擬似乱数生成系の検定方法に関する調査報告書

－Lempel-Ziv 圧縮検定について－

２００４年１月

東京理科大学

金子 敏信

疑似乱数生成系の検定方法に関する調査報告書

－Lempel-Ziv 圧縮検定について－

東京理科大学理工学部電気電子情報工学科

金子敏信

平成１６年１月

疑似乱数生成系の検定方法に関する調査報告書

－Lempel-Ziv 圧縮検定について－

１．はじめに

統計的手法により、乱数列の性質を解析する方法の一つとして、ＮＩＳＴのＳＰ８００－２２がある。発表以来、日も浅く、特に、次の２種の検定に関し疑問点が学会等で示されている。

・ 離散フーリエ変換
・ Lempel-Ziv 圧縮検定

本調査研究では、Lempel-Ziv 圧縮検定を中心にその問題点を調べ、ＣＲＹＰＴＲＥＣ疑似乱数生成系評価ＷＧ用資料としてまとめた。

２．ＮＩＳＴ　ＳＰ８００－２２による乱数生成器の検定

Special Publication 800-22（以後、SP800-22 と略）[1][2]は、米国商務省標準局（ＮＩＳＴ）が公開している暗号アプリケーションの為の乱数と擬似乱数の統計試験ツール及びドキュメントである。SP800-22 では、16 種類 189 個の検定法が採用され、全て２元乱数系列を検定対象としている。評価結果は、乱数列の検定毎の p-value で与えられ、その一様性と比率の観点から良い乱数生成器としての合格基準が示されている。ここで、p-value とは、真の乱数生成器が検定対象の乱数系列よりも乱数らしからぬ系列を生成する確率である。検定は、次のように行われる。

[乱数生成器検定法]

Step1(乱数列の検定)：複数の乱数系列（1000 本程度を推奨）を用意し、各系列に対し、真の乱数生成器がそれよりも乱数らしからぬ系列を発生する確率を評価し、p-value とする。推奨例では、1%棄却即ち、p-value≧0.01 の乱数を“良い”乱数と判断する。

Step2(乱数生成器の検定)：各乱数系列の p-value を標本データとして、その一様性及び棄却されたデータの比率を評価し、乱数生成器の合否を判断する。

Step1 の検定における、p-value への変換は、検定法毎に異なる。真の乱数をその検定法に通したとき、与えられた検定対象よりも乱数らしからぬ振る舞いをする確率を評価し、それを p-value とする。

Step2 は、各検定に対し同一であり次の判断基準が適用される。

p-value の一様性では、[0,1]の区間を 10 等分し、区間毎の p-value の頻度分布を χ^2 検定する。真の乱数を生成する乱数生成器であれば、この区間内に p-value の値が一様に分布すると考え、対象系列数を m、各区間の頻度を Fi として、検定量

$$\chi^2 = \sum_{i=1}^{10} \frac{(F_i - m/10)^2}{m/10}$$

を評価する。この検定量に対する p-value が 0.0001 以上ならば一様性が合格。

棄却率を P'とすれば、真の乱数列においてもこの確率で棄却される。乱数列の棄却本数を二項分布と考え、検定系列に対し棄却された系列の比率が

$$P' \pm 3\sqrt{\frac{P'(1-P')}{m}}$$

の範囲に入っていれば、乱数生成器は、p-value 比率で合格。

[検定法の問題点の指摘]

SP800-22 は、提案以来、何回かの改訂をへている。ここでは、ドキュメントに関し、2001 年 5 月 15 日版[1] について、これまでに指摘されている問題点を述べる。

CRYPTREC プロジェクトの報告書[3]では、

・合否判定に用いる p-value の閾値の定め方の合理的理由が述べられていない。

・離散フーリエ変換検定（Discrete Fourier Transform Test）と Lempel-Ziv 圧縮検定(Lempel-Ziv Compression Test)に関し、予想以上の不合格が発生すること

の指摘がある。また、文献[4]でも、離散フーリエ変換検定と Lempel-Ziv 圧縮検定の不合格の問題及びドキュメントとソースプログラムの内容の不一致に関し、指摘されている。

これら問題に対する、対応として、文献[5]では,離散フーリエ変換検定に関し、フーリエ変換された値を二値系列に変換する時の閾値が SP800-22 では、理論的に誤っていること。また、p-value に変換するとき、仮定する分散が、実験的に異なっていることが指摘し、改定法が示されている。また、文献[6]でも離散フーリエ変換に関し、同様の指摘と、Lempel-Ziv 圧縮検定に関し、部分列数の平均値及び分散について検討がなされ、実験値として、それらの値を定めると共に、検定の後半部分（Step2（乱数生成器の検定)）の一様性検定を Lempel-Ziv 圧縮検定用に、実験値に基づき変更する提案がなされている。

本報告では、SP 800-22 における Lempel-Ziv 圧縮検定を調査し、検定の前半部分(Step1.(乱数列の検定))に関し、検討を加えた。なお、Step2 に関する検討は、Lempel-Ziv 圧縮検定のみに関し、実験値に基づきこの部分を変更する妥当性がない為行っていない。

３．Lempel-Ziv 圧縮検定法

Lempel-Ziv 圧縮法は、与えられた情報系列を増分分解法により部分列に分解し、その部分列を符号化する情報源符号化法である[7]。その平均符号長Ｌは、漸近的に情報源符号化の限界であるエントロピーに一致する。情報元系列の長さ n が十分長ければ、Lempel-Ziv 圧縮後の系列長 N は、$H(S)$を情報源エントロピーとして、

$$N = nH(S) \qquad (1)$$

に漸近することが証明されている。この時、部分列の総数は、

$$W_{\infty} = \frac{nH(S)}{\log_2 n} \qquad (2)$$

となる。理想的にランダムな２元情報源として、シンボルが等確率で生起する無記憶情報源を考える。エントロピーは、最大となり $H(S)=1$である。乱数列を Lempel-Ziv 圧縮した時、十分大きな n に対し符号化後の系列長 $N=n$ ならば、よい乱数であり。$N<n$ ならば圧縮可能であり、何らかの偏り（シンボル生起確率の偏り又は記憶）のある乱数列と見なす事ができる[8]。この理想的にランダムな情報源に比べ、検定対象の部分列総数が少ないことは、圧縮可能であることを表す、というのが、Lempel-Ziv 圧縮検定の基本思想である。

増分分解法では、系列の先頭から眺め、未出現の部分列が出た場合、新しい部分列として区切って行く。二元系列 011011100・・・であれば、0 | 1 | 10 | 11 | 100 |・・・と区切ることになる。長さ n の理想的にランダムな二元乱数列を、部分列に区切ったときの部分列の総数を$W(n)$とするならば、その平均$E[W(n)]$は

$$\lim_{n\to\infty} E[W(n)] = \frac{n}{\log_2 n} \qquad (3)$$

となる。また、$W(n)$の分散を$\sigma[W(n)]$とすれば、正規化確率変数ｚは、十分大きな n に対し

$$z = \frac{W(n) - E[W(n)]}{\sigma[W(n)]} \Rightarrow N(0,1) \qquad (4)$$

と、標準正規分布に従う事が示されている[9]。分散については、文献[10]より

$$\sigma^2[W(n)] \approx \frac{n\{C + \delta(\log_2 n)\}}{\log_2^3 n} \qquad (5)$$

が知られている。ここで、$C = 0.26600\cdots$、$\delta(\cdot)$は、$|\delta(\cdot)| < 10^{-6}$なる微少項である。

乱数らしからぬ系列は、部分列の総数が少なくなり、圧縮可能な系列であり、この標準正規分布の片側に着目して、その確率を求めれば p-value となる。しかし、NIST は、$n=10^6$程度では、（３）式、（５）式の近似精度が悪いとして、SHA-1 を使った乱数生成器及び Blum-Blum-Shub 乱数生成器の発生乱数を元に、実験的に、$n=10^6=100$ 万ビットに対し

平均値　$\mu = 69588.20190000$

分散　　$\sigma^2 = 73.23726011$　　（６）

として、検定プログラムを供給し、$n = 10^6$での使用を推奨している。なお、プログラムソースでは、n=10 万、20 万、40 万、60 万、80 万ビットに対しても実験的に平均値及び分散を定めている。

検定系列に対し、求められ部分列数をW_{obs}とするならば、p-value は

$$p - value = \frac{1}{2} erfc(\frac{\mu - W_{obs}}{\sqrt{2\sigma^2}})$$　　（７）

と計算される。ここで、（補）誤差関数であり

$$erfc(z) = \frac{2}{\sqrt{\pi}} \int_z^{\infty} e^{-u^2} du$$　　（８）

である。

４．部分列数の平均値と分散の理論的解析

Lempel-Ziv 圧縮アルゴリズムに関し、NIST のドキュメントでは、述べられていない理論的解析として、文献[11]がある。長さｎの二元系列を増分分解したときの部分列数$W(n)$の k 次モーメント$E[W(n)^k]$は、$k \geq 1$に対し、次式で与えられる。

$$E[W(n)^n] = x(n)^k \left(1 + O\left(\sqrt{\frac{\log n}{n}}\right)\right) + O\left(\frac{n^{k-1}}{\log^{k-1} n}\right)$$　　（９）

ここで、$x(n)$は次式である。

$$x(n) = \frac{nH(S)}{\log_2 n}\left(1 + \frac{\log\log n}{\log n} + \frac{A - \log H(S)}{\log n} + O\left(\frac{(\log\log n)^2}{\log^2 n}\right)\right)$$　　（１０）

ここで、$A \approx 1.5297$、$H(S)$は、エントロピーである。

これらの式に、$H(S)$=1 を代入し、部分列数の平均値を求めれば

$$E[W(n)] = \frac{n}{\log n}\left(1 + \frac{\log\log n}{\log n} + \frac{A}{\log n}\right) + O\left(\frac{n(\log\log n)^2}{\log^3 n}\right)$$　　（１１）

となる。分散$\sigma^2(W(n))$も、（９）、（１０）式から求められそうであるが、計算するとオーダー項となってしまい、現在までに、理論解析として判っているのは、（１１）式の平均と（５）式の分散である。これらに、n=100 万を代入するならば、オーダー項を絶対誤差項として無視して

$E[W(10^6)] = 64888.912$　　（１２）

$\sigma^2[W(10^6)] = 33.59365$

となる。なお、参考の為、（１１）式の絶対誤差項を n=100 万に対し評価すれば

$$\left[\frac{n(\log\log n)^2}{\log^3 n}\right]_{n=10^6} = 2353.6198 \qquad （１４）$$

であり、必ずしも小さくない。また、NIST のドキュメントによる式で、同じく計算すれば

$$E[W(n)]_{n=10^6} = \left[\frac{n}{\log n}\right]_{n=10^6} = 50171.66594 \qquad （１５）$$

である。

５．部分列数の平均値及び分散の実験的検証

５．１　部分列総数の平均値

部分列数の平均値 $E[W(n)]$ の理論値として、NIST のドキュメントによる（３）式及び、Louchard らの結果から得られる（１１）式がある。有限の n に対し、これらの式の妥当性を実験的に調査した。実験対象としては、NIST の検定プログラムに付属の SHA-1 を用いた擬似乱数生成器（G Using SHA-1）及び XOR を用いた。乱数列 800 万ビットまでの結果を図１に示す。これは、各ビット長の擬似乱数系列 250 本の平均である。NIST のドキュメントの式に比べ、Louchard らの近似式がより実験結果に近い値を示しているように見える。しかし、その誤差の値は大きい。系列長に対し、誤差を示せば、図２である。

系列長 n=100 万ビットにおいて、SHA-1 のデータで比較すれば

理論値　　$E[W(10^6)] = 64888.912$

実験値(SHA-1)　$\mu = 69586.640$

であり、誤差＝4697.725751 となっており（１４）式で評価した絶対誤差項が無視できない係数を持っていると考えられる。この誤差を系列長に対し、線形近似したものを、グラフ中に示しておいた。誤差項は、系列長 100 万から 800 万ビットに対し、ほぼ線形に変化している。

理論式の（５）式の平均値及び分散が正しいと仮定するならば、この SHA-1 の実験値は、系列長 $n=10^6$ において、$\mu = E[W(10^6)] + 810\sigma[W(n)]$ であり、SHA-1 による乱数生成器が理想乱数から極端に離れた系列を出力している事になる。他の乱数生成器についても、この理論平均と実験値を比較する為、NIST の検定プログラムに付随する 11 種類の乱数生成器について、100 万ビットの乱数系列 250 本を発生させ、そのときの部分列数 W(n)の最大値と最小値を求めた。結果を図３に示す。

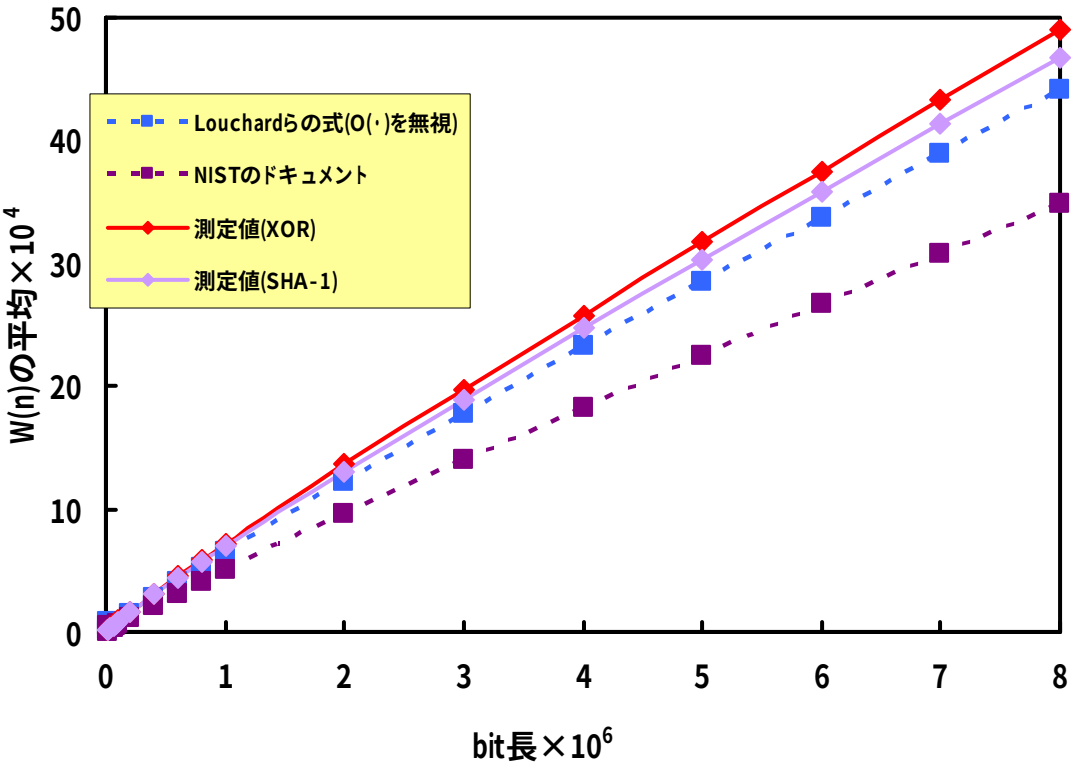

図１．W(n)の平均値と系列長(理論値と測定値)

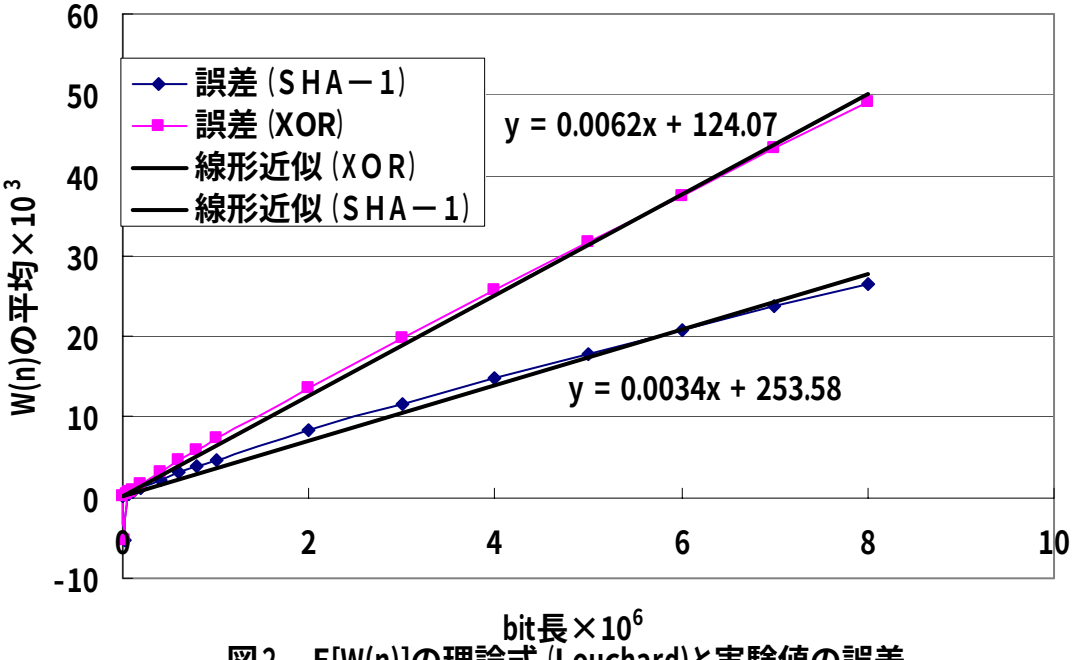

図２．E[W(n)]の理論式（Louchard)と実験値の誤差

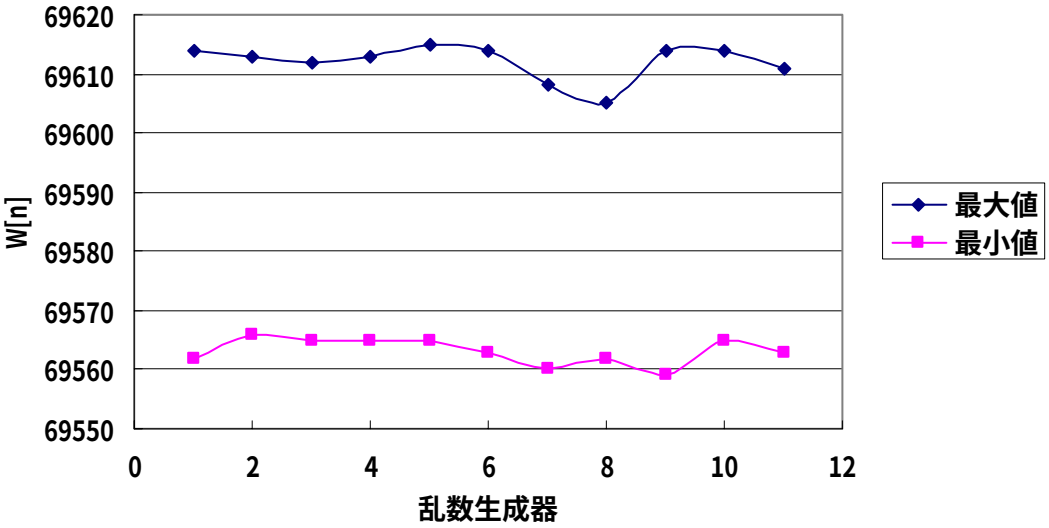

図３．W(n)の最大値と最小値

乱数生成器

1: G-U SHA-1　　2: Linear Congruential,　　3: Blum-Blum-Shuf
4: Micalli Schnor　　5: Modular Exponentiation　6: Quadratic Congruential I
7: Quadratic CongruentialⅡ　8: Cubic Congruential　9: XOR
10: ANSI X9.17　　11: G-U DES

各乱数生成器の発生した系列に対し、部分列数は 69550 から 69620 の範囲に収まっており、理論値の $E[W(10^6)] = 64888.912$ が、いかにかけ離れた値であるかが判る。

理論式の分散を NIST による実験値で置き換えたとしても、同様であり、Louchard の E[W(n)]の理論式は、本検定法の理論的根拠として採用するには、系列長 n=100 万 bit 近辺において誤差が大きすぎると判断できる。

５．２　部分列総数の分散

部分列数の分散 $\sigma^2[W(n)]$ の理論値として、(５)式が知られている。有限の n に対し、これらの式の妥当性を実験的に調査した。実験対象としては、NIST の検定プログラムに付属の SHA-1 を用いた擬似乱数生成器（G Using SHA-1）及び XOR を用いた。乱数列 800 万ビットまでの結果を図４に示す。これは、各ビット長の擬似乱数系列 250 本の平均である。分散について、式（５）の理論値と実験値では、約２倍程度の開きが見られる。このグラフにおいて、系列長 200 万ビット程度までは、滑らかな曲線であるが、それ以上に関し、不規則な動きが見られる。その理由として、系列数 250 本の妥当性も考え、G Using SHA-1 について、初期値を変えた系列 250 本を 10 通り採取

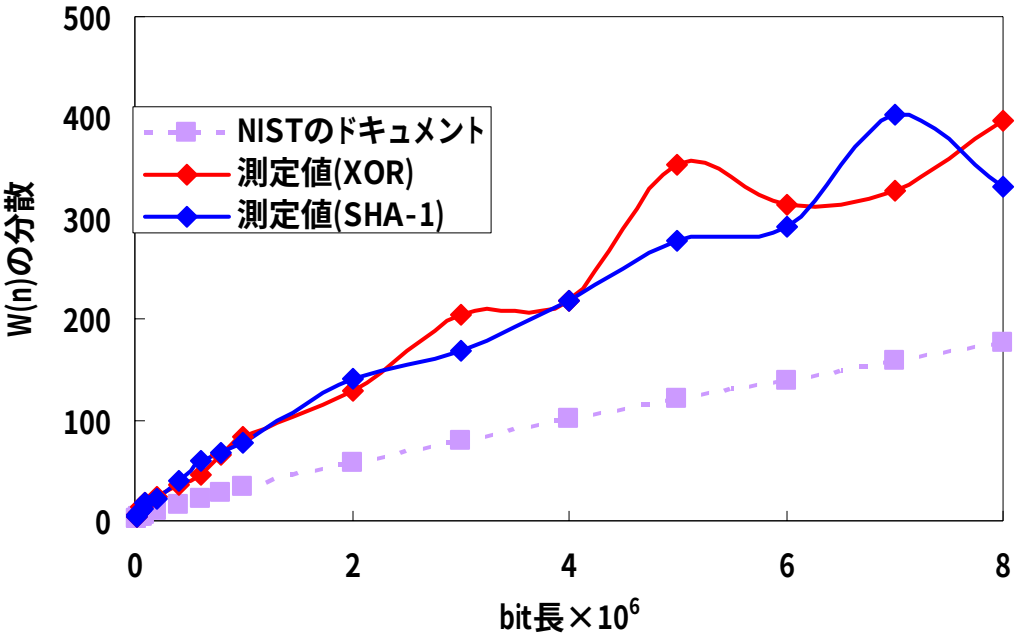

図４．　部分列数 W(n)の分散

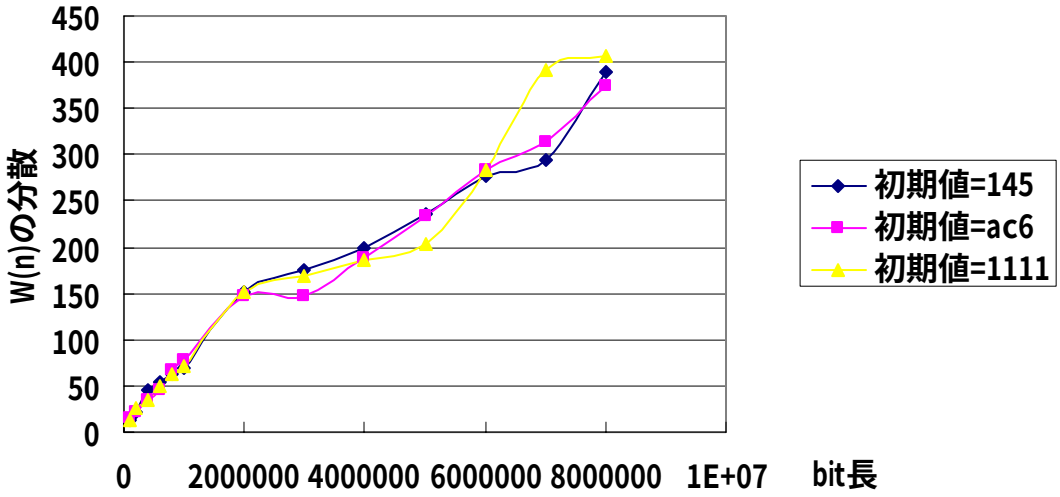

図５．　SHA-1で初期値をいろいろ変えて分散を測定

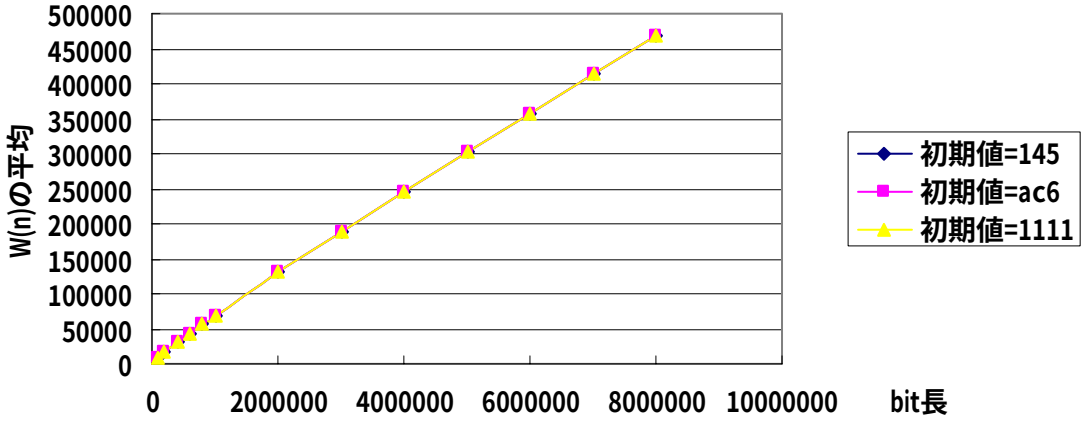

図6． SHA-1で初期値をいろいろ変えて平均を測定

し、W(n)の各の平均と分散を調べた。その一部を、図５，６に示す。平均値については、系列長に依存した一定の値になっているが、分散については、系列長 200 万ビット以上で、初期値にも依存する傾向が見られる。シンボルの生起確率が 1/2 の場合の特有の不安定現象なのか又は違う原因なのか、これからの検討課題である。

以上のように、W(n)の分散に関し、理論値と実験値の開きは大きく、実用的な乱数検定法として、（５）式の理論値を使うのは、適切でないと判断する。

６．各種の平均値及び分散を適用した乱数検定

Lempel-Ziv 圧縮検定の部分列数 W(n)の平均値及び分散に関し、各種の理論値及び実験値が知られている。これらを、系列長 n=100 万ビットの本検定で用いた時、検定結果にどのような影響を与えるのかを考察する。

今まで知られている部分列数の平均及び分散の組み合わせを使い、NIST の検定プログラムのソースを書き換え、５種類の Lempel-Ziv 圧縮検定法を実装した。そのパラメータを表１に示す。方式Ａは NIST のソース、方式Ｂは NIST のドキュメント,、方式Ｃは文献[6]、方式Ｄは、NIST のプログラムに付属の 11 種類の疑似乱数生成器について得られた平均値及び分散を平均したもの、方式Ｅの平均値は Louchard の理論値（（１２）式）を使用した。表の最右列には、それらの平均及び分散を使用したとき p-value＝0.01 となる部分列数を示す。実験値に基づく方式である方式Ａ、Ｃ、Ｄは何れも、この数は、ほぼ、69568 である。Lempel-Ziv 圧縮検定においては、片側確率で p-value を評価している為、平均値及び分散が変わっても、乱数らしからぬ系列と判断される閾値は、ほぼ同じである。特に、殆ど根拠無く選んだ 11 種類の疑似乱数生成器の平均を使った方式Ｄと、他の実験値方式の方式Ａ、Ｃで、その閾値が、ほぼ同じであることは、

興味深い。また、理論値方式である方式Ｂ、Ｅのその閾値は、図３の結果に照らし、通常の乱数生成器から得られる部分列数と遙かに隔たった値を示している。

表1　系列長n＝100 万ビットに対する部分列の統計量

方式	平均値E[W(n)]		分散σ²[W(n)]		p<0.01 の W(n)
A	NIST ソース	69588.2019	NIST ソース	73.23726011	69568.11
B	NIST ドキュメント	50171.6659	NIST ドキュメント	33.59365	50158.19
C	文献[6]Kim	69588.09	文献[6]Kim	75.57433652	69567.86
D	11 種の PRNG 平均	69587.6393	11 種 PRNG 平均	72.84555249	69567.77
E	文献[11]Louchard	64888.912	NIST ソース	73.23726011	64869

実際に、これら検定方式を６種類の疑似乱数生成器の出力系列の検定に適用した。結果を表２に示す。前述の考察を反映した結果である。

表2．検定結果

方式	G U SHA-1		Linear Congruential		Quadratic Congruential I		Quadratic Congruential II		Cubic Congruential		XOR	
	一様性	比率	一様性	比率	一様性	比率	一様性	比率	一様性	比率	一様性	比率
A							X					
B	ー	ー	ー	ー	ー	ー	ー	ー	ー	ー	ー	ー
C												
D												
E	ー	ー	ー	ー	ー	ー	ー	ー	ー	ー	ー	ー

空白→合格　X→不合格　ー→エラー

このように、Lempel-Ziv 圧縮検定においては、現在、知られている理論値を元に、p-value 評価をすることは、方式Ｂ、Ｅを使った検定結果で見られるように、実際的ではない。一方、この検定法の性質から、乱数の評価が片側分布で行われている為、経験値を元に、評価基準の平均値と分散を決めても、実際には、乱数らしからぬ系列として判断する部分列数の閾値一つの自由度しかなく、少なくとも、比率検定においては、適当に平均値と分散を決めた方式Ｄと他の実験値方式の方式Ａ，Ｃの本質的な違いは無い。

p-value の一様性に、関しては平均値と分散の両方の値が原理的に関係するはずであり、方式Ａと方式Ｃ，Ｄでは違いが見られる。しかしながら、文献[6]で指摘されているように、平均値と分散を実験値で適切に定めて、p-value に変換しても、p-value の一様性が、観測されないとの報告もある。部分列数の分布は、正規分布に漸近する（（４）式）ことは証明されているものの、実用的な乱数系列長で、そのように近似できるか否か疑問が残る。

７．まとめ

乱数生成器検定法である NIST SP800-22 の Lempel-Ziv 圧縮検定を中心にその問題点を調べた。結果をまとめれば以下である。

- NIST SP800-22 では、理想的な乱数についての部分列数の平均値及び分散を、SHA-1 を使った乱数生成器及び Blum-Blum-Shub 乱数生成器の発生乱数を元に、実験的に定めている事
- 現在、判っている理論式で計算される平均値及び分散は、実用的な乱数列長において、実際の乱数生成器のものと遙かに異なっている事
- 本検定の性質から、理想乱数の場合の平均値と分散を、深い考察なく定めても、乱数らしからぬ系列と判断される閾値は、ほぼ、同じであり、p-value の比率検定には影響しない事
- 実用的な乱数列長において、p-value の一様性に関し、疑問が示されている事

以上を、総合するに Lempel-Ziv 圧縮検定には、さらなる理論的解析の進展が期待されている状態であり、CRYPTREC において、乱数生成器検定用のミニマムセットを選択する際に、この Lempel-Ziv 圧縮検定を採用することは勧めない。

参考文献

[1] NIST, Special Publication 800-22:,”A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications”, (available at http://csrc.nist.gov/rng/DP800-22b.pdf

[2] NIST, Special Publication 800-22:,“NIST Statistical Test Suite”,(available at http://csrc.nist.gov/rng/sts-1.5.tar, http://csrc.nist.gov/rng/rng2.html)

[3] 情報処理振興事業協会、通信放送機構:,”暗号技術評価報告書(2002 年度版) CRYPTREC Report 2002”,平成 15 年 3 月

[4] 東芝:,”電子政府情報セキュリティ技術開発事業　擬似乱数検証ツールの調査開発調査報告書“ 、 14 情経第 1684 号 平成 15 年 2 月

[5] 濱野、佐藤、石川：”離散フーリエ変換を用いた乱数検定”、防衛庁技術研究本部技法（発行予定）

[6] 金、梅野、長谷川：“ NIST のランダム性評価テストについて ”、電子情報通信学会技術研究報告,ISEC　(2003.12)

[7] J.Ziv and A.Lempel:, “A Universal Algorithm for Sequential Data Compression,” IEEE Trans.on IT, Vol.23, pp.337-343, (1977)

[8] J.Ziv:, ”Compression, test of randomness, and estimating the statistical model of

individual sequence", in SEQENCES, R.Capocelli, Ed. New York, Springer-Verlag, 1990, pp.366-373

[9] D.Aldos and P.Shields:, "A Diffusion Limit for a Class of Randomly-Growing Binary Trees", Probability Theory and Related Fields, 78, pp.509-542, (1988)

[10] P.Kirschenhofer, H.Prodinger, and Szpankowski:, "Digital Search Trees Again Revisited: The Internal Path Length Perspective", SIAM Journal on Comp. 23, pp.598-616, (1994)

[11] G.Louchard and W.Szpakowski: "On the Average Redundancy Rate of the Lempel-Ziv Code", IEEE Trans.IT, Vol.43, No.1,pp.2-8,(1997)

[12] G.Louchard and W.Szpankowski: "Average Profile and Limiting Distribution for a Phrase Size in the Lempel-Ziv Parsing Algorithm", IEEE Trans.IT Vol.41, No.2, (1995)