数理解析研究所講究録
1011巻 1997年 58-63

有限乱数列の定義と擬似乱数生成

東京大学大学院工学系研究科　伏見正則 (FUSHIMI Masanori)

1. はじめに

擬似乱数の生成法や統計的検定結果について論じた論文や書物は数知れないほど多い。しかし、これらの論文や書物では、ごく一部の例外を除いて、“乱数とは一体何であり、記述している生成法によって生成される数列は、乱数列の定義とどう関係するのか”についてはほとんど述べられていない。統計的検定に関しても、ad hocなテストが多数提案されていて、それらのいくつかを選んで、生成される数列の1周期分のごく一部分について検定を行っているだけであり、数列の全容を解明するには程遠い。

一方、“乱数とは一体何であるか”という多くの人が抱く根源的な疑問に対する答えを出そうとする試みもいくつかあった。有名なKnuthの本[4]には、これらの試みに基づく乱数の定義がいくつも載っている。しかも、そのどれもが満足すべきものではないとKnuthは述べている。数学の書物ならば、ひとつの概念に対する定義は通常ひとつであるのとはいちじるしく対照的である。これは、擬似乱数生成という分野が数学ではなくて、Knuthの本の題名どおり、まさにArtであることを示しているものといえよう。

真の一様乱数が持っている重要な性質のひとつに、多次元均等分布という性質がある。$\{X_t : t = 1, 2, \cdots\}$ を b 個の値を等確率でとる一様乱数列 (無限列) とすると、任意の自然数 k に対して

$$\Pr\{X_t = v_1, \cdots, X_{t+k-1} = v_k\} = 1/b^k \qquad \text{for all } t \tag{0}$$

が成り立つ。ここで $v_1, \cdots, v_k$ は、乱数が取りうる b 個の値のうちの任意のものである。

擬似乱数には必ず周期性があるので、もちろん上記の性質は成り立たない。(0) が特定の k に対して成り立つとき、この系列は k 次均等分布をするという。できるだけ大きな k に対して k 次均等分布をする擬似乱数が良い擬似乱数であるというのがひとつの考え方であろう。ただし、(0) における確率の意味については、次のように解釈するのがふつうであろう: 無限列の場合は相対頻度の極限、擬似乱数の場合は1周期分の相対頻度。

つぎに乱数をランダムサンプリングに使う場合のことを考えてみよう。確率論の初等的な教科書には、つぼの中に球を入れておいて取り出すというモデルを使って、復元抽出と非復元抽出の違いを説明しているものがある。非復元抽出のサンプルはもちろんi.i.d. とはならないが、同じ番号の球が多数つぼの中に入っていて、取り出す球の個数が比較的少なければ、近似値にはi.i.d. であると見なしてもよいであろう。乱数表あるいは擬似乱数のような有限列を利

用する場合は、非復元抽出に相当するので、有限列全部を見たときに、同一の数が多数含まれていることが重要となる。

2. 有限乱数列のランダムネス

有限長の乱数列の定義を与えようとする試みは、Kolmogorov, Martin-Löf, Chaitin 等によって行われた。これは、与えられた有限数列を生成するチューリングマシンのプログラムの最短の長さをもってこの数列の乱数度と定義し、乱数度の高いものほど真の乱数であるとするものであるが、この定義に沿うような乱数生成法を考案するのは、ほとんど不可能である。

Kolmogorov は、乱数列の定義に関していくつかの論文を発表しているが、上記のような趣旨の定義を与える前に発表した論文に [5] がある。これは、von Mises のコレクティフの概念に基づくもので、ごく大ざっぱに言えば、与えられた有限列から種々の抽出規則に従って選んだ部分列における数の出現頻度の分布に安定性がある場合に、もとの有限列を乱数列であると認めようというものである。そして Kolmogorov は、抽出規則の数が多くなければ、この意味での有限乱数列が存在することを証明している。しかし、数列の構成法は示していない。

Kolmogorov は部分列の抽出規則の例として 6 つの規則を挙げている。しかし、これだけの少数の抽出規則のもとでも頻度分布が安定な数列を構成するのは至難のわざである。そこで、ここでは彼が示した抽出規則のうちでもっとも単純な次の規則だけを取り上げてみよう：“数列の偶数番目の数だけを取り出す。” これは、きわめて単純ではあるが、応用上は大変に重要な抽出規則である。例えば待ち行列のシミュレーションでは、偶数番目の乱数は客の到着間隔を定めるのに使い、奇数番目の乱数は客に対するサービス時間を定めるのに使う、というような使い方がよく行われる。この場合には、偶数番目 (あるいは奇数番目) を取り出して得られる数列が良い乱数列となることが望ましい。

上記の抽出規則をもう少し一般化すると、$\{x_t : t = 1, 2, \cdots\}$ をもとの数列としたとき、これから n 番目ごとの要素を抽出して得られる数列 $\{x_{nt} : t = 1, 2, \cdots\}$ が種々の n に対して良い乱数列となることが望ましいということになる。

例えば古典的な線形合同法

$$x_t = ax_{t-1} \qquad (\mathrm{mod}\ M) \tag{1}$$

の場合には、乗数 a だけでなく、$a^n (\mathrm{mod}\ M)$ も種々の n に対して良い乗数であること（ふつうは、スペクトル検定に合格すること）が望ましい。

また、線形合同法の行列 - ベクトル版

$$X_t = AX_{t-1} \qquad (\mathrm{mod}\ M) \tag{2}$$

の場合には、$A^n (\mathrm{mod}\ M)$ が種々の n に対して良い乗数行列となることが望ましい。

Tausworthe系列やGFSR系列の場合には、同様な問題がかなり複雑な様相を呈する。詳細はFushimi[2]で論じられているが、次節でその概要を述べる。

3. GFSR系列の部分別のランダムネス

ガロア体GF(2)上の原始多項式

$$f(z) = 1 + c_1 z + c_2 z^2 + \cdots + c_p z^p \tag{3}$$

を特性多項式とする漸化式

$$a_t = c_1 a_{t-1} + c_2 a_{t-2} + \cdots + c_p a_{t-p} \qquad (\mathrm{mod}\ 2) \tag{4}$$

によって生成される系列$\{a_t\}$のことをシフトレジスタ系列あるいはM系列と呼ぶ。ただし、$c_p = 1$であり、初期値$a_1, a_2, \cdots, a_p$は、すべて0とならないかぎり、任意に定めてよい。M系列の周期は$T = 2^p - 1$であり、nをTと互いに素な自然数とすると、系列$\{a_{nt} : t = 1, 2, \cdots\}$の周期も$T$である。(後者の系列は$n$が2の整数べき乗ならば、(3)の原始多項式を特性多項式とするM系列であり、そうでなければ、(3)とは異なるp次の原始多項式を特性多項式とするM系列である。)

M系列のk個組$(a_t, a_{t+1}, \cdots, a_{t+k-1})$が1周期$(t = 1, 2, \cdots, T)$にわたってとるパターンについては、$k \leq p$ならば、$(0, 0, \cdots, 0)$が$2^{p-k} - 1$回、その他のあらゆるパターンが$2^{p-k}$回現れることが知られている。したがって、$k$が$p$に近くなければ、1周期中に同じパターンが何度も現れることになり、非復元抽出を行っても、一様分布に近いサンプルが得られるための一つの必要条件を満たしている。

Tausworthe系列およびGFSR系列はM系列から次のようにして構成されるlビットの2進小数の系列である。

Tausworthe系列:

$$x_t = 0.a_{nt+1} a_{nt+2} \cdots a_{nt+l}$$

GFSR系列:

$$y_t = 0.a_t a_{t+r} a_{t+2r} \cdots a_{t+(l-1)r}$$

両系列の間の関係については、伏見[1]を参照されたい。

本稿の目的のためには、これらを一般化した次の系列も考える。

一般化Tausworthe系列[2]:

$$x'_t = 0.a_{nt+j(1)}a_{nt+j(2)}\cdots a_{nt+j(l)}$$

ここで、$\{j(1), j(2), \cdots, j(l)\}$ は $\{1, 2, \cdots, l\}$ に対して任意の置換を施したものである。
一般化 GFSR 系列 [2]:

$$y'_t = 0.a_{t+\tau_1}a_{t+\tau_2}\cdots a_{t+\tau_l}$$

これらの系列の k 次均等分布については、次の定理が成立する。

定理 $y'_t (0 \le t \le k-1)$ に含まれる M 系列 $\{a_t\}$ の kl 個の要素が GF(2) で線形独立ならば、系列 $\{y'_t\}$ は k 次均等分布をする。系列 $\{x'_t\}$ についても、また n が T と互いに素ならば、$\{y'_{nt}\}$ および $\{x'_{nt}\}$ についても同様である。

ここで、GF(2) 上の線形独立性というのは、次の意味である。漸化式 (4) を繰り返し使えば、任意の t に対して a_t は次のように $a_1, a_2, \cdots, a_p$ の一次結合で書ける。

$$\begin{aligned} a_t &= c_1a_{t-1} + c_2a_{t-2} + \cdots + c_pa_{t-p} \qquad (\mathrm{mod}\ 2) \\ &= c_1(c_1a_{t-2} + c_2a_{t-3} + \cdots) \\ &+ c_2(c_1a_{t-3} + c_2a_{t-4} + \cdots) \\ &+ \cdots \\ &= \cdots \\ &= e_1^{(t)}a_1 + e_2^{(t)}a_2 + \cdots + e_p^{(t)}a_p \end{aligned}$$

このようにして、各 a_t に対して唯一の重みベクトル

$$\mathbf{e}_t = (e_1^{(t)}, e_2^{(t)}, \cdots, e_p^{(t)})$$

が定まる。上記の線形独立性というのは、これらの重みベクトルの GF(2) 上における独立性のことである。

この定理からただちに導かれる結論は、上記の四つの系列が k 次均等分布をし得る最大の k は $m = \lfloor p/l \rfloor$ であるということである。そこで系列 $\{x'_{nt} : t = 1, 2, \cdots\}, n \in N$, が m 次均等分布をするように順列 $\{j(1), j(2), \cdots, j(l)\}$ を定めることをわれわれの目標とする。ただし、N は周期 T と互いに素な自然数の集合の部分集合であり、実用的な観点からすれば、比較的小さい自然数の集合とするのが普通であろう。

Nの要素の数$|N|$が大きい場合には、どのような順列を用いても上記の目標を達成できないことがありうる。そこで目標を少し弱めて、上位のl'ビットだけに注目すれば、m次均等分布をするならば満足することとして、そのようなl'がなるべく大きくなる順列を探すこととする。

この問題は次のような集合被覆問題として定式化できる。まず、

$$\{x'_{nt} : 1 \leq t \leq p, \qquad n \in N\}$$

に含まれるすべてのM系列の要素に対して上記の重みベクトルを計算し、それらを$\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_{lm}$とする。つぎに、成分が0または1である$p$次元ベクトルを

$$\mathbf{z} = (z_1, z_2, \cdots, z_p)$$

とし、次の整数線形計画問題を考える。

$$\min\ z_0 = z_1 + z_2 + \cdots + z_p$$

$$\text{s.t.}\quad \mathbf{e}_i \cdot \mathbf{z} \geq 1, \quad 1 \leq i \leq lm$$

これは集合被覆問題と呼ばれている典型的なNP困難な問題であり、pが大きいときには厳密な解を求めることは困難であるが、近似解を求める方法はいろいろ知られている。求められた近似解において$z_j = 0$となるjが$j_1, j_2, \cdots, j_q$で、他のjに対しては$z_j = 1$であったとしよう。このとき、順列$\{j(1), j(2), \cdots, j(l)\}$を

$$j(1) = j_1, j(2) = j_2, \cdots, j(q) = j_q$$

となるように定めると、上記のl'はq以上となる。

4. まとめ

乱数列の定義に関する研究と、乱数生成に関する研究との間のギャップを埋める努力の必要性を述べた。そのためのごく小さな試みとして、系統抽出(等間隔抽出)によって得られる数列に対しても多次元均等分布が保証されるような乱数生成アルゴリズムを設計することが重要であることを指摘した。具体例として、GFSRアルゴリズムを改善する方法を述べた。他のアルゴリズムについても、このような観点から検討してみることが重要であろう。

なお、このようなビットの入れ換えを実際に行うのは初期値の設定時のみであり、以後は一般化GFSR系列が満たす漸化式を使って系列を生成すれば、特別の操作を一切行うことなく、ビットの入れ換えを行った系列が自然に得られることを注意しておく。

参考文献

[1] 伏見正則: M系列に基づく乱数発生法に関する相反定理とその応用, 情報処理学会論文誌 **24**(1983), 576-579.

[2] Fushimi, M.: Designing a uniform random number generator whose subsequences are k-distributed, *SIAM J.Comput.* **17**(1988), 89-99.

[3] Fushimi, M. and Tezuka, S.: The k-distribution of generalized feedback shift register pseudorandom numbers, *Comm. ACM* **26**(1983), 516-523

[4] Knuth, D.E.: *The Art of Computer Programming, Vol.2: Seminumerical Algorithms*, 2nd Ed. Addison-Wesley, Reading, Mass., 1981.

[5] Kolmogorov, A.N.: On tables of random numbers, *Sankhya* **A25**(1963), 369-376.