蒙特卡罗模拟中相关变量随机数序列的产生方法

文德智† 卓仁鸿 丁大杰 郑慧 成晶 李正宏

(中国工程物理研究院核物理与化学研究所, 绵阳 621900)

(2011 年 8 月 5 日收到; 2012 年 6 月 14 日收到修改稿)

蒙特卡罗模拟有时需要对多维相关随机变量进行模拟抽样. 本文介绍基于 Choesky 因子线性变换 - 非线性变换产生具有指定边缘分布和相关系数的多维相关随机变量抽样序列的一般方法, 给出一种简单易行的高效数值实现途径和一些模拟结果. 模拟结果表明, 该方法产生的各随机变量抽样序列间具有预期要求的相关性, 并能通过指定边缘分布 Kolmogorov-Smirnov 非参数假设检验. 对该方法应用中的一些限制问题进行了讨论.

关键词: 蒙特卡罗方法, 伪随机数, 相关随机变量, 抽样

PACS: 02.50.Ng

1 引 言

自蒙特卡罗方法于 20 世纪 40 年代作为一种独立的统计模拟方法被提出以来, 已在多个研究领域得到了广泛的应用 [1−8]. 由具有已知分布的随机变量总体中抽取简单子样, 在蒙特卡罗方法中占有非常重要的地位 [1]. 对于多维随机变量的蒙特卡罗方法模拟问题, 往往假定各随机变量是相互独立的, 因此对由各已知分布随机变量总体中抽取产生满足独立性和均匀性要求的伪随机数序列子样的方法研究较为充分深入. 在某些科学问题中, 需要对多维具有相关性的随机变量进行模拟, 首先要从各已知边缘分布总体中抽取子样产生具有指定相关系数 (或协方差矩阵) 的多个随机数序列. 对多维正态分布的相关变量随机数抽样序列的产生方法, 已有较多文献介绍, 基于协方差矩阵 Cholesky 因子分解的线性变换方法被认为是最好的一种方法 [9−13]. 但由于线性变换不能保证非正态分布随机变量的分布规律不发生改变, 因此对于多维非正态分布相关随机变量的抽样无法通过线性变换产生. 一些重要的蒙特卡罗专著中较少见到关于非正态分布相关随机变量抽样序列模拟的深入讨论 [11−16]. Li 和 Hammond, Ronald 和 Conover, Ronald 和 James, Charles 等对多维非正态分布相关随机变量抽样的一般方法进行过探讨 [17−20]. Li 和 Hammond 的方法基于线性变换 - 非线性变换两步变换法, 从原理上看实现过程比较容易控制, 在具体实现方法上采用的是 Newton-Raphson 迭代, 需要计算比较复杂的函数导数. Ronald 等的方法基于 Spearman 相关系数矩阵的调整, 数值实现过程较难控制. Charles 介绍的方法基于快速发展中的联接函数 (copula) 理论, 并对部分同概率分布的二维相关随机变量进行过模拟抽样. 另有一部分文献仅涉及一些特殊分布相关随机变量伪随机数序列的抽样方法 [21,22]. 金畅、夏尊铨对相关随机变量抽样序列模拟作过一些理论介绍, 但对实现方法没有深入讨论 [23]. 本文介绍基于两步变换法的多维相关随机变量随机数序列模拟的一般方法, 并给出一种简单易行的高效实现技术途径和一些模拟结果.

2 相关随机变量模拟方法

2.1 正态分布相关随机变量模拟方法

由于正态分布具有在线性变换下保持分布规

† E-mail: wdzwdz518@gmail.com

律不变的独特性质, 因此多维相关正态分布随机变量抽样序列可以通过协方差矩阵的 Cholesky 因子对独立正态分布随机变量抽样序列进行线性变换来产生 [9—13]. 多维正态分布相关随机变量随机数序列产生方法是两步变换法产生多维非正态分布随机变量随机数抽样序列的基础, 为便于应用并与 Matlab 等通用数学软件的向量和矩阵运算规则协调一致, 本文在此一并给出, 数学表示上与文献 [9—13] 略有不同.

令 $X_{m\times n}=(X_1,X_2,\cdots,X_n)$ 为需要产生的 n 维相关正态分布随机变量抽样序列, 其中每个变量 $X_i(i=1,2,\cdots,n)$ 均为 m 个随机数抽样序列构成的列向量 (蒙特卡罗模拟中, m 是各变量需要进行的模拟抽样总次数, 一般情况下 m 远大于 n), 其均值为 $\mu=(\mu_1,\mu_2,\cdots,\mu_n)$, 协方差矩阵为

$$v_{n\times n}=\begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n}\\ v_{21} & v_{22} & \cdots & v_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ v_{n1} & v_{n2} & \cdots & v_{nn}\end{pmatrix}, \qquad (1)$$

则产生 n 维相关正态分布随机变量抽样序列 X 的步骤如下.

步骤 1 产生 n 维独立标准正态分布随机变量抽样序列 $Y_{m\times n}=(Y_1,Y_2,\cdots,Y_n)$, 每个变量抽样序列随机数个数为 m;

步骤 2 将协方差矩阵 $v_{n\times n}$ 进行 Cholesky 分解 (实际存在的随机变量抽样序列的协方差矩阵 $v_{n\times n}$ 总是对称、正定的 [24], 这种分解总是可行的), $v_{n\times n}=C_v^{\mathrm{T}}\cdot C_v$, 其中 $C_{v,n\times n}$ 为上三角矩阵;

步骤 3 进行线性变换, 令

$$X_{m\times n}=Y_{m\times n}\cdot C_{v,n\times n}+\mu_{m\times n}. \qquad (2)$$

产生已知相关系数矩阵的相关正态分布随机变量抽样序列也是比较容易的. 实际上, 如果需要产生的 n 维相关正态分布随机变量抽样序列的相关系数矩阵为

$$R_{n\times n}=\begin{pmatrix} 1 & r_{12} & \cdots & r_{1n}\\ r_{21} & 1 & \cdots & r_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ r_{n1} & r_{n2} & \cdots & 1\end{pmatrix}. \qquad (3)$$

对 $R_{n\times n}$ 进行 Cholesky 分解 $R_{n\times n}=C_r^{\mathrm{T}}C_r$, 则

$$X'_{m\times n}=Y_{m\times n}\cdot C_{r,n\times n}. \qquad (4)$$

为具有给定相关系数矩阵 $R_{n\times n}$ 的标准正态分布, 令

$$X_i=\sigma_i\times X'_i+\mu_i,\quad i=1,2,\cdots,n, \qquad (5)$$

也可得到标准偏差为 σ_i 和平均值为 μ_i 的相关正态分布随机变量抽样序列 X_i, 各 X_i 仍然具有相关系数矩阵为 $R_{n\times n}$ 的相关性.

2.2 非正态分布相关随机变量模拟方法

2.2.1 两步变换法

产生具有指定相关系数矩阵 $R_{Z_iZ_j}$ 和边缘分布的 n 维相关随机变量抽样序列 Z 的两步变换法途径示于图 1[17], 其中, Y 为独立标准正态分布随机变量抽样序列; V 为相关正态分布随机变量抽样序列, 它具有下文给出方法计算得到的 $R_{V_iV_j}$ 构成的相关系数矩阵.

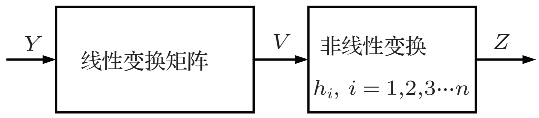

图 1 产生相关非正态分布随机变量抽样序列的两步变换法

第一步变换是由独立正态分布随机变量抽样序列 Y 变换到具有需要相关系数矩阵 $R_{V_iV_j}$ 的相关正态分布随机变量抽样序列 V. 选择相关正态分布随机变量作为中间变换变量是因具有需要相关系数矩阵的相关正态分布随机变量抽样序列容易通过 2.1 节所述方法实现.

第二步变换操作是对各相关正态分布随机变量抽样序列 (各列向量) 进行非线性变换, 其中 Z 的各变量可以指定具有不同的分布函数, 已知 Z 的各变量分布的方差则 Z 的相关系数矩阵和协方差矩阵可以相互换算得到. 利用该非线性变换, 可以根据要求产生的非正态分布相关随机变量 Z 的相关系数矩阵确定需要的相关正态分布随机变量 V 的相关系数矩阵. 事实上, V 的相关系数矩阵通常不同于 Z 的相关系数矩阵.

该方法的具体实施步骤如下.

步骤 1 确定由正态分布随机变量 V_i 到要求产生的随机变量 Z_i 非线性变换 h_i;

步骤 2 根据 $R_{Z_iZ_j}$ 和非线性变换确定相关正态分布随机变量抽样序列 V 的相关系数矩阵 $R_{V_iV_j}$;

步骤 3 通过 Cholesky 分解 $R_{V_iV_j}$ 确定线性变换矩阵;

步骤 4 产生独立的正态分布随机变量抽样序列 Y;

步骤 5 通过线性变换产生相关正态分布随机变量抽样序列 V;

步骤 6 由非线性变换 h_i 将正态分布随机变量 V_i 变换到要求产生的随机变量 Z_i, 构成随机变量抽样序列 Z.

可以看出, 相关非正态分布随机变量抽样序列产生所特有步骤是这里的步骤 1、步骤 2 和步骤 6, 而步骤 6 在步骤 1 完成后很容易实施.

2.2.2 非线性变换

假定要求产生的随机变量抽样序列 Z 的各变量 Z_i 概率密度函数 $f_{Z_i}(z_i)$ 已知, 则存在单调函数, 使得

$$\int_{-\infty}^{v_i} f_{V_i}(v_i)\mathrm{d}v_i = \int_{-\infty}^{z_i} f_{Z_i}(z_i)\mathrm{d}z_i = F_{Z_i}(z_i),\quad i = 1, 2, \cdots, n, \tag{6}$$

其中 $f_{V_i}(v_i)$ 为需要的正态分布中间变量的概率密度函数. 两个积分结果实际上是相应分布概率密度函数的累加分布函数. 为使变换简化, 可以让所需正态分布中间变量 V 的概率密度函数 $f_{V_i}(v_i)$ 具有"0" 期望值和单位方差的标准形式, 其累加分布函数可以用标准误差函数来表示, 因此有

$$z_i = F_{Z_i}^{-1}\left[\frac{1}{2}(1 + \mathrm{erf}(v_i))\right]. \tag{7}$$

(6) 式实际上定义了从标准正态分布到所要求概率分布随机变量的非线性变换, 即

$$z_i = h_i(v_i) = F_{Z_i}^{-1}\left[\frac{1}{2}(1 + \mathrm{erf}(v_i))\right]. \tag{8}$$

文献 [17] 中将标准正态分布的累加分布函数等同表示为标准误差函数, 与目前概率统计学中通常定义的标准误差函数不一致, (7) 式和 (8) 式对此作了修正.

2.2.3 相关系数阵 $R_{V_iV_j}$ 的确定方法

下面导出任何一对输入随机变量 V_i, V_j 和对应输出随机变量 Z_i, Z_j 的相关系数 $R_{V_iV_j}$ 和 $R_{Z_iZ_j}$ 间的关系, 其中 $i, j = 1, 2, \cdots, n, i \neq j$. 文献 [17] 在推导时假设了随机变量 Z_i, Z_j 的期望值为 "0", 为避免应用于非 "0" 期望分布时疏忽带来错误, 本文在一般情况下进行推导. 设输出随机变量 Z_i, Z_i 的期望值为 μ_{Z_i}, μ_{Z_j}, 标准偏差 $\sigma_{Z_i}, \sigma_{Zj}$, 则 Z_i 和 Z_i 相关系数可以表示为

$$\begin{aligned} R_{Z_iZ_j} &= \frac{E\left[(Z_i - \mu_{Z_i})\left(Z_j - \mu_{Z_j}\right)\right]}{\sigma_{Z_i}\sigma_{Z_j}} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{(h_i(v_i) - \mu_{Z_i})}{\sigma_{Z_i}} \frac{(h_j(v_j) - \mu_{Z_j})}{\sigma_{Z_j}} \\ &\quad \times f_{V_iV_j}(v_i, v_j)\mathrm{d}v_i\mathrm{d}v_j, \\ &\quad i, j = 1, 2, \cdots, n,\ i \neq j, \end{aligned} \tag{9}$$

其中 h_i 和 h_j 为对 V_i 和 V_j 的非线性变换, $f_{V_iV_j}(v_i, v_j)$ 为二元正态分布的联合概率密度函数, 其中 V_i, V_j 的边缘分布均为期望值为 "0", 方差为 "1" 的标准正态分布, 相关系数为 $R_{V_iV_j}$. 可以看到, 一旦所有非线性变换确定, $R_{Z_iZ_j}$ 仅依赖于联合概率密度函数 $f_{V_iV_j}(v_i, v_j)$, 最终决定于未知参数 $R_{V_iV_j}$. 可将 (9) 式进一步写为

$$\begin{aligned} & R_{Z_iZ_j} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\left(F_{Z_i}^{-1}\left[\frac{1}{2}(1 + \mathrm{erf}(v_i))\right] - \mu_{Z_i}\right)}{\sigma_{Z_i}} \\ &\quad \times \frac{\left(F_{Z_j}^{-1}\left[\frac{1}{2}(1 + \mathrm{erf}(v_j))\right] - \mu_{Z_j}\right)}{\sigma_{Z_j}} \\ &\quad \times \frac{1}{2\pi\sqrt{1 - R_{V_iV_j}^2}} \\ &\quad \times \exp\left[-\frac{\left(v_i^2 - 2\cdot R_{V_iV_j}\cdot v_iv_j + v_j^2\right)}{2\left(1 - R_{V_iV_j}^2\right)}\right]\mathrm{d}v_i\mathrm{d}v_j, \\ &\quad i, j = 1, 2, \cdots, n,\ i \neq j. \end{aligned} \tag{10}$$

实际上, (10) 式把相关系数 $R_{V_iV_j}$ 和 $R_{Z_iZ_j}$ 联系起来, 其中 $R_{Z_iZ_j}$ 已知而 $R_{V_iV_j}$ 未知.

在积分可以解析运算的情况下 $R_{V_iV_j}$ 可以容易地得到, 如 Z_i, Z_j 均为均匀分布时,

$$R_{V_iV_j} = 2\cdot\sin\left(\frac{\pi}{6}R_{Z_iZ_j}\right). \tag{11}$$

如要产生的 Z_i, Z_j 均为 σ 参数相同的对数正态分布, 文献 [23] 给出了 $R_{V_iV_j}$ 的理论解析计算公式, 但该式的推导存在错误, 依据该式不能产生所需的抽样序列. 正确的计算公式如下:

$$R_{V_iV_j} = \frac{\ln(R_{Z_iZ_j}\times\left[\exp(\sigma^2) - 1\right] + 1)}{\sigma^2}. \tag{12}$$

通常 (10) 式积分不易处理, 必须用数值方法. 文献 [17] 给出了 (10) 式基于 Newton-Raphson 迭代数

值解法, 该方法需要知道 (10) 式关于 $R_{V_iV_j}$ 的导数, 即使采用数值计算也是比较麻烦的. 本文给出另一种具体确定方案.

在标准偏差 σ_{Z_i}, σ_{Z_j} 一定的情况下, $R_{V_iV_j}$ 和 $R_{Z_iZ_j}$ 是由 (10) 式的单值函数确定的, 而 $R_{V_iV_j}$ 和 $R_{Z_iZ_j}$ 的取值范围均在 $[-1,1]$ 区间. 因此, 我们可以计算 $R_{V_iV_j}$ 在 $[-1,1]$ 区间系列的值对应的 $R_{Z_iZ_j}$, 通过拟合插值反过来确定要求的 $R_{Z_iZ_j}$ 下对应的所需 $R_{V_iV_j}$. 由于 $f_{V_iV_j}(v_i,v_j)$ 是二元标准正态分布的联合概率密度函数, 数值计算中两积分的积分上下限选择 $[-4,4]$ 就足够精确了. 这种方法通过数值计算实施起来要容易方便、快捷得多.

3 相关随机变量模拟示例和一些结果

3.1 指定 $R_{Z_iZ_j}$ 对应的 $R_{V_iV_j}$ 的数值计算

计算了产生多种概率分布组合要求相关随机变量的 $R_{Z_iZ_j}$ 所需相关标准正态分布的 $R_{V_iV_j}$ 值, 表 1 给出了要求产生的 Z_i, Z_j 为均匀分布和对数正态分布时对应的 $R_{V_iV_j}$ 的数值方法计算结果和理论解析计算结果.

表 1 产生相关系数为 $R_{Z_iZ_j}$ 随机变量序列所需的 $R_{V_iV_j}$ 值

$R_{Z_iZ_j}$	需要的相关标准正态分布的 $R_{V_iV_j}$ 值			
	Z_i, Z_j 服从均匀分布		Z_i, Z_j 服从参数 $\sigma=1$ 的对数正态分布	
	数值计算	理论值	数值计算	理论值
−1	−1.000	−1	不合理	不合理
−0.9	−0.90826	−0.90798	不合理	不合理
−0.8	−0.81379	−0.81347	不合理	不合理
−0.7	−0.71707	−0.71674	不合理	不合理
−0.6	−0.61837	−0.61803	不合理	不合理
−0.5	−0.51796	−0.51764	不合理	不合理
−0.4	−0.41612	−0.41582	不合理	不合理
−0.3	−0.31311	−0.31287	−0.72477	−0.72461
−0.2	−0.20923	−0.20906	−0.42116	−0.42107
−0.1	−0.10476	−0.10467	−0.18858	−0.18853
0	0	0	0	0
0.1	0.10476	0.10467	0.15863	0.15857
0.2	0.20923	0.20906	0.29556	0.29539
0.3	0.31311	0.31287	0.41603	0.41574
0.4	0.41612	0.41582	0.52359	0.52314
0.5	0.51796	0.51764	0.62077	0.62011
0.6	0.61837	0.61803	0.7094	0.70851
0.7	0.71707	0.71674	0.79086	0.78973
0.8	0.81379	0.81347	0.86621	0.86484
0.9	0.90826	0.90798	0.93625	0.9347
1	1	1	1	1

数值计算结果表明, 如果要求产生随机数抽样序列的相关随机变量中的某个变量 Z_i 指定是均匀分布, 则产生该变量抽样序列的标准正态分布中间变量 V_i 与产生其他变量 Z_j 抽样序列的标准正态分布中间变量 V_j 的相关系数 $R_{V_iV_j}$ 不依赖于该均匀分布随机变量 Z_i 分布区间的起始点和宽度. 如果要求产生的两个随机变量 Z_i, Z_j 抽样序列服从相同 σ 参数的对数正态分布时, 其 $R_{V_iV_j}$ 结果只与 σ 和要求的相关系数 $R_{Z_iZ_j}$ 有关, 与参数 μ 无关. 而理论分析和计算也可说明这一点. 对某些要求的概率分布随机变量 Z_i, Z_j 及 $R_{Z_iZ_j}$, 数值计算和理论分析计算都得不到合理的 $R_{V_iV_j}$ 值, 比如绝对值以不可接受的误差范围大于 1, 或者可能出现复数.

3.2 相关随机变量的模拟抽样示例结果

第一个例子是要求产生具有如下相关系数矩阵的相关均匀分布 Z_1, Z_2, Z_3, 其中 Z_1, $Z_2 \sim U(-1,1)$, $Z_3 \sim U(0,4)$:

$$R_{Z_1Z_2Z_3} = \begin{pmatrix} 1 & 0.8 & -0.8 \\ 0.8 & 1 & -0.375 \\ -0.8 & -0.375 & 1 \end{pmatrix}.$$

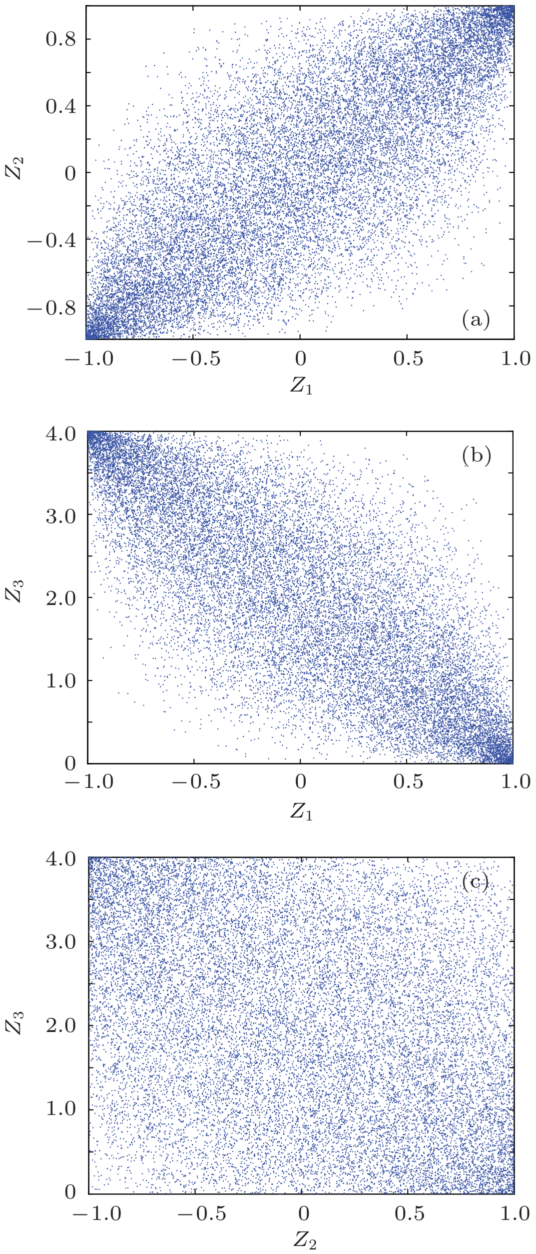

图 2 Z_1, Z_2, Z_3 两两关系散点图 (a) Z_1-Z_2 ($R_{Z_1Z_2} = 0.8$); (b) Z_1-Z_3 ($R_{Z_1Z_3} = -0.8$); (c) Z_2-Z_3 ($R_{Z_2Z_3} = -0.37$)

一个抽样总数为 200000 的示例结果给出的 Z_1, Z_2, Z_3 相关系数矩阵为

$$R_{Z_1Z_2Z_3} = \begin{pmatrix} 1 & 0.7994 & -0.7998 \\ 0.7994 & 1 & -0.3728 \\ -0.7998 & -0.3728 & 1 \end{pmatrix}.$$

图 2 为各抽样数为 20000 时绘制的 Z_1, Z_2, Z_3 之间两两关系散点图. 从图 2 中可以看出各随机变量分布区间和它们之间大致的相关关系.

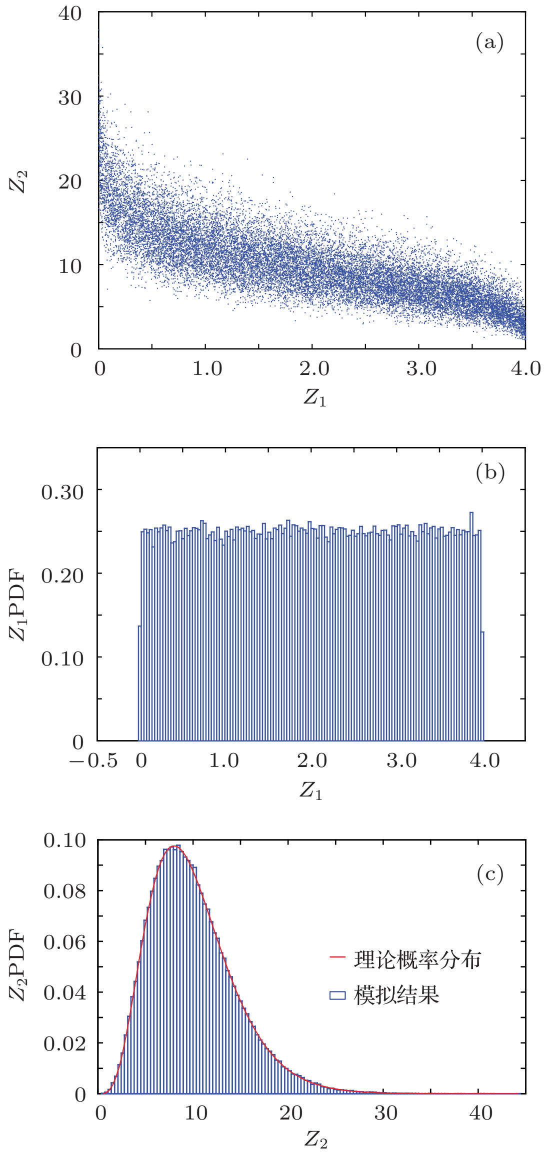

图 3 Z_1, Z_2 的抽样序列关系和归一化统计直方图 (a) Z_1, Z_2 抽样序列散点图; (b) (0, 4) 均匀分布 Z_1 抽样; (c) 自由度为 10 的 χ^2 分布抽样

第二个例子是产生要求相关系数为 -0.8 相关分布 Z_1, Z_2, 其中 $Z_1 \sim U(0,4)$, $Z_2 \sim \chi^2(10)$. 抽样总数为 200000 的一个示例结果给出的 Z_1, Z_2 相关系数为 -0.8017. Z_1, Z_2 的关系散点图和统计

直方图示于图 3. 其中图 3(a) 的 Z_1, Z_2 抽样数各为 20000, 图 3(b), (c) 的统计直方图是在抽样数各为 200000 时得到的, 图 3(c) 中的曲线为 $\chi^2(10)$ 分布的概率密度理论曲线. 从图中可以看出产生的随机变量抽样序列服从指定概率分布.

对两个例子产生的 Z_i 均进行了 Kolmogorov-Smirnov 检验, 所有产生的随机变量抽样序列均通过了服从预期分布的非参数假设检验.

4 讨 论

如表 1 所示, 对某些要求产生的相关随机变量分布, 存在一定相关系数 $R_{Z_iZ_j}$ 范围得不到合理的对应相关标准正态分布的 $R_{V_iV_j}$ 值. 在这种情况下, 无法通过本文两步变换法产生具有所要求相关关系的指定分布随机变量抽样序列. 这是本方法受到限制的情况之一.

本方法能够成功的充分必要条件是相关系数矩阵 $R_{V_iV_j}$ 是非负定的. 由于任意一组实际存在的随机变量抽样序列的协方差矩阵是对称、非负定的 [24], 而相关系数矩阵可以看作是方差为 1 的标准化协方差矩阵, 因此相关系数矩阵也是对称、非负定的. 要产生具有指定相关关系的正态分布, 如果要求相关关系是合符逻辑的, 即其相关系数矩阵是对称、非负定的, Cholesky 分解总是可行的, 总可以按 (4) 式产生所要求的相关正态分布. 因此, 如果要求的相关系数矩阵无法进行 Cholesky 分解, 即不是非负定的, 那么可以认为, 不存在这样一组相关关系的相关正态分布随机变量.

然而如果要产生具有指定相关关系的多维任意非正态分布随机变量抽样序列, 情况就要复杂一些. 对于二维任意非正态分布相关随机变量抽样序列的产生, 只要能够从要求的相关系数 $R_{Z_iZ_j}$ 得到合理的对应相关标准正态分布的 $R_{V_iV_j}$ 值, 两步变换法总是可以实施, 得到要求相关关系的任意非正态分布随机变量抽样序列 Z_1, Z_2. 对于三维以上任意非正态分布相关随机变量抽样序列的产生, 则有可能发生这样的情况: 要求的非正态分布随机变量相关系数 $R_{Z_iZ_j}$ 矩阵是对称、非负定的, 对每一对要求相关系数 $R_{Z_iZ_j}$ 也能够得到合理的对应相关标准正态分布的 $R_{V_iV_j}$ 值, 然而, $R_{V_iV_j}$ 构成的矩阵却不是非负定的. 例如, 要求的产生在 $[-1, 1]$ 均匀分布、相关系数 $R_{Z_iZ_j}$ 矩阵为

$$\begin{pmatrix} 1 & -0.4 & 0.2 \\ -0.4 & 1 & 0.8 \\ 0.2 & 0.8 & 1 \end{pmatrix}$$

的三维随机变量抽样序列 Z_1, Z_2, Z_3, 由于该矩阵是正定的, 具有这样的相关性关系的均匀分布随机变量及其抽样序列 Z_1, Z_2, Z_3 应该是存在的, 但非线性变换要求的标准正态分布的相关系数 $R_{V_iV_j}$ 矩阵是

$$\begin{pmatrix} 1 & -0.4162 & 0.20923 \\ -0.4162 & 1 & 0.81379 \\ 0.20923 & 0.81379 & 1 \end{pmatrix}.$$

该相关系数矩阵却不是非负定的, 由本两步变换方法也无法产生要求分布和相关关系的随机变量抽样序列. 这是本方法应用的另一个会受到限制的情况.

另外, 由于数值计算舍入误差和其他误差的原因, 也可能出现由非线性变换要求产生的标准正态分布中间变量的相关系数 $R_{V_iV_j}$ 矩阵无法进行 Cholesky 分解的情况, 与前面所述情况不同, 这种情况下 $R_{V_iV_j}$ 矩阵的行列式值接近于 “0”, 可以进行微扰调整补救, 以使得 Cholesky 分解能够进行, 并进一步产生与预期要求相关系数阵接近的随机变量抽样序列 [25].

关于随机变量相关性的研究, 目前较前沿的课题为 copula 函数理论 [20,26,27]. 基于 copula 函数理论的方法可能解决本方法受到限制的情况, 对于三维以上不同分布的相关随机变量抽样序列的产生方法值得研究.

5 结 论

重新推导并给出了产生具有任意指定概率分布和相关系数 (或协方差) 矩阵的多维数相关随机变量抽样序列方法新的数学表示. 相关正态分布随机变量抽样序列产生基于相关系数 (或协方差) 矩阵 Cholesky 因子的线性变换, 相关非正态分布随机变量抽样序列通过线性变换 - 非线性变换产生, 其中各非正态分布随机变量可以具有不同分布函数. 对确定产生具有规定相关系数的非正态分布随机变量抽样序列所需相关标准正态分布随机变量的相关系数的通用公式进行了推导完善, 提出了一种

简单易行的高效数值实现途径. 给出了部分计算结果和相关非正态分布随机变量抽样序列模拟抽样结果, 产生的序列通过 Kolmogorov-Smirnov 检验服从指定分布并具有要求的相关关系. 讨论了两步变换法应用的限制条件, 在计算得到的所需相关正态分布随机变量抽样序列的相关系数不合理时, 或从多维非正态分布相关随机变量相关系数得到的相关正态分布相关系数矩阵不符合非负定要求时, 该方法不适用. 研究基于 copula 函数理论对于三维以上不同分布相关随机变量抽样序列的产生方法可能有助于该问题的解决, 是作者有关本问题的未来研究方向.

感谢上海公共研发服务平台提供服务和帮助.

[1] Pei Lu cheng 1989 *Computer Stochastic Simulation* (Changsha: Hunan Science and technology Press) p1 (in Chinese) [裴鹿成 1989 计算机随机模拟 (长沙: 湖南科学技术出版社) 第 1 页]

[2] Xu S Y 2006 *Monte Carlo Method and its Application in Nuclear Physics Experiment* (2nd Ed.) (Beijing: Atomic Energy Press) p1 (in Chinese) [许淑艳 2006 蒙特卡罗方法在实验核物理中的应用 (第二版) (北京: 原子能出版社) 第 1 页]

[3] Ivan T D 2008 *Monte Carlo Methods for Applied Scientists* (Singapore: World Scientific Publishing Co. Pte. Ltd.) p1

[4] Peter J 2002 *Monte Carlo methods in finance* (Chichester: John Wiley & Sons, Inc.) p1

[5] Cox M G, Siebert B R L 2006 *Metrologia* **43** S178

[6] Matthew N O S 2009 *Monter Carlo Methods for Electromagnetics* (Boca Raton: CRC Press Taylor & Francis Group)P1

[7] Landau D P, Binder K 2000 *A Guide to Monte Carlo Simulations in Statistical Physics* (2nd Ed.) (New York: Cambridge University Press) p1

[8] Ferguson D M, Siepmann J I, Truhlar D G 1999 *Monte Carlo Methods in Chemical Physics* (New York: John Wiley & Sons, Inc.) p1

[9] Moonan W J 1957 *J. Amer. Statist. Ass.* **52** 247

[10] Box G E P, Muller M E 1958 *Ann. Math. Statist.* **29** 610

[11] Paul B, Fox B L, Linus E S (Translated by Yang Weigao) 1991 *A guide to Simulation* (Beijing: Science Press) p186 (in Chinese) [(美) 布雷特利等著杨惟高等译 1991 模拟导论 (北京: 机械工业出版社) 第 186 页]

[12] Xu Z J 1985 *Monte Carlo Method* (Shanghai: Shanghai Science and Technology Press) p132 (in Chinese) 徐钟济 1985 蒙特卡罗方法上海: 上海科学技术出版社第 132 页)

[13] Zhu B R 1987 *Introduction to Monte Carlo Method* (Ji-nan: Shandong University Press) p108 (in Chinese) [朱本仁 1987 蒙特卡罗方法引论 (济南: 山东大学出版社第 108 页]

[14] Niederreiter H 1992 *Random Number Generation and Quasi-Monte Carlo Methods*(Philadelphia: Society for Industrial and Applied Mathematics)P161

[15] Rubinstein R, Kroese D P 2008 *Simulation and the monte carlo method* (2nd Ed.)(Hoboken: John Wiley & Sons, Inc.) p65

[16] Kalos M H, Whitlock P A 2008 *Monte Carlo Methods* (2nd Ed.) (Weinheim: WILEY-VCH Verlag GmbH & Co.) p35

[17] li S T, Hammond J L 1975 *IEEE Transactions on Systems: Man, and Cybernetics* **SMC-5** 557

[18] Ronald L I, Conover W J 1982 *Communications in Statistics-Simulation and Computation* **11** 311

[19] Ronald L I, James M D 1982 *Communications in Statistics-Simulation and Computation* **11** 335

[20] Charles N H 1999 *Risk Analysis* **19** 1205

[21] Chen J T 2005 *European Journal of Operational Research* **167** 226

[22] Michael F 1999 *Communications in Statistics-Simulation and Computation* **28** 785

[23] Jing C 2005 *M. S. Dissertation* (Dalian: Dalian University of Technology) p31 (in Chinese) [金畅 2005 硕士学位论文 (大连: 大连理工大学) 第 31 页]

[24] Wang Z K 1976 *Probability theory and its application* (1st Ed.) (Beijing: Science press) P105 (in Chinese) [王梓坤 1976 概率论基础及其应用 (第一版)(北京: 科学出版社) 第 105 页]

[25] Salter M J, Ridler N M, Cox M G 2000 *Technical Report CETM 22* (Teddington: National Physical Laboratory) p14

[26] Nelsen R B 2006 *An introduction to Copulas* (2nd Ed.) (New York: Springer) p1

[27] Zhang Y T 2002 *Statistical Study* **4** 48 (in Chinese) [张尧庭 2002 统计研究 **4** 48]

Generation of correlated pseudorandom variables in Monte Carlo simulation

Wen De-Zhi† Zhuo Ren-Hong Ding Da-Jie Zheng Hui
Cheng Jing Li Zheng-Hong

(*Institute of Nuclear Physics and Chemistry, China Academy of Engineering Physics, Mianyang* 621900, *China*)

(Received 5 August 2011; revised manuscript received 14 June 2012)

Abstract

Correlated pseudorandom variables with prescribed marginal distribution functions sometimes are required in simulation such as in Monte Carlo studies. In this paper, we present a general procedure and a simple but effective numerical approach to generating correlated random variables sampling sequence with prescribed marginal probability distribution functions and correlation coefficient matrix based on linear transformation-nonlinear transformation with Choesky factor. Some simulation results are reported. Simulation results show that the collections of random numbers generated by the presented procedure have desired correlations and pass the Kolmogorov-Smirnov non-parametric hypothesis test of specified marginal distribution. Some restrictions on the application of this method are discussed.

Keywords: Moto Carlo method, pseudorandom numbers, correlated random variables, sampling

PACS: 02.50.Ng

† E-mail: wdzwdz518@gmail.com