第五章 蒙特卡罗方法及其应用

5.1 蒙特卡罗方法

5.2 蒙特卡罗方法应用

——投资项目风险分析

5.1 蒙特卡罗方法

5.1.1 引言

Monte Carlo模拟的应用:

自然现象的模拟:

宇宙射线在地球大气中的传输过程;

高能物理实验中的核相互作用过程;

工程实验与数值分析:

利用Monte Carlo方法求积分

经济管理问题分析:

利用Monte Carlo方法分析投资项目的风险,
研究金融衍生产品定价和风险估算等

系统仿真讲义下载网址
http://its.asom.sjtu.edu.cn/

5.1 蒙特卡罗方法

5.1.1 引言

- Monte—Carlo 方法的基本思想是当实验次数充分多时，某一事件出现的频率近似于该事件发生的概率，即

$$\frac{v}{N} \approx p \text{（当 N 充分大时）}$$

式中 p—某一事件发生的概率;

N—实验次数;

v—在 N 次实验中该事件出现的次数。

- 当所求解的问题是某种事件发生的概率或某一随机变量的数学期望，或其它数字特征时，通过实验方法可以得到该事件发生的样本频率或样本均值；当实验次数足够多时，通过统计推断，可以获得样本参数代表总体参数的置信度，或置信区间。

5.1 蒙特卡罗方法

5.1.1 引言

Monte Carlo方法:

亦称统计模拟方法，statistical simulation method

→利用随机数进行数值模拟的方法

Monte Carlo名字的由来:

- 是由Metropolis在二次世界大战期间提出的：Manhattan 计划，研究与原子弹有关的中子输运过程;
- Monte Carlo是摩纳哥（monaco)的首都，该城以赌博闻名

Nicholas Metropolis (1915-1999)

Monte-Carlo, Monaco

5.1 蒙特卡罗方法

5.1.1 引言

Monte Carlo算法的主要组成部分

✓概率密度函数(pdf)— 必须给出描述一个物理系统的一组概率密度函数;

✓随机数产生器—能够产生在区间[0,1]上均匀分布的随机数

✓抽样规则—如何从在区间[0,1]上均匀分布的随机数出发,随机抽取服从给定分布函数的随机变量;

✓模拟结果记录—记录一些感兴趣的量的模拟结果

✓误差估计—必须确定统计误差（或方差）随模拟次数以及其它一些量的变化;

✓减少方差的技术—利用该技术可减少模拟过程中计算的次数;

✓并行和矢量化—可以在先进的并行计算机上运行的有效算法

5.1 蒙特卡罗方法

5.1.2 Boffon投针实验

● Boffon 在平面上绘制相距均为 $2a$ 的平行线束，在一定高度向平面上随机投下一枚长度为 $2l$ 的钢针，投针结果有两种，即与平行线相交或不相交，为了避免钢针同时与两根相邻平行线相交。

● 设 $a > l > 0$，取 M—钢针的中点；y—M 点与最近平行线之间的距离；

φ —钢针与平行线之间的夹角（$0° \le \varphi \le 180°$）。

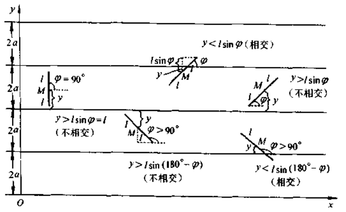


图5－1 Boffon随机投针实验结果

5.1 蒙特卡罗方法

5.1.3 样本量的确定

➢保证 Monte—Carlo 方法的精度，需要一定的样本量支持

若每次投针实验的结果用随机变量 $X_n(\omega)$ 表示，由于每次投针实验的结果与以前各次投针结果无关，故 $X_1(\omega)$，$X_2(\omega)$，……为独立同分布随机变量，

$$X_n(\omega)=\begin{cases}1, & 若\ \omega=\{相交\} \\ 0, & 若\ \omega=\{不相交\}\end{cases}$$，令 $P\{X_n(\omega)=1\}=p$ ，$P\{X_n(\omega)=0\}=q$

且 $p+q=1$

当进行一系列 n 次投针实验时，则 $\{X_n(\omega),\ n=1,2,\cdots\cdots\}$ 为伯努里过程。

相应的数学期望和方差分别为 $\mathrm{E}[\mathrm{X_n}]=\sum_{k=0}^{1} kP\{X_n=k\}=p$，

$$\mathrm{V}[\mathrm{X_n}]=E[X_n^2]-(E[X_n])^2=p-p^2=pq$$

5.1 蒙特卡罗方法

5.1.2 Boffon投针实验

● 钢针与平行线相交的条件为

$y \le l\sin\varphi$ （5－1）

● 钢针与平行线相交的概率为 p，则

$p=\frac{1}{\pi a}\int_0^{\pi} l\sin\varphi d\varphi=\frac{2l}{\pi a}$ （5－2）

投针实验中相交的频率接近相交的概率

令 N—投针次数；v—相交次数，则 $p \approx \frac{v}{N}$

$\frac{v}{N} \approx \frac{2l}{\pi a} \rightarrow \pi \approx \frac{2lN}{av}$ （5-3）

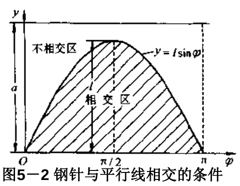


图5—2 钢针与平行线相交的条件

5.1 蒙特卡罗方法

5.1.3 样本量的确定

设共作 N 次投针实验。定义

$$v_N(\omega)=\begin{cases}0, & N=0 \\ X_1(\omega)+X_2(\omega)+\cdots+X_N(\omega), & N=1,2,\cdots\end{cases}$$

为在 N 次投针实验中出现相交的次数，显然 $v_N(\omega)$ 是一个随机变量。由于投针实验属于伯努里实验，故有

$$\begin{aligned}E[v_N]&=E[X_1(\omega)+X_2(\omega)+\cdots+X_N(\omega)]\\&=E[X_1(\omega)]+E[X_2(\omega)]+,\cdots,+E[X_N(\omega)]\\&=Np\end{aligned}$$

$$\begin{aligned}V[v_N]&=V[X_1(\omega)]+V[X_2(\omega)]+\cdots+V[X_N(\omega)]\\&=Npq=Np(1-p)\end{aligned}$$

显然，投针实验的精度取决于 $V\left[\frac{v_N}{N}\right]$，

$$V\left[\frac{v_N}{N}\right]=\frac{1}{N^2}Np(1-p)=\frac{1}{N}p(1-p) \quad （5—5）$$

5.1 蒙特卡罗方法

5.1.2 Boffon投针实验

•历史上几位学者的实验结果

Wolf (1853) N=5000, $\hat{\pi} \approx 3.1596$

Smith (1855) N = 3204, $\hat{\pi} \approx 3.1553$

FOX (1894) N = 1120, $\hat{\pi} \approx 3.1419$

Lazzarini (1901) N=3408, $\hat{\pi} \approx 3.1415929$

裴鹿成和张孝泽利用计算机进行投针实验，N=50 × 10^4，$\hat{\pi} \approx 3.1422$

由中心极限定理，当 $\frac{v_N}{N}$ 为独立同分布随机变量，且 $\sigma \ne 0$ 时，N 充分大时，统计量 $z=\frac{(v_N/N)-p}{\sqrt{V\left[\frac{v_N}{N}\right]}} \sim N(0,1)$，因此有 $P=\{|z| \le z_{1-\alpha/2}\}=1-\alpha$，式中：$1-\alpha$ —置信度；

$z_{1-\alpha/2}$ ——置信度为 $1-\alpha$ 的正态分布临界值。则 $P\left\{\left|\frac{\left(\frac{v_N}{N}-p\right)}{\sqrt{V\left[\frac{v_N}{N}\right]}}\right| \le Z_{1-\alpha/2}\right\}=1-\alpha$

如果要求实验的精度 $\left|\frac{v_N}{N}-p\right|$ 达到 0.001，且对实验的置信程度达到 95%，即

$\alpha=0.05$，由正态分布数值表可查得 $z_{0.975}=1.96$，故 $Z_{1-\alpha/2}\sqrt{V\left[\frac{v_N}{N}\right]}=0.001$

取 $l=a$ 这一极端情况，即针的长度与平行线间距相等，则

$$p=\frac{v_N}{N}=\frac{2l}{\pi a}=0.6366 \rightarrow V\left[\frac{v_N}{N}\right]=\frac{1}{N}p(1-p)=\frac{0.2313}{N} \rightarrow N=\frac{1.96^2 \times 0.2313}{0.001^2}$$

即，进行 88.8 万次投针实验，将能以 95% 的置信程度，达到 0.001 的精度来估计 π 的真值。

5.1 蒙特卡罗方法

5.1.3 样本量的确定

➢ 计算机进行投针实验模型

①设 y 和 φ 都是随机变量，y 的取值范围为(0, 1)，φ 的取值范围为(0, π)；

②每次实验的结果，可以分别从 $y_i \sim U(0,a)$ 和 $\varphi_i \sim U(0,\pi)$ 中产生均匀分布的随机数，并根据针与平行线相交的条件 $y_i \le l\sin\varphi_i$ 来判别该次投针是否在相交区内（见图 5—2）

③若满足相交条件则置 $v=v+1$，并计算 $\hat{\pi}=\frac{2li}{av}$；

④若 $|\pi-\hat{\pi}|<\varepsilon$（$\varepsilon$ 为要求达到的精度），则计算停止。如果不满足相交条件或达不到精度要求，则返回产生随机数步骤②。

5.1 蒙特卡罗方法

5.1.4 Monte—Carlo方法的收敛性

➢Monte—Carlo 的误差项表明，其收敛速度和精度与系统的结构和样本元素所在空间(变量维数)无关，因而对于多变量复杂系统具有根强的适应性。

例，求得 S 维中的任一数域 D_s 上的积分

$$I=\int_{D_s}\cdots\int g(x_1,x_2\cdots x_s)dx_1dx_2\cdots dx_s$$

不论 D_s 的形状如何特殊，只要能给出描述 D_s 的几何条件，如最大值、最小值、边界条件等，用 Monte—Carlo 方法总可以通过实验方法得到

$$\hat{I}_{DN}=\frac{D_s}{N}\sum_{i=1}^{N}g(x_1^{(i)},x_2^{(i)},\cdots x_s^{(i)})$$

式中 $\hat{I}_{DN}$ 是 I 的估计值，$(x_1^{(i)},x_2^{(i)},\cdots x_s^{(i)})\in D_s$ 是第 i 次实验中各变量的随机取值(按均匀分布在 D_s 上取值)。按强大致定理和中心极限定理，当 N 充分大时，$\hat{I}_{DN}$ 将以概率 1 收敛到 I，并以 $(1-\alpha)\%$ 的置信度达到规定的精度。

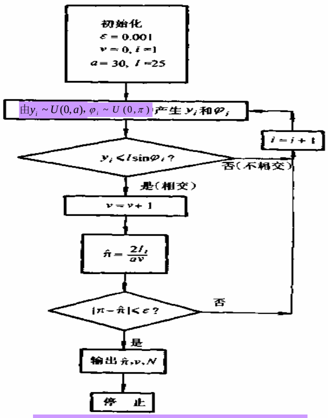


图5—3 计算机投针实验流程图

5.1 蒙特卡罗方法

5.1.4 Monte—Carlo方法的收敛性

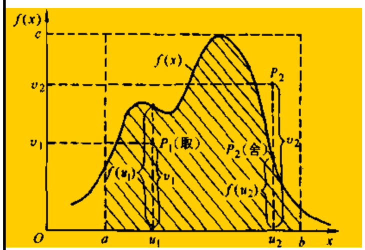


图5—4 蒙特卡罗方法求定积分的值

➢算法步骤：

①置 $i=0$，$v=0$，$N=1000$(设样本量为 1000)；

②置 $i=i+1$，并在区间 (a, b) 和 $(0, c)$ 中分别产生均匀分布的伪随机数 u_i，v_i；

③计算 $f(u_i)$，若 $v_i \le f(u_i)$，该点为“取”，置 $v=v+1$，转步骤④，若 $v_i > f(u_i)$，该点为“舍”，转步骤④；

④若 $i<N$ 返回步骤②，若 $i=N$，转步骤⑤；

⑤置 $\hat{I}=\frac{v}{N}(b-a)c$，停止。

5.1 蒙特卡罗方法

5.1.4 Monte—Carlo方法的收敛性

通常 Monte—Carlo 方法采用大量重复抽样和统计推断来估计系统的参数。样本均值 $\overline{X}_N$ 和样本方差是 E[X]和 v[X]的无偏估计。

由柯尔莫哥洛夫(Kolmogorov)强大数定理有 $P\{\lim_{N\to\infty}|\overline{x}_N - E[X]|=0\}=1$，即平均值 $\overline{X}_N$ 以概率 1 收敛到 E[X]，但样本量 N 必须充分地大。当样本量有限时，$\overline{X}_N$ 和 E[X]之间会存在差异，按中心极限定理有：

$$p\left\{|\overline{x}_N - E[X]|<\frac{Z_{1-\alpha/2}\sigma}{\sqrt{N}}\right\}=1-\alpha$$

即 $|\overline{x}_N - E[X]|<\frac{Z_{1-\alpha/2}\sigma}{\sqrt{N}}$ 以概率 $(1-\alpha)\%$ 成立。或 $\overline{X}_N$ 收敛到 E[X]的速度为 $0(N^{-\frac{1}{2}})$。

如果定义 Monte—Carlo 方法的误差 ε，即 $\varepsilon=\frac{Z_{1-\alpha/2}\sigma}{\sqrt{N}}$，误差 ε 由 σ 和 $\sqrt{N}$ 决定。当 σ 不变时，要提高精度 10 倍，样本量 N 需要增大到 100 倍。

5.2 蒙特卡罗模拟方法应用

—投资项目的不确定性分析

5.2.1 蒙特卡罗风险评估的基本步骤

（1）确定各风险变量

（2）估计风险子变量的分布特征

（3）建立随机模拟函数。利用Excel函数功能，对各风险变量（或子变量）按其分布特征设定随机模拟关系式。然后通过设定的Excel财务关系式，运用其中的“模拟运算”功能，随机模拟计算经济评价指标的若干个值，并对模拟结果加以统计分析。

5.2.2 应用实例—投资项目的不确定性分析

例 某工程投资项目的几个主要参数都具有不确定性，根据专家的意见和调查分析,这些参数的分布和特征值如下：

1.初始投资　正态分布，期望值和标准差分别为 **50 000**万美元和**1 000**美元；

2.研究期(项目的寿命周期)　均匀分布，最短**10**年，最长**14** 年；

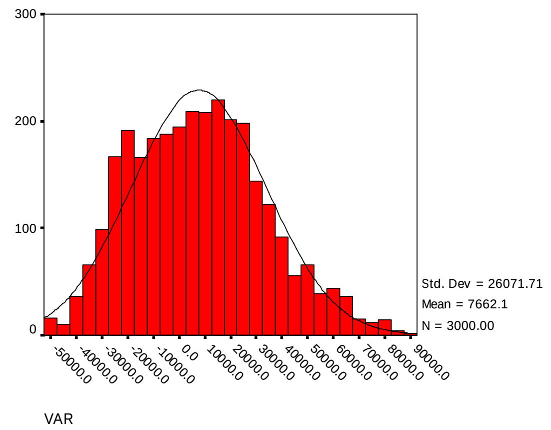
3. 年销售收入　离散分布，有三种可能：

年收入(美元)	概率
35 000	0.4
40 000	0.5
35 000	0.1

4.　年经营成本(包括税收等支出)正态分布,期望值和标准差分别是**30 000**美元和**2000** 美元。

5. 基准贴现率***(MARR)***　确定为**10%**。

试用以上参数模拟该投资项目的净现值***(PW)***的分布和特征值。



解　由于净现值与所给的参数有相当复杂的函数关系，无法用分析的方法求净现值的分布和特征值。可用电子表格进行模拟。根据模拟输出的样本，进行统计拟合，按净现值的分布、估计的期望值和标准差给出净现值*PW*≤*0* 的概率和置信区间。

MARR	10%								
			期望值	标准差					
		初始投资	$50,000	$1,000		年收入期望	概率	累计概率	
		年经营成本	$30,000	$2,000		$35,000	0.4	0.4	
						$40,000	0.5	0.9	
		寿命期	最短	最长		$45,000	0.1	1.0	
			10	14					
模拟次序	标准正态1	初始投资	(0,1)均匀	研究期	(0,1)均匀	年收入	标准正态2	年经营成本	净现值PW
1	-0.7903	49210	0.64101	13	0.9589	45000	-0.7894	28421	68556
2	-0.7871	49213	0.73543	13	0.2876	35000	(0.5730)	28854	-5556
3	-1.3376	48662	0.63459	13	0.7506	40000	1.1416	32283	6153

估计经济指标值在某一范围时的概率

➢经济指标取值的概率分布服从正态分布

$$F(X)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{X}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$$

令$z=\frac{t-\mu}{\sigma}$，上式可化为标准正态分布函数

$$F(X)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{(X-\mu)}{\sigma}}e^{-\frac{u^2}{2}}du=\Phi(\frac{X-\mu}{\sigma})$$

令$Z=\frac{X-\mu}{\sigma}$，由标准正态分布表 ，可直接查出$X<X_0$的概率值

$$P(X<X_0)=P(Z<\frac{X_0-\mu}{\sigma})=\Phi(\frac{X_0-\mu}{\sigma})$$

第11行单元格(背景有阴影的行)的函数和赋值

单元格	函数或赋值	说明
B11	=NORMSINV(RAND())	产生标准正态分布的随机数
C11	=(D3+ E3*B11)	产生初始投资的随机数(正态)
D11	=RAND()	产生(0，1)均匀分布的随机数
E11	= ROUND(D7+D11*($E $7－$D $7)，0)	产生寿命期，10～14均匀分布，取整
F11	=RAND()	产生(0，1)均匀分布的随机数
G11	=IF(F11<=I$4，G$4，IF(F11<=I$5，G$5，G$6))	产生离散分布的年销售收入随机数
H11	= NORMSINV (RAND())	产生标准正态分布的随机数
I11	=(D4十E4*H11)	产生经营成本的随机数(正态)
J11	=－CIl－PV(Bl，E11，G11－I11)	计算并输出净现值

经济指标在某一范围时的概率计算公式

- 方案经济指标 X 小于等于某一取值 X_0时的概率为

$$P(x\le X_0)=P(Z<\frac{X_0-E(X)}{\sigma})$$

- 方案经济指标 X 大于某一取值时的概率：

$$P(x>X_0)=1-P(x\le X_0)$$

- 方案经济指标 X 的取值在 X_1～X_2之间时的概率为

$$P(X_1<x\le X_2)=P(Z<\frac{X_2-E(X)}{\sigma})-P(Z<\frac{X_1-E(X)}{\sigma})$$

$$
\begin{aligned}
P(NPV > 0) &= 1 - P(NPV \le 0) \\
&= 1 - P(Z \le \frac{0 - \bar{X}}{\sigma}) \\
&= 1 - P(Z \le \frac{0 - 7662}{26072}) \\
&= 1 - P(Z \le -0.2939) \\
&= 1 - 0.6141 = 0.3859
\end{aligned}
$$