A-020

# 乱数度測定器 RMT テストにおける高乱数度判定基準の再考
# Reconsidering Criteria of the RMT-test as a Tool to Measure the Randomness

三賀森 悠大† 楊 欣† 田中 美栄子†
Yuta Mikamori Xin Yang Mieko Tanaka-Yamawaki

## 1. はじめに

乱数度とは，数列の数の並び方の予測や再現が如何に難しいかを示すものである．乱数度が高いほどその数列にはランダム性があり，良い乱数と見なされる．乱数検定法として良く知られた手法に JIS[1]や NIST[2]等があるが，いずれもかなり乱数度の高い数列に適用することを前提にしたものであり，乱数度の比較的低いデータの乱数度測定には適さない．また，データ形式や長さに強い制限がついているため，形式の異なるデータに対して異なる手法を使う必要があるという欠点がある．そこでもっと簡単にいろいろな種類のデータの乱数度を測定できる新たな手段として，我々は，ランダム行列理論(Random Matrix Theory：RMT)に基づく RMT テストを提案し，定性評価で直観的な乱数度判定を行うと共に，定量評価で数値化することによって，乱数度の近いデータ間の微妙な差異を見分けることができることを，疑似乱数 2 種と物理乱数 3 種を用いて大量のデータ処理の結果としての統計的評価に基づき示し，この結果に基づいて，良い乱数であるかどうかを判定する基準を定めることに成功した[3][4]．加えてその結果を NIST 乱数検定による結果と比較し，両手法の基準の検討を，様々な乱数度を持つ数列を人工的に作成することにより行なった[5]．しかしそこでデータとして使用した数列は規則列をシャッフルする回数を変化させて作ったものであり，その結果各数字の度数がすべて均一の場合に限られていた．擬似乱数列や物理乱数列のように乱数度がかなり高いことが自明なデータの度数が常に均一であるとは限らないことを考慮すると，各数字の度数にばらつきのある乱数列に対しても検証する必要がある．そこでデータとして用いる数列の範囲をもっと一般的なものに広げて，RMT テストの乱数度評価基準値の再検討を行った．

† 鳥取大学大学院工学研究科情報エレクトロニクス専攻

## 2. RMT の概要

本稿では L. Laloux, P. Cizeaux, J. Bouchaud, M. Potters[6], V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, H. E. Stanley[7][8]等により株式市場に応用された文脈に基づいて，N 個の等長(長さ L とする）時系列間の内積を成分とする，相関行列の固有値分布を求め，Q=L/N>1 を定数パラメータとして L→∞，N→∞の極限で RMT から導かれた固有値分布の理論式と比較することで，ランダム性の尺度とする．ここに現れるパラメータは

$$Q = \frac{L}{N} \tag{1}$$

のみであり， 固有値 λ の分布の最大値$\lambda_+$と最小値$\lambda_-$は

$$\lambda_\pm = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}} \tag{2}$$

を使って，固有値分布は以下の式で表される．

$$P_{RMT}(\lambda) = \frac{Q}{2\pi\lambda}\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)} \tag{3}$$

## 3. RMT テストの流れ

### 3.1 乱数データの扱い方

データ長 L の乱数列を N 個用意する．本研究では予め長い乱数列を生成しておき，図 1 のようにデータ長 L ごとに区切って N 分割する．

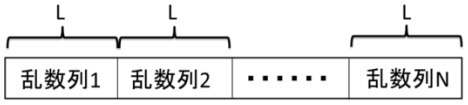


図 1. 乱数列の分割方法

### 3.2 相関行列作成

前節で用意した N 個の数列を各行に順に並べて N 行 L 列のデータ行列 $A$ を得る．この行列の i 行 j 列要素はもとの乱数列の(i-1)×L+j 番目の数字となる．

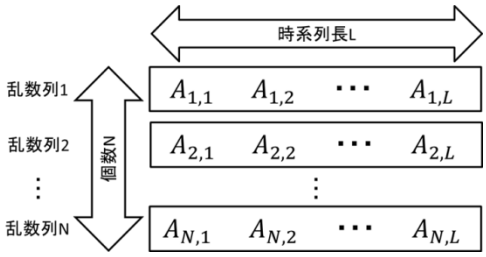


図 2. 乱数列データの並べ方

次に，この行列を行ごとに

$$g_{i,j} = \frac{A_{i,j} - \langle A_i \rangle}{\sqrt{\langle {A_i}^2 \rangle - \langle A_i \rangle^2}} \tag{4}$$

によって平均 0，分散 1 となるよう正規化し，N 行 L 列のデータ行列

$$\mathrm{G} = \begin{pmatrix} g_{1,1} & \cdots & g_{1,L} \\ \vdots & \ddots & \vdots \\ g_{N,1} & \cdots & g_{N,L} \end{pmatrix} \tag{5}$$

を作成する．相関行列 C は G とその転置行列の積として

$$\mathrm{C} = \frac{1}{L} G G^T \tag{6}$$

を使って求められ，N 行 N 列の対称行列となる．

### 3.3 定量評価

長さ 100 万のデータ列を N 本に等分割してそれらの内積から相関行列を作成する．先行研究[3][4]により N=500 以上であれば十分に RMT 公式が使えることが分かっているので，今回は，N=500，L=2000 の条件のもとで乱数度評価を行い，モーメントの理論値と実測値との誤差により，乱数度を数値で判定する．

最初に，相関行列 C を k 乗し，その対角要素の平均をとることにより

$$m_k = \frac{1}{N}\sum_{i=1}^{N} (C^k)_{i,i} = \frac{1}{N}\sum_{i=1}^{N} {\lambda_i}^k \tag{7}$$

によって k 次モーメントの実測値$m_k$を求める．次に，k 次モーメントの理論値を

$$\mu_k = \int_{\lambda_-}^{\lambda_+} \lambda^k P_{RMT}(\lambda) d\lambda \tag{8}$$

によって計算する．誤差は，式(7)，式(8)の結果を用いて

$$\text{誤差}(\%) = \left(\frac{m_k}{\mu_k} - 1\right) \times 100 \tag{9}$$

により数値化する[3][4]．これは，RMT が完全にランダムと見なす理論からどの程度ずれているかを表すもので，誤差の値が大きいものほど乱数度が低く，逆に誤差が 0%に近いものほど乱数度が高いと判断する．

## 4. NIST 乱数検定

RMT テストの結果を別の検定法の結果と比較することによりその精度を評価する．本稿では比較対象として，米国国立標準技術研究所（NIST）で開発された NIST SP 800-22 を用いる．これは，複数の検定法からなる米国標準の統計的乱数検定であり，0 と 1 からなる ASCII 形式の乱数データを読み込むことで，連の検定やランダム偏差検定など，15 種類の検定をまとめて行うことができる．NIST のホームページにソースコードが提供されており[2]，暗号の分野で広く使用されている．15 種類の検定においては，Proportion 評価，すなわち乱数列の全サンプル中，その検定に合格したサンプル数の比率によって合否を判断する．この比率が一定基準以上に達していれば，総合的に検定に合格したと判断される．

なお，先行研究により，NIST 乱数検定は RMT テストの比較対象として有用であることが分かっている[5].

## 5. 検証用乱数データの作成

### 5.1 乱数列生成目的

NIST 乱数検定により良い乱数として見なされる基準を探るにあたり，RMT テストで求めた乱数度と照合して解析を行う為，様々な乱数度の数列データを用意したい．検定に必要なデータ量の乱数を作成する際に人間の手間を省く為，ランダム性が極めて低い規則的な数列データをシャッフルさせることにより，徐々に乱数度を高くしつつ，2 種類の評価の比較を行う．コンピュータによる生成及びシャッフル作業により，元の規則的な数列から研究の目的に合った乱数列を高速で用意することが可能になる．

なお，NIST 乱数検定の条件に合わせる為に，検証に用いる乱数データは長さ 100 万で，0 と 1 から構成されるものとし，統計的に有意な結果を得る為に 55 サンプル用意する[9][10].

### 5.2 初期数列

シャッフルを開始する前の規則的な数列として，先行研究[5]と同様に 0 と 1 がそれぞれ 50 個ずつ交互に並べられ

た数列を用意し，さらに 0 を一定数で 1 に変換することで，度数の異なるものを用意する．本稿では 0 と 1 の度数の差が 100，200，300 である初期数列をそれぞれ用意した．

シャッフル方法も先行研究と同様，数列データの全要素の中から 2 要素をランダムに選び，それらの順番を入れ替える．

## 6. 実験

### 6.1 先行研究の結果及び基準値の定め方

まず，先行研究で得た，0 と 1 の度数が全て一定の場合の検証結果を表 1 に示す．ここで誤差は，式(9)によって求めた数値であり，55 サンプルの平均を表す．実験方法は，シャッフル回数 100 万～500 万回範囲で 10 万回ずつ区切り，それらの時点での数列全 41 種類を RMT テスト及び NIST 乱数検定にかけ，両者の結果を比較するものとする．

表 1. 先行研究の結果[5]

| RMTテスト誤差(%) | NIST合格率 | RMTテスト誤差(%) | NIST合格率 | RMTテスト誤差(%) | NIST合格率 |
|---|---|---|---|---|---|
| 29582.88 | 5/15 | 0.31 | 15/15 | 0.20 | 15/15 |
| 3803.87 | 7/15 | 0.30 | 14/15 | 0.20 | 15/15 |
| 572.09 | 7/15 | 0.29 | 15/15 | 0.19 | 15/15 |
| 101.80 | 7/15 | 0.28 | 14/15 | 0.19 | 15/15 |
| 22.27 | 7/15 | 0.28 | 15/15 | 0.18 | 15/15 |
| 5.79 | 10/15 | 0.28 | 14/15 | 0.18 | 15/15 |
| 1.60 | 13/15 | 0.26 | 15/15 | 0.18 | 15/15 |
| 0.60 | 14/15 | 0.25 | 15/15 | 0.14 | 15/15 |
| 0.38 | 14/15 | 0.24 | 14/15 | 0.12 | 14/15 |
| 0.36 | 15/15 | 0.22 | 14/15 | 0.11 | 15/15 |
| 0.35 | 15/15 | 0.21 | 14/15 | 0.10 | 15/15 |
| 0.34 | 15/15 | 0.21 | 15/15 | 0.10 | 15/15 |
| 0.34 | 14/15 | 0.21 | 14/15 | 0.10 | 15/15 |
| 0.32 | 15/15 | 0.20 | 14/15 | | |

表 1 の表示形式は，シャッフル回数ごとの数列の乱数度(誤差)を RMT テストによって求め，値が大きい順にソートしている．そして，その値が求められた数列の NIST 乱数検定での 15 種類中の合格率を右に示している．RMT テストの結果で，NIST 乱数検定の合格率が 14/15～15/15 と安定するようになる点をしきい値として，良い乱数と見なされる誤差基準値を定めている．表 1 では，RMT テストで求められた誤差が 0.60%以下の所で NIST 乱数検定の合格率が安定して遷移していることが分かる為，0.60%を基準値と定めている．次節でも同様の方法で，良い乱数と見なされる誤差基準値を定めていく．また，表 1 において，合格率 14/15 の乱数列全てにおいて，不合格と判断されたのは「重なりのないテンプレート適合」検定であった．

### 6.2 度数ごとの検証結果

次に，5.2 節で用意した 3 種類の乱数列について検証を行なった．先行研究で 0 と 1 の度数が一定であったのに対し，本稿ではさらに，度数が異なることによって，基準値に違いが現れるのかを確認することを目的とする．表 2 に，3 種類の乱数列の検証結果を示す．

表 2. 3 種類の比較結果

| 0 と 1 の度数差 | 誤差基準値(%) |
|---|---|
| 100 | 0.45 |
| 200 | 0.41 |
| 300 | 0.58 |

度数差が均一の場合と同様に，RMT テストにおいてシャッフル回数に応じて誤差の値が一定値以下で遷移し，その時に NIST 乱数検定における合格率も 14/15 以上で遷移していることが分かった．

また，0 と 1 の度数の差がより多い場合についても実験を行なった．度数の差は，5,000 及び 10,000 に設定している．その検証結果を表 3，表 4 に示す．

表 3. 度数差 5,000 の場合の結果

| RMTテスト誤差(%) | NIST合格率 | RMTテスト誤差(%) | NIST合格率 | RMTテスト誤差(%) | NIST合格率 |
|---|---|---|---|---|---|
| 27523.88 | 4/15 | 0.30 | 11/15 | 0.20 | 11/15 |
| 3601.35 | 5/15 | 0.29 | 11/15 | 0.20 | 11/15 |
| 558.85 | 5/15 | 0.28 | 10/15 | 0.19 | 11/15 |
| 102.25 | 4/15 | 0.28 | 11/15 | 0.18 | 10/15 |
| 22.22 | 6/15 | 0.27 | 11/15 | 0.17 | 11/15 |
| 5.79 | 17/15 | 0.26 | 10/15 | 0.16 | 10/15 |
| 1.71 | 9/15 | 0.26 | 10/15 | 0.15 | 11/15 |
| 0.54 | 9/15 | 0.25 | 10/15 | 0.13 | 11/15 |
| 0.43 | 12/15 | 0.24 | 11/15 | 0.13 | 9/15 |
| 0.39 | 11/15 | 0.24 | 11/15 | 0.12 | 10/15 |
| 0.34 | 11/15 | 0.23 | 11/15 | 0.07 | 11/15 |
| 0.31 | 10/15 | 0.22 | 11/15 | 0.06 | 11/15 |
| 0.31 | 11/15 | 0.22 | 10/15 | 0.01 | 11/15 |
| 0.31 | 11/15 | 0.22 | 10/15 | | |

表 4. 度数差 10,000 の場合の結果

| RMTテスト誤差(%) | NIST合格率 | RMTテスト誤差(%) | NIST合格率 | RMTテスト誤差(%) | NIST合格率 |
|---|---|---|---|---|---|
| 26445.12 | 2/15 | 0.33 | 7/15 | 0.21 | 9/15 |
| 3488.67 | 4/15 | 0.32 | 5/15 | 0.21 | 6/15 |
| 535.07 | 3/15 | 0.32 | 9/15 | 0.21 | 6/15 |
| 96.47 | 3/15 | 0.32 | 6/15 | 0.20 | 6/15 |
| 21.81 | 3/15 | 0.32 | 8/15 | 0.15 | 7/15 |
| 5.89 | 6/15 | 0.31 | 7/15 | 0.15 | 7/15 |
| 1.78 | 4/15 | 0.31 | 7/15 | 0.14 | 7/15 |
| 0.41 | 6/15 | 0.30 | 6/15 | 0.14 | 6/15 |
| 0.40 | 5/15 | 0.29 | 9/15 | 0.14 | 9/15 |
| 0.37 | 7/15 | 0.28 | 6/15 | 0.14 | 7/15 |
| 0.36 | 6/15 | 0.28 | 6/15 | 0.12 | 9/15 |
| 0.35 | 6/15 | 0.27 | 7/15 | 0.12 | 7/15 |
| 0.34 | 6/15 | 0.23 | 6/15 | 0.07 | 7/15 |
| 0.34 | 7/15 | 0.21 | 7/15 | | |

0 と 1 の度数の差が小さい場合と比べ，乱数度の向上に関わらず, 全体的に NIST 乱数検定における合格率が低く，それぞれの最高合格率は度数差 5,000 の場合で 12/15, 度数差 10,000 の場合で 9/15 となっていることが分かる．

## 7. 考察

度数差が 5,000 や 10,000 と大きい場合，RMT テストによって求められた乱数度はシャッフル回数に応じて向上している点で, RMT テストの結果の出方は度数差が小さい場合と類似しているが，シャッフル回数 100 万～500 万回の全ての乱数列において合格率が低いことが前章より分かる．この場合の NIST 乱数検定の結果を調査した結果，度数一定の場合と同様に,「重なりのないテンプレート適合」検定で不合格と判定される場合があるほか,「一次元度数」検定,「累積和」検定,「連」検定が必ず不合格と判定されていることが判明した．そのうち,「重なりのないテンプレート適合」検定においては，検定に使用するテンプレートのビット数が 9 の場合のみであった為，結果が 9 ビットのテンプレートのみに依存していたことが考えられる．後者 3 種類の検定は，度数差が大きい数列では合格基準に達することができないと考えられる．その為，NIST 乱数検定で合格する為の乱数列の度数には一定の差を境界線とした限界があると考えられる．

## 8. おわりに

本稿では，先行研究に続いて度数差を考慮に入れて検証を行なった．6.2 より，度数差が過度に大きくなるにつれて NIST 乱数検定における合格率が悪くなることが確認できるが，それに対し，RMT テストにおける乱数度の低下は見られなかった．そこで NIST 乱数検定で合格する為の数列の度数差の限界を探求することが今後の課題となる．

### 参考文献

[1] 日本規格協会, “JIS Z 9031 乱数発生及びランダム化の手順”，2001 年改正.

[2] NIST: http://csrc.nist.gov/groups/ST/toolkit/rng/documentation_software.html

[3] 田中美栄子，糸井良太，楊欣，“ ランダム行列理論を用いた乱数度評価法の提案”，情報処理学会論文誌数理モデル化と応用 Vol.5, pp.1-8, 2012.

[4] X. Yang, R. Itoi, M. Tanaka-Yamawaki, “Testing Randomness by Means of Random Matrix Theory”, Progress of Theoretical Physics, Supplement, Vol.194, pp.73-83, 2012.

[5] 三賀森悠大，楊欣，糸井良太，田中美栄子，“RMT テストの性能検証～NIST 乱数検定との比較～”，情報処理学会論文誌数理モデル化と応用, Vol. 6, pp.57-63, 2013.

[6] L. Laloux, P. Cizeaux, J. Bouchaud, M. Potters, "Noise Dressing of Financial Correlation Matrices", Physical Review Letters, Vol. 83 , pp.1467-1470, 1998.

[7] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, H. E. Stanley, Physical Review Letters, Vol. 83, pp. 1471-1474, 1999.

[8] V. Plerou, P. Gopikrishnan, B. Rosenow, L. Amaral, H. Stanley, "Random Matrix Approach to Cross Correlation in Financial Data", Physical Review E, Vol. 65 no.066126, 2002.

[9] 情報処理振興事業協会 セキュリティセンター, “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”，pp. 1-45,（平成 15 年 2 月）

[10] A. Rukhin, J. Soto, J. Neckvatal, M. Smid, E. Barker, S. Leigh, M. Levenson, M. Vangel, D. Banks, A. Heckert, J. Dray, and S. Vo, “Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications” , pp. 5-1 - 5-8, (April 2010)