记录（x_i，y_i）落入积分区间一次。

假设在 **N** 次实验以中，落入 Ω 内总次数为 **n**，那么量

$$\bar{I} = \frac{n}{N}$$

近似等于随机点落入 Ω 内的概率，即

$$I \approx \frac{n}{N}$$

假如所需计算积分 $I=\int_a^b f(x)dx$，其中 a,b 为有限数，被积函数 $f(x)$ 有界，并用 M 和 m 分别表示其最大值和最小值。作变换 $x=a+(b-a)x^*$， $f^*(x^*)=\frac{1}{M-m}[f(a+(b-a)x^*-m]$，此时

$$I=(M-m)(b-a)\int_0^1 f^*(x)dx+m(b-a)$$

且有 $0 \le f^*(x^*) \le 1, x \in [0,1]$，即转化为上面讨论过的情况。

4．2 平均值法

任取一组相互独立、同分布的随机变量 $\{\xi_i\}$， ξ_i 在**[a,b]**内服从分布率 $p(x)$，令 $p^*(x)=\frac{f(x)}{p(x)}$，则 $\{p^*(\xi_i)\}$ 也是一组相互独立、同分布的随机变量，而且

$$E\{p^*(\xi_i)\}=\int_a^b p^*(x)p(x)dx=\int_a^b f(x)dx=I$$

由强大数定理

$$P\left(\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}p^*(\xi_i)=I\right)=1$$

若记 $\frac{1}{N}\sum_{i=1}^{N}p^*(\xi_i)=\bar{I}$，则 $\bar{I}$ 依概率 **1** 收敛到 I，平均值法就是用 $\bar{I}$ 作为 I 的近似值。

假如所需计算积分为 $I=\int_a^b f(x)dx$，其中被积函数在**[a,b]**内可积，任意选择一个有简单办法可以进行抽样的概率密度函数 $p(x)$，使其满足条件：

- $p(x)\neq 0,$ 当 $f(x)\neq 0$ 时（ $a\le x\le b$ ）
- $\int_a^b p(x)dx=1$

记 $p^*(x)=\begin{cases}\dfrac{f(x)}{p(x)}, p(x)\neq 0,\\ 0, \quad \mathrm{p(x)}=0.\end{cases}$ 则所求积分为

$$I=\int_a^b p^*(x)p(x)dx$$

因而 **Monte Carlo** 近似求积算法为：

step1： 产生服从分布律 $p(x)$ 的随机数 $x_i, i=1,2,\dots,N$；

step2： 计算均值 $\bar{I}=\frac{1}{N}\sum_{i=1}^{N}p^*(\xi_i)$，即有 $\bar{I}\approx I$。

如果 a,b 为有限数，那么 $p(x)$ 可以取为均匀分布：

$$p(x)=\begin{cases}\frac{1}{b-a}, a\le x\le b\\ 0, \text{其它}\end{cases}$$

此时 $I=(b-a)\int_a^b f(x)\frac{1}{b-a}dx$，具体的算法为：

step1： 产生**[a,b]**上的均匀随机数 $x_i, i=1,2,\dots,N$；

step2： 计算均值 $\bar{I}=\frac{(b-a)}{N}\sum_{i=1}^{N}f(x_i)$，且有 $I\cong\bar{I}$。

4．3 多重积分的计算

根据上面定积分的 **Monte carlo** 计算方法，我们可以很容易地将它推广到$n(n\geq 2)$重积分的近似计算。为了避免重复，在此只给出具体的二重和三重积分计算公式。

- 二重积分的计算

$$\iint_D f(x,y)dxdy \cong \frac{|D|}{N}\sum_{i=1}^{N} f(x_i,y_i)$$

其中，(x_i,y_i)是平面区域 **D** 中的均匀分布的 **N** 个独立选取的随机点列，$|D|$表示区域 **D** 的面积。

- 三重积分的计算

$$\iiint_\Omega f(x,y,z)dxdydz \cong \frac{|\Omega|}{N}\sum_{i=1}^{N} f(x_i,y_i,z_i)$$

其中，(x_i,y_i,z_i)是空间区域Ω中的均匀分布的 **N** 个独立选取的随机点列，$|\Omega|$表示区域的Ω体积。很显然，该方法特别适合于多重积分，特别是积分区域比较复杂的重积分计算。

4．4 误差估计

Monte Carlo 求积方法常常以某个随机变量θ的简单子样$\theta_1,\theta_2,\cdots,\theta_N$的算术平均值$\frac{1}{N}\sum_{i=1}^{N}\theta_i=\bar{\theta}$作为积分$I$的近似值，由强大数定理知道，如果随机变量序列$\theta_1,\theta_2,\cdots,\theta_N$相互独立、同分布、期望值存在，那么当$N\to\infty$时，$\bar{\theta}$以概率 **1** 收敛到$I$。按照中心极限定理，只要随机变量序列$\theta_1,\theta_2,\cdots,\theta_N$相互独立、同分布、数学期望存在，且具有有限标准差$\sigma\neq 0$，那么，当当$N\to\infty$时，随机变量

$$Y=\frac{\bar{\theta}-I}{\sigma/\sqrt{N}}$$

渐进地服从标准正态分布$N(0,1)$，即

$$P(Y<t_\alpha)\cong\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t_\alpha}e^{-\frac{x^2}{2}}dx$$

因此，对任何$t_\alpha>0$，

$$P(|Y|<t_\alpha)=P\left(|\bar{\theta}-I|<\frac{t_\alpha\sigma}{\sqrt{N}}\right)\cong\frac{2}{\sqrt{2\pi}}\int_{0}^{t_\alpha}e^{-\frac{x^2}{2}}dx=1-\alpha$$

这表示，不等式$|\bar{\theta}-I|<\frac{t_\alpha\sigma}{\sqrt{N}}$成立的概率近似地等于$1-\alpha$。其中当$\alpha$很小时，$\alpha$称为显著水平，$1-\alpha$称为

置信水平。因此 **Monte Carlo** 方法的误差ε可以写为

$$\varepsilon = \frac{t_\alpha \sigma}{\sqrt{N}}$$

由此可以确定，对于给定的ε，**N** 的值应取为（在概率意义下的统计估计值）

$$N = \frac{t^2{}_\alpha \sigma^2}{\varepsilon^2}$$

但是，在一般意义下，我们无法知道理论值σ，所以在实际计算中，只能用标准差的近似估计值$\bar{\sigma}$来代替σ。具体地说，对于随机投点法，$\bar{\sigma} = \sqrt{\bar{I}(1-\bar{I})}$；对于平均值法，$\bar{\sigma} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} p^{*2}(x_i) - (\frac{1}{N}\sum_{i=1}^{N} p^{*}(x_i))^2}$ 。

5．库存管理问题

某种商品进货价格为 **a** 元，出售价格为 **b** 元，假设对该商品每天早晨进货配齐 **n** 个，每日顾客相互独立地到来，平均每日 **m** 人，且服从普阿松（**Poisson**）分布 $P(k)=\dfrac{m^k e^{-m}}{k!}$，其中 **k** 为每日到来的人数。问当 **a=2**、**b＝3**、**m=10** 时，每天早晨该商品备齐多少个可得到最大利润。

假设购买此商品的顾客每人只购一个，而且备齐的商品如果当天卖不出去则不可在第二天出售，若用 $a(n,k)$ 表示顾客为 k 人时的日利润，则

$$a(n,k)=\begin{cases} kb-na, & k<n \\ nb-na, & k\ge n \end{cases}$$

利用 **MATLAB** 给定的 **Poisson** 随机数产生函数 **poissrnd**，生成含有 **N** 个元素随机数序列 $\{k_i\}_{i=1}^{N}$，这样平均利润为 $\dfrac{\sum_i^N a(n,k_i)}{N}$。下面给出了 **MATLAB** 程序：

```
function y=lirun(a,b,m,n,N)
% N 为产生随机数的个数
totol=0; %totol 为总利润
rand('state',sum(100*clock));
randnumber=poissrnd(m,N,1); %产生 Poisson 随机数
for i=1:N
    if randnumber(i)<n
        totol=totol+randnumber(i)*b-n*a;
    else
        totol=totol+n*b-n*a;
    end
end
y=totol/N; %y 为平均利润
```

实际上，上述是用 **Monte Carlo** 方法计算出的结果，当然，我们也可以利用如下方法精确计算每日的平均利润。即平均利润 $A(n)$ 为

$$A(n)=\sum_{k=0}^{\infty} a(n,k)P(k)=\sum_{k=0}^{n-1} (kb-na)P(k)+\sum_{k=n}^{\infty} (nb-na)P(k)$$

由于 $\sum_{k=0}^{\infty} P(k)=1$，上式可以进一步表示为

$$A(n)=\sum_{k=0}^{n-1} (k-n)bP(k)+n(b-a)$$

MATLAB 程序如下

```
function y=pp(a,b,n,m)
summ=n*(b-a);
for k=0:n-1
summ=summ+(k-n)*b*poisspdf(k,m); %poisspdf 为 Poission 分布律
end
```

```
y=summ;
```

其结果为

n	6	7	8	9	10	11	12	13	14	15
平均利润	**5.6880**	**6.1840**	**6.7280**	$\boxed{\mathbf{6.8160}}$	**5.5840**	**5.0000**	**4.0800**	**2.8480**	**1.4000**	**-0.3120**
精确值	**5.6700**	**6.2796**	**6.6189**	$\boxed{\mathbf{6.6205}}$	**6.2467**	**5.4976**	**4.4073**	**3.0326**	**1.4392**	**-0.3104**

从上表可以看出，当每天进货 **9** 个时，利润最大，此时精确值和利用 **Monte Carlo** 方法得到的结果相同，但具体的利润有些差异。

【练习与思考】

1. 利用 Monte Carlo 方法，模拟掷骰子各面出现的概率。
2. 编写一个福利彩票电脑选号的程序。
3. 用 Monte Carlo 求积方法，计算定积分 $\int_0^{\pi} e^x \sin^2 x dx$ 的近似值，并分别就不同个数的随机点数比较积分值的精度。
4. 用 Monte Carlo 求积方法，计算二重积分 $\int_1^2 dx \int_2^3 e^{x-y} dy$ 的近似值，并分别就不同个数的随机点数比较积分值的精度。
5. 使用 Monte Carlo 方法求曲面 $x^2+y^2+z^2 \le 1$ 所围成的体积。
6. 炮弹射击的目标为一椭圆 $\frac{x^2}{120^2}+\frac{y^2}{80^2}=1$ 所围成的区域的中心，当瞄准目标的中心发射时，受到各种因素的影响，炮弹着地点与目标中心有随机偏差。设炮弹着地点围绕目标中心呈二维正态分布，且偏差的标准差在 x 和 y 方向均为 100 米，并相互独立，计算炮弹落在椭圆形区域内的概率。

7. 某企业生产易变质的产品，每天生成的产品必须当天售出，否则就会变质。该产品的单位成本为 **2.5** 元，单位产品的售价为 **5** 元。市场对该产品的每天需求量是一随机变量，且服从正态分布 **N(135,** 22.4^2) 。企业为了避免存货过多而造成损失，拟从以下两个方案中选出一个较优的方案。

方案 **1**：按前一天的销售量作为当天的货存量；

方案 **2**：按前两天的平均销售量作为当天的货存量。

试用 **Monte Carlo** 方法确定该选择以上哪一种方案？