論　文

擬似乱数生成法の考察*

——乗算型合同式法のパラメタ選択と検定——

栗　田　良　春**

Abstract

Various methods of pseudo-random number generation have been proposed, discussed and criticized considerably since 1950. However, it can be said that method of generation has not been yet established, because of lack of clear definition of finite pseudo-random sequence and detailed description of its structure.

Concerning congruential method which, I think, has been most widely used, some of its properties have been disclosed recently.

The purpose of this paper is to show a procedure of choosing the suitable parameters of this congruential method. This procedure is based on such properties as hyper-plane structure and correlation coefficients over the entire period, taking account of programming advantage. Also, the value of the parameters chosen through this procedure and results of statistical tests of generated sequence are given.

1. まえがき

決定論的動作がその特長のひとつである現在の計算機で，あるアルゴリズムにしたがって有限列を生成し，それを確率変数の実現値の列として代用するとき，この列を擬似乱数と呼ぶ．1950年頃より，このための生成アルゴリズムの提案，検討，批判が数多くなされて来たが，有限擬似乱数の定義あるいは満たすべき性質が明確にできないという理由によって，その生成法は未だ確立されていないようである．そして最も広く用いられ，検討されていると思われる合同式法については，特にこの数年間にその性質，規則性が明らかにされた．そこで，ここでは乗算型合同式法：

$$X_{n+1}=a\cdot X_n \bmod m \qquad (1.1)$$

について，その性質，規則性とプログラミング上の問題を考慮に入れて，2進計算機用のパラメタ a, m の選択の手続きを考察する．そのために，次のような点を主としてとり上げる．

i）周期
ii）演算 mod の所要時間
iii）レジスタの桁数の有効利用
iv）超立方体中の超平面の枚数
v）周期全体の系列相関係数

更に，この手続きにしたがって選択されたパラメタの値およびその生成列の統計的検定結果についても報告する．

2. パラメタの選択基準

ここでは以下，まえがきでのべたパラメタの選択基準についてのべる．

i）周期をなるべく長くすること

乗算型合同式法(1.1)のもちうる最大の周期は $m-1$ であり，これは a が m に関する原始根の場合にのみ達成できる．そして原始根が存在するためには m は次の形でなければならない[1]．

$$m=2, 4, p^\alpha, 2p^\alpha$$

（ここに，p：奇素数，$\alpha=1, 2, \cdots$）

* Consideration of Pseudo-random Number Generators——Choice of Parameters of Multiplicative Congruential Methods and its Statistical Tests——by Yoshiharu KURITA (National Research Laboratory of Metrology).

** 計量研究所第2部

もちろん，周期を最大にすることはそれ程強い要請ではなく，mの大きさと生成列の使用する長さの比の問題にすぎないと思われるが，ここではむしろ最大周期の場合に，後述のiv)，v)の計算が比較的容易にできることもあって，最大周期$m-1$をもつように，a, mを選ぶ．

ii) 演算 mod が速く実行できること

原理的にはこの演算は整数割算の余りをとればよいが，mをかなり大きく（$m \geq \sim 2^{35}$）とるのが望ましいという報告[2]に従えば，たとえば $a \simeq 2^{30}, m \simeq 2^{35}$ として 65 ビット÷35 ビットの整数割算が必要である．このような桁数の割算がハードウェアで用意されていない場合に，これを割算のプログラムによって実現することは速度の著しい低下を招く．そこで，

$$m = 2^{\beta} \pm 1 \qquad (2.1)$$

の形にして演算 mod を高速化することが考えられているが[3]，このフェルマー型あるいはメルセンヌ型の数はひとつの素数の巾乗にはなり得ない（すなわち，$2^{\beta} \pm 1 = p^{\alpha}$ とすると $\alpha = 1$）[4]．こうして原始根をもつmの候補となり得るのはフェルマー素数あるいはメルセンヌ素数だけであり，これらのうちで生成式に使用可能なものは $2^{31}-1$, $2^{61}-1$ しかない．そこで次に，

$$m = 2^{\alpha} - 2^{\beta} \pm 1 \qquad (2.2)$$

の形の素数を考える．これは (2.1) の形に比較すれば演算時間がかかるが，多倍長のシフトおよび多倍長加減算によって実現できる(附録 (833 頁参照))．この形の m の候補を素数の中から選ぶと，$2^{15}-511$, $2^{29}-3$, $2^{35}-31$, $2^{47}-127$ などが適当と考えられる．

iii) レジスタの有効利用

mの大きさは使用できるレジスタの桁をなるべく有効に使うのが，周期を長くするために望ましい．この意味で，32, 36, 48 ビット・レジスタでは，たとえばそれぞれ，$2^{31}-1$, $2^{35}-31$, $2^{47}-127$ などが考えられる．

iv) 超平面の問題

合同式法がもつ，かなり顕著な欠点として「超平面の問題」が知られている．これにはいろいろな表現があるが[2),5),6)]，Marsaglia は幾何学的に見て次のようにのべている．混合型あるいは乗算型による生成列 $\langle x_n \rangle$ を n 個ずつとって，$\pi_i = (x_i, x_{i+1}, \cdots, x_{i+n-1})$, $(i = 1, 2, 3, \cdots)$ を一辺 m の，n 次元立方体の中の点列と見なすと，これらは層をなして，何枚かの平行・等間隔な超平面群のいくつかの組をつくる (**Fig. 1** 参照)．Coveyou はこれを列のフーリエ変換によって解析しているが，結局，最も肌目の粗い超平面の枚数 ν_n (Coveyou の表現ではフーリエ変換によって現われる最小の波数) を求めることは，次の数論的最小値問題に帰着する：$S_1 + S_2 \cdot a + \cdots S_n a^{n-1} = 0 \bmod m$ を満足する，すべてが 0 ではない整数の n-組 $(S_1, S_2, \cdots, S_n)$ の中で $\nu_n = \min_{S_i \in Z} (S_1^2 + S_2^2 + \cdots + S_n^2)^{1/2}$ を求めること．この解法を Knuth[7] が詳細に論じている．2 次元については，$a < \sqrt{m}$, あるいは $a > m - \sqrt{m}$ であれば $\nu_2 \simeq a$ であることが容易にたしかめられる．この ν_n には上限 $\bar{\nu}_n$ があって，$m^{1/n}$ に比例することが知られている（たとえば，$m = 2^{31}-1$ とすると $\bar{\nu}_8 \simeq 21$)．したがってモンテカルロ法のひとつの守備範囲ともいうべき多次元の問題において，このような点の作り方は，その使用法によっては危険な場合も起こりうる．こうして ν_n をなるべく多くするように a を選ぶべきであるというひとつの選択基準が得られる．

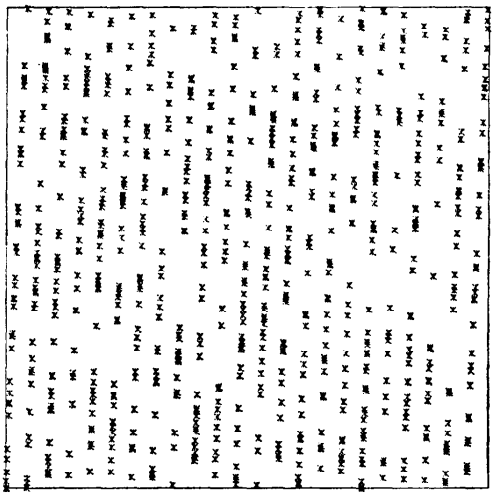

Fig. 1 Hyper-plane Structure (2-dim.)
$X(n) = 23 * X(n-1) \bmod (10**8+1)$
$X(0) = 1$, # points $= 859$

v) 周期全体の系列相関係数

パラメタ選択の次の基準となるものとして，周期全体の系列相関係数 C がある：

$$C = \frac{m \cdot \Sigma\, x(ax \bmod m) - (\Sigma\, x)^2}{m \Sigma\, x^2 - (\Sigma\, x)^2}$$

ここに Σ は x_n のとり得るすべての値についての和であり，それが $[1 : m-1]$ を尽くす場合に一般化されたデデキント和を用いて C を数値評価できる[7]．つまり，一般的には最大周期をもつ場合に，lag=1 の

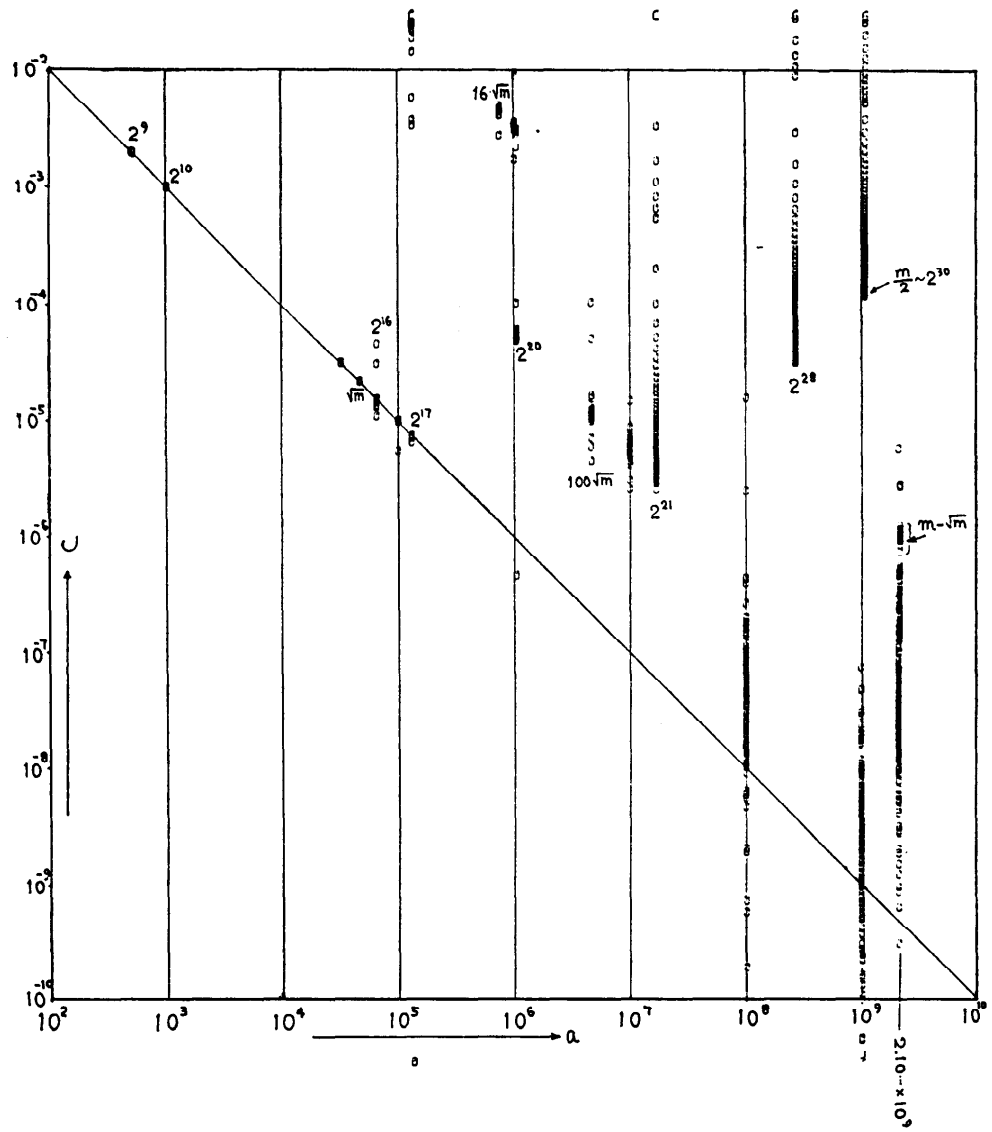

Fig. 2 Serial correlation applied over the entire period ($m=2^{31}-1$)

系列相関係数を計算できる*.

$m=2^{31}-1$ の場合の計算結果のグラフを **Fig. 2** に示す. 横軸 a は j から始まる引き続いた原始根を数十個ずつ, $j=2^9, 2^{10}, \sqrt{m}, \cdots, -\sqrt{m}$ などについてプロットしたものである. 縦軸は C の値で, いずれも log スケールである. これらが判ることは, a/m が小さな所では $1/a$ の直線の上に乗っているが, $a/m\to1$ にしたがって, C は大きくばらつき始める. このことより, m に近い所で a を注意深く探せば C の非常に低い原始根 a を見つけることができる筈である. 但し, これは周期全体の系列相関係数であって, 部分列についてのローカルな保証ではないことに注意する必要がある.

この他にも, いくつか選択の基準が考えられ, 提案

---

* (1) $x_{n+l}=a^l\cdot x_n \bmod m$
(2) a が原始根のとき a^l も原始根である条件は $(l, m-1)=1$ なることである.
これらのことより, 奇数の lag については計算できる可能性がある.

されているが[6),8),9)]，現在までに判っている性質は本質的には以上で尽くされていると思われる．

3. パラメタの選択手続き

以上の考察により，ここでは次のような手続き〔P1〕～〔P4〕によってパラメタ，a, m を選び出した．

〔P1〕: 使用する計算機のレジスタの桁数をなるべく有効に利用する m の大きさの上限を定める．

〔P2〕: この上限に近い所で(2.1)あるいは(2.2)の形の素数を探し，m とする．

〔P3〕: この m について，v)でのべた系列相関係数 C が低い値をもつような a の区間 I_m を探す．これは Fig. 2 で予想がつくようにかなり m に近い所であり，かつ $a<m-\sqrt{m}$ であるべきである．その区間 $|I_m|$ の長さをどの位にとるかは，以下の〔P4〕の計算が時間的に可能であるように決められる．

〔P4〕: この区間 I_m から原始根をすべて選び出す．すなわち，m に関する原始根の全体を $G(m)$ として $A=\{a|a\in G(m)\cap I_m\}$ なる A を作る．この A の各々を **Fig. 3** に示すような超平面の枚数による篩にかけて a を選び出す．これは，n 次元 $(2\leq n\leq 7)$ の a による最も粗い超平面の枚数 ν_n を $n=2$ から順に計算し，各 ν_n がすべて篩の目の粗さ δ_n よりも大きい場合だけ，その a を残す．

このような手続きによる計算結果を以下に示す．例として $m=2^{31}-1$ の場合についてのべる．この場合，$\#G(m)=\varphi(m-1)=534,600\ 000\simeq m/4.017$，ここに φ はオイラーの関数，であり，$\forall g\in G(m)\rightarrow m-g\notin G(m)$ である．$I_m=[2\ 100\ 000\ 000, 2\ 100\ 030\ 207]$ とすると，$\#A=7440$，このとき篩の目の粗さ δ_n を ν_n の 0.7 倍程度に選ぶと十数個の原始根が a の候補として残る．これらのうち，系列相関係数 C の最も低いものから3例を，他の m の場合と合わせて，**Table 1** に示す．計算時間は原始根の選択および ν_n, C の計算に多倍精度の整数四則演算（m の 3～4 倍の桁数）が必要であるので，16 bit＋16 bit の所要時間が 2 μsec 程度の計算機で，数時間を要した．

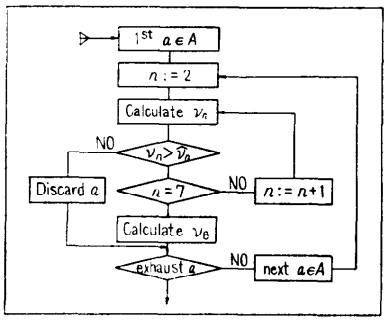

Fig. 3 Sieve by hyper-plane structure

また生成時の乗算時間を短くするために，a を m と同じく，$2^{\alpha}-2^{\beta}(\pm 2^{\gamma})$ の形の原始根を求め，ほぼ同様の手続きで選び出した a についての ν_n $(2\leq n\leq 5)$ および C の値を **Table 2**（次頁参照）に示す．a の形を限定したため，Table 1 に比べて，ν_n および C の値は多少悪くなっているが，多精度の乗算も時間がかかる場合には有効と思われる．

4. 統計的検定

ここまでは乗算型合同式法による生成式のパラメタだけから判る性質について検討して来た．使用目的を限定しない汎用の生成式の提示はここまでであるが，

Table 1 Minimum hyper-planes and serial correlation coefficents C of some selected a, m for $x_{n+1}=a\cdot x_n \bmod m$.

	a	2-dim.	3-dim.	4-dim.	5-dim.	6-dim.	7-dim.	8-dim.	C
(I)	$m=2^{47}-2^{7}+1$								
	126 903 398 710 863	1.17×10^{7}	46427	3142	681	198	99	41	-1×10^{-13}
	126 903 398 710 871	1.19×10^{7}	36765	2546	672	237	102	54	2×10^{-13}
	126 903 398 710 988	1.21×10^{7}	49521	3100	546	199	103	55	1×10^{-13}
	upper lim	1.27×10^{7}	58390	4096	831	254	141	83	
(II)	$m=2^{35}-2^{3}+1$								
	25 000 000 495	122 708	3208	393	118	48	32	19	5×10^{-10}
	25 000 001 926	152 216	2801	346	103	48	32	17	-3×10^{-10}
	25 000 007 056	146 587	3239	433	106	56	29	18	3×10^{-10}
	upper lim	199 920	3649	512	158	74	43	29	
(III)	$m=3^{31}-1$								
	2 100 005 341	43 487	1202	206	65	32	19	14	1×10^{-8}
	2 100 016 018	35 193	1155	223	70	34	19	13	9×10^{-8}
	2 100 017 008	44 643	1197	194	69	32	21	11	-1×10^{-8}
	upper lim	49 797	1448	256	91	47	29	21	

Table 2 Minimum hyper-planes and serial correlation coefficients C of some selected a, m for $x_{n+1}=a\cdot x_n \bmod m$ (form of a: $2^q+2^p\pm2^r$)

	2-dim.	3-dim.	4-dim.	5-dim.	C
(I) $m=2^{47}-2^7+1$					
$2^{36}-2^{12}$	8.4×10^6	4.7×10^4	2.3×10^3	5.6×10^2	-4.4×10^{-11}
$2^{44}-2^{28}+1$	8.9×10^6	4.9×10^4	2.1×10^3	5.5×10^2	-5.8×10^{-11}
upper lim	1.2×10^7	5.8×10^4	4.1×10^3	$8\ 3\times10^2$	
(II) $m=2^{35}-2^5+1$					
$2^{27}-2^{18}-1$	1.7×10^5	2.1×10^3	3.1×10^2	1.0×10^2	2.6×10^{-10}
$2^{30}-2^{16}-1$	1.3×10^5	2.9×10^3	3.4×10^2	1.1×10^2	-5.4×10^{-9}
upper lim	1.9×10^5	3.6×10^3	5.1×10^2	1.5×10^2	
(III) $m=2^{31}-1$					
$2^{23}-2^{14}+4$	3.1×10^4	1.2×10^3	1.8×10^2	6.6×10^1	1.3×10^{-7}
$2^{26}-2^{16}+2$	3.4×10^4	1.2×10^3	1.8×10^2	5.1×10^2	-5.4×10^{-9}
upper lim	4.9×10^4	1.4×10^3	2.6×10^2	9.0×10^2	

実際の使用にあたっては，その生成された列の使用する部分（列）についてのローカルな統計的検定を試みる必要があるといわれている．そこで上述の手続きによって選択された生成式を含む種々の生成列についての検定結果についてのべる．

検定法は数多く提案されているが，ここでは最も sensitive* といわれている[10]連の検定についてだけ報告する：列 $\{x_0, x_1, \cdots, x_{n-1}\}$ から n 個の長さ l の部分列 $B_k=\{x_{kl}, x_{kl+1}, \cdots, x_{(k+1)l-1}\}$, $(k\in[0:n-1])$ をとり出し，各 B_k について長さ1から5までの連および6以上の連の頻度 C_i $(i\in[1:6])$ をカウントする．たとえば，3,5,6,2,9,5,1,4,1,3 という長さ 10 の列については長さ 1,2,3 の上昇連がそれぞれ 1,3,1 個あると考えることにする．この C_i から

$$V=\Sigma(C_i-E(C_i))(C_j-E((C_j))\cdot A^{-1},$$

ここに A は C_i と C_j の共分散の期待値の行列，を求めると，V は l が大きいとき $d.f.6$ の χ^2-分布に従う（詳しくは Knuth[7] pp. 60～63 参照）．$m=2^{31}-1$, $a=2\ 100\ 005\ 341$, $x_0=1$, $l=4096$ の場合の例を **Table 3** に示す．第1，2行目の up, down の括弧内はこの列のはじめの l 個についての実測値であり，その下の行は引き続いた l 個についてのものである．こうして，一回の連の検定で引き続いた 128 個の部分列(長さ 4096)を調べ，128 個の χ^2-値を求めた．結局ひとつの列について 2^{19} の長さについて調べることになる．しかし，このままではその結果の良否が直ちには判り難いので，これら 128 個の χ^2-値は $d.f.6$ の分布に従うことから，その期待分布と実測分布をプロットしたものが **Fig. 4**（次頁参照）である．横軸が $d.f.6$ の χ^2-密度関数の積分の上限であり 20 までプロットしてある．滑らかな曲線が理論分布であり，折線が観測値である．数十種類の列について l も変化させてこの検定を試みた．その結果を次に記す：

Table 3 Observed and expecled frequency table of length of up and down run.

#45 ($M=2^{**}31-1$, $A=2100005341$)
BLOCK LENGTH: 4096

	Length of run	1	2	3	4	5	6≤	CHI**2
1	UP	(651	881	368	117	16	5)	0.618E+01
	DOWN	(673	890	375	100	16	6)	0.605E+01
2	UP	(706	864	372	95	27	5)	0.316E+01
	DOWN	(667	856	349	119	34	4)	0.860E+01
3	UP	(682	816	382	118	28	4)	0.401E+01
	DOWN	(720	828	384	113	22	1)	0.676E+01
4	UP	(707	867	372	106	18	4)	0.273E+01
	DOWN	(657	844	384	104	27	8)	0.607E+01
5	UP	(738	867	373	97	21	2)	0.839E+01
	DOWN	(639	812	404	112	24	8)	0.124E+02
6	UP	(698	874	367	110	18	3)	0.327E+01
	DOWN	(656	848	383	112	27	2)	0.400E+01
7	UP	(726	854	366	97	29	5)	0.598E+01
	DOWN	(670	823	372	121	31	4)	0.535E+01
8	UP	(662	881	400	88	14	1)	0.110E+02
	DOWN	(647	881	388	108	12	5)	0.953E+01
9	UP	(667	902	353	105	19	8)	0.106E+02
	DOWN	(656	877	388	97	22	4)	0.437E+01
10	UP	(692	872	374	92	28	5)	0.448E+01
	DOWN	(664	855	373	117	22	4)	0 181E+01
EXPECTED V.		683	853	375	108	23	4.8	

a. Table 1,2 に掲げた選択されたパラメタに関しては，$l=2^{11}, 2^{12}, 2^{13}$ について，長さ 2^{21} まで $(x_0=1)$ まで調べた結果，特に悪いものは見当らない．かなりよく一致している例および，比較的悪い例をそれぞれ Fig. 4 (a), (b) に示す．

b. $m/2$ に近い原始根 a を選ぶと短い連（特に長さ2）が多すぎることにより，適合度が非常に悪い．(Fig. 4 (c) 参照)．$m/4$, $m/8$ でもこの傾向は l によっては，僅かに認められる．

c. Knuth** が発表している一様性のテストで悪い結果を示す列については，この連の検定でも非常に悪い．また混合型には悪い例が認められ，これは部分列についてのフーリエ変換によるスペクトル・テストでも認められた[11]．

d. 超平面，周期全体の系列相関係数による結果とこの連の検定結果との相関は認め難いようである．

e. a が $\sqrt{m}$ あるいは $m-\sqrt{m}$ 附近の原始根であってもこの検定では異常は見出せない (Fig. 4 (d)参照)．

* 検定法 t が最も Sensitive であるとは t 以来の検定法で有意差が検出された列は，t でも有意差が検出されることが多いという意味である．

** 文献 7) page 40.

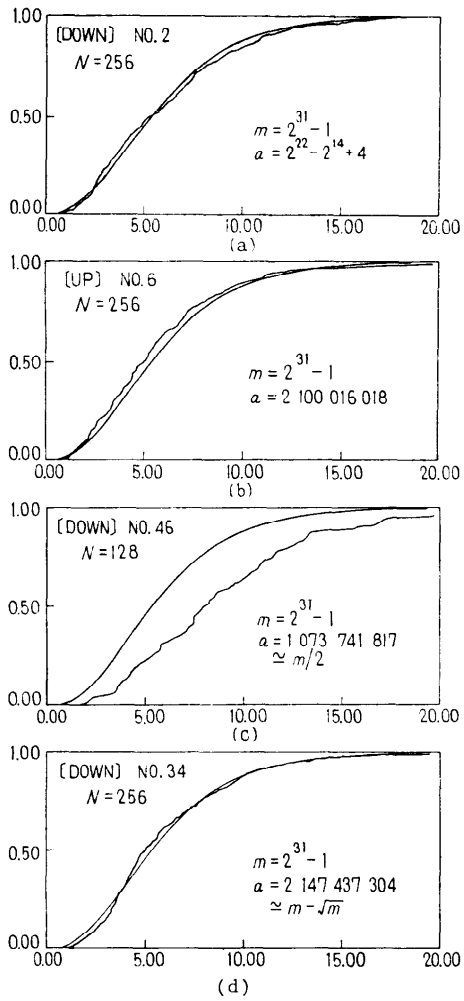

Fig. 4 Examples of observed distributions of runs up and down tests

5. 結　び

超平面構造（lattice 構造）の問題など，この10年間にあきらかにされた主な性質およびプログラミング上の問題も考慮に入れた，乗算型合同式法のパラメータの選択法のひとつを示した．擬似乱数という特質上，はっきりとした結論づけは困難であるが，少なくとも，問題となるいくつかの性質に対して解答が用意されているという意味で，そうでないものよりも使い易いはずである．

また，統計的検定についても明快な結論が出て来ないが，元来検定法というものが，理想的なランダムネスからの偏りをできるだけ鋭敏に結果に反映させるように仕組まれたものであることを考えると，通常の乱数の使用には充分耐えるのではないかと思われる．

謝辞．この研究にあたり，計量研究所第2部，森村正直・熱学計測課長および米田信夫・学習院大学教授に有益な示唆とはげましを頂いた．また日本アイ・ビー・エム，渋谷政昭氏にもいろいろとご教示を頂いた．ここに深く感謝いたします．

参考文献

1) ヴィノグラードフ著　三瓶，山中訳：整数論入門，p. 200，共立出版，東京 (1959).
2) R. R. Coveyou & R. D. MacPherson: Fourier Analysis of Uniform Random Number Generators, JACM, Vol. 14, No. 1, pp. 100～119 (1967).
3) W. H. Payne et al.: Coding the Lehmer Pseudo-random Number Generator, CACM, Vol. 12, No. 2, pp. 85～86 (1969).
4) W. Sierpiński: Elementary Theory of Numbers, Warszawa (1964).
5) G. Marsaglia: Random Numbers fall in the Planes, Proc. N. A. S., Vol. 61, pp. 25～28 (1968).
6) G. Marsaglia: The Structure of Linear Congruential Sequences, (∈Zaremba (ed.): Application of Number theory to Numerical Analysis, Academic Pr. (1972)).
7) D. E. Knuth: The Art of Computer Programming, Vol. 1. 2, p. 624, Addison-Wesley, (1971).
8) U. Dieter: Pseudo-Random Numbers: The Exact. Distribution of Pairs, Math. of Comp. Vol. 25, No. 116, pp. 855～883 (1971).
9) U. Dieter: Statistical Interdependence of Pseudo-Random Numbers Generated by the Linear Congruential Method, (∈ Zaremba (ed.): Application of Number theory to Numerical Analysis, Academic Pr. (1971).).
10) D. Y. Downham: The runs up and down test. Computer J. Vol. 13, pp. 373～377 (1970).
11) 栗田良春，森村正直：乱数の評価(I)～(IV)，第 12～15 回情報処理学会大学予稿集．

附　録

$a.x \bmod m$, $m=b^{\alpha}-b^{\beta}+1$ について，

$1\leq a\leq m-1$, $1\leq x\leq m-1$ のとき

$$Y=X_0 \bmod m,\ 1\leq X_0\leq (m-1)^2-1,$$

$$X_{j+1}=\lfloor x_j/b^{\alpha}\rfloor(b^{\beta}-1)+X_j \bmod b^{\alpha},\ j=0,1$$

とすれば　$y-X_2=0 \bmod m$　(1)

であって $\alpha\geq 2\beta>0$ であれば

$$X_2<2m \qquad (2)$$

である．

$\because\ X_{j+1}=X_j-\lfloor X_j/b^\alpha\rfloor(b^\alpha-b^\beta+1)$

したがって，(1)が成立．

$Z-1<\lfloor Z\rfloor\leq Z$ を用いて，

$$X_{j+1}\leq X_j(1-m/b^\alpha)+m-1.$$

ここで

$s=1-m/b^\alpha=(b^\beta-1)/b^\alpha$ とおくと

$s>0$ であって

$$X_{j+1}\leq sX_j+m-1.$$

したがって

$$X_2\leq s^2X_0+s(m-1)+(m-1)$$
$$\leq s^2(m-1)^2+s(m-1)+(m-1)-s^2.$$

また

$s(m-1)<b^\beta-1$ であるから

$$X_2<2m-(b^\alpha-b^{2\beta}+2+(1-m/b^\alpha)^2)$$

したがって，(2)が成立する． ☒

(昭和50年6月16日受付)