

ランダム行列理論を用いた乱数度評価法の提案

田中 美栄子^{1,a)} 糸井 良太¹ 楊 欣¹

受付日 2011年4月19日, 再受付日 2011年6月9日/6月21日/7月26日/10月27日,
採録日 2011年11月22日

概要: ランダム性の高い時系列の相関行列の固有値分布は, 次元 N と時系列長 L が無限大の極限で, その比 $Q = L/N$ のみにより表される簡単な関数となることがランダム行列理論 (RMT) により導かれる. 本稿ではこれを用いて新しい乱数度評価法を提案する. すなわち, 対象とする数列から作成した相関行列の固有値分布が RMT 公式に一致するか否かで乱数度を判定しようとするわけである. 図の目視によりデータの固有値分布と理論式とを比較した場合は, 代表的な機械乱数である線形合同法とメルセンヌ・ツイスタの生成する乱数列をいずれも「可」と判定するが, 乱数列の初期部分や数列の変化率のように乱数度を低下させた数列に適用すると, 目視だけでも乱数度の低さが検出できる. 本手法は可視化による直観性と分布の形状による特徴表現が可能な点により, 経済・社会・医療データ等への応用が期待される. この方法を本稿では「RMT法による定性評価」とよぶ. 一方, 機械乱数列のように乱数度の高い数列の乱数度を比較するためには, 特徴を数値化することが必要である. そこで, 固有値分布のモーメントと理論の比較を利用し, 「 k 次以下のモーメントが理論式に x パーセント以下の誤差で一致すれば乱数度が高い」という基準を定めることにより図の目視に頼らない基準を作ることができる. これを本稿では「RMT法による定量評価」とよぶ. 数値実験の結果, $k=6$ の場合に $x=5$ とする基準を採用した. この方法で線形合同法とメルセンヌ・ツイスタにより発生させた乱数列を比較すると, 個々の乱数列に対する SEED を変えた乱数のばらつきが大きく, この2種類の疑似乱数の優劣をつけることはできない. さらに, 物理的な方法で発生させた乱数列を同様の方法で評価した結果, WEB上でダウンロードできる3種類の物理乱数発生器の中で, 最新のものについては他の発生器によるものより誤差は小さいものの, 線形合同法やメルセンヌ・ツイスタに比べて乱数列間のばらつきが大きいため大きな誤差を生じる結果となる. したがって, 本手法の提案する定性評価で区別のつかないほど乱数度の高い数列であれば, 定量評価でも優劣がつけ難いことが分かった.

キーワード: ランダム行列理論, 乱数検定, モーメント解析, 固有値分布

Testing Randomness by Means of Random Matrix Theory

MIEKO TANAKA-YAMAWAKI^{1,a)} RYOTA ITOI¹ XIN YANG¹

Received: April 19, 2011, Revised: June 9/June 21/July 26/October 27, 2011,
Accepted: November 22, 2011

Abstract: Random matrix theory derives, at the limit of both dimension N and length of sequences L going to infinity, that the eigenvalue distribution of the cross correlation matrix between time series with high random nature can be expressed by a simple function of $Q = L/N$. Using this fact, we propose a new method of testing randomness of a given sequence. Namely, the randomness of a sequence passes the test if the eigenvalue distribution of the cross correlation matrix matches the random matrix theory. We first test the method on two popular random number generators, the Linear Congruential Generator (LCG) and the Mersenne Twister (MT), and show that both pass the test. The advantage of our method is to visualize the randomness of the data sequences, which is an essential factor for social, economic/financial or medical applications. We call this visual method as the ‘Qualitative evaluation of the RMT test’. In order to discriminate subtle differences between pseudo-random numbers, we need quantify the test. For this purpose, we introduce the moment method. We derive the theoretical formula for the moments of the eigenvalue distributions based on the RMT formula as a function of the parameter Q , and then compare them to the corresponding moments directly computed from the actual eigenvalues. The error is thus defined as the deviation of the ratio, of the actual moment over the theoretical moment, from one. We can set the criterion of randomness by ‘the randomness of a sequence is high if the moments up to the k -th order is less than x percent’, without a help of visual representation of figures. We call this moment method as the ‘Quantitative evaluation of the RMT test’. Applying this method on the sequences generated by LCG and MT, we found that those two are indistinguishable due to the fluctuation between SEEDs. In other words, both generators pass the test if we choose $x = 5$ for $k = 6$. We have also applied this method on three physical random numbers generated in the physical process and confirmed that all the three pass the test, although one of them (Tokyo ED) performs better compared to the rest. However, the comparison of randomness between the physically generated random numbers and the pseudo-random generators is difficult due to the fluctuation of data from the physical process and the dependence of SEEDs for the pseudo-random generators. We conclude that the randomness of a sequence is indistinguishable by the ‘Quantitative evaluation of the RMT test’ if it is indistinguishable by the ‘Qualitative evaluation of the RMT test’.

Keywords: RMT-test of randomness, moment analysis, eigenvalues of random matrix, Marcenko-Pastur distribution

1. はじめに

ランダム行列理論 (Random Matrix Theory: RMT と略) [1] はランダム性の高いデータからランダム部分を抜き取り, 相関の強い部分を残すことによって, 株価のようなランダム性の高い時系列間の相関から主成分を抽出するために使うことができる [2], [3]. 本研究はそこに着目し, ランダム行列理論 (RMT) により求められた時系列相関行列の固有値分布に基づく主成分分析法 (RMT-PCA) [4] を擬似乱数の乱数度評価法に応用できるかどうかを実証するとともに, 本手法の持つ, 可視化による直感性と分布の形状による特徴表現が可能な特徴を生かして, 経済・社会・医療データ等のランダム性の評価に用いることを提案する.

2. ランダム行列理論

本稿で用いるのは Plerou らにより 2002 年に株式市場に応用された [2] 文脈に基づき, 時系列の相関行列の固有値分布をランダム行列理論から導かれる式と比較する方法である. 以下に手法を概説する. 時系列長 L の無作為なデータ N 個の内積を成分とする相関行列を作成し固有値を求める. これは乱数列を図 1 のように各列に並べて L 行 N 列

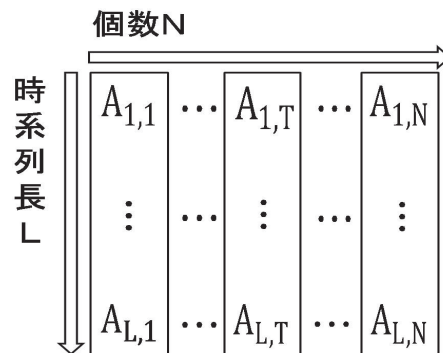


図 1 データの形
Fig. 1 Data type.

のデータ行列 A を作ると $A^T A$ が相関行列となるのでこれを対角化すると定式化してもよい. このような相関行列の固有値経験分布は, データがランダム列であれば $N \rightarrow \infty$, $L \rightarrow \infty$ でその統計性によらず $Q = L/N$ のみに依存する簡単な関数となり, 次の式 (1)–(3) で表せる [5], [6].

$$P_{RMT}(\lambda) = \frac{Q}{2\pi\lambda} \sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)} \quad (1)$$

$$\lambda_{\pm} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}} \quad (2)$$

$$Q = \frac{L}{N} \quad (3)$$

3. 提案手法の手順

(1) データを準備する

¹ 鳥取大学大学院工学研究科情報エレクトロニクス専攻
Department of Information and Electronics, Graduate
School of Engineering, Tottori University, Koyama-cho,
Tottori 680-8550, Japan

a) mieko@ike.tottori-u.ac.jp

- 擬似乱数でデータを生成する
本文のデータとしては線形合同法 (Linear Congruential Generator: LCG と略) [7], およびメルセンヌ・ツイスタ (Mersenne Twister: MT と略) [8] で作成する擬似乱数を用いる.
- 物理乱数を取得する
統計物理研究所のホームページ [9] から入手する物理乱数を使用する.
- データの成形
長さ L の時系列を N 個用意し, 図 1 のように並べる. 各列ごとに正規化して

$$g^{(i,j)} = \frac{A_{(i,j)} - \langle A_j \rangle}{\sqrt{\langle A_j^2 \rangle - \langle A_j \rangle^2}} \quad (4)$$

とし ($i = 1, 2, \dots, L, j = 1, 2, \dots, N$), これを行列 G

$$G = \begin{pmatrix} g_{11} & \cdots & g_{1N} \\ \vdots & \ddots & \vdots \\ g_{L1} & \cdots & g_{LN} \end{pmatrix} \quad (5)$$

とする.

- (2) 相関行列 C を作成する

$$C = \frac{1}{L} G^T G \quad (6)$$

ここで

$$C_{i,j} = C_{j,i} \quad (7)$$

対角要素

$$C_{i,i} = 1 \quad (8)$$

- (3) 相関行列 C の固有値を求める
(4) 乱数度を評価する

- 定性評価
固有値分布のヒストグラムと RMT 式 (1) を目視により比較する. 両者が一致すれば乱数度テストに合格したとする. この方法は直感的に見やすく, 形状の特徴を把握しやすいという利点がある.
- 定量評価
実データから計算した 6 次までのモーメントを式 (1) から計算したモーメントの理論値と比較することにより乱数度を判定する. 大量な実験結果をもとに乱数度に対しては誤差の絶対値は 5% を乱数度判定の基準と選定する (詳しくは 5 章に記述する).

4. 乱数度の評価

長い時系列を外観的に見た場合でどれだけランダムかということを, 本稿では乱数度という言葉で表現する. ここで, 擬似乱数発生器の LCG と MT により作成する時系列と物理乱数時系列の乱数度を評価する.

4.1 N と L の範囲の決定

実験パラメータとして使用するデータ長 L とデータ数 N の選択について述べる. 提案手法で使用する固有値分布式 (式 (1)), P_{RMT} は 1 個のパラメータ $Q (= L/N)$ のみに依存する. しかし, この式は N と L が無限大の極限という条件で導かれているため, N と L を大きくとる必要があり, また, 必要な自由度を確保するため, $N < L$ を満たす必要がある. どのくらい大きな N が必要かを調べるため $N = 200, 300, 400, 500$ のデータから得た固有値分布が理論式 P_{RMT} に一致するかどうかを $Q = 3$ の場合に調べた結果を図 2 に示す. 図 2 の下から上までの線は 1 次モーメントから 6 次モーメントの誤差平均値を示す. その結果 6 次以下のモーメントとその理論式との誤差の平均値は図 2 に示すように N を大きくするに従って減少し, $N = 500$ で 0.25% 以下になった. ここで, これらの結果から $N = 100$ から 500 まで増加するにつれて乱数列は P_{RMT} に収束していくと考えてよい.

4.2 LCG と MT の乱数度評価

擬似乱数生成法として最も一般的な LCG 法は次式で生成される.

$$X_{n+1} = (aX_n + b) \bmod M \quad (9)$$

ここでは次式のパラメータ [7] を持つ rand() 関数によって乱数を生成する.

$$a = 1103515245, b = 12345, M = 2147483648.$$

一方, MT [8] は $2^{19937} - 1$ の周期を持つ, 高次元に均等分布するデータが作成できる乱数発生器である. 製作者の HP にソースコードが配布されており, 本研究ではそれを使用して MT より乱数生成を行った. MT はもっと簡便に生成された乱数列を直接入手できる HP も用意されている [9]. 本実験では LCG と MT による擬似乱数を使用して提案手法の評価を行う.

4.2.1 定性評価による結果

2 種類の疑似乱数発生器 LCG と MT からそれぞれ生成したデータに対する定性評価の結果を図 3 ($N = 500, L = 1500$) に示す. 両方とも固有値分布が理論分布の範囲に一致し, 曲線の形もほぼ同一といえる. つまりいずれも乱数度は良好と判定できる.

4.2.2 定量評価による結果

上記の方法で LCG と MT いずれも乱数度が高いと判定されたが, 両者の乱数度を比較するためには定量的な基準が必要となる. ここでは, 固有値のモーメントを対応する理論値と比較することで定量評価を行う. まず, 固有値の k 次モーメントを実測値から計算し

$$m_k = \frac{1}{N} \sum_{i=1}^N x_i^k \quad (10)$$

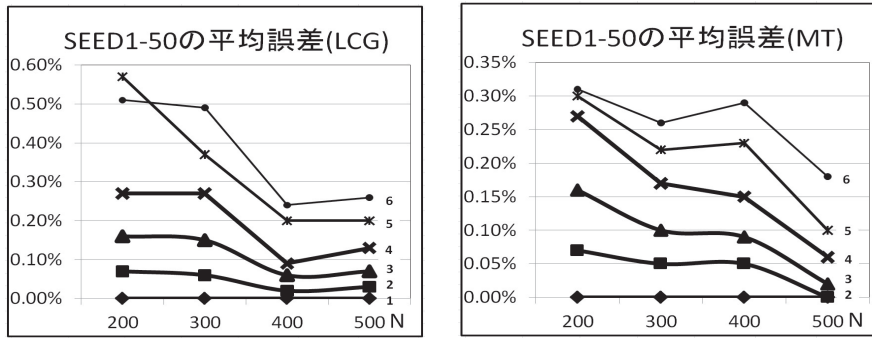


図 2 N によるモーメントの誤差変化図. LCG (左) と MT (右)

Fig. 2 Errors of moments according to N LCG (left) and MT (right).

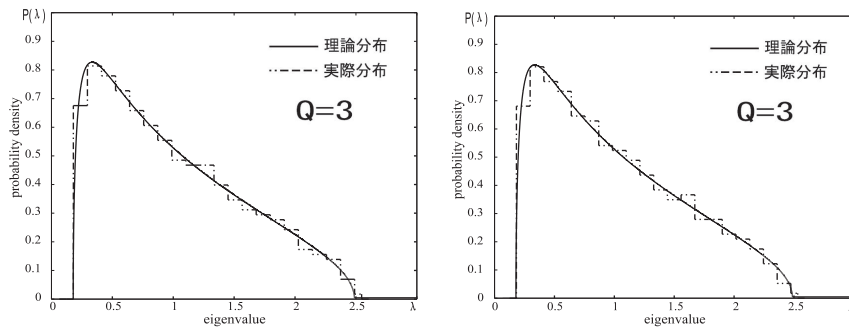


図 3 LCG (左) と MT (右) の評価例

Fig. 3 Example of LCG (left) and MT (right) passing the RMT-test.

対応する理論値を式 (1) を用いて以下のように計算する.

$$\mu_k = E(\lambda^k) = \int_{\lambda_-}^{\lambda_+} \lambda^k P_{RMT}(\lambda) d\lambda \quad (11)$$

6 次以下のモーメントは Q の関数として以下のようになる.

$$\mu_1 = 1 \quad (12)$$

$$\mu_2 = 1 + \frac{1}{Q} \quad (13)$$

$$\mu_3 = 1 + \frac{3}{Q} + \frac{1}{Q^2} \quad (14)$$

$$\mu_4 = 1 + \frac{6}{Q} + \frac{6}{Q^2} + \frac{1}{Q^3} \quad (15)$$

$$\mu_5 = 1 + \frac{10}{Q} + \frac{20}{Q^2} + \frac{10}{Q^3} + \frac{1}{Q^4} \quad (16)$$

$$\mu_6 = 1 + \frac{15}{Q} + \frac{50}{Q^2} + \frac{50}{Q^3} + \frac{15}{Q^4} + \frac{1}{Q^5} \quad (17)$$

式 (12)–(17) により計算した理論値からの実測値の誤差は以下のようになる.

$$error = m_k / \mu_k - 1 \quad (18)$$

6 次以下のモーメントに対し, LCG と MT の SEED=1–50 のモーメントの誤差平均値と標準偏差を Q = 3 と Q = 6 (N = 500) に対して計算すると表 1 と表 2 のようになる. ここで k = 1 の場合は固有の平均値でつねに 1 となることから, k = 2 以上のみを表に記し, 表記方式は「誤差平均値 (標準偏差)」とした. SEED や Q を変化させると, LCG および MT の乱数度も変わる. たとえば, Q = 3 の

表 1 SEED=1–50 の誤差の平均値と標準偏差

Table 1 Average and standard deviation of errors (SEED=1–50).

k	LCG(Q = 3)	MT(Q = 3)	LCG(Q = 6)	MT(Q = 6)
2	-.0003(.0009)	.0000(.0010)	-.0001(.0005)	-.0003(.0006)
3	-.0007(.0023)	.0001(.0026)	-.0003(.0014)	-.0007(.0016)
4	-.0013(.0039)	.0004(.0045)	-.0006(.0026)	-.0012(.0030)
5	-.0020(.0058)	.0010(.0067)	-.0008(.0041)	-.0017(.0046)
6	-.0026(.0080)	.0018(.0091)	-.0010(.0059)	-.0021(.0065)

表 2 SEED=10001–10050 の誤差の平均値と標準偏差

Table 2 Average and standard deviation of errors (SEED=10001–10050).

k	LCG(Q = 3)	MT(Q = 3)	LCG(Q = 6)	MT(Q = 6)
2	-.0004(.0010)	-.0003(.0010)	-.0004(.0006)	-.0002(.0006)
3	-.0010(.0025)	-.0008(.0024)	-.0010(.0016)	-.0004(.0017)
4	-.0017(.0046)	-.0013(.0042)	-.0017(.0028)	-.0007(.0031)
5	-.0025(.0070)	-.0017(.0062)	-.0025(.0043)	-.0010(.0047)
6	-.0031(.0098)	-.0019(.0084)	-.0032(.0059)	-.0011(.0066)

場合, 表 1 では MT の誤差平均値は正, すなわち実測値が理論値より大きいものに対し, SEED の異なる表 2 では負, つまり実測値は理論値より小さくなっている. また, Q = 6 の場合, 表 1 では MT の誤差平均値が LCG より小さく, 表 2 ではその逆となっている. よって本手法では, LCG と MT の差を検出できるものの, SEED の違いによるデータのゆらぎのために, 単純にどちらかよいかを評価

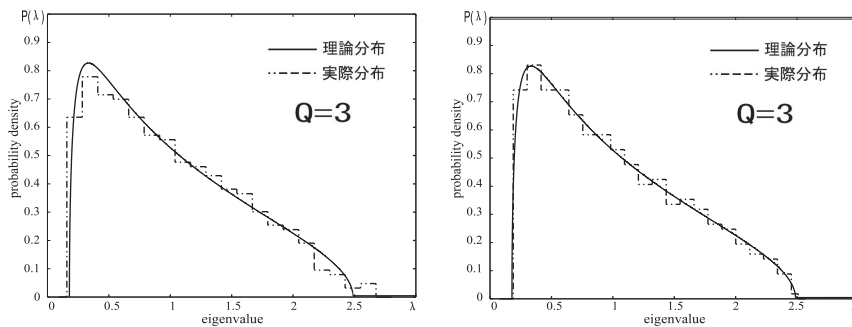


図 4 最初の 500 個のみを集めて評価した結果. LCG (左) MT (右)

Fig. 4 Evaluation of the initial part of LCG (left) and MT (right).

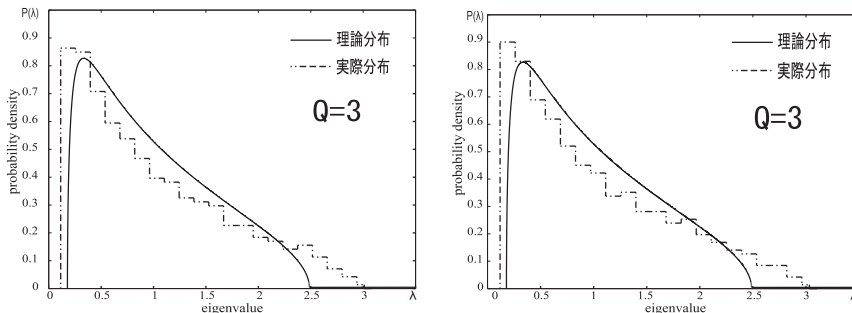


図 5 LCG (左) と MT (右) の $N = 500, L = 1500$ の対数収益評価結果

Fig. 5 Evaluation of log-return sequences of LCG (left) and MT (right).

できなかった。しかし、それぞれの SEED に対して各次モーメントの誤差の絶対値が 5% の範囲以内で変化していることを考慮すれば、5% 以内の誤差で一致する場合を乱数度が高いとする基準をとることにより、LCG も MT も乱数度テストに合格することになる。

4.2.3 初期乱数の乱数度評価

ここから、LCG と MT の初期乱数の乱数度を調べる。LCG で生成した乱数を異なる SEED に対して発生させた乱数の初期 500 個のみをつなげて用いる。そうすることにより、 500×500 相関行列を $L=1500$ について作成し、その定性評価結果を図 4 (左) に示す。固有値の分布がランダム行列理論より導かれる理論値 $[\lambda_-, \lambda_+]$ の範囲からはみ出しており、乱数度の低さを目視だけでも明らかに検出していることが分かる。同じ条件で MT で生成したデータの最初の 500 個のみを使って評価したものを図 4 (右) に示し、対応する定量評価結果はまとめて表 3 に示す。LCG の 6 次以下モーメントの誤差は絶対値 5% 以上であるのに対し、MT の方は 1% 未満であり明らかに MT の初期乱数の乱数度が高いといえる。

4.2.4 データの対数収益化による乱数度低下の検出

株価を処理するとき価格そのものでなく、その変化率である「収益率」が常用される。プログラミング上、これは対数差をとることにより実現され、対数収益と呼ばれるが、この過程で特有の癖が時系列に付与される。この効果を LCG と MT で生成した擬似乱数に適用して乱数度を下げ、提案手法でその乱数度を検定した結果を $N = 500,$

表 3 最初 500 個のみを集めた初期乱数の定量評価

Table 3 Quantitative evaluation of initial parts of LCG and MT.

k	LCG(Q = 3)	MT(Q = 3)	LCG(Q = 6)	MT(Q = 6)
2	.0045	-.0018	.0057	-.0007
3	.0103	-.0042	.0141	-.0021
4	.0197	-.0064	.0251	-.0041
5	.0352	-.0083	.0392	-.0066
6	.0583	-.0099	.0571	-.0092

表 4 変化率をとって乱数度を下げた場合の評価結果. LCG (左), MT (右)

Table 4 Quantitative evaluation of log-return sequences of LCG and MT.

k	LCG(Q = 3)	MT(Q = 3)	LCG(Q = 6)	MT(Q = 6)
2	.1047	.1227	.0696	.0702
3	.2578	.3088	.1866	.1892
4	.4445	.5442	.3391	.3450
5	.6596	.8260	.5240	.5342
6	.9092	1.174	.7426	.7579

$L = 1500$ の場合について図 5 に示す。ランダム行列理論の許容範囲 $[\lambda_-, \lambda_+]$ から出ており、乱数度が低いといえる。6 次以下のモーメントによる評価結果を表 4 に示す。この数値からも対数収益をとることによる乱数度の低下が検出される。

また、表 5 の結果より、LCG と MT の擬似乱数の生成パターンを評価した。同様の手法で対数収益をとったもの

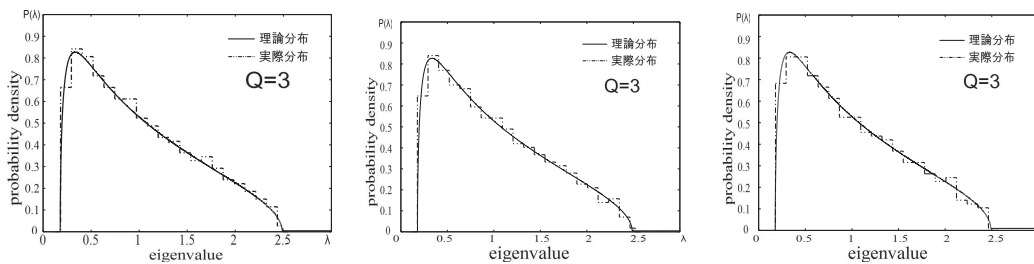


図 6 物理乱数の評価例

Fig. 6 Example of true random number passing the RMT-test.

表 5 対数収益の分布範囲と理論範囲の比較. LCG (左) と MT (右)

Table 5 Range of eigenvalues compared to the theoretical range (LCG (left) and MT (right)).

Q	RMT	LCG	LCG/RMT	MT	MT/RMT
2	2.83	3.43	1.21	3.43	1.21
3	2.30	2.79	1.21	2.80	1.21
4	2.00	2.40	1.20	2.41	1.21
5	1.78	2.15	1.21	2.15	1.21
6	1.63	1.97	1.21	1.96	1.20

表 6 物理乱数の誤差平均値と標準偏差の比較 (Q = 3)

Table 6 Quantitative evaluation of physical random number (Q = 3).

k	Toshiba	Hitachi	Tokyo ED
2	-0.0004(.0010)	-0.0004(.0011)	-0.0000(.0008)
3	-0.0010(.0027)	-0.0010(.0030)	-0.0002(.0021)
4	-0.0017(.0048)	-0.0017(.0053)	-0.0005(.0038)
5	-0.0025(.0073)	-0.0022(.0081)	-0.0008(.0057)
6	-0.0033(.0101)	-0.0027(.0113)	-0.0010(.0079)

をデータとした場合で、固有値分布範囲（固有値の最大と最小値の差）が得られた。対数収益をとることによる固有値分布の浸出範囲は、本研究の結果から、経験的に理論の分布範囲 ($RMT = \lambda_+ - \lambda_- = 4/\sqrt{Q}$) から 1.2 倍になると考えられる。

4.3 物理乱数の乱数度評価

物理乱数は本来規則性も再現性もないために予測は不可能である（例：サイコロを振るとき、今までに出た目から次に出る目を予測するのは不可能）。一方、擬似乱数は計算によって作るのだから、作り方が分かれば理論的には予測可能であり、また内部の初期値が分かれば、先に計算しておくこともできる。統計数理研究所ホームページから東芝製物理乱数、日立製作所製物理乱数と東京エレクトロン製物理乱数の 3 種類の物理乱数のデータを入手できる。本稿では 3 種類の物理乱数を使用しそれらのデータの乱数度を評価する。ここで、擬似乱数と同じ $Q = 3$ と $Q = 6$ ($N = 500$) の 3 種類の物理乱数の定性評価結果を例として示す（図 6）。図 6 の左から右まで順に東芝製、東京エレクトロン製と日立製の物理乱数を用いた結果である。定

表 7 物理乱数の誤差平均値と標準偏差の比較 (Q = 6)

Table 7 Quantitative evaluation of physical random number (Q = 6).

k	Toshiba	Hitachi	Tokyo ED
2	-0.0001(.0006)	-0.0004(.0010)	-0.0002(.0005)
3	-0.0004(.0016)	-0.0010(.0021)	-0.0005(.0013)
4	-0.0008(.0029)	-0.0015(.0035)	-0.0009(.0025)
5	-0.0012(.0045)	-0.0021(.0051)	-0.0012(.0038)
6	-0.0016(.0063)	-0.0028(.0069)	-0.0014(.0053)

表 8 50 回 (左) と 500 回 (右) 繰返し誤差の変化範囲 ($N = 500$, $L = 1500$)

Table 8 $Q = 3$ ($N = 500$) [min.:max.] of 50 (left) and 500 (right) independent tests.

k	LCG(50)	MT(50)	LCG(500)	MT(500)
2	[-0.0021, .0016]	[-0.0021, .0024]	[-0.0022, .0024]	[-0.0025, .0023]
3	[-0.0051, .0040]	[-0.0052, .0061]	[-0.0068, .0064]	[-0.0073, .0060]
4	[-0.0086, .0065]	[-0.0086, .0109]	[-0.0108, .0122]	[-0.0122, .0115]
5	[-0.0124, .0104]	[-0.0122, .0164]	[-0.0168, .0193]	[-0.0182, .0182]
6	[-0.0164, .0152]	[-0.0160, .0227]	[-0.0236, .0276]	[-0.0243, .0258]

量評価の結果（表 6, 表 7）から見るといずれも乱数度が高い。3 種類の中では、東京エレクトロン製のものが他の 2 種類と比較して一貫して乱数度が高い。

5. 定量評価基準について

乱数度が高い例（図 3）に対しては目視で乱数度が判別不可である。定性評価では両方共乱数度が高いとしかいえず、両者の乱数度を比較するためには定量評価が必要である。物理乱数のような乱数度が高い時系列に対して、初期値の異なる 50 種類の時系列の 6 次以下のモーメントの誤差は平均 5% 以下になり、500 種類の時系列を用いた結果の平均も大差ない結果となった（表 8）。また、図 4（左）のように 6 次モーメントの誤差が 5% 以上となるような時系列は目視だけでも乱数度が低いと判別できる。一方、価格時系列、気象データ、地球・宇宙環境の測定データ、人口データを含む、様々な社会データを想定し、対象とするデータにどの程度のランダムネスを期待するかによって数値基準を選ぶことが必要になる。

6. 提案手法の考察

乱数列の評価法はすでに多くある。代表的な例として、脇本 [10] の乱数列検定法があるが、乱数列の等確率性と無規則性の2点に対して統計的仮説検定を適用する方法である。ほかにも、米国のNIST FIPS14-2に基づく、乱数度評価法 (Mono bit, Runs test, Long runs) 等も一般的に使われるが、1回のテストで20000ビット以下という長さの上限や、データ形式にも強いという制限がある。これに対し、本手法は非常に長いデータ (数百万点) に対し、データ構造や表現によらず適用できることと、図の目視による比較によって直感的に判断可能な点にメリットがある。既存の検定法はデータの種類によってどの方法を使うかをまず決めなければならない。本手法は $Q > 1$ と $N > 300$ の条件を満たせばどんなデータにも適用可能である。本稿で提案する評価法は大量の乱数列をコンピュータ・プログラムにより自動処理できるため、数万を超えるような大量のデータにも適用可能である。

本提案手法のもう1つの利点は結果の視覚化である。モアレ縞法、および他の視覚的な評価法は定量化評価ができない。これに対し、本手法は視覚的評価とともに定量的な評価が可能な点において有用性が高いと考えられる。

7. 終わりに

本研究ではランダム行列理論を用いる新しい擬似乱数の乱数度評価法を提案した。これを使って乱数度を低下させた数列や、乱数度が低いと予想されるデータの乱数度を評価することができた。LCGとMTのそれぞれについて同じ初期値からの生成パターンを乱数度評価し、乱数度が良いという結果が出た。定量評価を行いMTとLCGの差が検出できた。しかし、MTとLCGはSEEDによるばらつきが大きいため、乱数度を比較できなかった。各初期値からの初期乱数部分を評価したところ、LCGの方がMTに比べて悪い評価結果を検出した。提案手法の性能をチェックするために乱数の対数収益をした時系列を用いると、乱数度低下を検出できた。また、LCGとMTで生成した乱数列をその変化率である対数収益列にすると、乱数度が低下することとその結果生じる固有値分布の特徴を検出できた。最後に、東芝、日立と東京エレクトロン製の3種類の物理乱数の乱数度を評価した。定性評価では3種類とも2つの擬似乱数と区別できないが、定量評価の結果は物理乱数の乱数度は一様性に欠け、擬似乱数よりも低い場合もあることを検出した。3種類の中では、東京エレクトロン製のものが他の2種類と比較して一貫して乱数度が高いことが分かった。

参考文献

- [1] Mehta, M.: *Random Matrices*, 3rd edition, Academic Press (2004).
- [2] Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L. and Stanley, H.: Random Matrix Approach to Cross Correlation in Financial Data, *Physical Review E*, Vol.65, 066126 (2002).
- [3] Laloux, L., Cizeaux, P., Bouchaud, J. and Potters, M.: Noise Dressing of Financial Correlation Matrices, *Physical Review Letters*, Vol.83, pp.1467-1470 (1998).
- [4] 田中美栄子, 木戸丈剛: ランダム行列との比較による株価日中変動の相関行列解析, FIT2010: 第9回情報科学技術フォーラム講演論文集, pp.153-156, 電子情報通信学会・情報処理学会 (2010).
- [5] Marcenko, V. and Pastur, L.: Distribution of Eigenvalues for Some Sets of Random Matrices, *Mathematics of the USSR-Sbornik*, Vol.1-(4), pp.457-483 (1994).
- [6] Sengupta, A. and Mitra, P.: Distribution of Singular Values for Some Random Matrices, *Physical Review E*, Vol.60, pp.3389-3392 (1999).
- [7] 松本 真: 有限体の擬似乱数への応用 (2004). 入手先 <<http://www.math.sci.hiroshima-u.ac.jp/~m-mat/TEACH/0407-2.pdf>>.
- [8] Matsumoto, M. and Nishimura, T.: Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudorandom Number Generator, *ACM Trans. Modeling and Computer Simulation*, Vol.8, pp.3-30 (1998).
- [9] Tamura, Y.: Random Number Library (2010). available from <<http://random.ism.ac.jp/random>>.
- [10] 脇本和昌: 乱数の知識, 森北出版 (1970).



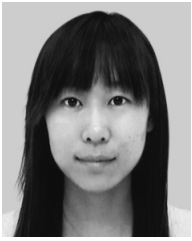
田中 美栄子 (正会員)

1950年生。1974年京都大学理学部卒業, 1979年名古屋大学大学院満期退学, 1983年Rochester大学博士課程修了 (Ph.D. in Physics). CCNY, SUNY, NASC, 椛山女学園大学, 宮崎大学工学部を経て, 現在, 鳥取大学大学院工学研究科情報エレクトロニクス専攻知能情報工学講座教授。主たる研究テーマは経済物理学, 複雑系科学。日本物理学会, I.E.E.E., 応用数理学会各会員。



糸井 良太

平成元年生。平成23年鳥取大学工学部知能情報工学科卒業。平成23年鳥取大学大学院工学研究科情報エレクトロニクス専攻博士後期課程入学。



楊 欣 (学生会員)

1984年生. 2009年中国吉林大学大学院工学研究科物流工学専攻修士課程修了. 2011年鳥取大学大学院工学研究科情報エレクトロニクス専攻博士後期課程入学. 乱数に関する領域に着目し, 研究している.