任意概率分布的伪随机数研究和实现

朱晓玲,姜 浩
(东南大学 计算机科学与工程学院,江苏 南京 210096)

摘 要:非均匀随机数在计算机仿真、信息安全、自动控制等领域有重要应用,但计算机系统中一般只提供均匀分布的随机数。介绍如何通过均匀分布的伪随机数来产生任意概率分布的伪随机数分理论和方法,包括反函数法、变换法和舍选法,并给出了舍选法的具体实现,最后通过实验结果进行检验。

关键词:任意概率分布;伪随机数;均匀概率分布
中图分类号:TP301.6;O212 **文献标识码**:A **文章编号**:1673-629X(2007)12-0116-03

Study on Pseudo-Random Number of Arbitrariness Probability Distributing and Its Implementation

ZHU Xiao-ling, JIANG Hao
(School of Computer Science & Engineering, Southeast University, Nanjing 210096, China)

Abstract: The arbitrariness probability distributing random number was greatly used in system simulation, information security and automatic controller. Introduced several methods for generating arbitrariness probability distributing random number through uniformly distributing random number offered by the computer system. The implementation of arbitrariness probability distributing random number generator was also given in this paper.

Key words: arbitrariness probability distributing; pseudo-random number; uniformly probability distributing

1 问题的提出

常用的计算机高级程序设计语言大多提供了产生在[0,1]区间内连续均匀分布的独立伪随机数 r 的函数。若将产生的随机数作简单的变换 $X = a + (b - a) * r$,就能得到在区间 $[a,b]$ 上均匀分布的随机数 X。如果与取整或舍入函数结合运用,还可以得到离散均匀分布的随机数。但在计算机仿真、信息安全、自动控制和生物系统识别等领域都需要用到按某些特定规律分布的非均匀随机数,如二项分布、泊松分布、指数分布和正态分布等。因此,如何生成各种概率分布的随机数就成为计算机技术在仿真等领域应用必须首要解决的问题。

文中主要介绍了任意概率分布的伪随机数各种生成方法,并采用舍选法实现了任意概率分布的伪随机数的生成,最后通过实验结果来验证该方法的可行性。

收稿日期:2007-02-10

作者简介:朱晓玲(1980-),女,广东湛江人,硕士研究生,研究方向为Petri网应用;姜 浩,副教授,研究方向为工作流应用研究。

2 均匀分布随机数的产生

均匀随机数是指理论上没有规律可循、在指定的区间内每个数的出现几率相等、无法根据之前的数来预测下一个数的数列。它是产生其它概率分布的随机数的基础和关键。至今应用最为广泛的均匀随机数生成器是基于Lehmer于1951年首先提出的线性同余伪随机数生成器(LCG,Linear Congruential pseudo random Generator),其迭代公式为[1]:

$$X_{i+1} = (a * X_i + c) \bmod m$$

其中,$a(0 < a < m)$、$c(0 \leqslant c < m)$ 和 m 三个参数分别叫做乘数、增量和模;$i = 0$ 时 X_0 称作种子。

若对任意的正整数 i 均有 $X_{i+T} = X_i$,则最小正整数 T 称为LCG的周期。在一个周期内 T 个模 m 的非负整数的取值是两两不同的,因此LCG的最大可能的周期必为 m。将任一周期内的 T 点数据对 m 归一化,即令 $R_j = X_j / m$,遂可得分布于(0,1)间的均匀随机数序列 $\{R\}$。

以下是C++语言中stdlib.h库的rand()函数实现代码[2]。此发生器为 LCG(a,c,m,X_0),其中 $a = 214013$,$c = 2531011$,$m = 2^{16}$,X_0 为由种子函数srand()定的初值。

```
Int _ cdecl rand(void)
{
# ifdef _ MT
_ ptiddata ptd = _ getptd();
return (((ptd - > holdrand = ptd - > holdrand * 214013L +
2531011L) > > 16) &0x7fff);
# else /* _ MT */
    return(((holdrand = holdrand * 214013L + 2531011L) > >
16) &0x7fff);
# endif /* _ MT */
}
```

可见,C++语言 stdlib.h 库 rand()函数产生 0～32767 范围内的整数,周期为 2^{16}。利用 rand()函数,能生成各种概率分布的随机数。

3　任意概率分布的随机数生成

有了均匀概率分布的伪随机数,就可以通过各种变换及映射关系来得到任意概率分布的伪随机数,主要的方法有反函数法、变换法和舍选法等。下面分别介绍这三种方法,其中重点介绍了舍选法,并采用舍选法实现了任意概率分布的随机数生成器。

3.1　反函数法

通过反函数法产生任意分布伪随机数的方法是最常用的方法之一,其原理是:已知[0,1]区间上均匀分布的伪随机数 r,将所需的概率分布的伪随机数函数 $F(x)$ 进行反变换,得到 $F(x)$ 的反函数 F^{-1},令 $X = F^{-1}(r)$,则 X 就是服从概率分布函数为 $F(x)$ 的伪随机数[3]。因此,只要知道所需概率分布函数的反函数,就可以从[0,1]均匀分布的伪随机数产生服从所需分布的随机数。

例　已知 $r \sim U(0,1)$,求服从指数分布的随机变量 X。

解:因为指数分布函数为:

$F(x) = 1 - e^{\lambda x}, x > 0$

所以,易求得 $F(x)$ 的反函数为:

$$F^{-1}(y) = \frac{1}{\lambda}\ln(1-y)$$

$r \sim U(0,1)$,令 $X = F^{-1}(r) = \frac{1}{\lambda}\ln(1-r)$,则 X 服从概率分布函数为 $F(x)$ 的指数分布。即 X 就是所求的随机变量。

3.2　变换法

变换法通过一个变换将一个分布的随机数变换成为不同分布产生的随机数,例如常用的线性变换能够把一个有限区间[a,b]上的分布变换到任意实数区间[u,v]上。对于每一个 X 值,都能够根据下式给出 Y 值:

$$Y = \frac{X(u-v)}{(b-a)} + u$$

变换法的典型例子是 Box_Muller 变换,它可产生精确的正态分布随机变量[4]。其变换式为:

$$Y_1 = \sqrt{-2\ln(X_1)}\sin(2\pi X_2)$$

$$Y_2 = \sqrt{-2\ln(X_1)}\cos(2\pi X_2)$$

X_1、X_2 是在区间[0,1]上均匀分布的随机变量,所得的 Y_1、Y_2 是相互独立的服从期望值 $\mu = 0$,方差 $\sigma^2 = 1$ 的正态分布的随机变量。

3.3　舍选法

用反函数法需要知道所求概率分布函数的反函数,当反函数不存在或难以求出时,反函数法便难以使用。这时可以考虑使用舍选法。

舍选法是冯·诺曼为克服反函数法和变换法的困难最早提出来的[5]。它的基本思想是:按照给定的分布密度函数 $f(x)$,对均匀分布的随机数序列{R}进行舍选。舍选的原则是在 $f(x)$ 大的地方,保留较多的随机数 r_i;在 $f(x)$ 小的地方,保留较少的随机数 r_i,使得到的子样本中 r_i 的分布满足分布密度函数的要求。

生成分布概率密度函数为 $f(x)$ 的伪随机数 X 的步骤如下[6]:

设随机变量 X 的概率密度函数为 $f(x)$,又存在实数 $a < b$,使得 $P(a < X < b) = 1$,如图 1 所示。

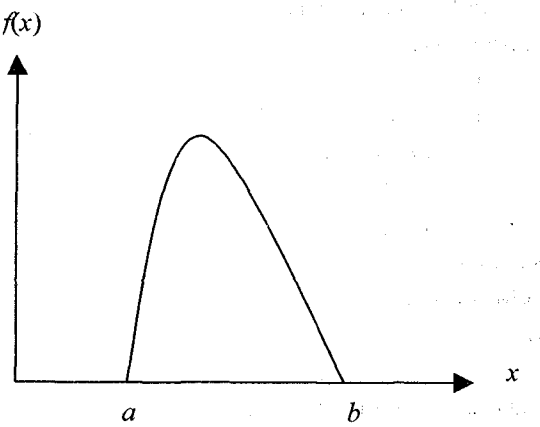

图 1　X 的概率密度函数分布图

(1) 选取常数 λ,使得 $\lambda f(x) \leqslant 1, x \in (a,b)$;

(2) 产生在[0,1]上均匀分布的随机数 r_1 和 r_2,令 $y = a + (b-a)r_1$;

(3) 比较 r_2 与 $\lambda f(y)$,若 $r_2 \leqslant \lambda f(y)$,则令 $x = y$;否则剔除 r_1 和 r_2,重返步骤(2)。

如此重复循环,产生的随机数序列 $x_1, x_2, \cdots, x_n$ 的分布由概率密度 $f(x)$ 确定。

若不存在有限区间(a,b),使$\int_a^b f(x)\mathrm{d}x = 1$,可选取有限区间$(a_1,b_1)$,使得$\int_{a_1}^{b_1} f(x)\mathrm{d}x \geqslant 1-\varepsilon$,其中,$\varepsilon$

是很小的正数，例如取 $a_1=\mu-3\sigma, b_1=\mu+3\sigma$，有 $\int_{a_1}^{b_1}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx>1-0.003$。对于区间 (a_1,b_1) 应用舍选法，仅会出现较小的系统误差。

4 应用实例

由于构造 petri 网仿真系统的需要，笔者根据 3.3 节所述的舍选法用 C++ 编制了多种常用概率分布的随机数发生器。下面给出程序中重要的函数并分别说明之。

```
/* 函数功能:产生一个在 min~max 范围内精度为 4 位小数的平均分布的随机数 */
double AverageRandom(double min,double max)
{
  int minInteger=(int)(min*10000);
  int maxInteger=(int)(max*10000);
  int randInteger=rand()*rand();
  int diffInteger=maxInteger-minInteger;
  int resultInteger=randInteger%diffInteger+minInteger;
  return resultInteger/10000.0;
}
/* 函数功能:产生指定概率密度分布的随机数 */
double AnyRandom(
      double(*functionPoint)(double,double,double),/* 指定的概率密度函数 */
      double min,double max, //区间边界
      double upper, //概率密度函数上界
      double param1,double param2) //概率密度函数的参数
{
  double dResult;
  double dScope;
  double dNormal;
  do
  {
    dResult=AverageRandom(min,max);
    dScope=AverageRandom(0,upper);
    dNormal=func(dResult, param1, param2);
  }while(dScope>dNormal);
  return dResult;
}
```

其中，AverageRandom 是一个均匀分布随机数发生函数，它产生区间[min, max]范围内的随机数。AnyRandom 是一个任意概率分布的随机数发生函数，它通过一个函数指针 functionPoint(第一个参数)获得指定的概率密度函数，根据舍选法的思想产生符合该概率密度函数分布的随机数。

限于篇幅，笔者在此只给出正态分布和指数分布随机数发生器的检验结果。对于正态分布随机数发生器，取 $\mu=15, \sigma=5.82$；对于指数分布随机数发生器，取 $\lambda=0.16$，用这两个随机数发生器各自产生区间[0, 30]的 50000 个样本，它们的统计图和概率密度曲线图如图 2～5 所示。

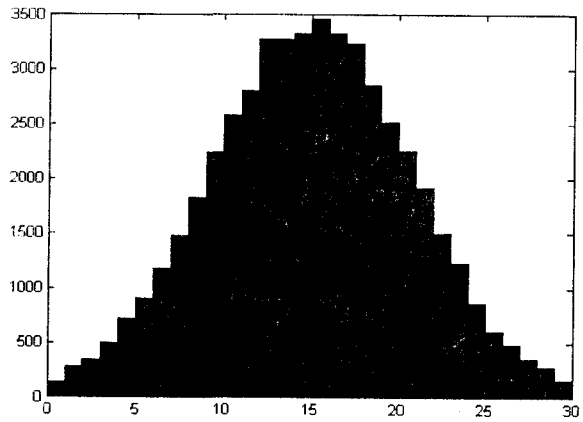

图 2 正态分布统计图

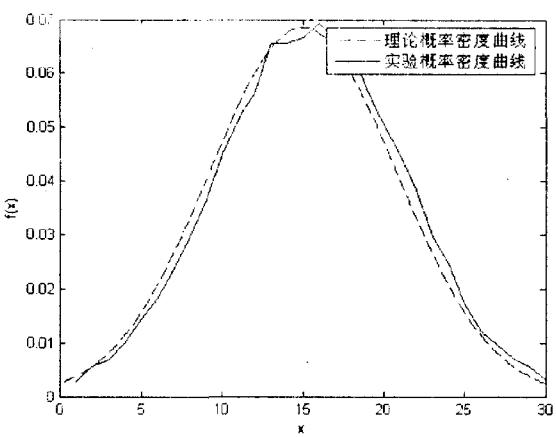

图 3 正态分布的理论与实验概率密度曲线比较

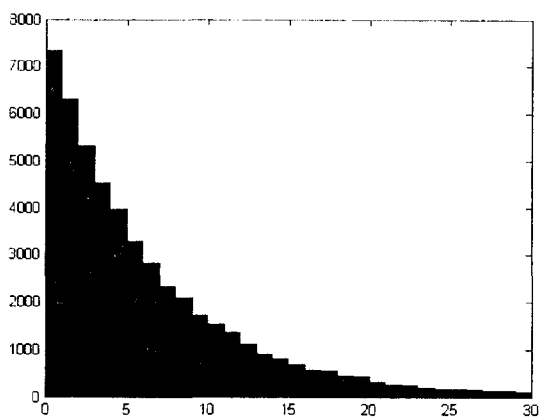

图 4 指数分布统计图

从图可知，用舍选法产生的正态分布和指数分布的随机数序列较好地吻合了理论概率密度曲线，具有较好的精确度。另外，舍选法实现简单，适用范围广，是产生任意概率分布随机数的优选方法。

(下转第 168 页)

4.5 装备维护和修理

大量复杂装备的维修一直是部队的棘手问题，特别是战时如何实现装备的快速维修更是决定战场胜负的要素之一。目前，装备说明书一般是文本和图形形式的，不方便于技术人员的维修。而将增强现实用于装备维修中，可以直接在实际设备中添加 3D 画面，一步一步地提示技术人员应该做什么以及如何做，方便装备的维修，极大提高装备保障的效率[9]。

4.6 协同工作

将增强现实应用于协同工作，可以允许多个用户终端协同活动，同时观看、讨论以及和虚拟物体交互。协同增强现实系统可以为多个用户能够建立一个共享的、可理解的虚拟空间，类似于他们所理解的自然空间。融入增强现实的协同工作所提供的协同工作环境，将在模拟推演、军事标绘等领域有着广泛的应用。

4.7 军用飞机

军方将增强现实用于飞行员座舱的显示，在飞行员座舱的前方玻璃上或者他们的头盔显示器上，将矢量图形叠到飞行员的视野中，不仅向飞行员提供导航信息，而且提供了包括敌方隐藏力量的增强战场信息。目前美国军方从事注册跟踪目标的研究工作，为飞机上装载的武器装备提供瞄准路径方面的增强信息。

5 总 结

增强现实是一个多学科交叉的领域，它包括计算机视觉、计算机图形学、传感学、网络和 GPS 等。作为一门新型的技术，目前增强现实的应用还处于实验室研究阶段，达到实用的增强现实系统还很少，增强现实的广泛使用还受到技术、用户界面和社会接受度等问题的限制。随着增强现实研究和应用的日趋成熟，必将在军事领域产生深远的影响。

参考文献：

[1] Milgram, Kishino. A taxonomy of Mixed Reality Visual Displays[J]. IEICE Trans Information Systems, 1994, E77-D (12): 1321-1329.

[2] Azuma R. Survey of augmented reality[J]. Teleoperators and Virtual Environments, 1997, 6(4):355-385.

[3] 齐 越，马红妹. 增强现实：特点、关键技术和应用[J]. 小型微型计算机系统，2004, 25(5): 900-903.

[4] 朱淼良，姚 远，蒋云良. 增强现实综述[J]. 中国图形图像学报，2004, 9(7):767-774.

[5] Caudell T. AR at boing[EB/OL]. 1990. http://www.ipo.tue.nl/homepages/mrauterb/presentation/HCI-history/tsld096.htm.

[6] Drascis D, Milgram P. Perceptual issues in augmented reality [C]// Proc. SPIE, Stereoscopic Displays Ⅶ and Virtual Systems Ⅲ. [s.l.]: SPIE Press, 1996:123-134.

[7] Holloway R L. Registration error analysis for augmented reality[J]. Presence: Teleoperators and Virtual Environments, 1997,6(4): 413-432.

[8] Julier S. Information Filtering for Mobile Augmented Reality [C]// Proc. Int'l Symp. Augmented Reality 2000 (ISAR 00). Los Alamitos, Calif.: IEEE CS Press, 2000.

[9] 柳祖国，李世其，李作清. 增强现实技术的研究进展及应用[J]. 系统仿真学报，2003, 15(2): 222-225.

(上接第 118 页)

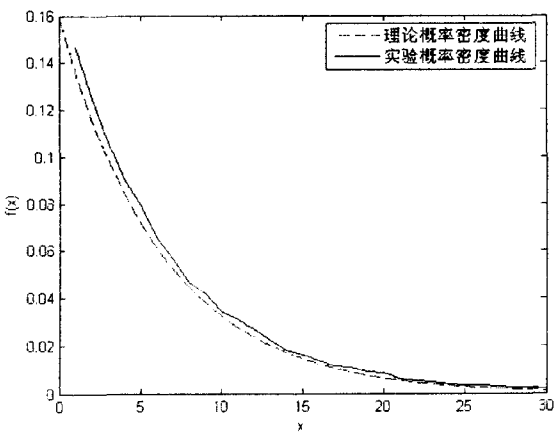

图 5 指数分布的理论与实验概率密度曲线比较

5 结 语

主要介绍了任意概率分布的随机数的产生方法，主要有反函数法、变换法和舍选法。其中，舍选法因其计算简单、适用范围广而得到了广泛的应用。详细描述了舍选法的实现步骤，给出了具体的实现例子，并通过实验结果验证舍选法的可行性。

参考文献：

[1] 张淑梅，李 勇. 计算机产生随机数的方法[J]. 数学通报，2006,45(3):44-45.

[2] 赵雪峰. 一种伪随机数生成算法的研究与实现[J]. 电脑学习，2005,12(6):25-26.

[3] 胡性本，刘向明，方积乾. 非均匀分布随机数的产生及其在计算机模拟研究中的应用[J]. 数理医药学杂志，2000,13(1):59-60.

[4] 肖化昆. 系统仿真中任意概率分布的伪随机数研究[J]. 计算机工程与设计，2005,26(1):168-171.

[5] 张艳红，吴 勇. 基于 Monte Carlo 方法的任意概率密度随机数字信号发生器设计[J]. 电子科技，2004,8:45-48.

[6] 李曙雄，杨振海. 舍选法的几何解释及其应用[J]. 数理统计与管理，2002,21(7):40-43.