并行机上的伪随机数发生器

魏公毅 （中科院计算中心） 刘擎宇 （核工业二院）

摘要:本文讨论产生并行随机数的乘同余法递推公式及并行随机数检验方法。在 *Transputer* 并行机上模拟并行随机数产生及检验,其结果令人满意。同时给出二个处理机统计模拟试验的例子,其加速比接近于 2。

1.引言

在计算机上用统计模拟方法(或称 Monte Carlo 方法)解决数学、物理和工程技术问题时,需要使用大量不同分布的随机变量抽样序列。[0,1] 区间上均匀分布随机变量是最基本,最简单,应用最广泛的一种随机变量,通常把[0,1]区间上均匀分布随机变量的抽样值称随机数。

用数学方法递推公式产生的随机数称伪随机数,伪随机数需要经过统计检验才能决定在实际中能否使用。随机数产生及其检验方法都非常重要,其详细内容可参看[1]。

随着并行机的出现,并行算法的研究得到了迅速地发展。经典 Monte Carlo 方法本身就具有很强的内在并行性,它非常适合在并行机上实现。为了在并行机上能实现统计模拟方法,首先就需要并行地产生随机数以供在并行机上完成模拟过程使用。许多作者研究了并行随机数算法,V.C.Bhavsar 和 L.A.Lambrou 讨论了三种方法,即由线性同余序列的非邻接子序列形成的并行随机数序列,伪随机树序列,多重切比雪夫混合变换序列[2,3]。

本文 2 将介绍乘同余法产生随机数的并行算法,3 讨论并行随机数的统计检验方法,4 讨论并行随机数产生及检验在 Transputer 上的具体实现。

2.并行随机数产生

在串行计算机上,人们普遍重视产生伪随机数的数学方法为乘同余法,其递推公式为

$$x_{i+1}\equiv\lambda x_i(\mathrm{mod}\ M) \qquad (1)$$
$$i=0,1,\cdots$$

其中 λ 为乘子,x_0 为初值,M 为模。注意,实际使用〔0,1〕区间上的随机数序列应为 $\{r_i=x_i/M\}$,类似情况不再重复说明。

在向量计算机上,可以构造维数 p 的乘同余向量递推公式

$$X_{i+1}\equiv AX_i(\mathrm{mod}\ M) \qquad (2)$$
$$i=0,1\cdots$$

其中乘子 $A\equiv\lambda^p(\mathrm{mod}\ M)$

初值 $X_0=(x_1,x_2,\cdots,x_p)^T$

而 $x_1,x_2,\cdots,x_p$ 是由(1)产生的前 p 个随机数。

按递推公式(2),产生如下随机数向量:

$X_1=(x_{p+1},x_{p+2},\cdots x_{2p})^T$

;

乘子 A 按下式计算:

$A\equiv\lambda^p(\mathrm{mod}\ M)$

$\equiv(\cdots\cdots((\lambda*\lambda)(\mathrm{mod}\ M)*\lambda)(\mathrm{mod}\ M)\cdots*\lambda)(\mathrm{mod}M)$

对于多处理机情况,假设 p 表示处理机的个数,这时,上述的随机数向量 X_i 的 p 个分量,按顺序分别与 p 台处理机的随机数序列相对应。实际上,每个处理机都使用同样的乘同余法随机数发生器,差别就在于初值不一样。若把式(2)按分量改写,则第 j(j=1,2,…p)个处理机的乘同余法递推公式为

$$y^j_{i+1}\equiv Ay^j_i(\mathrm{mod}\ M) \qquad (3)$$
$$i=0,1\cdots$$

式中 $y^j_0=x_j$ 表示第 j 个处理机随机数序列初值。这样就形成了 p 个并行随机数子序列。

3.并行随机数检验

在串行机上使用随机数序列之前,一般都要先经过各种统计检验才能知道随机数发生器的好坏。我们希望使用独立性、均匀性都满足要求的随机数发生器。统计软件包 SASD 提供了一个随机数检验程序系统 SUTEST,它包括十二类,二十七种不同的统计检验方法,共有六十一个统计量。它从均匀性、随机性、独立性和统计模拟计算等许多不同的方面;为随机数检验提供了完

整、配套的统计检验方法和有效可行的算法。这个程序系统使用灵活、方便,是对随机数进行统计检验的有力工具,详细情况见[4,5]。

对并行随机数序列的基本要求,应当与串行随机数序列一样。首先,在每个处理机上都要产生满足要求的随机数序列,为此在每个处理机上都要使用系统SUTEST进行检验。其次,还要做到各处理机随机数序列相关不显著,这就需要进行相关系数的显著性检验。

假设有随机变量 ξ 和 η,它们之间的线性相关系数为

$$\rho=\frac{COV(\xi,\eta)}{\sqrt{D\xi}\cdot\sqrt{D\eta}}$$

其中 cov(ξ,η)为变量 ξ 和 η 的协方差,Dξ 和 Dη 分别为变量 ξ 和 η 的方差。

有一组变量 ξ 和 η 的观测值序列 $\{\xi_i\}$和 $\{\eta_i\}$(i=1,2,…),其样本相关系数为

$$r=\frac{1/n\cdot\sum_{i=1}^{n}(\xi_i-\bar{\xi})(\eta_i-\bar{\eta})}{\sqrt{1/n\cdot\sum_{i=1}^{n}(\xi_i-\bar{\xi})^2}\cdot\sqrt{1/n\cdot\sum_{i=1}^{n}(\eta_i-\bar{\eta})^2}}$$

$$\bar{\xi}=1/n\cdot\sum_{i=1}^{n}\xi_i$$

$$\bar{\eta}=1/n\cdot\sum_{i=1}^{n}\eta_i$$

给出原假设 H。:$\rho=0$,构造统计量

$$F=\frac{r^2}{1-r^2}(n-2)\qquad(4)$$

当 $\rho=0$ 时,统计量(4)服从自由度为(1,n−2)的 F 分布。假设 Q 表示自由度为(1,n−2)统计量(4)的分布上侧概率值。

若 Q>0.05,则变量 ξ 和 η 线性相关不显著;

若 0.01<Q≤0.05,则变量 ξ 和 η 线性相关显著;

若 Q≤0.01,则变量 ξ 和 η 线性相关极显著。

4. 数值计算

并行随机数产生及检验可以在计算机上模拟实现。在 Transputer 计算机上产生并行随机数时,模拟处理机个数 p=3,每个处理机产生随机数个数 $N=10^5$。[4] 给出两种随机数产生方法,即广义乘同余法及素数模乘同余法。本文使用的是广义乘同余法:

$$y\equiv Ay_i(\mathrm{mod}\ 2^{32})$$

$$y_{i+1}=\begin{cases}y, & \text{当}y<2^{31}\text{时}\\ 2^{32}-y, & \text{当}y>2^{31}\text{时}\end{cases}$$

其中乘子 A=2000762195 (注意,在(1)中,$\lambda=1220703125$,x。=1)

3个处理机随机数初值: 1220703125

839070905

2000762195

SUTEST 系统输出 61 个统计量的检验结果:

处理机编号	检验显著数	检验极显著数
1	2	0
2	3	1
3	4	0

3 个相关系数的检验结果:

检验显著数	检验极显著数
0	0

由上述检验结果看出,用文中方法产生的并行随机数有较好的统计性质。

用统计模拟方法估计如下积分值:

$$\int_0^1 e^{x-1}dx\cong 0.632120558$$

在 Transputer 上使用 1,2 个处理机产生并行随机数,并应用随机投点法估计积分值。计算结果如下:

处理机个数	试验次数	误差	时间(秒)	加速比	效率
1	100000	0.001641	5.972		
2	100000	0.001249	2.989	1.998	0.999

这个例子使用二个处理机,其加速比接近于 2。如果使用更多个处理机将会有更高的加速比。

最后指出,若广义乘同余法失效,可以使用素数模乘同余法,并把全部序列平均分成为 p 段,以形成并行随机数序列。

参考文献:

[1] 中科院计算中心概率统计组编著,概率统计计算,科学出版社,1979.

[2] V.C.Bhavsar,Parallel algorithms for Monte Carlo Solutions of somelinear Operator problems, Dept. of Electrical Engg. , IndianInstitute of Technology, Bombay,ph.D.Thesis(Nov.1981),236pp.

[3] L.A.Lambrou,Pseudo−random number Sequences for Parallel Computers,School of Computer Science,University of New Brunswick,N.B.,M.Sc.Project Report,TR86−033(Feb.1986),192pp.

[4] 中科院计算中心,SASD 统计包(组合模块及部分子程序使用手册),1987.

[5] 张建中,随机数检验程序系统 SUTEST,计算物理,6,3,1989.