物理乱数の一様性の偏りと統計的ゆらぎに関する考察

岸本 俊祐 *

A Study for Biases and Statistical Fluctuations of Random Nnumbers Generated by Physical Means

Shunsuke KISHIMOTO*

About 1×10^9 random digits were generated by a simple electronic device. A statistical frequency test was carried out to check the uniformity of probability that each digit occured. The frequency tests were also applied to many parts of a series of the random digits to examine the local fluctuations of the uniformity.

If the physical device has defects, there are biases of the uniformity. On the other hand, even if a sequence of the random digits is an ideal one, it has a distribution of each digit biased by statistical fluctuations. This paper makes the relation between the former and the latter clear.

Key words: **Random number, Random number generation, Phisical random number**

1. はじめに

無規則な自然現象の“でたらめさ”を反映させて，ハードウェアで発生させる乱数，つまり，物理乱数は原理的に良質な乱数列をいくらでも発生できる可能性がある．そのため，発生方法は古くから研究され，一部は実用に供されている．

その発生原理は大きく分けて２通りあり，１つは無規則な自然現象が一定時間内に生起する数のでたらめさを利用する方法[1-5]，他の１つは無規則現象が生じる時間間隔のでたらめさを利用する方法である[6-8]．両方法とも一長一短があり決め手に欠けるが，前者の方がより簡単なハードウェアで発生が実現でき，また，乱数の発生確率に対する理論的見積もりも比較的容易に計算できるため，前者の方法で多くの試行がなされている．

筆者等も前者の方法で乱数を発生させる最も簡素な方法を試みてみた．その方法は，半導体ダイオードのショット雑音から得た頻度の高いランダムパルスの数をｎ進の高速電子カウンタでけたあふれさせながら一定時間計数し，計数値の最後の１けたをｎ進１けた整数乱数 $d:(d=0,1,\ldots,n-1)$，とするものである．この方法で発生できるそれぞれの乱数 d の発生確率を $P(d)$ とすると，カウンタに入力されるランダムパルス数の分布が正規分布か，もしくは，それに近い分布ならば，その分布の標準偏差を σ として，等確率 $1/n$ からのずれの最大相対誤差は

$$\epsilon_d = \left| nP(d) - 1 \right| \doteqdot 2 \cdot exp(-\frac{2\pi^2\sigma^2}{n^2}) \qquad (1)$$

と計算されている[9]．

実験条件を上手に設定して，ランダムパルス数分布の標準偏差 σ を 10 以上にすることは比較的容易である．そこで，ｎ $= 10$, $\sigma = 10$ として (1) 式を計算してみると，$\epsilon_d \doteqdot 5 \times 10^{-9}$ となり，このような簡素で原理的な発生方法でも理論的には十分一様性（等確率性）のよい 10 進整数乱数列が得られることがわかる．

しかし，このことが言えるのは，電子カウンタ等ハードウェアが設計通り理想的に動作した場合のことである．現実には，ディジタル回路素子のアナログ特性等，試行錯誤でしか対応できない要素もある．特に，電子カウンタを構成するフリップフロップ回路には，ノイズを含む入力信号に対して，０状態から１状態へ遷移する確率とその逆の方向へ遷移する確率が微妙に異なっているという不可避の欠点があり，それが，多くの乱数を発生したときに偶数乱数と奇数乱数の数の間に偏りを発生させることを前報で示した[10]．

一方，乱数列にはでたらめであるが故の部分的な“ゆらぎ”が本質的に存在し，十分多い乱数の中では各乱数出現の一様性は保たれるが，あまり多くない部分の中ではその出現度数に偏りが生じる．したがって，全個数があまり多くない乱数列の中では，ハードウェアや発生方法の欠陥による偏りが含まれていても，統計的なゆらぎによる偏りの中に埋もれてし

原稿受付　平成12年8月31日
*情報工学科

まって，陽に現れてこないということもありうる．そこで，ハードウェアの欠陥による一様性の偏りと統計的ゆらぎによる偏りが，どのような関係になっているか，ハードウェアによる偏りを持つ実際の乱数データを用いて調べてみた．

2. 乱数発生回路と実験条件

ハードウェアの構成を Fig.1 に示す．

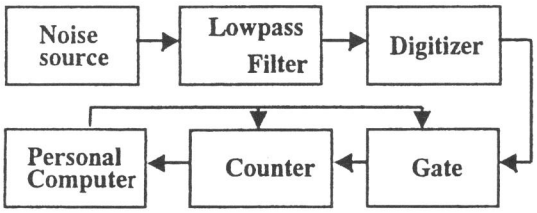

Fig. 1　Block diagram of a random digit generation.

・信号源

ノイズ源として，９Ｖのツェナー電圧を持つダイオードを用いた．このダイオードに適当な逆バイアス電流を流すことによって，10MHz 以上の高周波成分を含むショットノイズが得られる．そのノイズを含んだ電流を直接トランジスタのベースに流し込み増幅を行って，数 Vpp の電圧出力のノイズ信号を得た．

・ローパスフィルタ

ノイズを含んだ信号はディジタル回路に誤動作を生じさせやすい．今回の信号はノイズそのものなのでそのための配慮が必要である．その中で，特に極端に速く変化する高周波成分が問題となるので，遮断周波数が約 1MHz で減衰特性が -40dB/dec の２段バタワース型ローパスフィルタを通し，1MHz 以上の成分を取り除いた．また，信号源とフィルタを交流結合とし，ダイオードからの直流成分や極低周波成分とともに，1/f ノイズもその主要部分を取り除いた．

・ディジタイザ

これはゼロクロスコンパレータで，ローパスフィルタを通過したアナログのノイズ信号を適当な電圧のしきい値で２値化し，ランダムなディジタルパルス信号に変換する．

・ゲートとカウンタ，パソコン

カウンタは 10 進１けたの非同期型のものを用いた．ゲートやカウンタをパソコンで制御する．パソコンは内蔵のタイマーを使って 400 μ 秒ごとに自分自身に割り込みをかけ，ゲートを約 395 μ 秒間は開で，残りの約 5 μ 秒間が閉となるよう制御する．そして，その 5 μ 秒間でカウンタの計数値を読みとり，カウンタのリセット等を行う．読みとった計数値は順次ハードディスクに収録する．

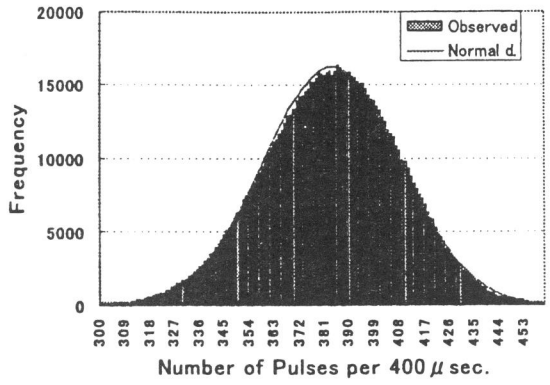

Fig. 2　The distribution of the random pulse number for 400 μ sec.

Fig.2 に実際カウンタに入力されるランダムパルス信号の数の分布を示す．これは，400 μ 秒間ずつ 10^6 回測定して集計したものである．この分布を正規分布と見なしたときの数の分布の平均値は $m = 383.6$，標準偏差は $\sigma = 24.5$ であった．ちなみに，このデータで (1) 式を計算してみると，$\epsilon_d \doteq 3.5 \times 10^{-52}$ となり，原理的には十分一様性のよい 10 進乱数が得られていることになる．なお，400 μ 秒はカウンタのけたあふれが十分起こりうるよう，試行錯誤で決めた値である。

以上の条件で 1×10^9 個の 10 進整数乱数を発生させ，パソコンのハードディスクに収録した．

3. 一様性の検定と統計的ゆらぎ

０から９までの 10 進１けたの数がでたらめに並んでいる整数型一様乱数の検定を考える．ハードウェアで発生した乱数の一様性を調べるには，その有効性が確立されている統計的「適合度検定（χ^2 検定）」がよく用いられる．それは，発生した乱数列の中で，０から９までの各乱数が一様に分布していると仮定して，実際の個数を集計した実測度数分布と，一様分布から期待される理論度数分布との適合度を χ^2 検定によって調べるものである．

話を一般化して，いま，ある分布を持つデータを k の階級に分けて集計したときの各階級の実測度数を n_i $(i = 1, 2, \ldots, k)$，それに対応する理論的期待

度数を f_i とする．計算すべき検定統計量は

$$\chi^2_{k-1} = \sum_{i=1}^{k} \frac{(n_i - f_i)^2}{f_i} \tag{2}$$

で与えられる [11]．各階級の実測度数と理論度数とのずれが大きいほど χ^2 値は大きくなる．χ^2 値は階級の数 k に依存するある確率分布（χ^2 分布）を持ち，その値が大きくなるほどその値が実現する確率は小さくなる（$k-1$ を自由度という）．有意水準とか危険率と呼ばれているある限界の値を設けておき，(2) 式から得られる χ^2 値がそれより大きくなれば，確率的にそのような事態は起こりにくいのに実際起こってしまっている → それは最初の仮定が間違っているからだ → 実測度数は理論度数に合っていない，と判断するのである．

さて，0から9までの各乱数がまったくでたらめに出現することによる度数の統計的ゆらぎはどの程度と見積もればよいであろうか．ある確率変数が平均値 m，標準偏差 σ の分布を持って出現するとする．確率論によれば，チェビシェフの不等式より，その分布がどのような任意の分布であっても，その確率変数の実現度数が $m \pm 3\sigma$ の外に出る確率は 1/9 より小であるから [12]，出現度数の期待値 m を中心に $\pm 3\sigma$ 程度が純粋な統計的ゆらぎの範囲と見積もることができる．従って，相対的なゆらぎの限界は

$$\delta \approx \left| \frac{\pm 3\sigma}{m} \right| = \frac{3\sigma}{m} \tag{3}$$

となる．いま，確率変数の出現度数分布が平均 m のポアソン分布のばあいは，$\sigma = \sqrt{m}$ であるから

$$\sqrt{m} \cdot \delta \approx 3 \tag{4}$$

となる．

4. 検定結果と考察

4.1 一様性の検定

今回新たに発生した 1×10^9 個の 10 進 1 けた乱数について，まず一続きの 10^4 個に対して，0から9までの各乱数の出現個数を集計した．その結果を Fig.3 に示す．各乱数は，全体の 1/10 ずつ，つまり，1000 個ずつ一様に分布するという仮定のもとで，(2) 式より χ^2 値を求めた．結果は χ^2=8.85であった．階級の数は 10 なので，自由度は 9 となる．自由度 9 の χ^2 分布における危険率 5 %の棄却域は $\chi^2_9 > 16.9$ なので，実測値はその領域に入っていない．したがって，仮定は破棄されず，統計的見地からは各乱数は等確率で一様に出現しているとみなしても差し支えない．各実現度数のでこぼこの程度も，系統的な偏りはなく，分布の平均を m，標準偏差を σ としたとき，$m \pm \sigma$ の中に実現度数のほぼ 7 割が入るという統計上の常識にも一致している．

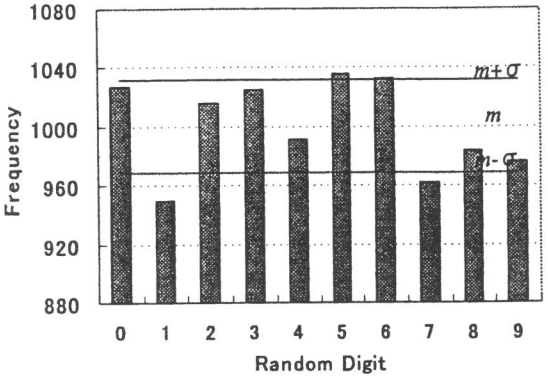

Fig. 3　Frequency histogram of the each digit. The grand total of the digits is 10^4.

同様に，乱数の全集計個数を 10^5，10^6 と順次増やしていき，10^9 まで χ^2 値をそれぞれ求めた．その結果を Fig.4 に示す．比較のため，この図にはハードウェアによる偏りのない（No biased）乱数によるデータも示してある．これは，前報 [10] で報告した改良乱数列を同様に集計したものである．全集計個数が 10^7 個までの χ^2 値は危険率 5 %の棄却域に入らず，度数分布も 10^4 の場合とほぼ同様な分布が得られた．しかし，10^8 個と 10^9 個の場合は χ^2 値は棄却域に入っており，10^9 個の場合は極端に大きくなる．この場合の度数分布を示すと，Fig.5 のようになっている．この図から，偶数乱数の度数の方が奇数乱数の度数よりも常に大きいという系統的な偏りが見られ，それが原因で χ^2 値が大きくなったものと思われる．この偏りが正にフリップフロップの動作の偏りによるものである．

Fig.3 や Fig.4 からわかるように，たとえ偶数乱数

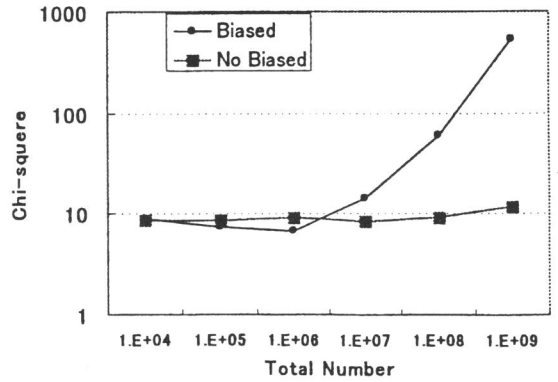

Fig. 4　The change of χ^2-values versus the grand total of the digits.

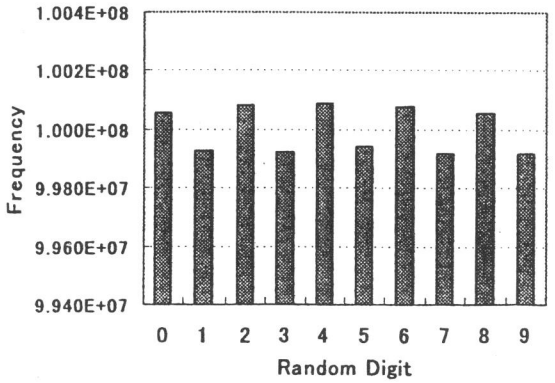

Fig. 5　Frequency histogram of the each digit. The grand total of the digits is 10^9.

と奇数乱数の数にFig.5のような系統的な偏りがあっても，乱数の数が 10^7 個程度までの一部を利用するのであれば，十分ランダムな乱数列として使用できることがわかる．ちなみに，今回発生した全乱数 10^9 個を 10^4 個ずつ 10^5 ブロックに分けて，各ブロックごとに χ^2 検定し，10^5 個の χ^2 値の集計度数分布を求めてみた．その結果をFig.6に示す．この図の実線は自由度9の χ^2 分布による理論的期待値で，各乱数の出現度数のゆらぎが純粋に統計的なものと仮定したときの分布である．一見して，実測度数分布は理論度数分布によく合っていることがわかる．念のため，

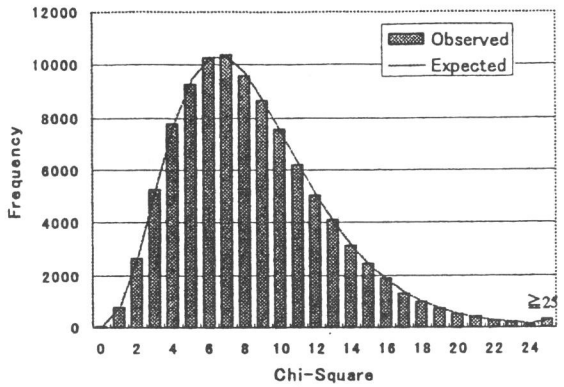

Fig. 6　Frequency distribution of χ^2-values for 1×10^5 blocks.

この実測度数分布と理論度数分布との適合度を自由度25の χ^2 検定で調べてみると，$\chi^2 = 25.3$ となった．自由度25の危険率5％の棄却域は $\chi^2_{25} > 37.6$ なので，検定上からも両分布はよく適合していることが確かめられた．なお，10^5 個の χ^2 値の内，自由度9の χ^2 分布の棄却域に入ったブロックの数は4936であった。これは，全体の4.9％にあたる。このことはブロックの大きさが 10^7 程度まで同様に成り立つ．したがって，乱数全体の中から 10^7 個程度を利用するのであれば，全体の乱数列のどの部分を切り出して使用してもまったく問題がないことがわかる．

4.2　統計的ゆらぎと偏り

0から9までの乱数の出現確率はどの乱数も同じで1/10であるが，実際の出現度数の分布はポアソン分布で近似できることが実験的に確かめられる．その証拠をFig.7に示す．いま，得られた物理乱数列

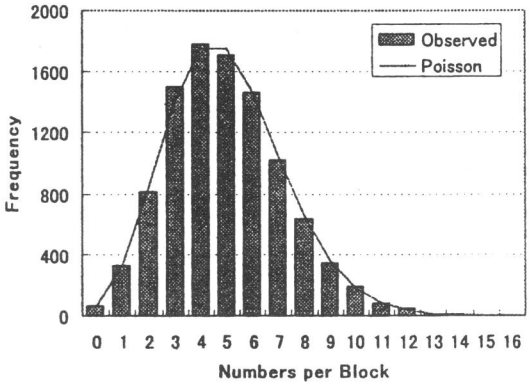

Fig. 7　Frequency distribution of the digit '5' in 50 random digits.

を50個ずつのブロックに区切っていき，その中で1つの乱数，たとえば“5”の個数がいくつ出現したかを集計する．出現確率からは，各ブロックの中に平均5個となる．10^4 ブロック分についての集計結果がFig.7である．この分布は平均が5のポアソン分布によく合っていて，今回の乱数データには(4)式を適用してもよいことがわかる．

全体の集計個数を，10^3，10^4，$\cdots$，10^8 として，それぞれについて(4)式の値，つまり，相対的な最大ズレの値を求めた．その結果をFig.8に示す．計算には各乱数の実度数の内，平均値からのずれが最も大きいものを最大ずれとして用いた．

この図からわかるように，純粋に統計的ゆらぎによる出現度数の偏り（相対偏差）は，ほぼ2のあたりを推移している．前節で，各乱数の平均出現度数からのずれの最大値を $\pm 3\sigma$ 程度と見積もったが，実際は $\pm 2\sigma$ 程度に収まっていることがわかる．各乱数の個数 m が 10^7 と 10^8 の場合は，明らかにその偏りからズレていて，統計的なゆらぎによる偏りとは別の偏りの存在を示唆している．これは，Fig.4の結果とよく一致している．

m が比較的小さいときは，乱数がランダムであるが故の統計的ゆらぎのため，実現度数の平均 m からのずれは相対的に大きなものとなる．そのため，ハードウェアの不具合による偏りはその中に埋没してしまって，表面に出てこないものと思われる．統計的

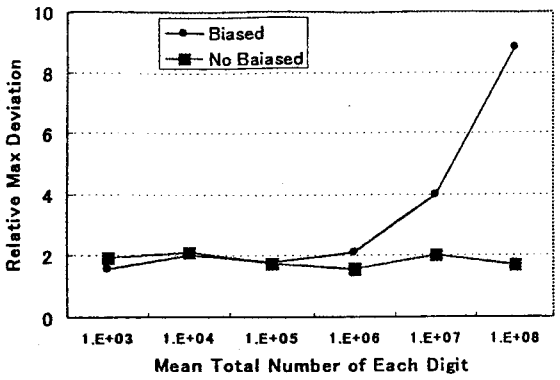

Fig. 8　The change of the relative maximum deviation : $\sqrt{m} \cdot \delta$ versus the mean total number of each digit : m.

ゆらぎによるずれは，正負がほぼ等しい確率で生じるので，m が大きくなってもずれの累積は正負が適当にキャンセルして，あまり大きくなることはない．統計論で言えば，その累積は $\sigma = \sqrt{m}$ に比例して大きくなる程度である．しかし，ハードウェアによる偏りは統計的ゆらぎによる偏りにくらべて絶対値が小さくても一定なので，その累積は $\sigma^2 = m$ に比例して大きくなる．その結果，Fig.8 のような結果になったものと思われる．

5.　おわりに

ハードウェアで発生する乱数は，ハードウェアが設計通り理想的に動作すれば，良質な乱数列をいくらでも発生できる反面，ハードウェアの特性がわずかでも歪んでいると，多量の乱数を発生したとき，発生数のでたらめさに系統的な偏りを生じさせることを示した．その典型例が，フリップフロップ回路の状態遷移確率が，ノイズを含んだ入力信号に対して，0状態から1状態への遷移とその逆方向への遷移ではわずかに異なっていることが，発生した乱数の偶数と奇数の間に数の偏りをもたらすという事実である．

一方，乱数には統計的ゆらぎによる発生数の偏りも存在し，この偏りがないものは乱数とは言えない．そこで，本来，あってはならないハードウェア原因の偏りと，当然なくてはならない統計的ゆらぎによる偏りが混在するとき，それが乱数列の出現度数の一様性にどう影響してくるかを，ハードウェアによる偏りを持つ実際の乱数列を用いて，実験的に明らかにした．

今日では，コンピュータによるシミュレーションや数値実験等，乱数が数多く使用されている．それらの成果を発表するとき，結果とともに実験条件として，使用した乱数の特性についても言及しておかねばならない．そのような場合，乱数の特性を考察するとき，この論文の内容が参考になれば幸いである．

参考文献

1) M.Ishida et al.: Random Number Generator, Ann. Inst. Stat. Math. **8**, (1956) 119-126

2) Y. A. Shreider: Monte Calro Method, Pergamon Press, (1966) 289

3) 石田正次，他: ダイオードノイズを利用した乱数発生装置，日立評論, **54** (1972) 894-898

4) 仁木直人: パーソナル・コンピュータのための物理乱数発生器，統計数理研究所彙報，**31**，1 (1983) 33-49

5) 岸本俊祐，他: ダイオードノイズを利用した物理乱数発生とその評価，信学会論文誌，Vol.J82-A, 11 (1999) 1704-1709

6) O. Miyatake et al.: Generation of Physical random numbers, Math. Jap. **20**, (1975) 207-217

7) S. Kishimoto et al.: A Generation method of physical random number sequence by using microcomputer, J. Japan Statist. Soc., **11**, 2 (1981) 169-173

8) S. Kishimoto: A simple generation method of physical random digits, J. Japanese Soc. Comp. Statist., **6** (1993) 67-73

9) 仁木直人: パーソナル・コンピュータのための物理乱数発生器，統計数理研究所彙報，**31**，1 (1983) 33-49

10) 岸本俊祐，他: ダイオードノイズを利用した乱数発生の問題点と改良，津山高専紀要，41 (1999) 91-96

11) 東京大学教養学部統計学教室:自然科学の統計学，東京大学出版会 (1992) 145-176.

12) 薩摩順吉: 確率と統計, 岩波書店，(1989) 46-47