数理解析研究所講究録
第498巻 1983年 153-171

M系列のL-tupleの weight distribution の偏りについて

計量研究所　栗田良春
(Y.KURITA)

§1　はじめに

擬似一様乱数の生成法として現在用いられている手法は，合同式法とM系列法の2種に大別できる．この合同式法は，D.H.Lehmer（1949）によって提案され，その後20年間，Coveyou-MacPherson（1967），Marsaglia(1968)等によってその本質的な規則性：「結晶構造をもつこと」が指摘される頃まで，殆ど唯一の方法であった．この頃，Tausworthe(1966）によりM系列を用いる方法が提唱され，その発生能率を改善する方法がLewis-Payne（1972）によって示された．合同式法についてはその欠陥の指摘，M系列法については初期値さえ予め用意しておけば合同式法に劣らない速さで発生できる方法の提示等の条件が重なって，後者の方法はTLP（Tausworthe-Lewis-Payne）列と名付けられ，広く用いられるようになった(と思われる)．

筆者は，これら種々の擬似乱数について統計的検定の実験

を試みるうち，一様性の分布に関してＴＬＰ列が異常な振舞いをしばしば示すことを見い出し，その原因がＭ系列の部分列における０と１の個数の分布の偏りにあることをつきとめた．この偏りは，現在までの実験では次数が数百までの３項原始多項式によるＭ系列では，統計的検定によるまでもないほど殆んど常に，同じ傾向をもつ偏りであり，一言でいえば「０の個数の多すぎる部分列があり，１についてはそうではない」と要約される．本報告はこの実験結果の一部を示すもので前半で［０,１］一様乱数としてのＴＬＰ列の偏りを，後半でその由来である２値（０,１）系列としてのＭ系列の部分列の重みの偏りについて述べる．

§２　ＴＬＰ列の一様性についての Kolmogorov-Smirnov 検定（ＫＳ test）

ある確率変数Ｘの分布関数Ｆ(ｘ)と予め与えられた連続な分布関数Ｆ₀(ｘ)との適合度を検定する手法のひとつに標記：ＫＳ testがある．これは，分布の型によらない，いわゆるノンパラメトリックな検定法で，具体的にはまず，Ｘからのｎ個のサンプルから標本分布関数を作る，
すなわち，昇順に並べたサンプル：$x_{r_1}, x_{r_2}, \cdots, x_{r_n}$ に対して，

$$F_n(x) = \begin{cases} 0 & : x < x_{r_1} \\ j/n & : x_{r_j} \leq x < x_{r_{j+1}} \ ; \ j = 1,2,\cdots,n-1 \\ 1 & : x_{r_n} \leq x \end{cases} .$$

この$F_n(x)$と$F_0(x)$の差を次の2量K_n^+, K_n^-で測る：

$$K_n^{\pm} = \sqrt{n} \ \max\{ \pm F_n(x) \mp F_0(x) \} .$$

この統計量$K_n^{\pm}$は共に次の分布関数に従う：

$$Pr(K_n^{\pm} \leq t/\sqrt{n}) = \frac{t}{n^n} \sum_{k=0}^{\lfloor t \rfloor} \binom{n}{k} (k-t)^k (t+n-k)^{n-k-1} \quad (1).$$

（ここに，$Pr(\cdot)$は事象・の起る確率）．

以上がKS testの概要である．

さて，[0,1]一様分布が期待される列$x_1, x_2, \cdots$からL-tupleを順次N個とり出す，すなわち，i番目（$1 \leq i \leq N$）のL-tupleとは$\{x_{(i-1)L+1}, x_{(i-1)L+2}, \cdots, x_{iL}\}$のことである．（尚，L，Nは以下この意味で用いる）．

このL-tupleに対してF_0を一様分布として上述のKS testを適用し．それぞれN個のK_L^+, K_L^-を得る．このN個のK_L^+は（1）式の分布に従うことが期待されるから，これをF_0として再びKS testを行いK_N^+, K_N^-を得る．K_L^-についても同様である．こうしていわば2重のKS testによって2組の(K_N^+, K_N^-)から列$\{x_i\}$の一様性の検定ができる．

さて，3項原始多項式：

$$X^p + X^q + 1 \qquad (2)$$

（p, q は以下この意味で用いる），によるTLP列の生成法は Lew., Pay.[72] に具体的に手続きが示されており，そこで提示されている一例：p＝98，q＝27 に対して上述のKS testを適用した結果，いくつかのLで非常に大きな分布の偏りが検出された．その1例を Fig.1 (a), (b) に示す*．これはL＝513，N＝512の場合の2段目のKS test の結果すなわち K_L^+ の実測分布を（1）式による F_0 と併せて plot したものである．　この図からN個のL-tuple の中には一様分布から大きく偏った分布をもつものがあり，更にこの図示例に限らず，偏りはいつもこの傾向をもつことから，すなわち，K_L^+ の実測分布は常に F_0 を下まわり，K_L^- のそれは逆であることから，非常に小さい平均値をもつL-tuple があることが判る．実際，この列を調べてみると，363番目のL-tuple は $K_L^+ = 2.84$（図の scale out 右端に相当）であって，平均

*）　この図中，(c), (d) は合同式法 $x_n = (2^{16}+3)x_{n-1} \bmod 2^{31}$ による列の検定結果で，$x_1 = 1$ として x_{100} からの列を使ったものであり，比較のために示した．尚，このパラメタは前述の結晶構造の点からは極めて性質が悪く，一般の使用には耐えない．

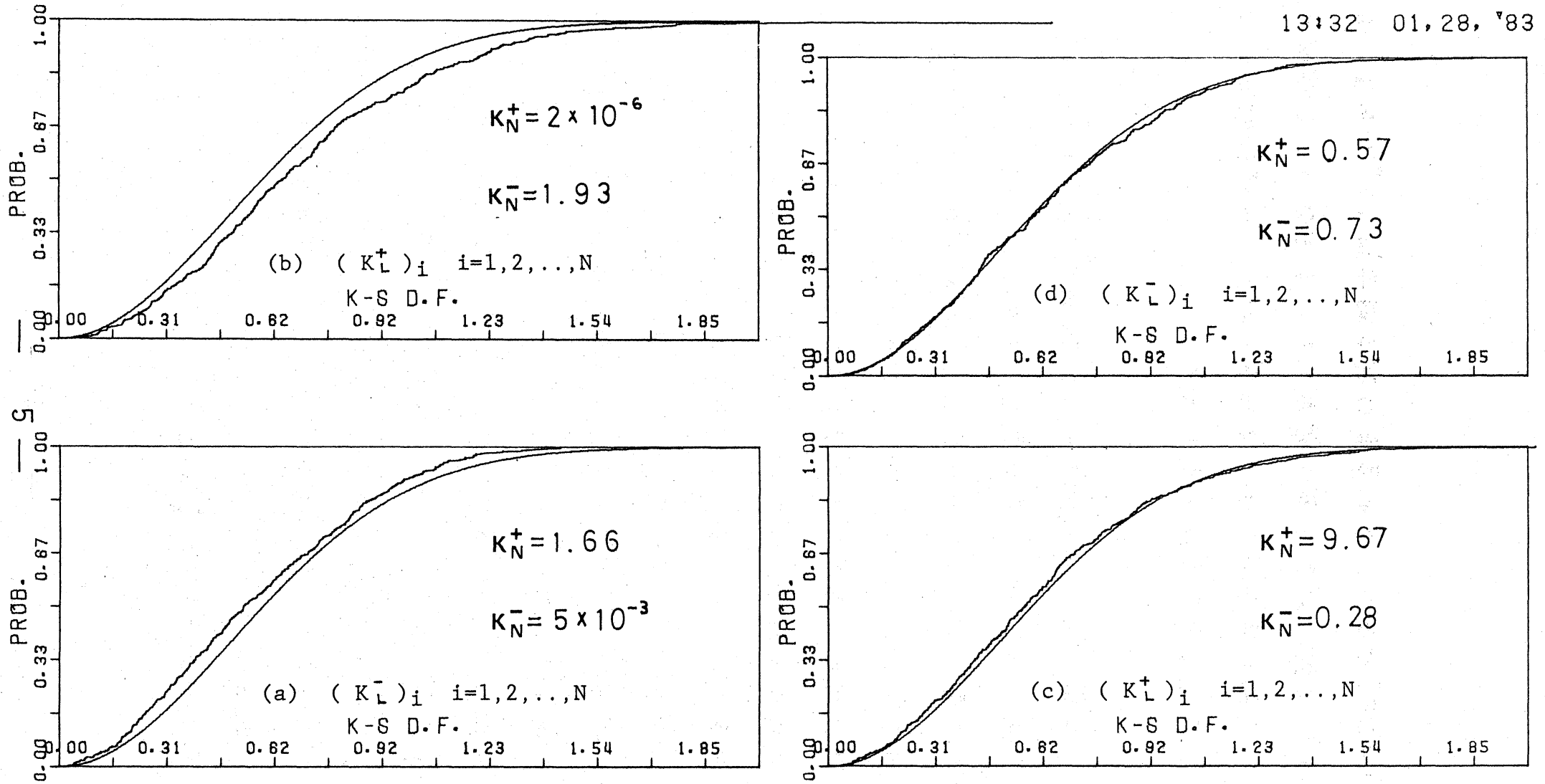


Fig.1 Observed distribution of KS-statistics (L=513,N=512)

(a),(b): TLP($X^{98} + X^{27} + 1$)

(c),(d): Congruential method ($X_n = (2^{16}+3) \cdot X_{n-1} \ \underline{\text{mod}} \ 2^{31}$)

値は 0.435 であり，この事象の起る確率は 10^{-7} である．勿論，10^{-7} で起る現象を偶々観測したにすぎないとも解釈できるが，次節の結果からは3項M系列のもつ一般的性質に由来すると考える方が妥当と思われる．

§3 「異常に軽い weight をもつ L-tuple が M系列にある」

前節で扱った [0,1] 一様乱数としてのTLP列 $x_1, x_2, \cdots$ の引続いた要素 x_i, x_{i+1} のMSBは(2)式に従う 0,1 を要素とするM系列 $a_1, a_2, \cdots$ の引続いた要素 a_i, a_{i+1} である．従って，前節の結果：「異常に平均値の低い L-tuple がある」とはそのもとになるM系列についていえば，「0が異常に多い L-tuple がある」ことである．M系列の L-tuple (L=513)では平均として L/2 = 256.5個の0が期待されるが，前述363番目の L-tuple のMSBを作るM系列の L-tuple には320個の0が観測される．

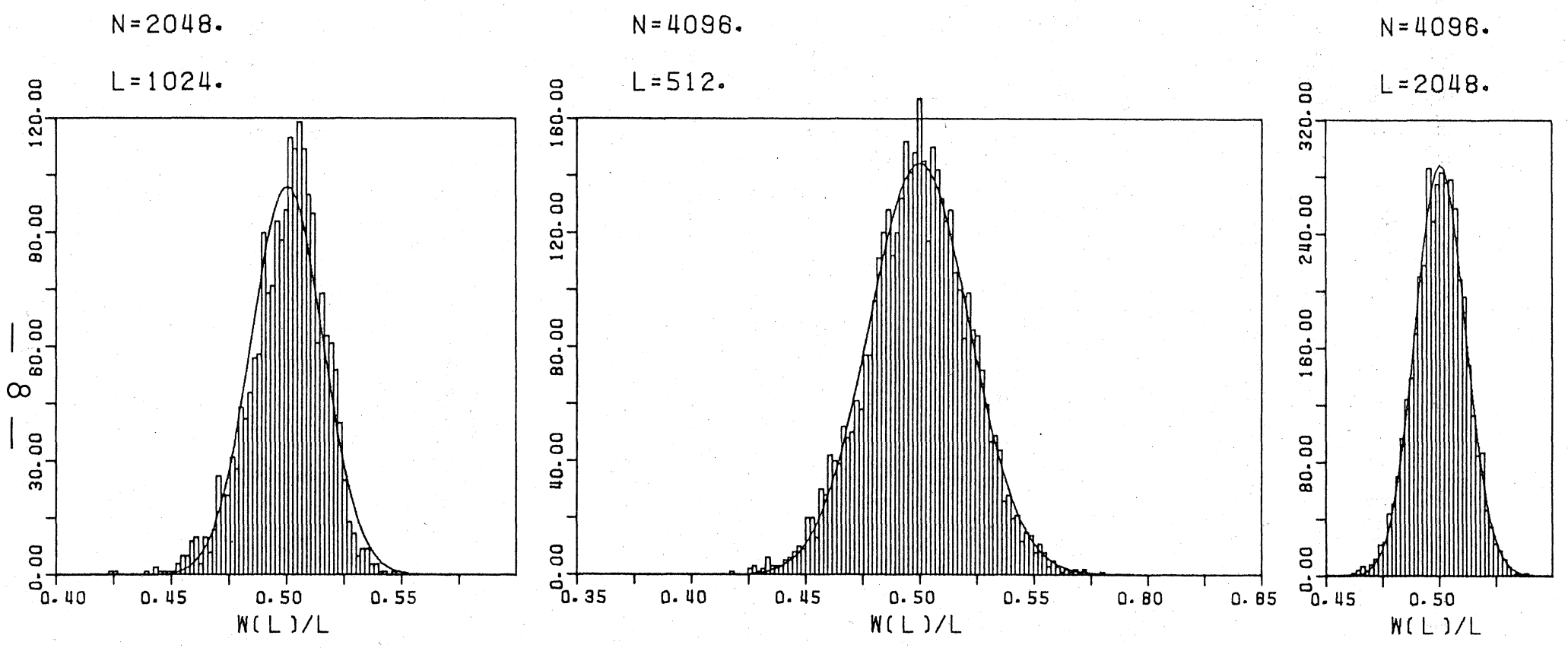
この現象は，種々の原始多項式(2)によるM系列でも見ることができる．まず，(0,1)2値系列の L-tuple の中の1の個数を weight; w(L) と呼ぶことにすると，M系列が乱数の実現として有効であるためには w(L) が2項分布に従うことが必要である：

$$Pr(w(L)=k)=\binom{L}{k}(1/2)^L \qquad (3).$$

そこで，$28 \leq p \leq 1279$ の範囲の17種の3項原始多項式，いくつかのL，Nの組合せ（$L \cdot N = 2^{21 \sim 23}$），数通りの初期値のパタンについて組合せ計250例について，期待分布(3)と実測頻度とのずれ，及び実測頻度の3次モーメント M_3 の値を調べた．（初期パターンは，$x_n = (2^{16}+3) \cdot x_{n-1} \bmod 2^{31}$，$x_1 = 1$ として $x_{I+1}, x_{I+2}, \cdots, x_{I+p}$ のMSBから与えた）．そのうちの3例をFig.2に示す．いずれも，縦軸は頻度，横軸はLで規格化した重み：$w(L)/L$，滑らかな曲線が(3)式による期待値，柱状図が実測ヒストグラムである．この図からも読みとれるが，観測した実測分布の殆んどは2項分布に比べて一般的に，「左の尾が長く，左肩は痩せ，右肩は太り，右尾はない」形をもち，$M_3 < 0$ である*．
実際，$p \leq 127$ については殆んどいつでも（q の値，初期パターンに拘らず），$p \geqslant 521$ についてはいくつかの場合，この形をもつ．

---

*）すなわち，これは本節の標題に示した「軽すぎるweightをもつL-tuple」があり，重すぎるL-tupleはないことを示しており，（勿論，1周期の中の0の個数と1の個数の差は1にすぎないから），生じたアンバランスは分布の左右の肩で相殺されている．



(a) X ** 89 + X ** 38 + 1

I=100, M_3 =-2302

(b) X ** 1279 + X ** 216 + 1

I=2700, M_3 =-143

(c) X ** 607 + X ** 273 + 1

I=100, M_3 =-1247

Fig. 2. Empirical weight distribution of (L-tuple)$_i$; i=1,2,..,N

§4 「軽い p-tuple に続く p-tuple は軽い」

前節でのべた偏りは，L/p が小さいと目立たなくなり，L > P/2 であるようなLでは分布の左尾が長くなる傾向が見られる．

本節では，このLをpに等しく選び，k番目の p-tuple と (k+1)番目の p-tuple の間の weight の相関を調べた結果を報告する．

原始多項式(2)によって生成されるM系列 $\{a_1, a_2, \cdots\}$ とは次の漸化式によって定義される列であるとしてよい．

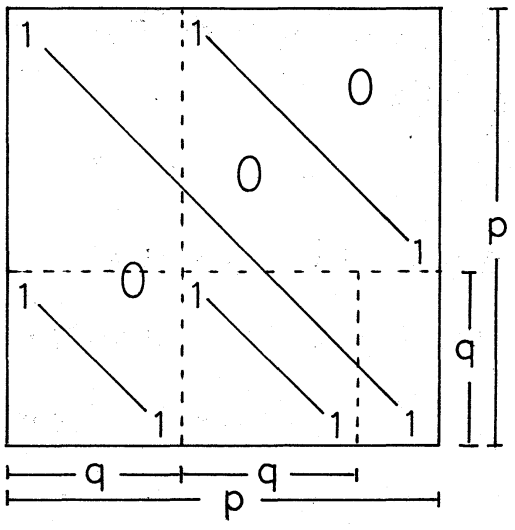
$$a_n = a_{n-p} + a_{n-p+q} \quad \underline{\mathrm{mod}}\ 2$$

そこで，a_n から始まる p-tuple $\{a_n, a_{n+1}, \cdots, a_{n+p-1}\}$ を転置し，縦ベクトル A_n と名付けると，

$$A_n = T \cdot A_{n-p} \quad \underline{\mathrm{mod}}\ 2 \qquad (4)$$

が成立つ．ここに，

$T =$



但し $p > 2q$.

さて，A_n の weight を $w(A_n)$ とすると，$w(A_n)$ と $w(A_{n-p})$ とは，乱数の実現という意味では独立であることが望ましいが，前節までの結果からはかなり強い相関をもつことが予想される．その理論的解析は(4)からできる筈であるが，それは別報に譲り，ここでは以下，その実験結果について述べる．

まず，M系列の1周期を考える．その中には p ビットのすべてのパターンが（すべてがゼロのパターンを除いて）一度ずつ現われ，その総数は 2^p-1 である．これらのうち，その weight が i であるパターンの数は ${}_pC_i$ である，（$1 \leqslant i \leqslant p$）．

次に，$i = w(A_{n-p})$，$j = w(A_n)$ であるような A_n，（$p+1 \leqslant n \leqslant 2^p+p-1$）の総数を h_{ij} とすると $H=(h_{ij})$ が weight の推移行列となり，その第 i 行の和，第 j 行の和はそれぞれ ${}_pC_i$，${}_pC_j$ であり，$w(A_{n-p})$ と $w(A_n)$ とが独立であると仮定すれば，h_{ij} の期待値：$h_{ij}^{(EX)}$ は

$$h_{ij}^{(EX)} = \frac{{}_pC_i \cdot {}_pC_j}{2^p-1} \qquad (5)$$

とかける．この期待値は，すぐに判るように，$i=j=p/2$ で最大値をとり，この点を中心として回転対称な分布である．

ここで，$p=15$，$q=4$ の場合の期待分布(5)を Table.1 に，その実測分布 $h_{ij}^{(EM)}$ を Table.2 に示す．

i → j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1*	.0	.0	.2	.6	1.4	2.3	2.9	2.9	2.3	1.4	.6	.2	.0	.0	.0
2*	.0	.3	1.5	4.4	9.6	16.0	20.6	20.6	16.0	9.6	4.4	1.5	.3	.0	.0
3*	.2	1.5	6.3	19.0	41.7	69.5	89.4	89.4	69.5	41.7	19.0	6.3	1.5	.2	.0
4*	.6	4.4	19.0	56.9	125.1	208.5	268.1	268.1	208.5	125.1	56.9	19.0	4.4	.6	.0
5*	1.4	9.6	41.7	125.1	275.2	458.7	589.7	589.7	458.7	275.2	125.1	41.7	9.6	1.4	.1
6*	2.3	16.0	69.5	208.5	458.7	764.5	982.9	982.9	764.5	458.7	208.5	69.5	16.0	2.3	.2
7*	2.9	20.6	89.4	268.1	589.7	982.9	1263.7	1263.7	982.9	589.7	268.1	89.4	20.6	2.9	.2
i 8*	2.9	20.6	89.4	268.1	589.7	982.9	1263.7	1263.7	982.9	589.7	268.1	89.4	20.6	2.9	.2
9*	2.3	16.0	69.5	208.5	458.7	764.5	982.9	982.9	764.5	458.7	208.5	69.5	16.0	2.3	.2
10*	1.4	9.6	41.7	125.1	275.2	458.7	589.7	589.7	458.7	275.2	125.1	41.7	9.6	1.4	.1
11*	.6	4.4	19.0	56.9	125.1	208.5	268.1	268.1	208.5	125.1	56.9	19.0	4.4	.6	.0
12*	.2	1.5	6.3	19.0	41.7	69.5	89.4	89.4	69.5	41.7	19.0	6.3	1.5	.2	.0
13*	.0	.3	1.5	4.4	9.6	16.0	20.6	20.6	16.0	9.6	4.4	1.5	.3	.0	.0
14*	.0	.0	.2	.6	1.4	2.3	2.9	2.9	2.3	1.4	.6	.2	.0	.0	.0
15*	.0	.0	.0	.0	.1	.2	.2	.2	.2	.1	.0	.0	.0	.0	.0

Table 1. 15-tuple weight transition matrix (expected value)

X ** 15 + X ** 4 + 1

i → j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1*	0	11	4	0	0	0	0	0	0	0	0	0	0	0	0
2*	4	7	8	48	32	6	0	0	0	0	0	0	0	0	0
3*	4	3	47	70	74	127	95	31	4	0	0	0	0	0	0
4*	4	7	62	81	232	289	248	231	148	52	10	1	0	0	0
5*	3	16	58	128	349	463	611	577	393	242	121	37	5	0	0
6*	0	21	52	202	386	662	964	968	808	529	264	110	34	5	0
7*	0	21	42	229	412	844	1155	1254	1128	763	390	148	40	8	1
8*	0	13	48	206	426	854	1180	1345	1136	731	356	116	22	2	0
9*	0	5	53	145	416	738	1008	1092	848	478	179	39	4	0	0
10*	0	1	46	96	332	540	704	644	416	178	42	4	0	0	0
11*	0	0	26	72	210	317	353	243	112	29	3	0	0	0	0
12*	0	0	8	52	100	131	104	47	12	1	0	0	0	0	0
13*	0	0	1	27	30	31	13	3	0	0	0	0	0	0	0
14*	0	0	0	8	4	3	0	0	0	0	0	0	0	0	0
15*	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0

Table 2. 15-tuple weight transition matrix (empirical value)

X ** 15 + X ** 4 + 1

このTable 1，2から，たとえば weight が 3である p-tuple の直後の p-tuple の weight が4である場合の頻度は19回期待されるが，実際には70回起り，実測値は51だけ正の方に偏っていることが判る． このような偏り，($h_{ij}^{(EM)} - h_{ij}^{(EX)}$) をすべての i, j について示したものが，Table 3である． ここでは，実測分布の期待値からの偏りの3次元起伏を多少とも視覚的に見るために，偏りが10以上の（実測値が期待値を大きく超えている）点は太線の枠で囲み，偏りの絶対値が10未満である所は四角で囲むことにする．

ここから読取れる特長について述べる前に，引続いてpを実用的な大きさに拡張した場合の偏りについて示す．
この場合には，1周期 $(2^p - 1)$ の中からいくつかの部分列を無作為にとり出し，推移行列Hを作る．具体的には，この系列の生成のためには pビットの初期値が必要であるが，合同乗算型生成法による乱数からの0，1をこれに充て，そこから出発して 10^4 個の p-tuple を収集する．次に初期値からの従属を避けるために，この初期値を更新し再び 10^4 個の p-tuple を集める，この過程を100回繰返し，計 $10^4 \times 10^2$ 個の p-tuple の推移を $h_{ij}^{(EM)}$ として蓄積する．
このようにして，$p = 89$，$q = 38$ の場合について，期待値からの偏り $(h_{ij}^{(EM)} - h_{ij}^{(EX)})$ を濃淡図として表わしたものが

i → j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1*	-.0	11.0	3.8	-.6	-1.4	-2.3	-2.9	-2.9	-2.3	-1.4	-.6	-.2	-.0	-.0	-.0
2*	4.0	6.7	6.5	43.6	22.4	-10.0	-20.6	-20.6	-16.0	-9.6	-4.4	-1.5	-.3	-.0	-.0
3*	3.8	1.5	40.7	51.0	32.3	57.5	5.6	-58.4	-65.5	-41.7	-19.0	-6.3	-1.5	-.2	-.0
4*	3.4	2.6	43.0	24.1	106.9	80.5	-20.1	-37.1	-60.5	-73.1	-46.9	-18.0	-4.4	-.6	-.0
5*	1.6	6.4	16.3	2.9	73.8	4.3	21.3	-12.7	-65.7	-33.2	-4.1	-4.7	-4.6	-1.4	-.1
6*	-2.3	5.0	-17.5	-6.5	-72.7	-102.5	-18.9	-14.9	43.5	70.3	55.5	40.5	18.0	2.7	-.2
7*	-2.9	.4	-47.4	-39.1	-177.7	-138.9	-108.7	-9.7	145.1	173.3	121.9	58.6	19.4	5.1	.8
8*	-2.9	-7.6	-41.4	-62.1	-163.7	-128.9	-83.7	81.3	153.1	141.3	87.9	26.6	1.4	-.9	-.2
9*	-2.3	-11.0	-16.5	-63.5	-42.7	-26.5	25.1	109.1	83.5	19.3	-29.5	-30.5	-12.0	-2.3	-.2
10*	-1.4	-8.6	4.3	-29.1	56.8	81.3	114.3	54.3	-42.7	-97.2	-83.1	-37.7	-9.6	-1.4	-.1
11*	-.6	-4.4	7.0	15.1	84.9	108.5	84.9	-25.1	-96.5	-96.1	-53.9	-19.0	-4.4	-.6	-.0
12*	-.2	-1.5	1.7	33.0	58.3	61.5	14.6	-42.4	-57.5	-40.7	-19.0	-6.3	-1.5	-.2	-.0
13*	-.0	-.3	-.5	22.6	20.4	15.0	-7.6	-17.6	-16.0	-9.6	-4.4	-1.5	-.3	-.0	-.0
14*	-.0	-.0	-.2	7.4	2.6	.7	-2.9	-2.9	-2.3	-1.4	-.6	-.2	-.0	-.0	-.0
15*	-.0	-.0	-.0	1.0	-.1	-.2	-.2	-.2	-.2	-.1	-.0	-.0	-.0	-.0	-.0

Table 3. 15-tuple weight transition matrix (empiric.- expect.)

X ** 15 + X ** 4 + 1

Fig.3である．この濃淡図は，偏りの最大値から最小値までを１０等分し，それに１０段階の濃淡を割当てることにより各点の偏りを表示している．（図中の TEST BAR 参照）．第５階調である一様な灰色で周辺部が覆われているが，これは最も偏りの少ない平野部であり，それより高い所は濃く，逆に低い（実測値が期待値に満たない所）は白色に近くなる．縦軸はℓ，横軸はjであり，$23 \leqq \ell, j \leqq 65$ の範囲が示されているが，これはそれ以外の領域では $h_{ij}^{EM} = 0$ であったことを意味する．

さて，Table３の $p=15, q=4$ の場合の偏りのパターンとFig.3の $p=89, q=38$ のそれと比較してみると，その「島構造」は全く同じであることが判る．この共通した島構造から次のような性質を抽出すことができる．尚，[…] はその直前の記述を補うためのもので，Fig.3の場合の具体例を示す．

(i) $w(A_{n-p}) < p/2$，$[34 \leqq w(A_{n-p}) \leqq 40]$．であれば，$w(A_n)$ も同じ範囲の値をとることが多く，$p/2$ を超えることは少ない．従って，このような A_{n-p}, A_n の組によって非常に小さな weight をもつ $(2p)$-tuple が多く現われる．更に，この場合 A_{n+p} 以下もここに入り，loop を作る可能性もある．もしこの loop を作ると

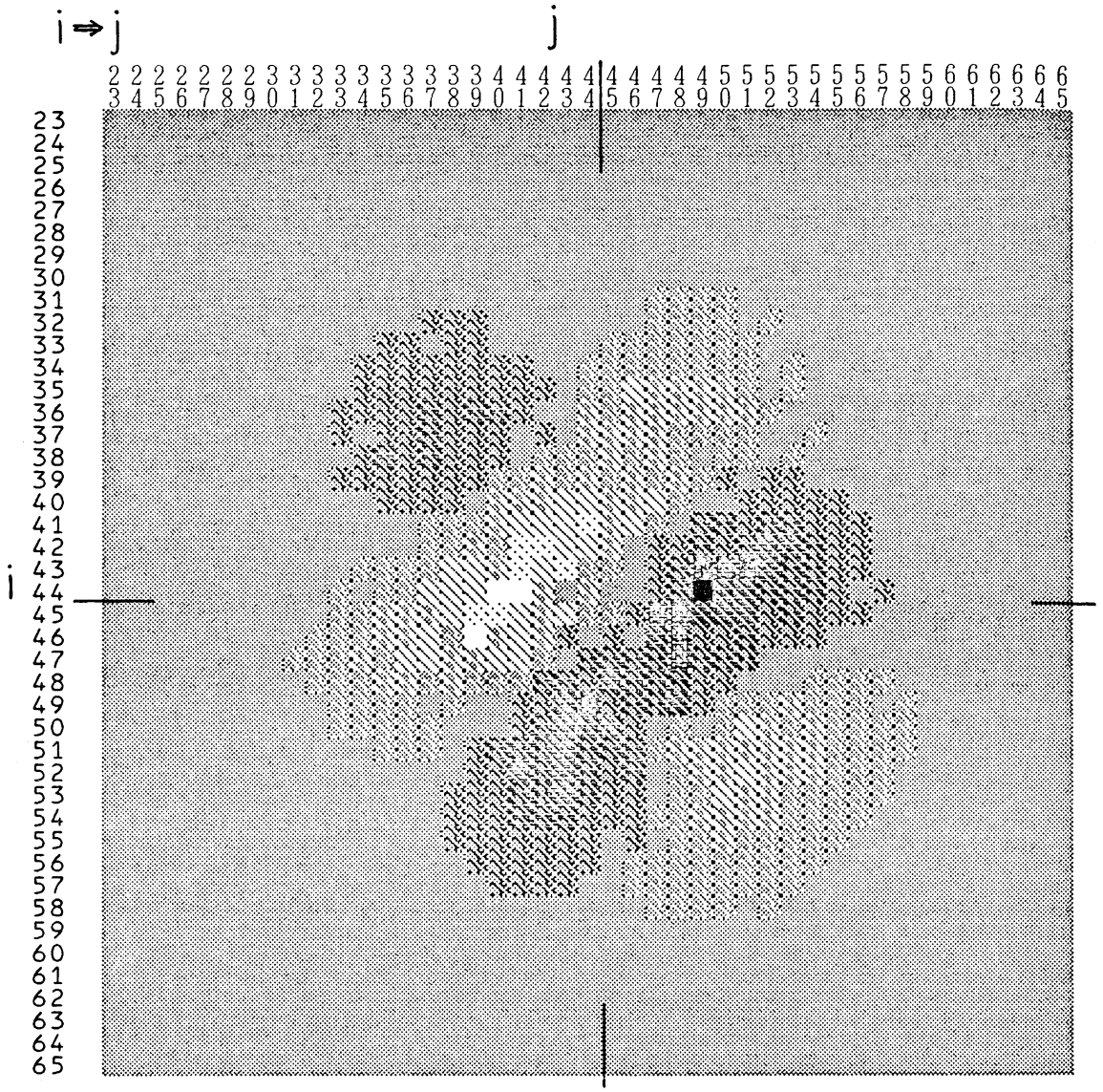


Figure 3. 89-tuple weight transition pattern (empiric.- expect.)

X ** 89 + X ** 38 + 1

TEST BAR.. 0 1 2 3 4 5 6 7 8 9 10

すれば，(3p)-tuple 以上で異常に軽い weight をもった部分列が出現するだろう．

(ii) $w(A_{n-p})$ が $P/2$ に等しいか僅か下まわる範囲［$40 \leq w(A_{n-p}) \leq 46$］では $w(A_n)$ が $P/2$ を超えることが多い．従ってこのような A_{n-p}, A_n の組による (2p)-tuple は p よりも少し大きい weight をもつことが多い．

(iii) $w(A_{n-p}) > P/2$［$47 \leq w(A_{n-p}) \leq 57$］であれば，$w(A_n) \lesssim P/2$［$38 \leq w(A_n) \leq 46$］となることが多くこれは相殺される方向であり，$w(A_n) > P/2$ となることは少ない．

(iv) このパターンは右下り対角線についてほぼ対称である．このことはM系列を逆向きに発生させても，(2)式についていえば相反の原始多項式によっても，偏りの傾向が不変であることを示す．従って，(4)式の T については $p > 29$ の形を考えれば充分である．

以上がこの偏りのパターンから読みとれる主な特長であって，前節までの結果とよく一致し，更により詳細な現象を伝えるものである．

この「島構造」は分布の形から予想されるように，p が大きくなるに従って不明瞭になる傾向をもつが，$p = 521; 607$ についてはその輪郭が多少滲み偏りがランダムになる程度で

保存される．実用上の限界と考えられる $p=1279$ については，更に不明瞭となるが，まだ構造は保存されていることが認められる．

また，この偏りの振幅の，期待値の最大値に対する割合，（$p=15, q=4$についていえば，Table 3の最大値と最小値の差：$173.3-(-177.7)=351$ のTable 1の最大値：1263.7に対する比：28%）は $89 \leq p \leq 1279$ では12～19%であり，特にpとqの値との関連は認め難い．むしろ，偏りの大きさも問題ではあるが，いつも同じ傾向に偏りをもつ，すなわち規則性をもつことが重大である．

§5 おわりに

3項原始多項式に従うM系列の部分列には0と1の偏りが見られることの実験結果を報告した．

GF(2)の上の加算演算は0と1を平均的に平等に出力するし，更に1周期については0の個数は1の個数より1だけ少ないだけで殆んど完全に均衡がとれている．こうして，いわば両極端では0と1の出現は対称であることが保証されているので，この実験報告，この両極端の中間では対称性が崩れている現象がある，は奇異に思われるが両極端の性質を否定するものではない．

その理論的解析については続報で報告したい．

尚，M系列のL-tupleについてLin.[68]その3次モーメントが0でなく，p,qの値によって大きくなりうることを既に示しており，本報告と密接に関連しているが，「M系列のL-tupleでは0が異常増殖したものがあり，イについてはそうはならない」という直接的・明確な指摘は筆者の知る限り見当らず，報告する次第である．TLP列を用いる場合には慎重な取扱いが必要と思われる．

文献

Lew.,Pay.(73) Lewis,T.G.and Payne,W.H. "Generalized feedback shift register pseudorandom number algolithm" J.ACM Vol.20(3),(1973),456-468

Lin.(68) Lindholm,J.H. "An analysis of the pseudo-randomness properties of subsequences of long m-sequences" IEEE Trans.Inform.Theory Vol IT-14 ,(1968),569-576