統計数理（1996）
第44巻 第1号 99–104

［原著論文］

M系列による滞在時間とハミング重みのシミュレーションからの予想

——GF(2)上のある種の原始多項式の倍数について——

大阪教育大学* 高 嶋 恵 三

（1995年3月 受付）

1. 序

擬似乱数はコンピューターの発達と共に広く利用されており，特に，線型合同法とM系列（maximum-length linearly recurring sequence, m-sequence）等が有名である．なかでも，M系列はその生成速度の高速性とk次均等分布性とにより，広く利用されるとともに理論面からも詳細な研究の対象とされてきた．M系列擬似乱数の理論は有限体 GF(2)とその拡大体の理論に帰着し，多くの理論的成果は一周期上での性質についてのものである．しかしながら，実際に利用されるM系列擬似乱数はその特性原始多項式の次数が大きい場合が普通であり，そのようなM系列の周期は事実上無限大に近く，一周期の一部分を利用するに過ぎず，理論的成果が適用しにくい．さらに，擬似乱数の一周期全体をとれば，それはランダムではなく，常に一周期の小部分を使用しなければならない．このため，周期全体の一部分に対して，その統計的ランダムネスを調べるための経験的検定は不可欠と言える．これは擬似乱数の周期全体の一部分に対する有効な理論が得られていない現状では避けられない問題である．

乱数に対する統計的検定としていろいろな検定方法が考案されているが，Takashima（1994）は一次元対称単純ランダムウォークの滞在時間の理論に基づく検定方法を提案し，原始3項式を特性多項式に持つM系列擬似乱数の検定を行ない，そのようなM系列擬似乱数は滞在時間検定において，極めて特徴的な統計的偏りを示すことを明らかにした．本報告では，符号理論における，ハミング重み（Hamming weight）と滞在時間（sojourn time）との関係を踏まえてM系列擬似乱数の滞在時間検定の検定結果について再考する．そして滞在時間検定において観測される原始3項式を特性多項式とするM系列の顕著な統計的偏りは特性原始3項式によるM系列だけでなく，伏見（1989）で提案された，3項式より派生する，多数項原始多項式によるM系列に共通する特徴である可能性が示唆されることについて，その種の多数項原始多項式の倍数に関する予想を交えて議論する．

2. ハミング重みと滞在時間の分布について

長さnの符号$\boldsymbol{x} = x_1\, x_2 \cdots x_n$，に対して$\boldsymbol{x}$のハミング重み$W_n(\boldsymbol{x})$は次で定義される：

$$W_n(\boldsymbol{x}) = \#\{k : x_k \neq 0,\ 1 \leq k \leq n\}.$$

特に，$x_k \in \mathrm{GF}(2)$の時は，$W_n(\boldsymbol{x}) = \sum_{k=1}^{n} x_k$である．

* 数理科学：〒582 大阪府柏原市旭ヶ丘4–698–1.

一方, 一次元ランダムウォーク $S_1, S_2, \ldots, S_n$ (単純対称ランダムウォークとは限らないものとする) の滞在時間 T_n の分布について, 次の結果が知られている (cf. Sparre Andersen (1953, 1954), Spitzer (1956, 1964)) :

定理. $\lim_{n\to\infty} P(S_n > 0) = \alpha$ のとき, ランダムウォークの時刻 n までの正の部分における滞在時間を $T_n = \sum_{k=1}^{n} I(S_k)$, $I(x) = 1$ $(x > 0), = 0$ $(x \leq 0)$, とするとき, T_n/n の極限分布は指数 α, $1-\alpha$ のベータ分布である.

特に, 単純ランダムウォーク S_n が, ビット列 $x_k (= 0, 1)$ から構成される場合, 即ち $S_n = \sum_{k=1}^{n}(2x_k - 1)$ の場合, $S_{2n} > 0 \Leftrightarrow W_{2n}(\boldsymbol{x}) > n$ であることに注意する. このことより, ベータ分布の指数 α はビット列のハミング重みが列の長さの1/2より大きい確率の極限値に等しいことがわかる.

3. M系列のハミング重みの理論分布について

Jordan and Wood (1973) は, M系列擬似乱数のハミング重みの分布が次のような係数で決定されることを示している.

x_k, $k \geq 1$, をM系列とし, $f(x)$ をその特性多項式で, $\deg f = r$ とする. 前節の記号で

$$P(W_m(\boldsymbol{x}) = k) = 2^{-m}\binom{m}{k}\frac{2^r}{2^r-1}\left(1 + \sum_{l=2}^{m} S_m^l \, F_m^k(l)\right),$$

ここで, $F_m^k(l)$ は2項係数によって決まる係数で f には無関係. また,

$$S_m^l = \sharp\{h : \deg h \leq m,\ l \text{ 項式},\ f \text{ は } h \text{ を割り切る}\}.$$

従って, ハミング重みの分布を決める量で f が関係するのは S_m^l だけである. また, Lindholm (1968) は $W_m(\boldsymbol{x})$ の3次のモーメントを S_m^3 によって具体的に与えている. なお, f が原始多項式であれば, $m < 2^r - 1$ であるかぎり, $S_m^2 = 0$ である.

4. 多数項原始多項式によるM系列擬似乱数の高速生成法について

伏見 (1989) は以下のように, 原始3項式から派生する多数項の原始多項式によるM系列の高速生成法を与えている :

p を奇数, $g(x) = x^p + x^q + 1$ を原始3項式とするとき, $G(x) = g(x^3)$ を因数分解して得られる p 次の原始多項式 f を特性多項式とするM系列 x_k, $k \geq 1$ は3項関係 $x_n = x_{n-3p} + x_{n-3q}$, $n \geq 3p$ を満たす. 従って, 初期値 $x_0, \ldots, x_{3p-1}$ を f によって定めておけば, f を特性多項式とするM系列擬似乱数 x_n はこの3項関係で高速に生成できる.

5. シミュレーションの結果について

原始3項式を特性多項式にもつM系列と前節で述べた, 伏見 (1989) による, 多数項原始多項式によるM系列に対する滞在時間検定等の結果について述べる.

Fig. 1, Fig. 2 は, M系列擬似乱数による滞在時間のシミュレーションのグラフである. 滑らかな曲線はベータ分布の密度関数を表し, 折れ線は滞在時間の経験分布を表している.ここで, ベータ分布の指数 α は

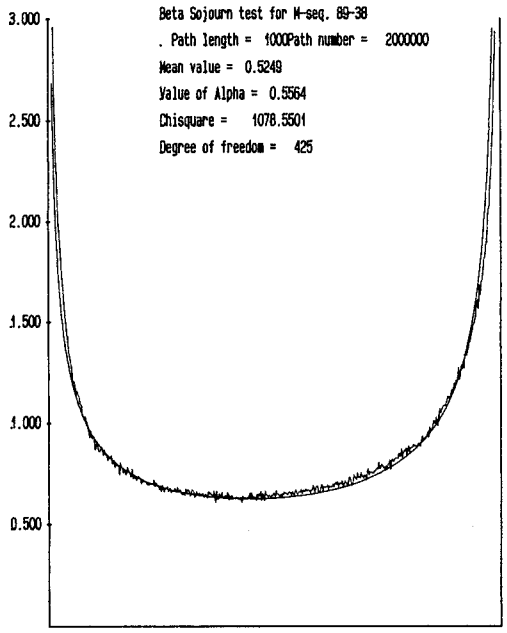
Beta Sojourn test for M-seq. 89-38
. Path length = 1000Path number = 2000000
Mean value = 0.5249
Value of Alpha = 0.5564
Chisquare = 1078.5501
Degree of freedom = 425
3.000
2.500
2.000
1.500
1.000
0.500

Fig. 1.

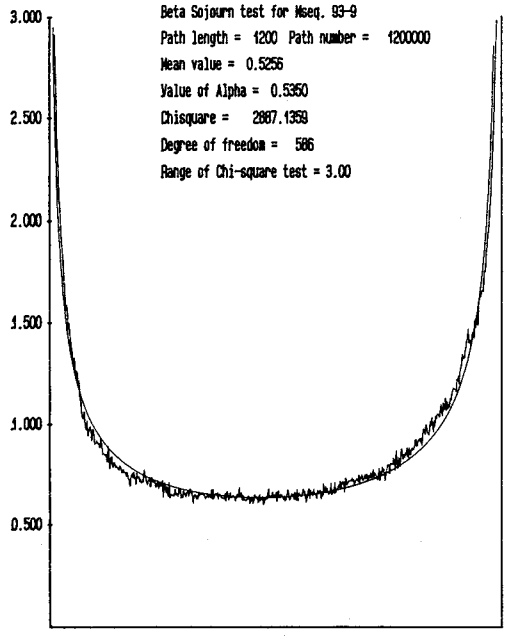
Beta Sojourn test for Mseq. 93-9
Path length = 1200 Path number = 1200000
Mean value = 0.5256
Value of Alpha = 0.5350
Chisquare = 2887.1359
Degree of freedom = 586
Range of Chi-square test = 3.00
3.000
2.500
2.000
1.500
1.000
0.500

Fig. 2.

$$\sum_k \left(\frac{N}{L}p(u_k)-f_k\right)^2 \Big/ \frac{N}{L}p(u_k)$$

が最小になるように選んだ．ただし，$u_k = k/L$，$2L$ はランダムウォークの見本関数の長さ，$p(x)$ はベータ分布の密度関数，f_k は事象 $\{T_{2L} = 2k\}$ の観測度数，N は見本関数の総数であり，和をとる k の範囲は $p(u_k) \leq 3.0$ を満たす範囲に限定している．これは，全ての k に対して2乗和をとると0と $2L$ の近傍での挙動が大きく影響して指数 α は1/2に極めて近い値を取り，経験分布の L での近傍の挙動をベータ分布の密度関数で十分に近似できないためである．また経験分布は0と $2L$ の近傍で $1/\sqrt{x}$ のオーダーで増大している．

Fig. 1 は原始3項式 $f(x) = x^{89}+x^{38}+1$，を特性多項式とするM系列擬似乱数による滞在時間の $2L = 1{,}000$，$N = 2{,}000{,}000$ によるシミュレーションの結果を示している．Fig. 1 では $\alpha = 0.5564$ であり，事象 $\{S_{2L} > 0\}$ の相対観測頻度は0.5249であった．

即ち，滞在時間の経験分布の L の近傍での挙動は，Sparre Andersen の定理より予想される $P(S_{2L} > 0)$ の値を指数にもつベータ分布より，より大きな指数をもつベータ分布に近い挙動を見せている．

原始3項式に基づくM系列がハミング重みに対して偏りを示すことについては，Lindholm (1968), Jordan and Wood (1973), 栗田 (1983) などを参照のこと．

Fig. 2 は原始3項式 $g(x) = x^{31}+x^3+1$，から伏見の方法で得られるM系列擬似乱数による，$2L = 1{,}200$，$N = 1{,}200{,}000$ の滞在時間のシミュレーションの結果である．このM系列の特性多項式は

$$x^{31}+x^{30}+x^{28}+x^{25}+x^{24}+x^{23}+x^{22}+x^{19}+x^{16}+x^{15}+x^{12}+x^{10}+x^7+x^6+x^5+x^4+x^3+x+1$$

である．Fig. 2 のグラフは，Fig. 1 の，原始3項式を特性多項式とするM系列と同様に，$k/L < 0.5$ で相対観測頻度が期待頻度を下回り，逆に $k/L > 0.5$ で上回るという特徴的な偏りを示している．なお，$\alpha = 0.535$，事象 $\{S_{2L} > 0\}$ の相対観測頻度は 0.5256 であった．経験分布に L の近傍でベータ分布を重ね合わせるとき，α の値は事象 $\{S_{2L} > 0\}$ の相対観測頻度より有意に大きくなる傾向が見られる．

6. 考察と予想

Fig. 1 のグラフの偏りとFig. 2 のグラフの偏りの類似性に着目して以下のような予想が考えられる．ただし，$g(x)$，$G(x)$，$f(x)$ は4節に準ずる：

$l \geq 3$ に対して，奇数 p が十分大きければ，あまり大きくない $m(\geq 3p)$ に対して，

$$\begin{aligned}&\#\{h : \deg h \leq m,\ h \text{ は } l \text{ 項式},\ f \text{は} h \text{を割り切る}\}\\ &\quad = \#\{h : \deg h \leq m,\ h \text{ は } l \text{ 項式},\ G \text{は} h \text{を割り切る}\}.\end{aligned}$$

ここで，m は任意に大きくすることは出来ないことに注意しておく．例えば，$l = 4$ の時，$m = 2^p$ では成り立たない（松本真氏による注意）．

$m = 6p$ の場合，Knuth (1981) の原始3項式の表 (p.29, 第1表) により，Wolfram 社の数式処理ソフト Mathematica を使用して確認した結果は以下のようである：

$l = 3$ に対して，$17 \leq p \leq 97$ の場合成立する．この場合，f を因数にもつ3項式 h は $G(x)$ と $G(x)^2$ のみである．

しかし，$p \leq 15$ の場合は成立しない原始3項式が存在する．例えば，$p \leq 11$ の場合にはすべての3項式に対して成立しないが，$p = 15$ の場合には（13次の原始3項式は存在しない），

$x^{15}+x^1+1$ に対しては成立するが，$x^{15}+x^4+1$，および $x^{15}+x^7+1$ に対しては成立しない．

$l=4$ の場合，$25 \leq p \leq 55$ に対しては成立する．$p \leq 23$ の時，成立しない原始3項式が存在する．例えば，$p \leq 11$ の場合にはすべての3項式に対して成立しない．$x^{15}+x^1+1$，$x^{17}+x^5+1$，$x^{23}+x^5+1$ に対しては成立するが，その他の 15次，17次，21次，23次の3項式に対しては成立しない．

$l=5$ の場合，計算時間が非常に長くなるため，$l=3$，4の場合のように多くの原始3項式について検証できていないが，$p \leq 29$ の時，成立しない原始3項式が存在する．例えば，17次の原始3項式のすべてと $x^{25}+x^3+1$ に対しては成立しない．一方，$x^{31}+x^3+1$ に対しては成立する．

従って，$g(x)=x^{31}+x^3+1$ に対しては，$l=3$，4，5の場合に上の予想は正しいことが確認できた．これは Fig. 2 のM系列の基になっている3項式である．

さらに，$g(x)=x^{521}+x^{32}+1$ に対して，$l=3$ の場合上記の予想が正しいことが確認された．これは，伏見（1989）で取り上げている原始3項式である．より大きな p の場合，より大きな l に対して，この予想が正しいことが期待されるので，このような大きな p の場合，生成されるM系列のハミング重みや滞在時間に関する統計的性質は3項式に基づくM系列に極めて近いことが予想される．また，M系列の特性多項式としては，なるべく次数の小さな3項式を割り切らない多項式を選ぶべきである，と考えられる．

謝　辞

本研究で扱った3項式の因数分解は，主にMathematicaで計算したが，一部は㈱ＮＴＴの植田広樹氏作成によるプログラムで確認した．植田氏に深く感謝する．九州大学の宗政昭弘氏は，$l=3$ に対するMathematicaによる計算結果を数式処理ソフトウェアMagmaを使用して追試して下さった．また，京都大学数理解析研究所の松本真氏は，著者の予想が m が大きすぎると成立しないことの理論的説明を教示して下さった．両氏に深く感謝する．なお本論文の一部は，統計数理研究所共同研究（4-共研-1，5-共研A-11，6-共研A-10）によるものである．また本論文を査読していただいた先生方の御助言により，本論文を改良することができました．先生方に感謝致します．

参　考　文　献

伏見正則 (1989).『乱数』, 東大出版会, 東京.

Jordan, H.F. and Wood, D.C.M. (1973). On the distribution of sums of succesive bits of shift-register sequences, *IEEE Trans. Comput.*, **C-22**, 400-408.

Knuth, D.E. (1981).『準数値算法／乱数』(渋谷政昭 訳), サイエンス社, 東京.

栗田良春 (1983). M系列の L-tuple の weight distribution の偏りについて, 数理解析研究所講究録, **498**, 153-171.

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

Sparre Andersen, E. (1953). On the fluctuations of sums of random variables, *Math. Scand.*, **1**, 265-285.

Sparre Andersen, E. (1954). On the fluctuations of sums of random variables II, *Math. Scand.*, **2**, 195-223.

Spitzer, F. (1956). A combinatorial lemma and its application to probability theory, *Trans. Amer. Math.* Soc., **82**, 323-339.

Spitzer, F. (1964). *Principles of Random Walk*, Springer, New York.

Takashima, K. (1994). Sojourn time test for maximum-length linearly recurring sequences with characteristic primitive trinomials, *J. Japanese Soc. Comput. Statist.*, **7**, 77-87.

A Conjecture from Simulations of Sojourn Times and Hamming Weights by m-sequences
—— On Multiples of a Certain Primitive Polynomials over GF(2) ——

Keizo Takashima
(Osaka Kyoiku University)

Lindholm (1968), Jordan and Wood (1973) discussed the distribution of Hamming weights of m-sequences and Kurita (1983) revealed statistical biases of distributions of Hamming weights of m-sequences with characteristic primitive trinomial. Takashima (1994) showed that m-sequences with characteristic primitive trinomial have evident statistical biases in sojourn time tests. Sparre Andersen (1953, 1954) proved that the limit distribution of sojourn times of 1-dimensional (not necessarily simple nor symmetric) random walks is a beta distribution with exponent α, $1-\alpha$ where α equals the limit probability that the random walk is positive.

In this paper, we discuss the relation between sojourn time tests and Hamming weights of m-sequences, and we conjecture the following:

Let $g(x)$ be primitive trinomial with sufficiently large odd degree p, $G(x) = g(x^3)$, and $f(x)$ a factor of $G(x)$ with degree p. The number of l-nomials $h(x)$ with degree $\leq m$, which can be divided by $f(x)$ is equal to the number of l-nomials $h(x)$ with degree $\leq m$, which can be divided by $G(x)$.

When p is small, or when m is too large, the above conjecture is not true. But the results obtained by using Mathematica show that the above conjecture holds for $l = 3$, $17 \leq p \leq 97$, and $m = 6p$.

Key words: Pseudorandom numbers, m-sequence, random walk, sojourn time, Hamming weight.