乱数の生成法

by T.Koyama

１．線形合同法の欠点

通常、擬似乱数の生成は、ほとんど線形合同法

$$I_{j+1} = (aI_j + c) \mod m \tag{1}$$

に基づき計算される。m は法で、a と c は正の整数である。この式が生成する整数列は、0以上 $m-1$ 以下で、その周期は最大で m である（周期が m の場合、0 から $m-1$ までの全ての整数が I_0 から I_{m-1} のどこかで必ず現れる）。I_0 は初め任意に与えるが、数列の出発点が異なるだけで擬似乱数列の本質的な性質は変化しない。以上から、実際に使用する乱数の個数 m_A と m の値は $m_A \le m$ であることが望ましい。しかし、ANSI C の規格では m の値は少なくとも RAND_MAX=32767 であると規定されているだけであるので、例えばモンテカルロ法で 10^6 個の点に乱数を配置した場合、同じ 32767 個の点を 30 回づつ置いただけとなってしまう（つまり、戻り値が 2 バイトのライブラリ乱数ルーチンは実際の科学技術計算には全て使い物にならない）。それでは m の値を大きくすれば問題が解決されるかと言うとそうではなく、線形合同法には呼出し毎の系列相関があるという欠点がある。つまり、もし一度に k 個の乱数（範囲は[0,1)）を用いて k 次元空間に点をプロットすると、これらの点は k 次元空間を埋め尽くすのではなく、いくつかの(k-1)次元平面上に並んでしまう。この平面の個数はたかだか約 $m^{1/k}$ 個である。さらに定数の m ,a ,および c を不用意に選ぶと平面の個数はさらに減少する。以上の点を踏まえて、現在、実用に耐えると考えられている擬似乱数生成法について以下説明する。

２．乗算合同法

基本的に乗算合同法は線形合同法で定数 $c = 0$ とした場合

$$I_{j+1} = aI_j \mod m \tag{2}$$

である。m と a の値を、

$$a = 7^5 = 16807, \quad m = 2^{31} - 1 = 2147483647 \tag{3}$$

とすることによって、性質の良い乱数列が生成されることが知られている（Park and Miller）。しかし、式(2)では a と $m-1$ の積が 32 ビット整数の最大値を越えるために、この点を改善する手法が Schrage によって提案されている。Schrage の方法は m の近似的素因数分解

$$m = aq + r, \qquad q = \llbracket m/a \rrbracket, \quad r = m \mod a \tag{4}$$

に基づき定式化される。$\llbracket\ \rrbracket$ は内部の有理数の整数部分を表す。もし r が小さく、特に $r < q$ ならば $0 < z < m-1$ の範囲の z について、$a(z \bmod q)$ と $r\llbracket z/q \rrbracket$ がともに 0～$m-1$ の範囲に含まれ次式が成立する。

$$\begin{aligned} az \mod m &= a(r \mod m) - r\llbracket z/q \rrbracket, && \text{for } a(r \mod m) - r\llbracket z/q \rrbracket \ge 0 \\ az \mod m &= a(r \mod m) - r\llbracket z/q \rrbracket + m, && \text{for } a(r \mod m) - r\llbracket z/q \rrbracket < 0 \end{aligned} \tag{5}$$

式(3)の値の場合、

$$q = 127773, \quad r = 2836 \tag{6}$$

と置かれる。$m = 2^{31} - 1 = 2147483647$ に対する、その他の $m(a,q,r)$ の組み合わせには、

$$\begin{aligned} m(a,q,r) &= 2147483647(48271,\ 44488,\ 3399) \\ m(a,q,r) &= 2147483647(69621,\ 30845,\ 23902) \end{aligned} \tag{7}$$

がある。また $m = 2^{31} - 1$ とは異なる m に対しては、

$$\begin{aligned} m(a,q,r) &= 2147483563(48271,\ 44488,\ 3399) \\ m(a,q,r) &= 2147483399(69621,\ 30845,\ 23902) \end{aligned} \tag{8}$$

がある。実際の計算では、この $m(a,q,r)$ を用いて、以下で説明する切混ぜを行い、かつ両者の差を利用して乱数列を求める。

３．切混ぜアルゴリズム

これは上記の乗算合同法に基づく乱数値を用いるが、出力を切り混ぜて低次の系列相関を除く方法である。つまり乱数列の j 番目の乱数 I_j を、そのまま j 番目の呼出しで用いるのではなく、ランダムな回数（たとえば平均 $j+32$ 番目）だけ遅れて使用する方法である。この切混ぜの方法は非常に有効な手法で、擬態的な計算手順は以下のようになる。

（１）まず、乗算合同法にて適当な初期値から数回（7 回程度で良い）乱数を計算し、その後の乱数から、数値を配列に格納する。例えば配列の要素数を 32 とした場合、乱数を配列 iv[0]～iv[31]に順次格納していく。

（２）具体的に取り出す乱数の初期値 I_0 を iv[0]とする。また切混ぜ位置の決定に用いるダミー変数を j とする。

（３）１番目の乱数を決めるために、I_0 =iv[0]の値を $1+[\![(m-1)/32]\!]$ にて割り、j を

$$j = \left[\!\!\left[\frac{I_0}{1+[\![(m-1)/32]\!]} \right]\!\!\right] \tag{9}$$

にて定義する。$1+[\![(m-1)/32]\!]$ は法を 32 分割した量である。したがって、j は 0～31 の値を取る。得られた j を用いて、１番目の乱数として I_1 =iv[j]を抜き出す。空いた iv[j]には（１）の乗算合同法をそのまま続けて生成させた乱数を代入する。また I_1 を用いて次の j をあらためて、

$$j = \left[\!\!\left[\frac{I_1}{1+[\![(m-1)/32]\!]} \right]\!\!\right] \tag{10}$$

と置く。この j から２番目の乱数として I_2 =iv[j]を選択する。以下、以上の繰り返しによ

って、乱数列を求めることが出来る。

４．乱数の検定

計算された乱数列が一様かつランダムな分布となているかの検定には、以下の方法が用いられる。

（１）一様性のテスト

k 次のモーメントを計算する方法である。すなわち、

$$\left\langle I^k \right\rangle = \frac{1}{N}\sum_{i=1}^{N} I_i^k \tag{11}$$

を数値計算する。理想的な一様分布を $P(I)$ とすると、式(8)は近似的に $P(I)$ 自身のモーメントに一致する。つまり、

$$\frac{1}{N}\sum_{i=1}^{N} I_i^k \cong \int_0^1 I^k P(I)dI + O(1/\sqrt{N}) \cong \frac{1}{k+1} \tag{12}$$

が成立する。この式が成り立つのであれば一様分布である。またこの式からのずれが $1/\sqrt{N}$ に従って変化するならばランダムな分布である。

（２）相関を計算する方法

得られた乱数列に対して

$$C(k) = \frac{1}{N}\sum_{i=1}^{N} I_i I_{i+k} \tag{13}$$

を計算する方法である。k は小さな値とする。乱数 I_i と I_{i+k} が同時確率分布 $P(I_i, I_{i+k})$ に従うとする。この時、この２つの数が互いに独立で一様であれば、

$$\frac{1}{N}\sum_{i=1}^{N} I_i I_{i+k} \cong \int_0^1\int_0^1 xy\, p(x,y)dxdy = \frac{1}{4} \tag{14}$$

が成立する。この式が成り立つのであれば乱数列は無相関である。またこの式からのずれが $1/\sqrt{N}$ に従って変化するならばランダムな分布である。

（３）視覚的に確認する方法

これは視覚的に

$$x_i = I_{2i}$$
$$y_i = I_{2i+1}$$

の２次元散布図 (x_i, y_i) を見る方法である。乱数列が一様かつランダムであれば、このプロットは均一かつランダムに２次元平面を埋めていくはずである。雲のような目で見て認識できるパターンが認められた場合には、乱数に規則性があると結論される。