NIST 乱数検定ツールの検定項目間対角化手法

岩崎　淳（情報工学部情報工学科）

Diagonalizing method among test items included in NIST randomness test tool

Atsushi IWASAKI （Department of Computer Science and Engineering, Faculty of Information Engineering）

Abstract

Randomness tests included in NIST test tool are not independent each other, i.e., the distributions followed by p-value of each test item are not independent. Even if each test item has no problem, the fact makes it difficult to derive the significance level of the test tool as whole items or to introduce a rational criterion for results of the tool. In this paper, we propose a method to solve this problem. The method transforms the distribution followed by each test item's p-value to the standard normal distribution and assumes that the joint distribution after transformation is a multidimensional normal distribution. Diagonalizing to the joint distribution, the distribution of each test item's p-vale becomes independent each other. In addition, we evaluate this method numerically.

Keywords：Random number, Randomness test, Diagonalization

1. はじめに

規則性のない（見いだせない）数列である乱数は、情報セキュリティ、暗号技術、モンテカルロ計算、裁判員選出などの抽選、検品などさまざまな分野で応用されている。その厳密な定義や求められる性質は分野に依存する。乱数の生成方法についても多種多様な研究がなされているが、ほとんどの場合においては、厳密な意味で定義通りかつ要求される性質をすべて満たす乱数を生成できない、または、生成していると保証できない。

そのため、実用上は様々な観点から生成方法や結果的に生成された乱数を評価する必要がある。その中の一つの方法が乱数検定と呼ばれる仮説検定である。帰無仮説は「与えられた数列は理想的にランダムである」であり、乱数そのものを評価対象にする。ここで、「理想的にランダムである」とは、n-bit の数列を評価する場合、「ありうる2^n本の数列の中から等しい確率で選ばれたとみなせる（あるいは、そのように考えたとして差し支えない）」という理解で良い。乱数検定はあくまでも実験的な評価方法に過ぎず、これに合格したからと言って乱数の性質の良さが証明されるわけではない。一方で、生成方法に依らず適用できるため汎用性が高いという利点がある。特に、情報セキュリティ・暗号の評価に欠くことのできないものであり、実際、現在の米国共通鍵暗号標準 AES(Advanced Encryption Standard)[1]の選定時には、乱数検定のテストセットである NIST SP800-22 乱数検定ツール[2]が使用された。

乱数検定自体は無数に作られうるもので、特にパラメータを変えれば簡単に数をそろえることができる。そのため、行うべき乱数検定をまとめたテストセットが提案されている。前述の NIST 乱数検定ツールも乱数検定のテストセットの一つであり、最も広く使われている。最新版の revision 1a で 15 種類 188 項目の検定で構成されているが、問題点が指摘されている検定も含まれており、一部は修正されたが[3][4]、未だに未修正の検定も残っている。また、個々の検定とは別に、テストセット全体としての問題も抱えている。含まれている検定項目間の関係に明確な知見がないことである。ここで検定項目間の関係とは、p 値の結合分布と同義である。例えば、「検定項目 A の p 値と検定項目 B の p 値には正の相関がある」（すなわち、「検定項目 A に通るなら検定項目 B にも通りやすい」）ということであれば、個々の検定の有意水準をαに設定したとして検定項目 A と B 両方で帰無仮説が採択される確率は$(1-\alpha)^2$にならない。テストセット全体としての有意水準を求める、または、設定した有意水準となるようにテストセット全体としての合否基準を定めることは困難となる。また、強い相関がみられるなら複数の検定を実施することは計算資源の無駄遣いと言えよう。NIST 乱数検定ツールにおいては、先行研究で検定項目間に概ね正の相関があることが報告されてきた[5][6]。すなわち、理想的な乱数が合格する検定項目数は、項目が互いに独立である場合に比べると多くなる傾向があることを意味している。

本稿では NIST 乱数検定ツールの検定項目間の非独立性の問題に対して、それを取り除くアプローチで取り組む。具体的には、複数の検定項目を通じて得られた p 値のセット

を変換し、独立な分布に従うようにすることを目指す。

2. NIST SP800-22 乱数検定ツール

NIST SP800-22 乱数検定ツールの最新版 revision 1a は以下の 15 種類の検定で構成されている：

1. Frequency Test
2. Frequency Test within a Block
3. Runs Test
4. Tests for the Longest-Run-of-Ones in a Block
5. Binary Matrix Rank Test
6. Discrete Fourier Transform Test
7. Non-overlapping Template Matching Test
8. Overlapping Template Matching Test
9. Maurer's "Universal Statistical" Test
10. Linear Complexity Test
11. Serial Test
12. Approximate Entropy Test
13. Cumulative Sums Test
14. Random Excursions Test
15. Random Excursion Variant Test

これら 15 種類の検定が実装されたサンプルプログラムが NIST から提供されている。サンプルプログラムでは、パラメータを変えるなどして 1 種類に対して複数の検定を実施するものも含まれており、総計 188 項目の検定が行われる。

NIST 乱数検定ツールは複数本の数列に対して検定を行う。すなわち、長さn-bit×m本という形で検定対象となる数列が与えられ、各検定で一本一本の数列に対して p 値が計算される。ただし、例外的に Random Excursions Test と Random Excursion Variant Test の 2 種 16 項目に関しては数列によっては検定を行わないというオプションがあり、m本の数列すべてに対して p 値が計算されるわけではない。

各検定項目の合否は、その検定で得られたすべての p 値を用いて、さらに以下の 2 通りの検定を行い判定される：

- [Proportion Test] 0.01を下回った p 値の個数が平均（$0.01m$）± 3 ×標準偏差（$\sqrt{(0.01)(0.99)m}$）の範囲内なら合格、そうでなければ帰無仮説を棄却する。
- [Uniformity Test] 0から1の区間を10個のビンに等分割し、各ビンに入っている p 値の個数をカイ二乗検定にかけ、その p 値が0.0001より大きければ合格、そうでなければ帰無仮説を棄却する。

理想的な場合において、Proportion Test の有意水準は約0.0026である。Uniformity Test は有意水準が0.0001と極めて小さいので、理想的な乱数を棄却することはめったに起こらない。

3. 提案手法

検定項目間の非独立性から生じる問題に対して解決法を考えていく。まず、「検定間の相関を取り除けば、（実用上）独立とみなせる」というざっくりとした仮定をおこう。一般論として非独立性を完全に取り除くことは難しい。しかし、相関を取り除くことは共分散行列の対角化により行えるので、この仮定により問題はずいぶん簡単になる。

なお、Random Excursions Test と Random Excursion Variant Test は例外的に扱いにくいため、これらを除いた 13 種 162 検定のみを考えることにする。また、各検定には全く問題はなく、帰無仮説の下での各検定の p 値の分布は完全に[0,1]一様分布であるとする。

〈3･1〉共分散行列の対角化

各検定項目に 1 から 162 までの番号を付けることにしよう。（具体的にはサンプルプログラムの表示順とする。）また、検定対象のm本の数列にも1からmまでの番号が振られているとする。検定項目jで数列iに対して求まった p 値を$p_{i,j}$と書くことにし、

$$P := \begin{pmatrix} p_{1,1} & \cdots & p_{m,1} \\ \vdots & \ddots & \vdots \\ p_{1,162} & \cdots & p_{m,162} \end{pmatrix}$$

と定義する。

検定項目iとjの p 値の共分散を$c_{i,j}$として、(i,j)-成分が$c_{i,j}$である162×162の行列Cを共分散行列とする。行列Cは当然実対称行列であるから、ある直交行列Lが存在して$L^{\mathrm{T}}CL$と対角化できる。

次に、先に定義したPとLを用いて

$$P' := L^{\mathrm{T}}P$$

と定義し、P'の各行を新たに各検定項目で得られた p 値のセットだと思いなおす。（決して各行に対応する検定があるわけではない。あくまでも、単に便宜的にそのように考えるだけである。）そうすると、各検定項目の p 値の分布は無相関になる。

〈3･2〉問題点

先に示した共分散行列の対角化法には以下のような問題がある。

- 帰無仮説の下での p 値の結合分布が未知なので、共分散行列Cを理論的に求めることは出来ない。
- 変換後の各検定（P'の各行）では p 値が一様分布にならない。

このうち一点目に関しては、実験的に得られる分布から計算される共分散行列で代用することでひとまず解決できよう。二点目に関しては以下のような方針で対処する：

1. p 値を連続な関数を用いて変換し、各検定の p 値の分布を標準正規分布にする。さらに「複数の検定での p 値の結合分布は多次元正規分布に従っている」との仮定をおく。
2. この状態で共分散行列を求め、それを対角化する直交行列を用いて p 値を変換する。多次元正規分布の性質から変換後の分布は、各周辺分布が独立な正規分布になっているはずである。
3. 最後に各周辺分布を白色化し、分布をステップ 1 の逆変換で[0,1]一様分布になおす。

ステップ 2 で各周辺分布が独立になるので、以上の手順の変換で各検定の結果は完全に独立になる。

〈3･3〉変換法

以上述べてきたことをまとめ、変換法を具体的に述べておこう。以下のような手順である：

1. 行列Pの各成分を
$$p \mapsto \mathrm{erf}^{-1}(2p-1)$$
と変換した行列Qを求める。ここで erf は誤差関数である。
2. 行列Qから共分散行列$\hat{C}$を計算する。
3. 共分散行列$\hat{C}$を対角化する直交行列$\hat{L}$を求め、
$$Q' \coloneqq \hat{L}^{\mathrm{T}}Q$$
とする。
4. 行列Q'の第i行の成分からその平均値μ_iと標準偏差σ_iを計算し、各(i,j)-成分$q_{i,j}$を
$$q_{i,j} \mapsto \frac{q_{i,j}-\mu_i}{\sigma_i}$$
と変換した行列Q''を求める。
5. 行列Q''の各成分を
$$q'' \mapsto \frac{\mathrm{erf}(q'')+1}{2}$$
と変換した行列P'を求める。(行列P'の各行が変換後の各検定項目の p 値となる。)

以上の手順で各検定が独立で、かつ、p 値が一様分布に従うようになる。置かれている仮定を確認しておくと、

- （変換前の）各検定項目での p 値の分布は[0,1]一様分布である
- 各検定項目の p 値の分布を標準正規分布に変換したとき複数の検定での p 値の結合分布は多次元正規分布になっている

の 2 点である。一点目に関しては、各種問題はあるけれども、そもそも各検定は p 値が一様分布になることを期待して（あるいは、目指して）設計されているので不自然な仮定とはいえないであろう。二点目の仮定は以下のように考えればよい。すなわち、検定項目間の非独立性は実用上のネックになっているけれども、もともと各検定項目は独立に近く、非独立性のほとんどは線形な相関構造としてとらえられる。そうすると、二点目の仮定も近似的に妥当といえるだろう。

4. 実験

提案手法の数値的な評価を行う。

〈4･1〉実験 1

メルセンヌツイスタ[7]を用いて数列を作成し、実験を行った。NIST は検定対象となる数列のビット長 $n = 1000000$、検定本数$m = 1000$を乱数検定ツール使用時の推奨値としている。これを踏まえて、まず $n = 1000000$、$m = 1000000$として検定の実施、提案手法による変換を行い、変換後の結果を各セット1000本ずつの1000セットに分割して評価した。また、共分散行列を対角化する直交行列の選び方には任意性が残るが、変換後の固有値が大きさの順で対角行列に並ぶように選ぶことにする。

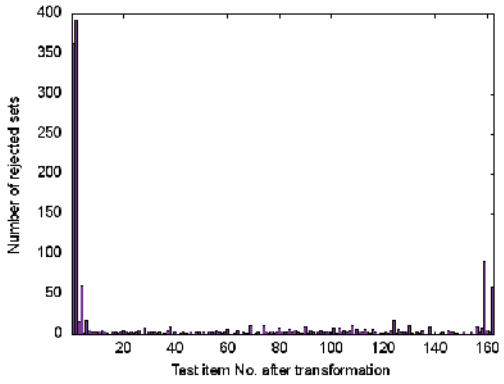

図 1　変換後の各検定項目における棄却セット数

Fig. 1.　Number of sets rejected by each test item after transformation.

図 1 は変換後の各検定項目で Proportion Test により棄却されたセット数を示している。大きな固有値に対応する検定項目と小さな固有値に対応する検定項目（両端の検定項目）で誤差（第一種過誤）の範囲を大きく超えて棄却されているのが見て取れる。提案手法でおかれている 2 つの仮定が（少なくとも厳密には）満たされていないしわ寄せがきているものと解釈できる。逆に言うと、変換前は多くの検定項目に亘って存在していた問題点を、変換により両端の検定項目に集めることができているともいえる。

〈4･2〉実験 2

実験 1 を踏まえると、提案手法による変換を行った後で、両端の検定項目を取り除けば有意義な検定結果を取り出せるものと期待できる。パラメータは実験 1 と同様にし、メルセンヌツイスタ[7]を用いた検定と変換を 2 回、AES[1]を CTR モードで使用して 2 回行った。その後、両端から検定を 1 つずつ（すなわち、ワンステップで 2 つ）削除していき挙動を

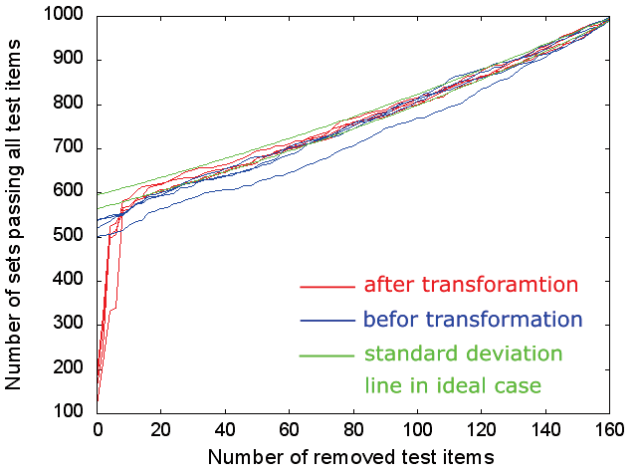

図 2　取り除いた検定項目数と全項目に合格したセット数の関係

Fig. 2.　Relationship between number of removed test items and number of sets passing all test items.

調べた。

図 2 は削除した検定項目数と残った検定項目にすべて合格したセット数を表している。変換後はおおよそ、10 から 20 程度の検定項目を取り除けば理想的な場合に一致していることが見て取れる。

変換前においては、検定項目の並びに特段の意味はないので、取り除く検定項目を両端から順に選ぶことに意味はないかもしれない。その意味においては、この実験の結果から変換後の優位性を主張するのはアンフェアであろう。しかしながら、ではどういう順番で取り除けばよいのかは不明であるからして、提案手法は順番を明確にできたという点に意義がある。

5. まとめ

検定により得られた p 値の集合を変換することで、検定項目間の非独立性を取り除く方法について提案した。提案手法においては、正規分布を介することで、変換後の p 値の分布を一様分布にすることができる。数値的な評価によると変換後の分布は理想的な場合からずれている。これは途中に置かれている仮定が厳密には満たされていないことの影響であると考えられる。しかしながら、変換後の検定項目をいくつか削除することで問題は解決しうる。

謝辞

本研究は、平成 29 年度研究費（新任スタートアップ支援）の助成を受けた。

（平成３０年８月３０日受付）

文　　献

（1） National Institute of Standards and Technology, “Specification for the ADVANCED ENCRYPTION STANDARD (AES)”, FIPS-Pub.197 (2001)

（２） National Institute of Standards and Technology, “A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications,” Special Publication 800-22 Revision 1a (2010).

（3） S. Kim, K. Umeno and A. Hasegawa, “On the NIST Statistical Test Suite for Randomness”, Technical report of IEICE, ISEC2003-87 (2003)

（4） K. Hamano and T. Kaneko, “Correction of overlapping template matching test included in NIST randomness test suite”, IEICE transactions on fundamentals of electronics, communications and computer sciences, Vol. 90, No. 9, pp. 1788-1792 (2007)

（5） D.Lihua, Z. Yong and J. Ligang, “Study on the Pass Rate of NIST SP800-22 Statistical Test Suite”, in: Proc. of 2014 Tenth International Conference on Computational Intelligence and Security, pp. 402-404 (2014)

（6） A. Iwasaki, “Analysis of NIST SP800-22 focusing on randomness of each sequence”, JSIAM Letters, Vo. 10, pp. 1-4 (2017)

（7） M. Matsumoto and T. Nishimura, “Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator”, ACM TOMACS-Special issue on uniform random number generation, Vol. 8, No.1, pp.3-30 (1998)