

ランダム行列理論を用いた乱数度測定法の  
開発とその実データへの応用

鳥取大学大学院工学研究科  
情報エレクトロニクス専攻

楊 欣

# 目 次

序章	～乱数検定ツールの研究現状および本文の構成～	1
第 1 章	ランダム行列理論	7
1.1	ランダム行列理論の紹介	7
1.2	ランダム行列理論の発展と応用	10
	第 1 章のまとめ	11
第 2 章	RMT-PCA	14
2.1	RMT-PCA の紹介	14
2.2	RMT-PCA の研究現状	14
	第 2 章のまとめ	16
第 3 章	乱数度検定法 RMT テストの提案	18
3.1	背景	18
3.2	RMT テストの定性評価法	20
3.3	RMT テストの定量評価法	22
3.3.1	moment 分析法	23
3.3.2	$\chi$ 二乗検定	24
	第 3 章のまとめ	25
第 4 章	RMT テストによる乱数度検定	28
4.1	RMT テスト と擬似乱数	28
4.2	RMT テスト と物理乱数	35
4.3	他の検定法との比較	39
	第 4 章のまとめ	43
第 5 章	RMT テストの実データへの応用	45
5.1	RMT テストにより Hash 関数安全性の判定	45
5.2	RMT テストにより株投資の研究	47
5.2.1	株価変動ランダム性	48
5.2.2	株価変動のランダム性と株投資	48
5.3	株価時系列の特徴を分析するため使った時系列の考察	62
	第 5 章のまとめ	66

第6章 結 章 .....	68
謝 辞.....	70
参考文献.....	71
付 表 .....	75
2007年の東証1分足から求めた乱数度順位(全) .....	75
2008(Jan.-Aug.)のデータによる乱数度順位(全).....	78
2009年の東証1分足から求めた乱数度順位(全) .....	81
分析対象とした銘柄 .....	84
English Summary .....	91
学会誌掲載論文リスト .....	94

# 目次

図 1.1 WIGNER 半円分布 .....	8
図 1.3 円則の固有値分布 .....	10
図 2.1 RMT-PCA により実データのランダム成分と主成分を分けられる .....	15
図 3.1 データの形 .....	20
図 3.2 RMT テスト定性評価の例 (左:合格, 右:不合格) .....	22
図 4.1 LCG の評価例 .....	29
図 4.2 MT の評価例 .....	29
図 4.3 最初の 500 個のみを集めて評価した結果 LCG(左)MT(右).....	31
図 4.4 LCG(左)と MT(右)の N=500,L=1500 の対数収益評価結果.....	33
図 4.5 LCG の対数収益評価結果:重複有り(左)と重複なし(右) .....	34
図 4.6 TOSHIBA 製の物理乱数の定性評価例.....	36
図 4.7 HITACHI 製の物理乱数の定性評価例 .....	36
図 4.8 TOKYO-ELECTRON 製の物理乱数の定性評価例.....	37
図 4.9 長さ 12 万の数列は 5%の定量評価基準に適用 .....	39
図 4.10 長さ 12 万の数列は 5%の定量評価基準に適用できない.....	39
図 4.11 シャッフル回数により K 次モーメントの誤差変化.....	42
図 4.12 シャッフル回数を増えると NIST の検定項目の合格数を増加.....	42
図 5.1 MD5 と SHA-1 の出力データ .....	46
図 5.2 MD5 (左) と SHA-1 (右) の定性評価結果.....	47
図 5.3 LCG の乱数度 (左), 株価ティックデータの乱数度 (右) .....	48
図 5.4 ティックデータの補足 .....	50
図 5.5 最適なパラメータは Q=4 .....	51
図 5.6 乱数度最高株(9504)の翌年収益は日経平均株価より安全であり, 日経平均は乱数度最低株(7201)より安全である .....	53
図 5.7 乱数度上位 5 位銘柄の翌年収益は乱数度最低株(7201)より安全である .....	54
図 5.8 乱数度最高の 9506 が乱数度最低の 8604 より翌年の対数収益が高く安全であることを示す.....	56
図 5.9 乱数度上位 5 位銘柄の内 6728 以外の銘柄の翌年収益は乱数度最低株(7201)より安全である.....	57
図 5.10 乱数度最高株 9508 が最低株 8058 より安全性が高い.....	58
図 5.11 乱数度上位 5 位銘柄の翌年収益は乱数度最低株(8058)より良いである.....	58
図 5.12 乱数度前 5 位 (左) と乱数度下 5 位 (右) の経年変化.....	60
図 5.13 日経平均株価(2007-2009).....	61
図 5.14 対数収益時系列の RMT テストの定性評価結果(6503: 左, 6841: 右).....	63
図 5.15 アップ/ダウン時系列の RMT テストの定性評価結果(6503: 左, 6841: 右).....	64

# 表目次

表 3.1	現存乱数度検定法の制限.....	19
表 3.2	NIST と DIEHARD 採用された検定項目 .....	19
表 4.1	SEED=1-100 の誤差の平均値と標準偏差 .....	30
表 4.2	MT と M-P 理論分布の度数表の例 .....	30
表 4.3	最初 500 個のみを集めた初期乱数の定量評価.....	32
表 4.4	変化率を取って乱数度を下げた場合の評価結果 LCG(左)MT(右).....	33
表 4.5	対数収益の分布範囲と理論範囲の比較 LCG(左)と MT(右).....	33
表 4.6	対数収益を取ったデータの重複部分を削除したデータの評価結果(Q=3) .....	34
表 4.7	MT_LOG と M-P 理論分布の度数表の例 .....	35
表 4.8	物理乱数と擬似乱数の乱数度比較(N=500, Q=3) .....	37
表 4.9	物理乱数と擬似乱数の乱数度比較(N=500, Q=6) .....	38
表 4.10	50 回(左)と 500 回(右)繰り返し誤差変化範囲(N=500, L=1500).....	38
表 4.11	対数収益を取ったデータの NIST 乱数評価法での評価結果.....	40
表 4.12	擬似乱数(LCG)及び物理乱数 3 種類の NIST 検定結果.....	41
表 4.13	物理乱数の対数収益を取った数列の NIST 検定結果 .....	41
表 5.1	MD5 と SHA-1 の比較 .....	46
表 5.2	MD5 と SHA-1 の出力データのランダム性.....	47
表 5.3	毎年の分析対象 .....	50
表 5.4	2007 年の東証 1 分足から求めた乱数度順位.....	52
表 5.5	2008 年の東証 1 分足から求めた乱数度順位.....	55
表 5.6	2008(JAN.-AUG.)のデータによる乱数度順位.....	55
表 5.7	2009 年の東証 1 分足から求めた乱数度順位.....	57
表 5.8	関連株分類 .....	59
表 5.9	分類された株の特徴.....	60
表 5.10	TOPIX の対前日騰落率の下落記録(-2010 年末).....	62
表 5.11	株価ティックデータの「株価時系列」の例(1 分毎).....	63
表 5.12	‘表 5.11’の例により作成した「アップ/ダウン時系列」の例.....	63
表 5.13	‘表 5.11’の例により作成した「株価の変動量時系列」の例.....	64
表 5.14	株価の対数収益時系列と変動量時系列の RMT テストの評価結果は一致である .....	65

# 序章

## ～乱数検定ツールの研究現状および本文の構成～

乱数とは規則性のない予測不可能な数である。出現する数の統計的な特徴により一様乱数や正規乱数などの種類がある。乱数を発生させる方法には、物理的手段で発生させる方法(物理乱数)と算術式により発生させる方法(擬似乱数)がある。物理乱数とはランダムな物理現象や特定のハードウェアを利用するもので、無相関で周期性が無く発生確率が均等である等の特徴がある。擬似乱数とは一定の算術式を用いて発生させるもので、無相関性、長周期性、非線形性（予測不可能性）などの性質が要求される[1]。コンピュータを使った乱数発生には普通、線形合同法などの擬似乱数発生アルゴリズムが用いられるが、このような擬似乱数には一般に周期が存在するため、完全な乱数とは言えない。その意味で一番良い乱数は自然に発生するノイズである。乱数のランダム性に要求される精度のレベルは目的毎に異なり、該当分野に十分なランダム性が有るか否かを判断する様々のランダム性検定法が提案されている。

例えば、擬似乱数のランダム性を評価する検定ツールとしては日本の JIS コード[2]があり、シミュレーション用の乱数に対する検定法としてはオーストリアで開発された PLAB [3]がある。またスイスで開発された ENT[4]は乱数生成器や圧縮アルゴリズムなどの評価法として良く使われるものである。一方、暗号分野では、鍵生成器から出力を検定したり、ブロック暗号からの出力を検定する評価法としてカナダで開発された Crypt-X[5]やアメリカで開発された NIST-SP 800-22[6] と DIEHARD[7]が良く知られている。

しかし、これらの殆どに共通する問題として、ツールごとに採用している検定の種類や数が異なり、使いにくいという難点がある。例えば、伏見著の“乱数”[8]には、7種類の検定項目が採用されているが、日本工業規格(JIS)による乱数の統計検定ツールでは9種類であり、米国商務省標準局 NIST(National Institute of Standards and Technology)が開発した乱数検定法 NIST Special Publication 800-22 では 16 種類、NSF により開発された DIEHARD 検定法は 18 種類の検定方法を採用している。このうち暗号用乱数を検定する際によく使われる NIST-SP 800-22 と DIEHARD を例にとると、検定される乱数の形式制限があり、バイナリ形式のデータと 0 と 1 からなる ASCII 形などしか検定できない。一回に検定する乱数列長にも制限があって自由に設定できない。NIST-SP 800-22 は長さ 100 万の、0 と 1 からなる一様

乱数を検定するために下記の 16 個の検定法[6]

1. 1次元度数検定
2. ブロック単位の頻度検定
3. 連の検定
4. ブロック単位の最長連検定
5. 2値行列ランク検定
6. 離散フーリエ変換検定
7. 重なりのないテンプレート適合度検定
8. 重なりのあるテンプレート適合度検定
9. Mauren のユニバーサル統計検定
10. Lempel-Ziv 圧縮検定
11. 線形複雑度検定
12. 系列検定
13. 近似エントロピー検定
14. 累積和検定
15. ランダムエクスカーション検定
16. 変形ランダムエクスカーション検定

を用いるよう設定されているし、DIEHARD は長さ 80Mbit から 88Mbit を対象にしたものであり、これより短い数列には適用できないなどの強い制限がついている。

一方、カナダで開発された Pierre L'Ecuyer and Richard Simard による一連の乱数検定プログラム TestU01 は、各種の統計的検定がグループにまとめられているが、テスト時間が長いという欠点があり、例えばメルセンヌツイスターで発生させた擬似乱数の場合は 15 時間以上もかかってしまう [9]。梅野等による Ransure [10]は、特定の暗号アルゴリズムに依存しない自動暗号評価テストであり、暗号、金融工学、ゲームなど乱数を用いるアプリケーション全てに対して適用可能なことや幅広い応用範囲を持つなどの利点があるが、NIST と同様、15 個程度の検定項目がある上、高価(52 万)なのが難点である。

また、統計的な乱数検定は、真にランダムな系列の持つべき種々の性質の一部を保障するのみであり、各種検定に合格しただけでは検定対象の乱数列の性質を全てクリアしたとは言えないことに注意すべきである。例えば、暗号で用いる乱数列を評価する場合は統計

的な評価だけでは不十分であり、乱数生成アルゴリズムの理論的な評価も同時に行う必要がある[11].

以上のような背景の中で我々が提案し開発してきた『RMT テスト』は、ランダム行列理論で導かれている公式を基準として用いることにより、時系列のランダム性を測定する方法である。近年のデジタル化により、気象、医療、金融、国勢調査など様々な分野で膨大なデータの取得・保存が可能になったことで大容量データが蓄積されている情勢において、本手法はビッグデータ解析向きの単純で有効な乱数度計測法として広範囲の応用が見込める有用な手法である。本手法は Ransure のように複雑な検定項目を持たず、理論基礎が明確で単一の評価基準を持つ上、多様なデータ形式に対応でき、必ずしも乱数度の高くない実データにも適用可能なため、従来、複雑系として研究されてきた分野のほぼ全域に対して利用可能性が期待される。

RMT テストの理論基礎として、ランダム行列理論 (Random Matrix Theory: RMT と略) を用いた主成分分析法 (Principal Component Analysis: PCA と略)、RMT-PCA があり、膨大なデータ群に対して相関の大きい成分とランダム成分を分ける手法として注目を集めている[12-18]。RMT テストは時系列の無相関性を用いてそのランダムネスを測るもので、理論的に明快なのが利点である。以下に RMT テスト[19-23]を概説する。時系列から計算した相関行列の固有値分布と、対応する RMT 理論式との類似度によって測定する。まず、「定性評価」として、与えられた時系列の相関行列の固有値分布を、理想的な乱数列に対して成立する RMT から導かれる分布式と比較することで直感的に表現できる形に定式化したうえで、乱数度の近いデータの比較を数値的に行えるようにモーメントの差を利用する「定量評価」を用いることで実現する。

RMT テストは、ランダム性が十分高く良い乱数と見做せる時系列間の乱数度を比較出来るだけでなく、乱数とは見做せないほど乱数度の低い時系列に対しても同様に適用できてその乱雑さの度合いを可視化・数値化できる点において、多方面の分野への応用を期待でき、多分野で蓄積されつつある膨大かつ様々な形式のデータへの応用が考えられる。

従来の乱数度検定法と比較して、ここで提案する RMT テストの利点は三点ある。一点目は可視化手法であること、即ち、乱数度が非常に低い数列に対しては、RMT テストの定性評価を用いることで、直感的に見てすぐ判断できる検定法であり、複数の様々な検定を併用する必要がない。二点目は、データ型を問わない点であり、従来の乱数度検定アルゴリ



ズム、米国の NIST などのように、対象とするデータの長さやデータ型（実数・整数・二進数など）に強い制限がなく、実数・整数・二進数などと同じアルゴリズムで使える。本論文では、擬似乱数や物理乱数の実数列や 01 数列などを含む大量のデータに適用することにより RMT テストの有効性を検証した。三点目として、RMT テストは乱数度の高い時系列ばかりでなく、定量評価結果の大小により、乱数度の低い実データのランダム性の差異をも比較できる。

本論文ではこのような 3 大利点を生かして実データへの応用を行った結果を、ハッシュ関数の安全性判定と高収益株の抽出の 2 例について詳説する。暗号学的ハッシュ関数安全性を確保するためには、ハッシュ値の無規則性、予測不可などランダムな性質が必要である。RMT テストを用いて、よく使われる暗号学的ハッシュ関数 MD5 と SHA-1 のランダム性を比較検定し、SHA-1 の方が MD5 に比べてランダム性が高い、つまり、安全性が高いということを実証した。また、高収益株の抽出への応用については、2007 年から 2009 年にかけての東証株価の各株価をデータとして用い、高頻度株価変動の乱数度を測定することで、各株価の乱数度とその株式の収益率の推移との間に一定の関連性が見られることを見出した。それは、「乱数度の高い株の方の収益性が高い」という経験則で表現できる。これは実験で使った 2007 年から 2009 年にかけて東証株価が下がり相場であったという特殊な条件にのみ適用できるものである可能性もあるが、少なくとも株式市場が下がり相場になる際は、乱数度が高い株に投資すると収益が高いことをデータが示していると言える。この仮説は今後多くのデータを用いて検証することにより、株選定の指標として乱数度を利用できる可能性を開く端緒になると考えられる。

以下に本論文の構成を述べる。第 1 章は RMT テストの基礎となる、ランダム行列理論について簡単にまとめる。ランダム行列理論の主要な型の中から、時系列の自己相関行列の固有値分布である Maruchenko Pastur 分布(M-P 分布と略)を重点的に記述する。これは RMT テストの理論基礎として重要である。引き続いて、第 2 章で RMT テストの原型となる主成分分析手法、RMT-PCA について述べる。このような理論的基礎に基づいて提案された RMT テストの「定性評価法」および「定量評価法」について第 3 章で詳説する。第 4 章は実数列または  $\{0,1\}$  で作成した整数列の擬似乱数と物理乱数を RMT テストにより検定し、RMT テストの有用性を検証する。第 5 章では RMT テストを実データへ応用し、ハッシュ関数の安全性判定と高収益株の抽出について述べる。

注釈及び参考文献：

- [1] 乱数について：<http://sehermitage.web.fc2.com/crypto/random.html>
- [2] JIS:乱数発生及びランダム化の手順，JIS Z 9031(2010),JSA Web Store order No.911023999
- [3] L. Hannes: PLAB a system for testing random numbers, Proceedings of the International Workshop Parallel Numerics'94, pp.89-99 (1994)
- [4] ENT: A pseudorandom number sequence test program(online), <http://www.fourmilab.ch/random/>
- [5] The Information Security Research Centre at Queensland University of Technology, "Crypt-X," (<http://www.isrc.qut.edu.au/resource/cryptx/>)
- [6] NIST, Special Publication 800-22, "A Statistical Test Suite For Random and Pseudorandom Number Generators for Cryptographic Applications," 2001
- [7] Dieharder: A random number test suite version 3.31.1 (online),<http://www.phy.duke.edu/rgb/General/dieharder.php>.
- [8] 伏見正則, "乱数," 東京大学出版
- [9] 谷口礼偉, 64 ビット整数の乗算とシフトによる非再帰的な擬似乱数の生成と検定, 統計数理研究所共同研究レポート 210, 乱数の応用指向性評価とその周辺 (2008)
- [10] 梅野 健, 決定論的カオスと「RanSure」による乱数の特性評価, 統計数理研究所共同研究レポート 210, 乱数の応用指向性評価とその周辺 (2008)
- [11] 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」(2003)
- [12] M.L. Mehta: Random Matrices,3rd edition, Academic Press (2004)
- [13] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral and H. E. Stanley: Universal and Non-Universal Properties of Cross-Correlations in Financial Time Series, Physical Review Letter, Vol.83, 1471-1474 (1999)
- [14] A.M. Sengupta and P.P.Mitra: Distribution of singular values for some random matrices, Physical Review E, Vol.60, p.3389 (1999)
- [15] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral and H.E. Stanley: Random matrix approach to cross correlations in financial data, Physical Review E, p.066126 (2002)
- [16] L. Laloux, P. Cizeaux, J.-P. Bouchaud and M. Potters: Physical Review Letters, Vol.83, pp.1467-1470 (1999)
- [17] 山本 敦史 , 木戸 丈剛 , 田中 美栄子 , 高石 哲也: 株式市場の主成分追跡～RMT-PCA による四半期間の主要業種抽出～, 情報処理学会研究報告：数理モデル化と問題解決(MPS), 2012-MPS-88(17)
- [18] 村上和正, 安本典生, 家富洋：ランダム行列理論を用いた気象データ相関抽出とその有効性の検討, 素粒子論研究, Vol.117, No.5,pp.116-117 (2009)
- [19] 楊 欣, 糸井 良太, 田中 美栄子：RMT 公式を用いた乱数度評価法の提案, 情報処

理学会研究報告：数理モデル化と問題解決(MPS), 2011-MPS-83(2)

[20] X. Yang, R. Itoi and M. Tanaka-Yamawaki: Testing randomness by means of RMT formula, Intelligent Decision Technologies, Smart Innovation, Systems and Technologies, Vol.10, pp. 589-596 (2011)

[21] X. Yang and M. Tanaka-Yamawaki: Testing randomness by means of Random Matrix Theory, 2011 Kyoto Workshop on NOLTA, pp.1-1 (2011)

[22] X. Yang, R. Itoi and M. Tanaka-Yamawaki: Testing randomness by means of Random Matrix Theory, Progress of Theoretical Physics Supplement, No.194, pp. 73-83 (2012)

[23] M. Tanaka-Yamawaki, X. Yang and R. Itoi: Moment approach for quantitative evaluation of randomness based on RMT formula, Intelligent Decision Thechnologies, Smart Innovation, Systems and Technologies, Vol.16, pp. 423-432 (2012)

# 第1章 ランダム行列理論

まず, RMT テストの理論基礎の一つランダム行列理論を紹介し, 行列サイズを非常に大きくしていった場合, 行列の種類により, 固有値分布は, Wigner の半円則, 円則, Tracy Widom 則, Marchenko-Pastur 則などに従う分布となる. RMT テストは Marchenko-Pastur 則を採用する.

## 1.1 ランダム行列理論の紹介

ランダム行列(Random Matrix)とは, 行列要素  $A_{i,j}$  がなんらかの確率法則あるいは確率分布に従う確率変数として与えられると仮定する行列モデルである[1]. ランダム行列は, 1928年に統計学者 John Wishart [2]により多変量の共分散を求める行列である  $X X^T$  (または  $X X^*$ ) により構成されるウィシャート行列が導入され, 1950年代に Eugene Wignerにより,  $N \times N$  実対称行列 (あるいはエルミート行列) の固有値や固有値の間隔の分布の統計的性質, それらの普遍性やその要因などの研究がおこなわれ, 行列サイズを非常に大きくしていった場合の固有値の同時確率密度関数の極限分布や, 最大固有値と最小固有値の極限分布などが導かれた. 行列の種類により, 固有値分布は, Wigner の半円則, 円則, Tracy Widom 則, Marchenko-Pastur 則などに従う分布となる. また, 固有値間の間隔分布についても一般則が知られている[3-7].

### 1) Wigner 半円則

Wigner 半円分布は物理学者である Eugene Wigner になんで名付けられた分布である. データから作成した行列の大きさが無限大に近づくに連れ, 固有値分布の極限分布として現れる.

$$\rho_{sc}(\lambda) = \frac{1}{2\pi R^2} \sqrt{4R^2 - \lambda^2} \quad (1.1)$$

ここで,  $\lambda$  は固有値式(1.1)の適応範囲は  $-R < \lambda < R$ ,  $R^2$  はウィグナー行列の非対角要素の分散  $R^2 = E(|A_{i,j}|^2) (i \neq j)$ , wigner 分布の形は図 1.1 に示す.

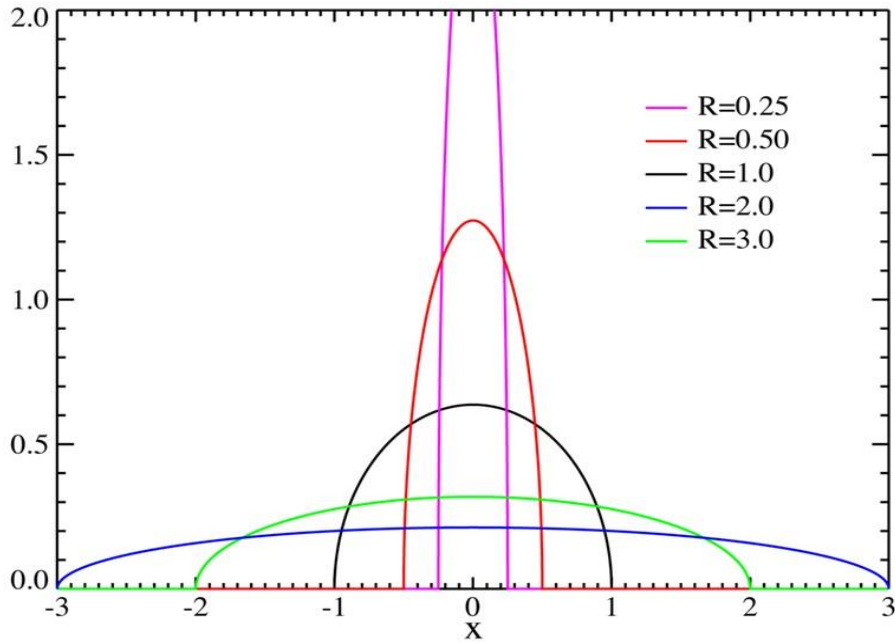


図 1.1 Wigner 半円分布

## 2) Marchenko-Pastur 則

長い時系列を同じ長さ  $L$  で切り分け、 $N$  個のデータを作成でき、データの  $L \times N$  行列により作成した  $N \times N$  の自己相関行列の要素を平均 0 分散 1 のように規格化し、次元  $N$  と時系列長  $L$  が無限大の極限で、固有値分布はその比  $Q=L/N$  のみにより表される簡単な関数式(1.2)となる。式(1.3)は式(1.2)の固有値  $\lambda$  の最大最小値を示し、Marchenko-Pastur(M-P と略)分布は図1.2に示す。

$$P_{RMT} = \frac{Q}{2\pi\lambda} \sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)} \quad (1.2)$$

$$\lambda_{\pm} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}} \quad (1.3)$$

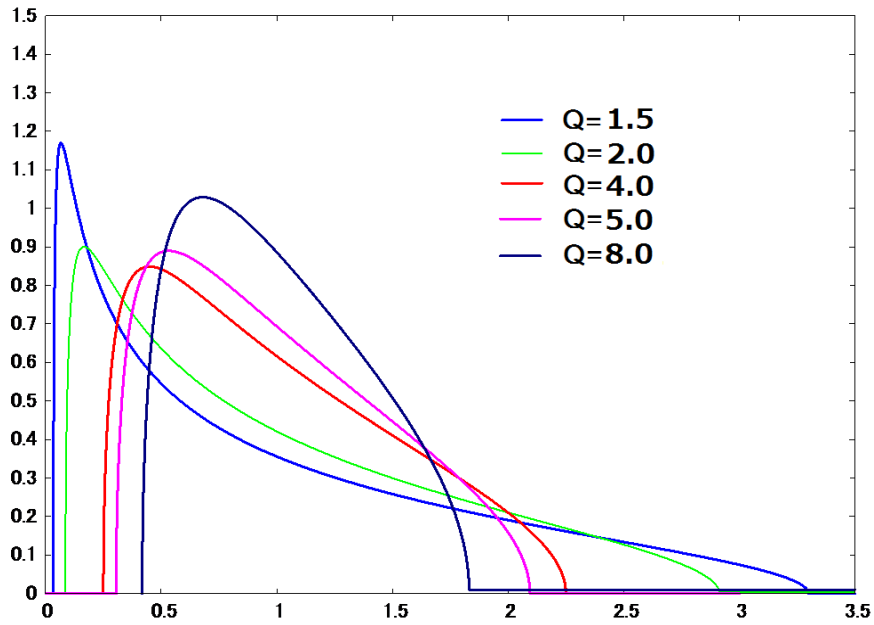


図 1.2 Maruchenko-Pastur 分布図

### 3) 円則

半円則の行列要素を平均 0 分散  $1/\sqrt{n}$  のように規格化すると、行列のサイズを非常に大きくしていく  $n \rightarrow \infty$  に従い固有値は複素平面上の単位円盤上で一様に分布するようになるというもの、図 1.3 に示す。円則と半円則、M-P 分布の区別が円則のはベルヌーイ・アンサンブルやジニブル・アンサンブルなどであり、固有値が実数分析対象となる複素平面の実軸上にのみ固有値が分布し、円則は当てはまらない。

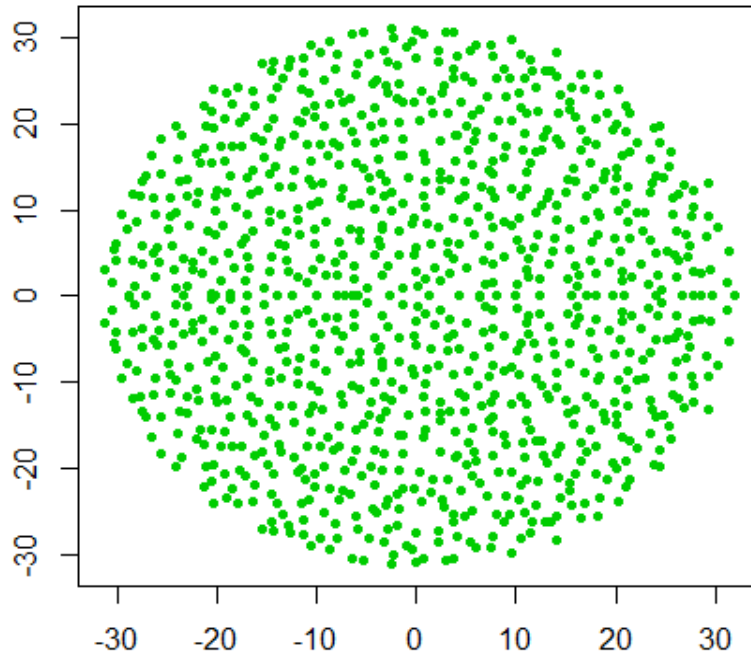


図 1.3 円則の固有値分布

#### 4) Tracy–Widom 分布

ランダム行列理論におけるガウシアン・アンサンブルの最大固有値分布, Tracy-Widom 分布 [8,9]と一致している. Tracy-Widom 分布には  $N \times N$  行列要素を平均 1, 分散  $N/2$  のように規格化し, アンサンブルの種類によって何種か存在するがある. 他の三種類により複雑である.

本稿では, 実データの固有値分布とランダム行列理論の理論分布と比較することを利用するため, そのうちに, 使いやすさを考え, 形状が一番区別しやすいのは半円則と M-P 則である. また, 図 1.1 と図 1.2 を比較すると, 図 1.1 の  $R$  をちょっと変化しても形が全く異なり, 図 1.2 は  $Q$  の増加に沿って分布曲線の形状はあまり変わらない, 1 に向いて収束していく. そこで, 本稿では, ランダム行列理論の一番使いやすいの M-P 分布を使用する.

## 1.2 ランダム行列理論の発展と応用

ランダム行列理論はメゾスコピック系の物理学や量子重力[11,12],統計力学[13], 生物学や医学[14-16], 経済物理学[17,18]を含む様々な分野へ広範な応用を持つ普遍性の高い理論である. 近年, デジタル化が進み膨大なデータの取得・保存が可能になった. 大容量データを様々な目的に則して解析するための, 有効な手法としてランダム行列理論が今, 広範囲

の分野で求められている。地震時系列[19]，気象データ[20]，金融データ[21,22]，人口分布調査[23]，交通状況管理[24]など，次々と研究対象を広げている。特に，株式市場で取引される株価銘柄間の同時刻相関行列の固有値問題に適用してトレンド業種を追う手法は，日次終値を用いた良好な結果[24,25]に触発されて更に高頻度の tick 価格時系列へと応用を広げている[26]。

## 第 1 章のまとめ

本章では RMT テストの理論基礎のその 1 であるランダム行列理論について述べた。ランダム行列理論の 4 つの主要則を紹介し，RMT テスト Wishart 型行列である相関行列の固有値分布である Marchenko-Pastur 則を採用する理由を説明した。



## 注釈及び参考文献：

- [1] 永尾 太郎：ランダム行列の基礎，東京大学出版会(2005)
- [2] E.P. Wigner: Ann. Math., Vol. 67, pp. 325-327 (1958)
- [3] A. Edelman and N.R. Rao: Random Matrix Theory, Acta Numerica, pp. 1-65. Cambridge University Press, Cambridge (2005)
- [4] G. Akemann, G.M. Cicutta and L. Molinari: Compact support probability distributions in random matrix theory, Phys Rev E Stat Nonlinear Soft Matter Phys, Vol.59,No.2, pp.1489-1497(1999)
- [5] A. ZEE: Law of addition in random matrix theory, Nucl Phys B, Vol.474, No.3, pp.726-744(1996)
- [6] R.E. Prange: Quasiclassical Random Matrix Theory, Phys Rev Lett, Vol.77, No.12, pp.2447-2450 (1996)
- [7] A. Lakshminarayan, N.R. Cerruti, and S. Tomsovic: Classical diffusion and quantum level velocities: Systematic deviations from random matrix theory, Phys. Rev. E Vol.60, No.4, pp. 3992–3999 (1999)
- [8] C. A. Tracy and H. Widom: Level-spacing distributions and the Airy kernel, Comm. Math. Phys. Vol.159, No.1, pp.151-174 (1994)
- [9] C. A. Tracy and H. Widom: On orthogonal and symplectic matrix ensembles, Comm. Math. Phys. Vol.177, pp.727-754 (1996)
- [10] V. Dotsenko: Universal randomness, Phys.-Usp, Vol.54, No.3, 259, doi:10.3367/UFNe.0181.201103b.0269 (2011)
- [11] 金澤拓也，クォーク物質とランダム行列理論，原子核研究，Vol.55, No.2, pp.22-33 (2011)
- [12] E. Brezin and S. Hikami: Spectral form factor in a random matrix theory, Phys Rev E Stat Nonlinear Soft Matter Phys, Vol.55, No.4, pp.4067-4083 (1997)
- [13] B. Dietz and F. Haake: Random matrix theory as statistical mechanics, Euro phys Lett, Vol.9, No.1, pp.1-6 (1989)
- [14] F. Luo: Application of random matrix theory to biological networks, Phys Lett A, Vol.357, No.6, pp.420-423 (2006)
- [15] P.L. Yang, R. Li and J. Chen: Application of random matrix theory to microarray data for discovering functional gene Modules of hepatocellular carcinoma, Shengwu Wuli Xuebao, Vol.25, No.3, pp.192-202 (2009)
- [16] R. Li, P.L. Yan and J. Chen: Application of random matrix theory to identification of lung cancer gene networks, Wuli Xuebao, Vol.58, No.10, pp.6703-6708 (2009)
- [17] 相馬 亘: 経済物理学の実用化に向けて—ランダム行列理論からのアプローチ，応用数理，Vol.15, No.3, pp.239-253 (2005)
- [18] 家富洋: “経済物理学”景気循環の物理：ランダム行列理論と揺動散逸定理，パリティ，Vol.27,

No.6, pp.13-16 (2012)

- [19] 新井優太, 家富洋, 相馬亘: ランダム行列理論を用いた Hi-net 地震時系列データの解析, 日本物理学会講演概要集, 巻 66, 号 2 (2011)
- [20] 村上和正, 安本典生, 家富洋: ランダム行列理論を用いた気象データ相関抽出とその有効性の検討, 素粒子論研究, Vol.117, No.5, pp.116-117 (2009)
- [21] J. Daly, M. Crane and H.J. Ruskin: Random matrix theory filters in portfolio optimisation: A stability and risk assessment, *Phys A Stat Mech Appl*, Vol.387, No.16-17, pp.4248-4260 (2008)
- [22] Y. Nakayama and H. Iyetomi: Random matrix theory of dynamical cross correlations in financial data, *Prog Theor Phys Suppl*, Vol.179, 60-70 (2009)
- [23] Z.X. Xu, Y. Wang and H.B. Si: Analysis of urban human mobility behavior based on random matrix theory, *Wuli xuebao*, Vol.60, No.4 pp.040501.1-7 (2011)
- [24] J. Liu, D.P. Jin, W.Z. Zhang: Monitoring of link patters based on random matrix theory, *Qinghua Daxue Xuebao(Ziranxueban)*, Vol.50, No.1, pp.117-120 (2010)
- [24] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral and H.E. Stanley: Random matrix approach to cross correlations in \_nancial data, *Phys Rev E*, p.066126 (2002)
- [25] L. Laloux, P. Cizeaux, J.-P. Bouchaud and M. Potters: *Phys Rev Lett*, Vol.83, pp.1467-1470 (1999)
- [26] 山本 敦史, 木戸 丈剛, 田中 美栄子, 高石 哲也: 株式市場の主成分追跡~RMT-PCA による四半期間の主要業種抽出~, 研究報告数理モデル化と問題解決(MPS), 2012-MPS-88(17)

## 第2章 RMT-PCA

### 2.1 RMT-PCA の紹介

主成分分析法[1]は多くの変量が存在する場合、それら変量の相関構造から主成分を抜き出し、2次統計量である共分散行列の大部分をより少数の因子で理解することにある。主成分分析は多くの分野で用いられていて、直感的に分かりやすく有力な手法である。しかし、古くから指摘されている問題の1つは、主成分の抽出基準である。ランダム行列の固有値分布を知るには、モンテカルロ計算を繰り返す必要がある。そのため、大規模な相関行列を解析する場合や、大きさの異なる相関行列を頻繁に解析する場合には有用性を欠く。したがって、ランダム行列の固有値が従う分布関数を解析的に求められれば、それらの欠点を克服でき、主成分分析の発展に大きく貢献できる[2]。

1967年に Marcenko と Pastur は、ランダム行列の固有値が従う分布関数を解析的に導いた[3]。しかし、この結果がすぐに主成分分析に応用されることは無く、1999年になってようやく、Laloux 等[4]と Plerou 等[6]によって別々に、ランダム行列理論 (Random Matrix Theory: RMT と略) を用いた主成分分析法 (Principal Component Analysis: PCA と略)、RMT-PCA[9]を株価相関の研究に対して応用された。具体的にこれらの研究では、ニューヨーク証券取引所の S&P500 銘柄が作る相関行列の固有値分布とランダム行列理論から導かれた理論式を比較し、これら2つの分布がずれる部分に主成分があり、分布が重なる部分の成分はノイズであることを明らかにした。これらの論文を契機に、海外の多くの研究者が、株価相関の研究にランダム行列理論を応用した論文を発表している。

### 2.2 RMT-PCA の研究現状

Plerou 等は RMT-PCA を米国の S&P500 に応用し[6,7]、青山等[8]はその後日本の東証データについて、いずれも6-7年分をまとめて1データとして同手法に基づく解析を行うことにより、長い株価データを対数収益列に整形し、数列のランダム成分と主成分を分けることができた。我が研究室 (鳥取大学知能情報工学科の田中研究室) では、同じ手法で RMT-PCA を株価日中変動 (tick データ) の相関行列を解析した[9-12]。長期的な投資には長期的な、短期投資には短期的なデータの分析が必須であるため、6-7年分を1データとした長期変動の解析に対し、我々は2007年から2009年の3年間の3ヶ月毎を1データとした短期変動

の分析を, 手持ちの日本株式市場の東証 TOPIX500 の 2009 年 1-3 月の対数時系列に RMT-PCA を適用して, 相関行列の固有値を計算し, 固有値分布を求める. これを実データ固有値ヒストグラムとして図 2.1 に赤線で示す. これに重ねて第 1 章で述べた実データと同じ  $Q$  を持つ  $M-P$  の理論分布を緑線で描き両者を比較する. その結果, ランダム行列理論より導かれた式(1.3)の範囲に収まる固有値をランダム成分とみなし[13], 最大固有値 $\lambda_+$ より大きい成分はランダム成分ではないという意味で主成分とみなすのである. これは特にランダム性の高い株式市場価格の変動からランダム成分を取り除き, 特定の期間に, 業種単位で連動している銘柄の抽出(主成分)を行える点で有効性が高い. 株式市場における価格変動の予測を直接に行うよりも, 連動している銘柄群を抽出することで, 「A 社の株を売買する」とまで銘柄を特定できない場合でも, 「〇〇業種」と業種を選ぶ指標とできる.

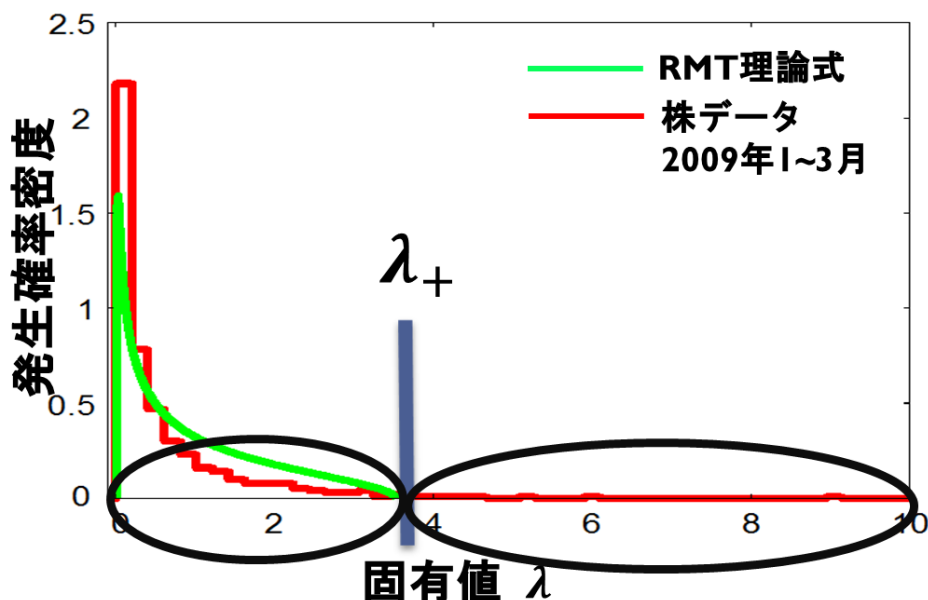


図 2.1 RMT-PCA により実データのランダム成分と主成分を分けられる

## 第 2 章のまとめ

本章では RMT テストの理論基礎の第 2 として, RMT-PCA, すなわちランダム行列理論 (Random Matrix Theory: RMT と略) を用いた主成分分析法 (Principal Component Analysis: PCA と略) 手法について簡略に述べた. これはビッグ・データ解析に於いて第一義的に有用なものであり, 膨大な量が日々蓄積されつつあるデータから相関の大きい成分をまず取り出すために, 金融分野における株価分析をはじめとして, 今や広範囲の分野に応用されつつある. 本稿では, plerou 等により 2002 年に株式市場に応用された文脈に基づき, 時系列の相関行列の固有値分布をランダム行列理論式と比較する方法を採用する. この方法について次の章で詳細に説明する.

## 注釈及び参考文献：

- [1] S. Wold, K. Esbensen, P. Geladi: Principal component analysis, *Chemometr Intell Lab Syst*, Vol.2, No.1-3, pp.37-52 (1987)
- [2] 相馬亘：ランダム行列理論に基づいた新たな相関分析手法，日本大学理工学部学術講演会予稿集， Vol.56, O-23, pp.1287-1288 (2009)
- [3] V.A. Marcenko and L.A. Pastur: Distribution of eigenvalues for some sets of random matrices, *Mathematics of the USSR-Sbornik*, Vol.1-(4), pp. 457-483(1994)
- [4] L. Laloux, P. Cizeaux, J. Bouchaud and M. Potters: Noise dressing of financial correlation matrices, *Phys Rev Lett*, Vol. 83 , pp.1467-1470 (1998)
- [5] M.O. Ulfarsson and V. Solo: Dimension estimation in noisy PCA with SURE and random matrix theory, *IEEE Trans Signal Process*, Vol.56, No.12, pp.5804-5816 (2008)
- [6] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral and H.E. Stanley: *Physical Review Letters*, Vol. 83, pp. 1471-1474 (1999)
- [7] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral and H.E. Stanley: Random matrix approach to cross correlation in financial data, *Phys Rev E*, Vol. 65 no.066126 (2002)
- [8] 青山秀明，他4名：経済物理学，共立出版，第5章 (2008)
- [9] 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，電子情報通信学会・情報処理学会，pp.153-156 (2010)
- [10] M. Tanaka-Yamawaki: Cross Correlation of Intra-day Stock Prices in Comparison to Random Matrix Theory, *Intelligent Information Management*, Vol. 3, pp.65-70 (2011)
- [11] M. Tanaka-Yamawaki, T. Kido and R. Itoi: Trend Extraction of Stock Prices in the American Market by Means of RMT-PCA, J. Watada et. al. (Eds.) , *Intelligent Decision Technologies*, SIST Vol. 10, pp. 637-646, doi:10.1007/973-642-22194-1 (2011)
- [12] 木戸丈剛，楊 欣，田中美栄子，高石哲弥:ランダム行列理論を用いた主成分抽出法による日本と米国の株式市場における主要セクタの変遷, *情報処理学会論文誌数理モデル化と応用* 4 卷 4 号, pp.104-110 (2011)
- [13] 山本敦史，木戸丈剛，田中美栄子，高石哲也:株式市場の主成分追跡～RMT-PCA による四半期間の主要業種抽出～，*研究報告数理モデル化と問題解決(MPS)*, 2012-MPS-88(17)

## 第3章 乱数度検定法 RMT テストの提案

### 3.1 背景

乱数[1]とは規則性のない予測不可能な数である。出現する数の統計的な特徴により一様乱数・正規乱数などの種類がある。通常、コンピューターでは完全な乱数ではなく、必要な範囲内で乱数と見なせる擬似乱数を用いる。よって、擬似乱数のランダム性を評価するため、乱数検定ルール JIS や、シミュレーション用の乱数に対する検定法である PLAB が提案された。また、乱数は情報セキュリティシステムにおいて、各種のプロトコルや秘密鍵の種として用いられ、その安全性は、システムの安全性に影響を持つ。乱数の安全性として、過去の乱数系列から未来の乱数系列を推定できないこと（前方予測不可能性）及び、乱数の種を推定できないこと（後方予測不可能性）が要求される。これらの予測不可能性を、発生した乱数系列から計算される統計量のランダム性を検定して判断する方法が乱数検定法である。乱数検定法で合格することは、安全な乱数であるための必要条件である。そこで、乱数生成器や圧縮アルゴリズムなどの評価法である ENT、暗号分野で利用される NIST-SP 800-22 や DIEHARD などが提案された[2-5]。しかし、ツールごとに採用している検定の種類や数が異なり、使いにくいという難点がある。

暗号用乱数を検定する際によく使われる NIST-SP 800-22 と DIEHARD を例にとると、表 3.1 に示すように、検定される乱数の形式制限があり、バイナリ形式のデータと 0 と 1 からなる ASCII 形などしか検定できない。一回検定する乱数列長も自由に設定できない、NIST は長さ 100 万の数列、DIEHARD は長さ 80Mbit から 88Mbit が必要であり、これより短い数列には適用できない。また、検定項目についても、表 3.2 に示すように NIST の検定項目は 16 個、DIEHARD は 18 個項目が採用されており、各項目がどのような比重で必要となるのかの理由も理論的に理解し難い。

表 3.1 現存乱数度検定法の制限

	NIST SP 800-22	DIEHARD
形式制限	バイナリ形式のデータと 0 と 1 からなる ASCII 形	0 と 1 からなる乱数列
乱数列長	1Mbit	80Mbit から 88Mbit 程度必要
項目個数	16 項目	18 項目

そこで、形式制限が弱い01形式と実数形式共に応用でき、ランダム性の評価基準が単一で、理論的にも理解しやすい乱数検定法が求められる。そのような乱数検定法として、ランダム行列理論に基づく主成分分析法(RMT-PCA)をベースにした「RMTテスト」を提案、検証し、実データへの応用を試行した結果を報告するのが本論文の主眼である。

表 3.2 NIST と DIEHARD 採用された検定項目

	NIST SP 800-22	DIEHARD
1	1 次元度数検定	バースデイ空間検定
2	ブロック単位の頻度検定	OPERM5 検定
3	連の検定	(31x31) の 2 値行列ランク検定
4	ブロック単位の最長連検定	(32x32) の 2 値行列ランク検定
5	2 値行列ランク検定	(6x8) の 2 値行列ランク検定
6	離散フーリエ変換検定	ビット列検定
7	重なりのないテンプレート適合検定	OPSO 検定
8	重なりのあるテンプレート適合検定	OQSO 検定
9	Maurer のユニバーサル統計検定	DNA 検定
10	Lempel-Ziv 圧縮検定	8 ビット中の文字数検定
11	線形複雑度検定	特定位置の 8 ビット中の文字数検定
12	系列検定	駐車場検定
13	近似エントロピー検定	最小距離検定
14	累積和検定	3DSPHERES 検定
15	ランダム偏差検定	スクイーズ検定



16	種々のランダム偏差検定	重なりのある和検定
17		連の検定
18		クラップス検定

### 3.2 RMT テストの定性評価法

本稿で用いるのはPlerou等[6,7]により株式市場に応用された文脈に基づき、時系列の相関行列の固有値分布をランダム行列理論式と比較する方法である。本稿では、時系列のランダム性を「乱数度」と定義した[8]。以下に手法を概説する。データとなる時系列 $S$ を整数長さ $L$ の正規化した数列を $N$ 個並べて $L$ 行 $N$ 列のデータ行列 $A$ を図3.1のように作ると、それぞれの各列を式(3.1)により、平均0, 分散1に正規化する。

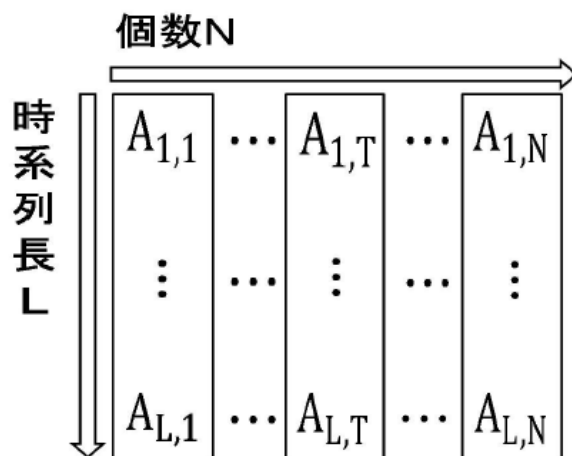


図 3.1 データの形

$$g_{i,j} = \frac{A_{(i,j)} - \langle A_j \rangle}{\sqrt{\langle A_j^2 \rangle - \langle A_j \rangle^2}} \quad (3.1)$$

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

$$G = \begin{bmatrix} g_{11} & \dots & g_{1N} \\ \vdots & \ddots & \vdots \\ g_{L1} & \dots & g_{LN} \end{bmatrix} \quad (3.2)$$

とする。

この $N$ 個の正規化時系列間の内積を成分とする相関行列 $C$ を式(3.3)により作成し、

$$C = \frac{1}{L} G^T G \quad (3.3)$$

ここで,

$$C_{i,j} = C_{j,i} \quad (3.4)$$

相関行列 $C$ の対角要素 $C_{i,i} = 1$ , 相関行列は対称行列でもある. 対称行列は直交行列 $V$ , 即ち $V^t = V^{-1}$ を満たす行列を使った相似変換 $V^{-1}CV$ により対角行列に変換できる. このような $V$ の各列は正方行列 $C$ の固有ベクトル

$$V_k = \begin{pmatrix} V_{k,1} \\ V_{k,2} \\ \vdots \\ V_{k,N} \end{pmatrix} \quad (3.5)$$

に対応し,  $C$ を掛けると元のベクトル $V_k$ に比例する.

$$CV_k = \lambda_k V_k \quad (3.6)$$

全部で $N$ 個の $k$ の値, それぞれに対して常識が成立し, 比例係数 $\lambda_k$ は固有値となる. 成分を頭わに書くと

$$\sum_{j=1}^N C_{i,j} V_{k,j} = \lambda_k V_{k,i} \quad (3.7)$$

となる. このような固有ベクトル $V_k$ は正規直交系を形成する. つまり, 各ベクトル $V_k$ は長さが1に規格化され,

$$V_k \cdot V_k = \sum_{n=1}^N (V_{k,n})^2 = 1 \quad (3.8)$$

異なる列 $k$ と1に対しては直交する.

$$V_k \cdot V_1 = \sum_{n=1}^N V_{k,n} V_{1,n} = 0 \quad (3.9)$$

これはクロネッカーのデルタを使って次のように書け

$$V_k \cdot V_1 = \delta_{k,1} \quad (3.10)$$

この右辺は $k=1$ ならば1, そうでなければ0であることを示す. 同時刻相関行列 $C$ は対称行列なのでJacobi回転アルゴリズムを繰り返すことにより, 固有値と固有ベクトルを求めること

ができる[9-11].

文献[7]で採用された方法は同時刻相関行列の固有値分布は時系列の長さを $L$ とすると, $N \rightarrow \infty, L \rightarrow \infty, Q \equiv L/N = \text{const.}$ の極限で式(1.2)となることが知られている.そこで,得られた固有値の中で,ランダム行列理論式に一致する,またはその範囲内に存在するものをランダム成分と見なし,時系列の固有値分布は理論分布から偏離する程度を乱数度と評価する.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,本稿では乱数度という言葉で表現する.

図3.2(左)に示す例のように,角のある線で示すヒストグラムと曲線で示すRMT理論式(データと同じ $Q$ に対応する式(1.2)のグラフ)の両者が一致すればRMTテストに合格したとし,図3.2(右)に示す例のように不一致ならばRMTテストに不合格と判断する.この定性評価[12,13]は直感的に見やすく,形状の特徴を把握しやすい利点がある.

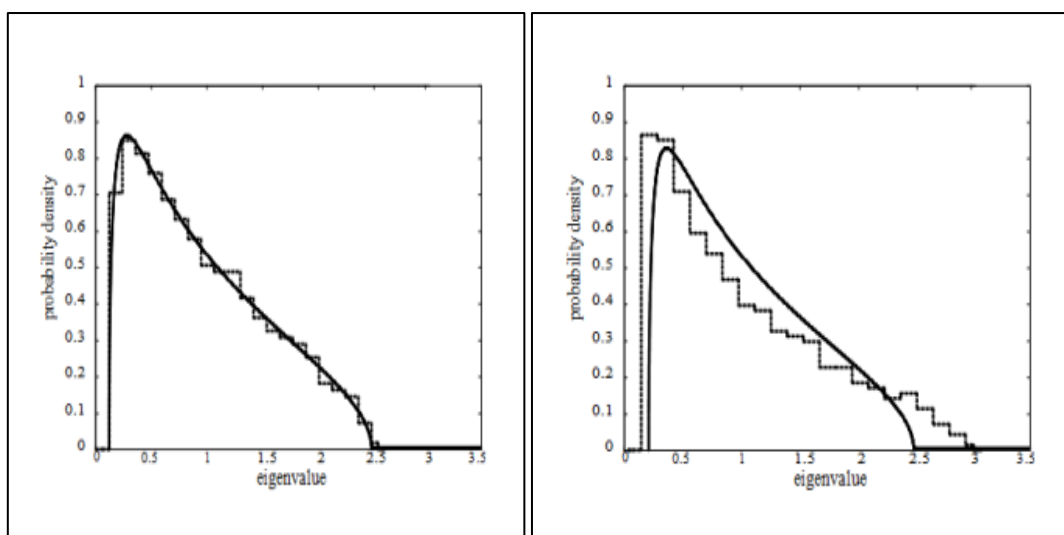


図 3.2 RMT テスト定性評価の例 (左 : 合格, 右 : 不合格)

### 3.3 RMT テストの定量評価法

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.

### 3.3.1 moment 分析法

等長時系列の内積を要素とする相関行列 $C$ の固有値分布の実測値と理論式(1)の $k$ 次モーメントをそれぞれ求めて両者の誤差で乱数度を数値化する．相関行列 $C$ の固有値分布の $k$ 次モーメントは $\lambda^k$ の平均値であり，次式で求められる．

$$m_k = \frac{1}{N} \sum_{i=1}^N \lambda_i^k \quad (3.11)$$

対応する理論値は固有値の理論分布式(1.2)から以下のように求められる．

$$\mu_k = E(\lambda^k) = \int_{\lambda_-}^{\lambda_+} \lambda^k P_{RMT}(\lambda) d\lambda \quad (3.12)$$

6次以下のモーメントは $Q$ の関数として以下ようになる．普段は4次まで計算し，我が研究室では6次モーメントの計算結果を示した．

$$\mu_1 = 1 \quad (3.13)$$

$$\mu_2 = 1 + \frac{1}{Q} \quad (3.14)$$

$$\mu_3 = 1 + \frac{3}{Q} + \frac{1}{Q^2} \quad (3.15)$$

$$\mu_4 = 1 + \frac{6}{Q} + \frac{6}{Q^2} + \frac{1}{Q^3} \quad (3.16)$$

$$\mu_5 = 1 + \frac{10}{Q} + \frac{20}{Q^2} + \frac{10}{Q^3} + \frac{1}{Q^4} \quad (3.17)$$

$$\mu_6 = 1 + \frac{15}{Q} + \frac{50}{Q^2} + \frac{50}{Q^3} + \frac{15}{Q^4} + \frac{1}{Q^5} \quad (3.18)$$

両者の誤差の逆数が乱数度となるが，当面，以下に定義する誤差

$$\text{誤差} = m_k / \mu_k - 1 \quad (3.19)$$

を用いて比較を行い，というのはRMTテストの定量評価として提案する[14-19]．

実際には相関行列 $C$ の $N$ 個の固有値を全て求めずとも，式(3.11)に必要なのは固有値の $k$ 乗の和であるから，これは $C^k$ の対角成分の和(trace)に等しく，対角化する前も後も不変である

から、 $k$  次モーメントは  $C^k$  の対角成分の和の  $1/N$  を取ることで簡単に求められる。特に行列サイズが大きい場合にはこの方法により計算時間が大幅に縮小できる[20]。即ち、 $N \times N$  対称行列  $C$  の固有値が縮退してないとして、直交行列  $T$  による対角化を行い固有値を求めたとする。以下、計算過程を説明する。

$$T^{-1}CT = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_N \end{bmatrix} \quad (3.20)$$

両辺を  $k$  乗することにより

$$(T^{-1}CT)^k = T^{-1}C^kT = \begin{bmatrix} \lambda_1^k & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_N^k \end{bmatrix} \quad (3.21)$$

となる。ここで、対角和(trace) $Tr$ をとり、 $1/N$ 倍すると、

$$\frac{1}{N}Tr(TC^kT^{-1}) = \frac{1}{N} \sum_{i=1}^N \lambda_i^k \quad (3.22)$$

となる。Traceの持つ性質、即ち任意の2つの行列  $X, Y$  について

$$Tr(XY) = Tr(YX) \quad (3.23)$$

であることを用いると、

$$m_k = \frac{1}{N} \sum_{i=1}^N (C^k)_{i,i} = \frac{1}{N} \sum_{i=1}^N \lambda_i^k \quad (3.24)$$

となるわけである。

### 3.3.2 $\chi$ 二乗検定

モーメント分析法は、 $\lambda \geq \lambda_+$  の部分に注目するが、乱数度を誤差絶対値により表現する有効性を検証するために、 $\chi$  二乗検定法を用い、 $\lambda < \lambda_+$  の部分の実データの固有値分布と理論分布の一致性を調査する。一致すれば、モーメント誤差のみで乱数度を評価する方法を有効と判断する。

$\chi$  二乗検定[21]とは帰無仮説がただしければ検定統計量が  $\chi$  二乗分布に従うような統計学検定法の総称である。次のようなものが含まれる。ピアソンの  $\chi$  二乗検定、一部の尤度比検定、イエイツの  $\chi$  二乗検定、マンテル・ヘンツェルの  $\chi$  二乗検定、累積  $\chi$  二乗検定と Linear

by Linear 連間  $\chi$  二乗検定. これらはいずれも :

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \quad (3.25)$$

という形の統計量を含み, “observed”は観測値という, “expected”は観測値から求められる期待値の推定量あるいは理論値を示す. 本稿では,  $\chi$  二乗検定のうち最も基本的かつ広く用いられる方法であるピアソンの  $\chi$  二乗検定を用いる. ピアソンの  $\chi$  二乗検定 2 つのタイプ適合度検定及び独立性検定があり, 「 $\lambda < \lambda_+$  の部分の実データの固有値分布と理論分布の一致性を調査する」ために, ここで, ピアソンの  $\chi$  二乗検定の適合度検定を使用する. 具体的な内容は次の節に説明する.

### 第 3 章のまとめ

本章では, まず RMT テストを提案する背景を紹介した. 次に, RMT テストを使用するため, 時系列の整形が必要である. 実データから自己相関行列の作成方法から, 固有値を算出する方法までを詳しく述べた. したがって, RMT テストの定性評価法を提案し, 乱数度高い例と低い例を明確した. 目視により乱数度を判断不可の数列に対して RMT テストの定量評価は適応できる. 定量評価ではモーメント法を使用する. モーメント分析法の理論知識を説明する上, 本研究は普段使う 4 次モーメントまでの結果だけではなく, 5, 6 次モーメントまでの計算式を示した. 他に, モーメント法の有用性を証明するために  $\chi$  二乗検定を併用する.

## 注釈及び参考文献：

- [1] 脇本和昌：乱数の知識，森北出版（1970）
- [2] L. Hannes: PLAB a system for testing random numbers, Proceedings of the International Workshop Parallel Numerics'94, pp.89-99 (1994)
- [3] ENT: A pseudorandom number sequence test program(online), <http://www.fourmilab.ch/random/>
- [4] NIST: A statistical test suite (online), [csrc.nist.gov/groups/ST/toolkit/rng/documentation-software.html](http://csrc.nist.gov/groups/ST/toolkit/rng/documentation-software.html) (2010.08.13)
- [5] Dieharder: A random number test suite version 3.31.1 (online), <http://www.phy.duke.edu/rgb/General/dieharder.php>
- [6] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral and H.E. Stanley: Scaling behavior in the growth of companies, Phys Rev Lett, Vol. 83, pp. 1471-1474 (1999)
- [7] V.Plerou, P.Gopikrishnan, B.Rosenow, L.A.N.Amaral and H.E.Stanley: Random matrix approach to cross correlations in financial data, Phys Rev E, pp.066126 (2002)
- [8] 楊 欣, 糸井 良太, 田中 美栄子：RMT 公式を用いた乱数度評価法の提案，研究報告数理モデル化と問題解決(MPS), 2011-MPS-83(2)
- [9] 木戸丈剛, 田中美栄子：ランダム行列理論との比較による米国株価日次変動のトレンド抽出，FIT2010:第9回情報科学技術フォーラム講演論文集，pp.157-162 (2010)
- [10] 小国力：ortran 95，C&JAVAによる新数値計算法—数値計算とデータ分析，サイエンス社 (1997)
- [11] 尾山壮一, 伊藤保昌, 上原政之: 行列を分割した場合の固有値問題とその主成分分析における変数増減への適用について II 固有値回転行列理論の主成分分析への適用，日本品質管理学会年次大会講演・研究発表要旨集, Vol.15, pp.46-49 (1985)
- [12] 楊欣, 糸井良太, 田中美栄子：RMT 公式を用いた乱数度評価法の提案，数理モデル化と問題解決研究報告，Vol.2012-MPS-83, No.2 (2011)
- [13] 田中 美栄子, 糸井 良太, 楊 欣：RMT 公式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)
- [14] X. Yang, R. Itoi and M. Tanaka-Yamawaki: Testing randomness by means of RMT formula, Intelligent Decision Technologies, SIST, Vol.10, pp.589-596 (2011)
- [15] 田中 美栄子, 糸井 良太, 楊 欣：RMT 公式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).
- [16] X. Yang, R. Itoi and M. Tanaka-Yamawaki: Testing randomness by means of random matrix theory, accepted for Progress of Theoretical Physics Supplement, Vol.194, pp.73-83 (2012)

- [17] 楊欣, 糸井良太, 田中美栄子: モーメント分析法によるランダム行列理論を用いた乱数度評価法の改良, 数理モデル化と問題解決研究報告, Vol.2012-MPS-88, No.16 (2012)
- [18] M. Tanaka-Yamawaki, X. Yang and R. Itoi: Moment approach for quantitative evaluation of randomness based on RMT formula, SIST, Vol.16, pp.423-432 (2012)
- [19] 楊欣, 三賀森悠太, 田中美栄子: 乱数度計 RMT テストの実データへの応用～ハッシュ値と Tick 株価～, 数理モデル化と問題解決研究報告, Vol.2012-MPS-91, No.33 (2012)
- [20] 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, No.1, pp.57-63 (2013)
- [21] W.G. Cochran: Some methods for strengthening the common  $\chi^2$  tests. Biometrics, Vol.10, pp.417-451 (1954)



## 第4章 RMT テストによる乱数度検定

秘密保持のために古くから使われている乱数表などがあるが、大量の乱数を必要とするときや多次元の計算をするときなど、コンピュータを用いてアルゴリズムに従い擬似乱数を発生させて使う。あるいはノイズ発生機のような物理現象を利用して物理乱数を発生させて用いる。本章では、ランダム性の高いことが知られている擬似乱数と物理乱数を使用し、RMT テストの有用性を検証する。

### 4.1 RMT テスト と擬似乱数

この節では、擬似乱数を用いて、RMTテストの有効性を実験する。分析対象は代表的な機械乱数である線形合同法(Linear Congruential Generator, LCGと略)とメルセンヌ・ツイスター(Mersenne Twister, MTと略)の生成する乱数列である。

擬似乱数生成法として最も一般的なLCG [1]法は次式で生成される。

$$X_{n+1} = (aX_n + b) \bmod M \quad (4.1)$$

ここでは次式のパラメータ[1]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[2] 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による乱数生成を行った。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。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。

#### 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パラメータを $Q=2,3,4,\dots,10$ まで設定し、 $N=100,200,\dots,500$ の五つを設定して実験を行った。本稿では( $N=500$ ,  $Q=3$ ) と ( $N=500$ ,  $Q=6$ ) の結果を例として出す。それぞれの定性評価の結果を図4.1と図4.2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言える。つまりいずれも乱数度は良好と判定できる。

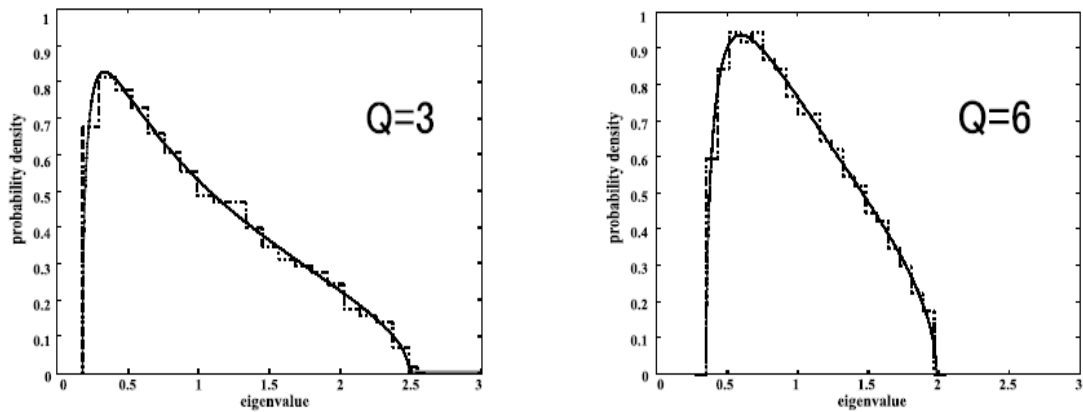


図 4.1 LCG の評価例

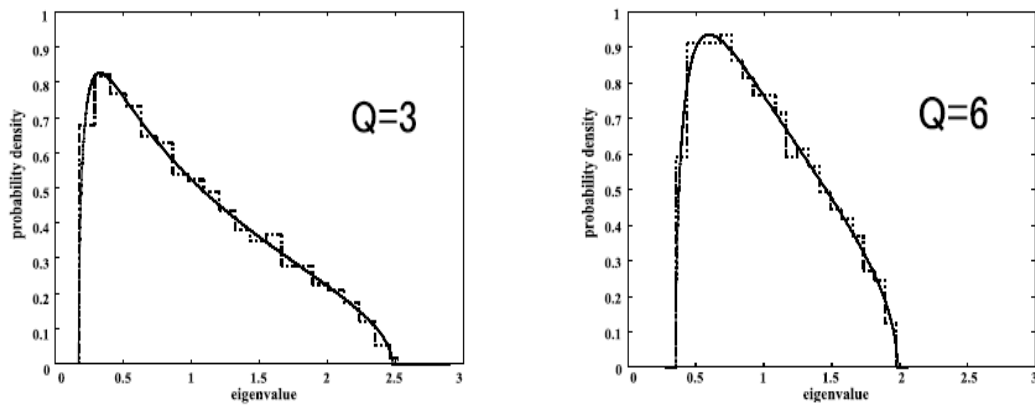


図 4.2 MT の評価例

## 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、前節で説明したRMTテストの定量評価も行った。6次以下のモーメントに対し、LCGとMTのSEED=1-100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4.1のようになる。ここで $k=1$ の場合は固有の平均値で常に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。表4.1に示すように、MTにより作成した乱数列の誤差平均値と標準偏差はいずれもLCGより小さい、つまり、MTの乱数度はLCGより高いということを分かった。

表 4.1 SEED=1-100 の誤差の平均値と標準偏差

k	LCG(Q=3)	MT(Q=3)	LCG(Q=6)	MT(Q=6)
2	-0.0004(.0010)	-0.0004(.0009)	-0.0003(.0006)	-0.0001(.0006)
3	-0.0010(.0026)	-0.0009(.0024)	-0.0008(.0016)	-0.0004(.0015)
4	-0.0018(.0047)	-0.0014(.0041)	-0.0014(.0028)	-0.0006(.0027)
5	-0.0027(.0072)	-0.0019(.0062)	-0.0020(.0043)	-0.0009(.0042)
6	-0.0036(.0100)	-0.0022(.0085)	-0.0026(.0060)	-0.0012(.0058)

次に、モーメント分析法は、 $\lambda \geq \lambda_+$ の部分に注目するが、ここで、ピアソンの $\chi^2$ 乗検定の適合度検定を使用し、 $\lambda < \lambda_+$ の部分の実データの固有値分布と理論分布の一致性を調査する。仮説は以下のようにする：

$H_0$ : 擬似乱数の固有値分布の $\lambda < \lambda_+$ 部分は M-P 分布と一致する。

$H_1$ : 擬似乱数の固有値分布の $\lambda < \lambda_+$ 部分は M-P 分布と一致しない。

同じ  $Q=3$  の LCG と MT の擬似乱数列を使用する。  $N=500$  ため 500 個の固有値を得る、 $Q=3$  で設定し固有値の範囲は  $(0.18 < \lambda < 2.49)$ 、表 4.2 に示す 1~47 は固有値を 0.05 で区切り 47 個の区間である。 よって、自由度を 46 に設定し、有意水準を 0.05 に設定する。 有意水準 0.05 自由度 46 の理論限界値は 62.83 と求められていて、これと比較し、実データの  $\chi^2 < 62.83$  なら、 $H_0$  を採用する。 表 4.2 に示す理論値は固有値の度数( $E$ )、MT の列は MT により作成した乱数列の固有値の度数( $O$ )である。

表 4.2 MT と M-P 理論分布の度数表の例

No.	理論値	MT	No.	理論値	MT	No.	理論値	MT	No.	理論値	MT	No.	理論値	MT
1	16	16	11	67	60	21	46	44	31	31	24	41	17	20
2	68	68	12	65	72	22	45	56	32	30	32	42	16	12
3	80	76	13	62	56	23	43	44	33	29	28	43	14	12
4	83	80	14	60	64	24	42	40	34	27	28	44	12	16
5	82	88	15	58	56	25	40	40	35	26	32	45	10	8
6	80	80	16	56	52	26	39	40	36	25	28	46	7	8
7	77	76	17	54	60	27	37	36	37	23	20	47	3	8
8	75	76	18	52	52	28	36	36	38	22	20			
9	72	72	19	50	48	29	34	32	39	20	16			
10	70	72	20	48	44	30	33	32	40	19	20			

式 4.2 のように、MT の  $\chi^2$  を計算し、その結果  $\chi_{MT}^2 = 25.55 < 62.83$ 、つまり、 $H_0$  を採用し、 $\lambda < \lambda_+$  の部分の実データの固有値分布と理論分布は一致である。一方、モーメント分析法により  $\lambda > \lambda_+$  の部分と理論分布の誤差により乱数度を判断できる。

$$\chi_{MT}^2 = \sum \frac{(O - E)^2}{E} \quad (4.2)$$

### 3) 初期乱数の乱数度評価

擬似乱数発生器により生成する乱数は最初生成する部分の乱数度が低いとよく言われるため、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個のみを繋げて用いる。そうすることにより、 $500 \times 500$  相関行列を  $L=1500$  について作成し、その定性評価結果は図4.3(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わかる。同じ条件でMTにより生成したデータの最初の500個のみを使って評価し、図4.3(右)の結果が出た、表4.3 ( $N=500$ ) に示す定量化結果から見るとLCGの6次以下モーメントの誤差の絶対値5%以上でMTの方は5%以内、これよりMTの初期乱数の乱数度が高いと言える。

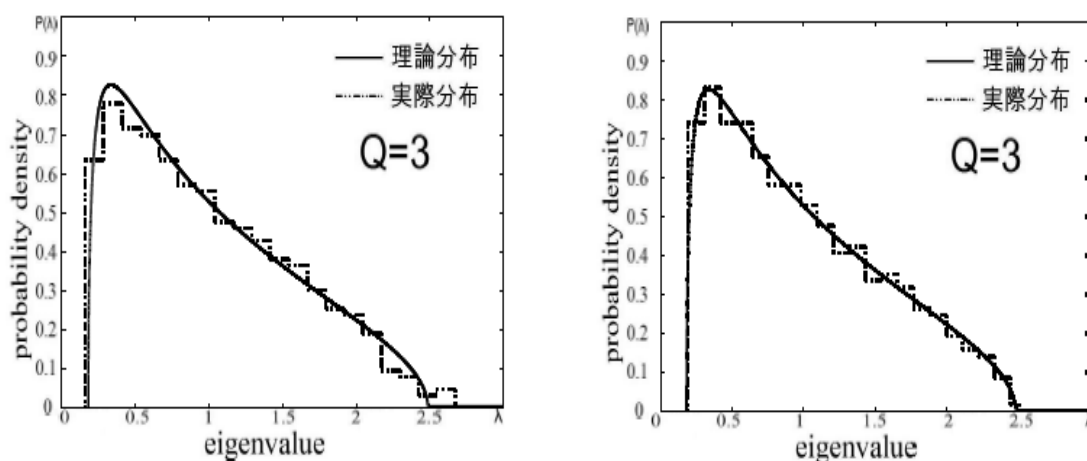


図 4.3 最初の 500 個のみを集めて評価した結果 LCG(左)MT(右)

表 4.3 最初 500 個のみを集めた初期乱数の定量評価

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

#### 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。ある時系列  $P(p_1, p_2, \dots, p_L)$  を式(4.3)による対数収益を取る、新しい時系列  $R(r_1, r_2, \dots, r_{L-1})$  を作成できる。プログラミング上、これは対数差を取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。

$$r_i = \ln p_{i+1} - \ln p_i \quad (4.3)$$

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  $N=500$ ,  $L=1500$  の場合について図4.4に示す。ランダム行列理論の許容範囲  $[\lambda_-, \lambda_+]$  から出ており、乱数度が低いと言える。6次以下のモーメントによる評価結果を表4.4に示す。この数値からも対数収益を取ることによる乱数度の低下が検出される。

また、表4.5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取ったもの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。対数収益を取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( $RMT = \lambda_+ - \lambda_- = 4/\sqrt{Q}$ )からの1.2倍になると考えられる。

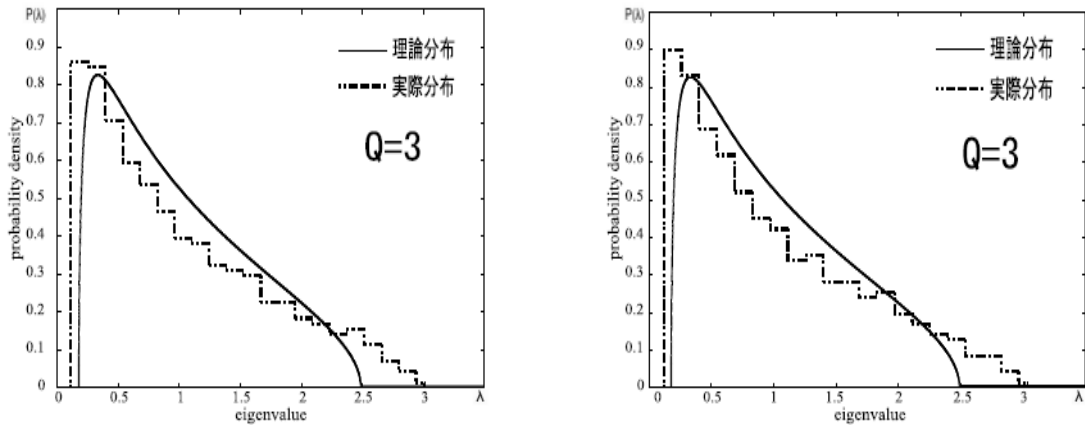


図 4.4 LCG(左)と MT(右)の N=500,L=1500 の対数収益評価結果

表 4.4 変化率を取って乱数度を下げた場合の評価結果 LCG(左)MT(右)

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

表 4.5 対数収益の分布範囲と理論範囲の比較 LCG(左)と MT(右)

Q	RMT( $4/\sqrt{Q}$ )	LCG_ln	LCG_ln/RMT	MT_ln	MT_ln/RMT
2	2.82	3.43	1.22	3.43	1.22
3	2.30	2.79	1.21	2.80	1.22
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
7	1.51	1.81	1.20	1.82	1.21
8	1.41	1.70	1.21	1.70	1.21
9	1.33	1.60	1.20	1.60	1.20
10	1.26	1.50	1.19	1.49	1.18

一方、時系列を対数収益とる際に、重複使用される項がある、それらの重複計算部分を削除し、LCGで生成した擬似乱数のランダム性は戻った (図4.5右) . 表4.6のように、重複

有り と 重複なし の対数収益時系列のRMTテストの定量評価結果を示す。重複なしの時系列に対し、6次モーメントの誤差絶対値は5%以下となった。

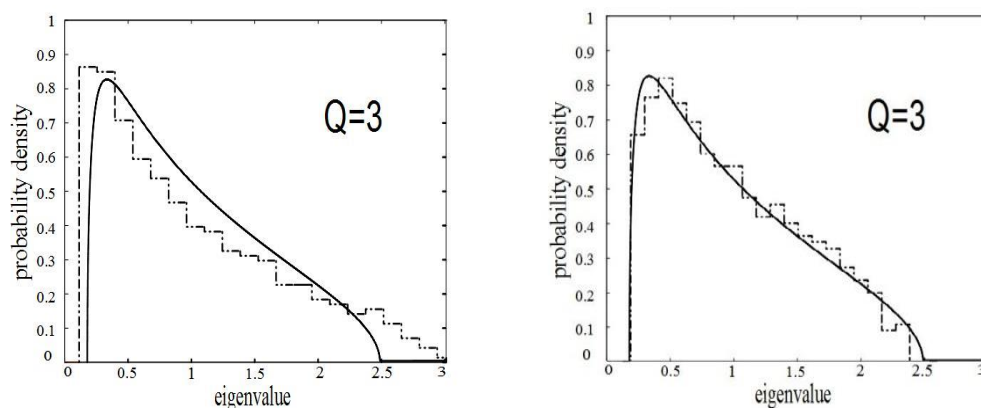


図 4.5 LCG の対数収益評価結果：重複有り（左）と重複なし（右）

表 4.6 対数収益を取ったデータの重複部分を削除したデータの評価結果(Q=3)

k	重複有り		重複無し	
	LCG	MT	LCG	MT
2	.1268	.1230	-.0013	-.0016
3	.3211	.3089	-.0039	-.0042
4	.5692	.5434	-.0072	-.0076
5	.8732	.8282	-.0112	-.0115
6	1.2416	1.1702	-.0156	-.0159

前文で MT により作成した乱数度高い時系列の固有値分布の  $\lambda < \lambda_+$  の部分を  $\chi$  二乗検定に合格した。つまり、 $\lambda < \lambda_+$  の部分とも M-P 分布の自己相関行列の固有値理論分布に一致する。ここで、乱数度低い例、MT による作成した時系列の対数収益をとった数列を対象としても分析を行った。MT の対数収益時系列の RMT テスト結果は  $\lambda \geq \lambda_+$  の部分に注目し、定性評価結果(図 4.4)では理論曲線からはみだす部分をはっきり見える。一方、RMT テストの定量評価結果は 6 次モーメントの誤差 (表 4.4) は元数列 (表 4.1) の誤差絶対値の 10 倍以上である。ここで、同じようにピアソンの  $\chi$  二乗検定の適合度検定を使用し、乱数度低い例の固有値分布の  $\lambda < \lambda_+$  の部分と理論分布の一致性を調査する。 $\chi$  二乗検定法を仮説は以下のようにする：

$H_0$ : MT の対数収益時系列の固有値分布の  $\lambda < \lambda_+$  部分は M-P 分布と一致する。

$H_1$ : MT の対数収益時系列の固有値分布の  $\lambda < \lambda_+$  部分は M-P 分布と一致しない。

パラメータなどの設定は前文と同じように設定した，実データの  $\chi^2 < 62.83$  なら， $H_0$  を採用する．表 4.7 に示す理論値は固有値の度数( $E$ )，MT\_ln の列は MT により作成した乱数列の対数収益列の固有値の度数( $O$ )である．図 4.4 に示すように， $0.5 < \lambda < 2$  の範囲のヒストグラムは理論曲線の下にあり，マイナスの誤差が出る．モーメント分析法ははみだす部分を評価しプラスの誤差を算出する．しかし，マイナス部分の誤差は算出しにくいから，今回の提案では考慮しなかった．今後の課題として研究していく．

表 4.7 MT\_ln と M-P 理論分布の度数表の例

No.	理論値	MT_ln	No.	理論値	MT_ln	No.	理論値	MT_ln	No.	理論値	MT_ln	No.	理論値	MT_ln
1	16	52	11	67	56	21	46	40	31	31	28	41	17	8
2	68	96	12	65	56	22	45	36	32	30	28	42	16	24
3	80	92	13	62	52	23	43	40	33	29	24	43	14	16
4	83	88	14	60	48	24	42	36	34	27	28	44	12	16
5	82	84	15	58	56	25	40	32	35	26	16	45	10	12
6	80	80	16	56	48	26	39	28	36	25	28	46	7	16
7	77	72	17	54	44	27	37	32	37	23	20	47	3	12
8	75	72	18	52	44	28	36	32	38	22	16			
9	72	68	19	50	36	29	34	28	39	20	24			
10	70	64	20	48	36	30	33	28	40	19	20			

式 4.2 のように，MT の  $\chi^2$  を計算し，その結果  $\chi^2_{MT} = 548.61 > 62.83$ ，つまり， $H_1$  を採用し， $\lambda < \lambda_+$  の部分の実データの固有値分布と理論分布は一致しない，つまり，モーメント法の結果と一致し乱数度良い時系列を対数収益化した列の乱数度が低い。

## 4.2 RMT テストと物理乱数

乱数は本来規則性も再現性も無いために予測は不可能である（例：サイコロを振る時，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。よって、予測不可能の利点を持つ物理乱数の乱数度もRMTテストにより評価した。従って、統計数理研究所ホームページ[3]から東芝製物理乱数(Toshiba)、日立製作所製物理乱数(Hitachi)と東京エレクトロニクス製物理乱数(Tokyo ED)等三種類の物理乱数のデータを手に入れた。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図4.6、図4.7と図4.8に示す。定量評価では擬似乱数と比較するために、ここで、物理乱数の三種類と擬似乱数の二種類の独立的な100サンプルを使用し、各次モーメントの誤差平均値と標準偏差を比較する。結果は $Q=3$ を表4.8に、 $Q=6$ を表4.9に示す。

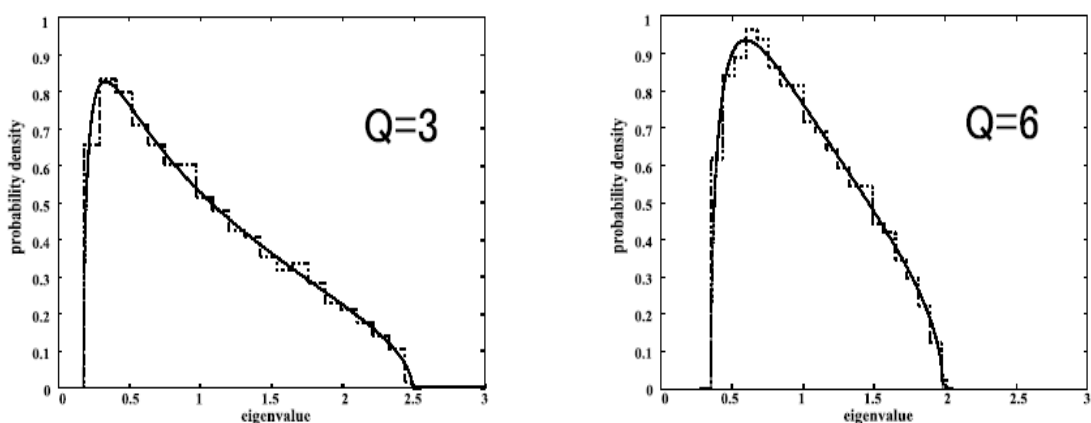


図 4.6 Toshiba 製の物理乱数の定性評価例

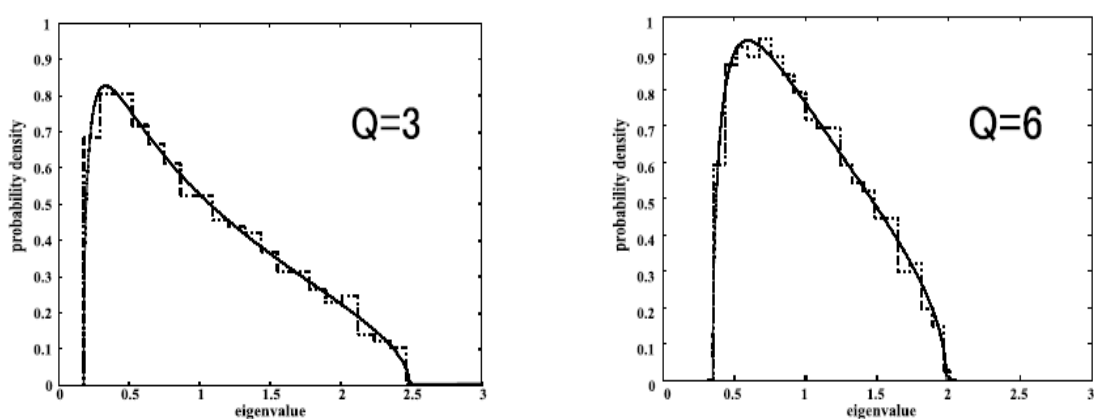


図 4.7 Hitachi 製の物理乱数の定性評価例

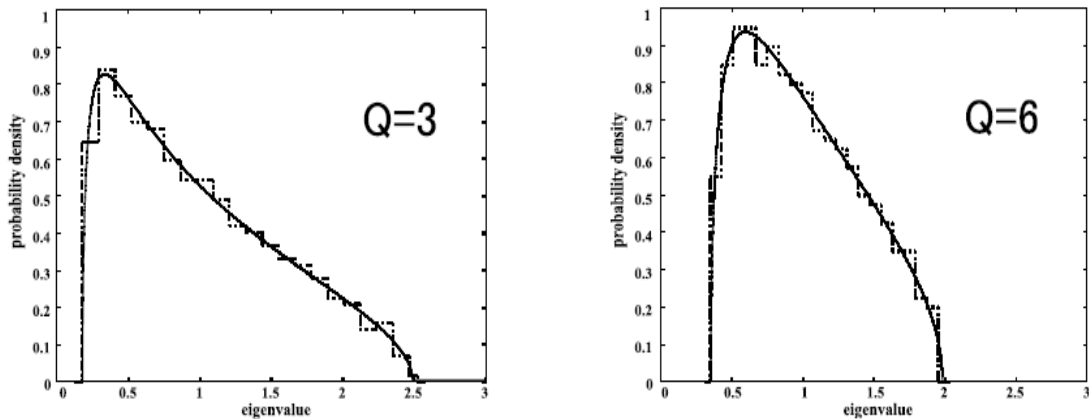


図 4.8 Tokyo-Electron 製の物理乱数の定性評価例

表 4.8 を見ると、長さ 75 万( $N=500, Q=3$ )の乱数列の各次モーメントの誤差平均値と標準偏差は擬似乱数の MT と物理乱数の Tokyo ED の乱数度が同じぐらい、他の三つより高い、つまり、乱数度一番高い物理乱数発生器は Tokyo ED 製である。擬似乱数の LCG と物理乱数の Toshiba 製、Hitachi 製を比較しても、同じぐらいであった。また、乱数列をその 2 倍 150 万( $N=500, Q=6$ )にすると、結果は表 4.9 に示す。その結果、Tokyo ED の乱数度を MT より高い、Toshiba 製と Hitachi 製の物理乱数の乱数度は LCG より高くなった、つまり、物理乱数のばらつきは乱数を長くなると小さくなる。また、RMT テストの定量評価基準を決定するために、モーメントの次数を決める必要である。表 4.8, 4.9 を分析すると、5 種類の乱数列の 2 次モーメントの誤差平均値は同じ程度、比較しにくい。また、ゴシック字の部分を目すると、3,4,5 次モーメントとも同じ誤差平均値などを出るから、6 次モーメントの誤差が一番比較しやすいである。よって、定量評価基準のモーメントの次数を 6 に設定した。

表 4.8 物理乱数と擬似乱数の乱数度比較( $N=500, Q=3$ )

k	LCG	MT	Toshiba	Hitachi	Tokyo ED
2	-0.0004(.0010)	-0.0004(.0009)	-0.0004(.0010)	-0.0004(.0010)	-0.0004(.0009)
3	-0.0010(.0026)	-0.0009(.0024)	-0.0011(.0026)	-0.0011(.0025)	-0.0009(.0025)
4	-0.0018(.0047)	-0.0014(.0041)	-0.0019(.0046)	-0.0019(.0044)	-0.0015(.0044)
5	-0.0027(.0072)	-0.0019(.0062)	-0.0026(.0070)	-0.0028(.0066)	-0.0019(.0067)
6	-0.0036(.0100)	-0.0022(.0085)	-0.0033(.0096)	-0.0037(.0092)	-0.0021(.0093)

表 4.9 物理乱数と擬似乱数の乱数度比較(N=500, Q=6)

k	LCG	MT	Toshiba	Hitachi	Tokyo ED
2	-0.0003(.0006)	-0.0001(.0006)	-0.0002(.0006)	-0.0002(.0006)	-0.0002(.0005)
3	-0.0008(.0016)	<b>-0.0004</b> (.0015)	-0.0005(.0015)	-0.0006(.0016)	<b>-0.0004</b> (.0014)
4	-0.0014(.0028)	-0.0006(.0027)	-0.0008(.0027)	-0.0011(.0028)	-0.0007(.0026)
5	-0.0020(.0043)	<b>-0.0009</b> (.0042)	-0.0012(.0042)	-0.0016(.0043)	<b>-0.0009</b> (.0040)
6	-0.0026(.0060)	-0.0012(.0058)	-0.0015(.0058)	-0.0020(.0060)	-0.0010(.0056)

表 4.10 50回(左)と500回(右)繰り返し誤差変化範囲(N=500, L=1500)

k	LCG(50)	MT(50)	LCG(500)	MT(500)
2	[-.0021,.0016]	[-.0021,.0024]	[-.0022,.0024]	[-.0025,.0023]
3	[-.0051,.0040]	[-.0052,.0061]	[-.0068,.0064]	[-.0073,.0060]
4	[-.0086,.0065]	[-.0086,.0109]	[-.0108,.0122]	[-.0122,.0115]
5	[-.0124,.0104]	[-.0122,.0164]	[-.0168,.0193]	[-.0182,.0182]
6	[-.0164,.0152]	[-.0160,.0227]	[-.0236,.0276]	[-.0243,.0258]

データ長75万の乱数列に対し、6次モーメントの誤差が3%以下という基準を取れば、上記5種類の乱数データが全て合格することが分かった。特に、表4.10のように、サンプル数を50に設定するとも、500までに設定するとも、誤差の変動範囲は変化無しである。また、乱数列の長さ制限を緩和するため、 $N=200, 300, 400$  ( $Q=3, \dots, 10$ ) の擬似乱数と物理乱数の6次モーメントの誤差を調査した。その結果、有意水準 $\alpha=0.05$ の時に $N=200, L=600$  ( $Q=3$ ) としても、6次モーメントの誤差の絶対値を5%以下に抑えることができることが判明した(図4.9)。これに基づき、RMTテストの定量評価基準を「6次モーメントが理論式に5%以下の誤差で一致すれば乱数度が高い」をRMTテストの定量評価基準とする。しかし、乱数列の長さが12万以下の乱数列にはこの基準を適用できない(図4.10)。例えば、 $N=100$  ( $Q=3$ ) の乱数列の場合は、6次以下のモーメントの理論式に対する誤差が10%以下ではあるが5%を越える値となってしまう。そこで、提案手法の定量基準の制限条件としてデータ全長が12万点以上となる。

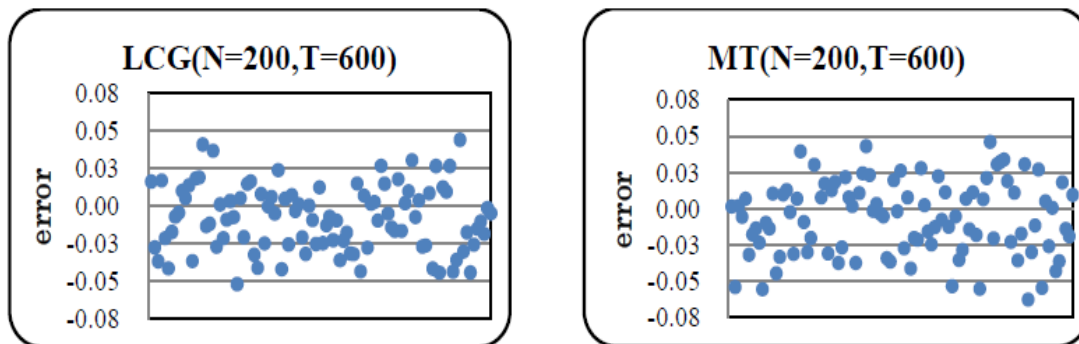


図 4.9 長さ 12 万の数列は 5%の定量評価基準に適用

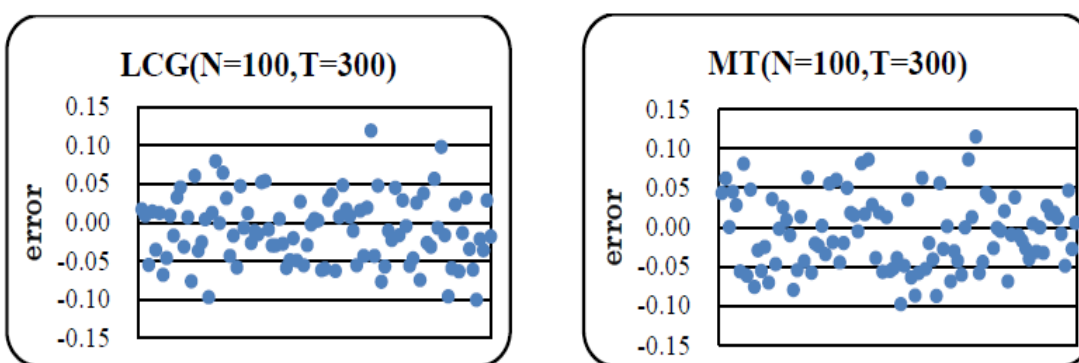


図 4.10 長さ 12 万の数列は 5%の定量評価基準に適用できない

### 4.3 他の検定法との比較

米国商務省標準局National Institute of Standards and Technology (NIST) が開発した乱数検定法は、本来、乱数を暗号として使用できるかを判定するためのものであるが、既存手法として確立しているので提案手法との比較を行う。公開当初は16方式計189の検定項目で構成されていたが、2010年でマイナーアップデートした「sts-2.1.1」版検定法は15個テスト項目を用いて、乱数の全体的な様子を見るようになっている。これを用いて提案手法で乱数度の高かったLCGデータは、NIST乱数判定法で評価すると、15項目全部に合格した。一方、LCGデータの対数収益を取ったデータは提案手法で乱数度が低く出たが、これをNISTで評価すると、15項目中6項目に合格した。この結果を表4.11に示す。つまり4割しかテスト項目をクリアしなかったことになる。これらの結果から、提案手法とNIST検定法の結果は矛盾しない。

しかし、NISTの評価法[4]はデータの形式の制限がある、実数を評価できない上、データ長が100万以上でなければならないという欠点がある。これに対して、提案手法ではそこまでの強い制限はない。評価基準を5%とした場合、長さ12万以上ならば使用出来る。また、

LCGの初期乱数のデータをNISTで評価すると合格してしまうが、提案手法では不合格となることから、NISTより提案手法の検定基準のほうが厳しいと言える。また、提案手法は定性評価において視覚化の利点がある上に、定量評価も可能である。これに対して他の可視化手法、例えばモアレ縞法等は定量化評価が出来ない。これに対し、本手法は視覚的評価と共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。

表 4.11 対数収益を取ったデータの NIST 乱数評価法での評価結果

テスト項目			結果
1	Frequency (Monobit) Test	一次元度数検定	○
2	Frequency Test within a Block	ブロック単位の度数検定	○
3	Runs Test	連の検定	×
4	Test for the Longest Run of Ones in a Block	ブロック単位の最長連検定	×
5	Binary Matrix Rank Test	行列ランク検定	○
6	Discrete Fourier Transform (Spectral) Test	離散フーリエ変換検定	×
7	Non-overlapping Template Matching Test	重なりのないテンプレート適合検定	×
8	Overlapping Template Matching Test	重なりのありテンプレート適合検定	×
9	Maurer's "Universal Statistical" Test	Maurer のユニバーサル統計検定	×
10	Linear Complexity Test	線形複雑度検定	○
11	Serial Test	系列検定	×
12	Approximate Entropy Test	近似エントロピー検定	×
13	Cumulative Sums (Cusum) Test	累積和検定	○

14	Random Excursions Test	ランダム偏差検定	×
15	Random Excursions Variant Test	種々のランダム偏差検定	○

また、RMT テストに合格した擬似乱数及び物理乱数の NIST の検定結果は表 4.12 に示す [5]。誤差は RMT テストの 6 次モーメントの誤差を使用する。LCG 乱数列に対し、NIST の不合格項目は重なりがないテンプレート適合検定のみであり、暗号として使えないだが、乱数度は高いと言える。物理乱数の三種類はどちらでも NIST 検定に対数収益をとることによって乱数度を低くした乱数列の物理乱数の比較結果を表 4.13 に示す。一次元度数検定、ブロック単位の度数検定、行列ランク検定、線形複雑度検定と累積和検定の五つしか合格しない。擬似乱数の LCG と同じように、元々乱数度が高い物理乱数の対数収益を取った後で、余分な相関が出るから、ランダム性を失い、NIST 検定に不合格である。RMT テストの結果と一致する。

表 4.12 擬似乱数(LCG)及び物理乱数 3 種類の NIST 検定結果

乱数の種類	NIST 乱数検定合格率
LCG	14/15
Hitachi	15/15
Toshiba	15/15
Tokyo ED	15/15

表 4.13 物理乱数の対数収益を取った数列の NIST 検定結果

乱数の種類	NIST 乱数検定合格率
Hitachi	5/15
Toshiba	5/15
Tokyo ED	5/15

RMTテストの性能を検証するために、三賀森、楊[5]は規則的な数列をシャッフルして作成した乱数データをRMTテスト及びNIST乱数検定で検証した。実験用データは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する。この時、データ長100万、かつ0と1のそれぞれの度数が全て均一の規則的な数列が構成される。このような規則を持つデータをシャッフルする回数を増えることにより乱数度が上がる。図4.11のように、シャッフル回数を増

えることにより乱数度が上がった。シャッフル回数を200万以上だったら、乱数度の変化は安定した。比較するために、NISTの評価結果を図4.12に示す。RMTテストと同じようにシャッフル回数を増えることにより、NIST検定の合格項目を増え、シャッフル回数を200万以上だったら、乱数度の変化は安定した。

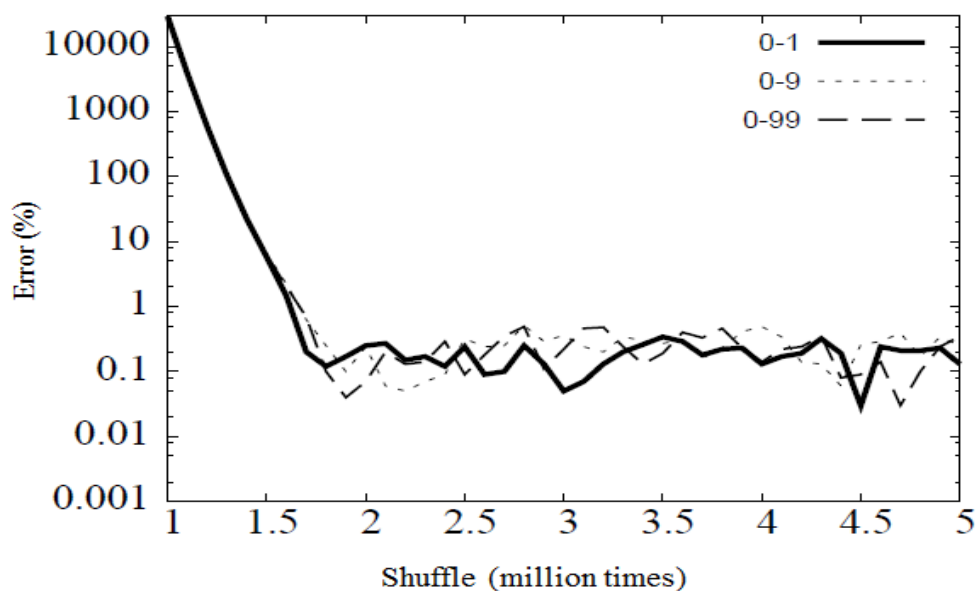


図 4.11 シャッフル回数により k 次モーメントの誤差変化

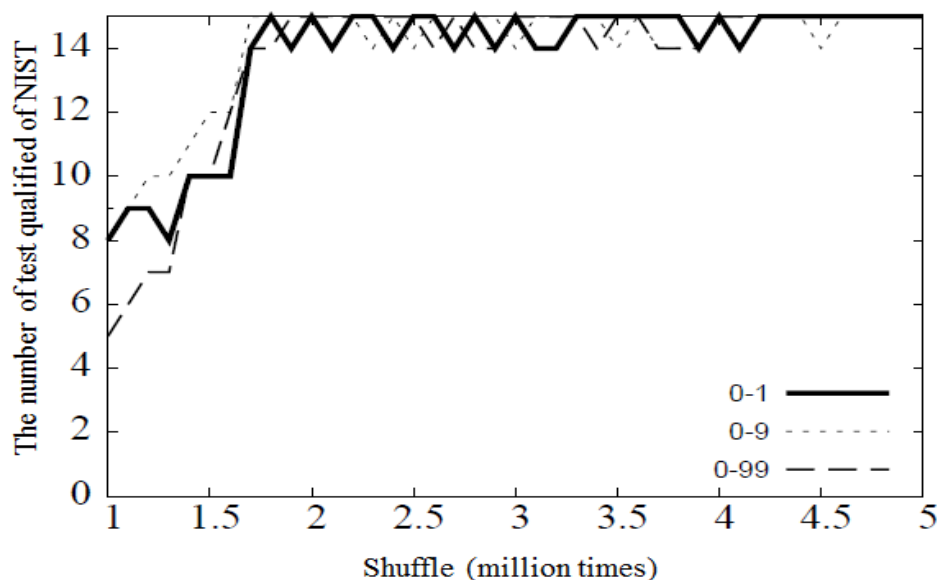


図 4.12 シャッフル回数を増えると NIST の検定項目の合格数を増加

## 第4章のまとめ

本章では提案された乱数度測定法RMTテスト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、乱数度が高いという結果が出た。また、定量評価を行えることにより、MTとLCGの差が検出できた。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。一方、東芝、日立と東京エレクトロン製の3種類物理乱数の乱数度を評価した。定性評価では3種類の物理乱数と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わかった。最後に、RMTテストと暗号分野でよく使われる米国のNIST検定法と比較し、RMTテストの性能を検証した。



注釈及び参考文献：

- [1] 松本 真：有限体の疑似乱数の応用(2004)  
<http://www.math.sci.hiroshima-u.ac.jp/m-mat/TEACH/0407-2.pdf>
- [2] M. Matsumoto and T. Nishimura: Mersenne Twister: A Dimensionally Equidistributed Uniform Pseudorandom Number Generator, ACM Transactions on Modeling and Computer Simulation, Vol.8, pp.3-30 (1998)
- [3] Y. Tamura: Random Number Library (2010). <http://random.ism.ac.jp/random>
- [4] 吉田等明, 村上武, 川村暁：NIST SP800-22rev.1aによる疑似乱数の検定に関する考察，電子通信学会技術研究報告， Vol.112, No.301, pp.13-18 (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, No.1, pp.57-63 (2013)

## 第5章 RMT テストの実データへの応用

### 5.1 RMT テストにより Hash 関数安全性の判定

ハッシュ関数[1]とは、可変長の（一般的には比較的長い）ビット列を、固定長の短い（160bitや256bit等）ビット列に変換関数である。式では $H : \{0,1\}^* \rightarrow \{0,1\}^n$ と書ける。ここで $H$ がハッシュ関数を表し、 $\{0,1\}^*$ は有限ビット列の集合、 $\{0,1\}^n$ は $n$  bitのビット列の集合を表す。入力 $x \in \{0,1\}^n$ に対して $H(x)$ を $x$ のハッシュ値と呼ぶ。ハッシュ関数を用いて、任意長のキーを固定長のメッセージへ変換することにより隠す。暗号的ハッシュ関数は情報セキュリティ分野で様々な利用されている。ハッシュテーブルのインデックス、フィンガープリント、重複データの検出などの用途がある。暗号的ハッシュ関数はハッシュ関数の一種である。偶然または意図的なデータの改竄によって、そのハッシュ値も変化する。

理想的な暗号的ハッシュ関数は次の4つの特性を持つ：

- ①与えられたメッセージに対してハッシュ値が容易に計算できる
- ②ハッシュ値から元のメッセージを得ることが事実上不可能であること
- ③ハッシュ値を変えずにメッセージを改竄することが事実上不可能であること
- ④同じハッシュ値をもつ二つのメッセージを求めることが事実上不可能であること

乱数列の特性は規則無、予測不可能などがある。ハッシュ値列はランダム列なら特性②③④を満足できる。よって、RMTテストによりハッシュ値列のランダム性を評価し、乱数度高いハッシュ値列を持つ暗号的ハッシュ関数の安全性が高い。

使用した乱数列は暗号的ハッシュ関数MD5とSHA-1の出力データ。暗号的ハッシュ関数は多数存在するが、その多くは脆弱性が判明し、使われなくなっている。2004年8月、当時よく使われていたハッシュ関数（SHA-0, RIPEMD, MD5など）の弱点が判明した。このことから、これらのハッシュ関数から派生したアルゴリズム、特にSHA-1とRIPEMD-128の長期的なセキュリティに疑問が投げかけられた。2009年現在、最も広く使われている暗号的ハッシュ関数はMD5[2]とSHA-1[3]である。MD5とSHA-1の比較結果を表5.1に示す。MD5とSHA-1の構造は同じであり、SHA-1の公開日はMD5より新し、また、SHA-1の方は $2^{80}$ 回の計算により一回の衝突がある可能性があり、MD5の $2^{64}$ より長い。また、MD5は既に破られているため、安全性が高いと言えるのはSHA-1である。

ここで、ハッシュ関数の出力データのランダム性が高ければ暗号的安全性も高いと考

え、手法としてMD5とSHA-1の出力データのランダム性を評価し、実験を行う前にSHA-1の安全性が高いことが既知なので、出力データのランダム性も高いと予想した。MD5とSHA-1の出力データ（ハッシュ値）は図5.1に示す。定量評価基準の長さ制限は12万ため、3750個ファイルのMD5のハッシュ値と3000個ファイルのSHA-1のハッシュ値でそれぞれの長さ12万の数列を取得でき、 $N=200$ 、 $Q=3$ と設定し、10サンプルの6次モーメントの誤差の絶対値は両方とも5%以下で乱数度が高い（表5.2）。定性評価の一例を図5.2に示す。また、SHA-1の各次モーメントの誤差の絶対値は小さく、従って、予想の通り、MD5よりSHA-1の出力データのランダム性が良いという結果が分かった。そこで、提案手法の有用性をチェックできた。

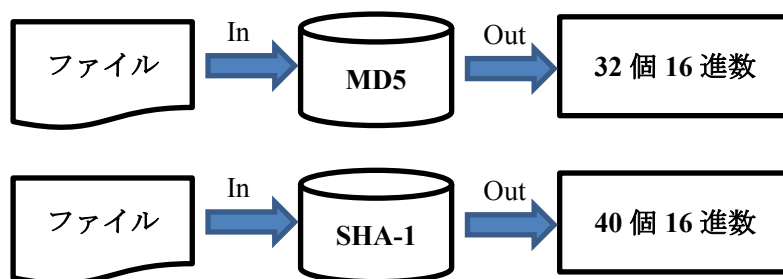


図 5.1 MD5 と SHA-1 の出力データ

表 5.1 MD5 と SHA-1 の比較

	MD5	SHA-1
最初公開	1992	1995
構造	Merkle-Damgård construction	Merkle-Damgård construction
ハッシュ値形式	16 進数	16 進数
ハッシュ値長	128bit	160bit
データブロック	512bit	512bit
衝突までのステップ数	64	80

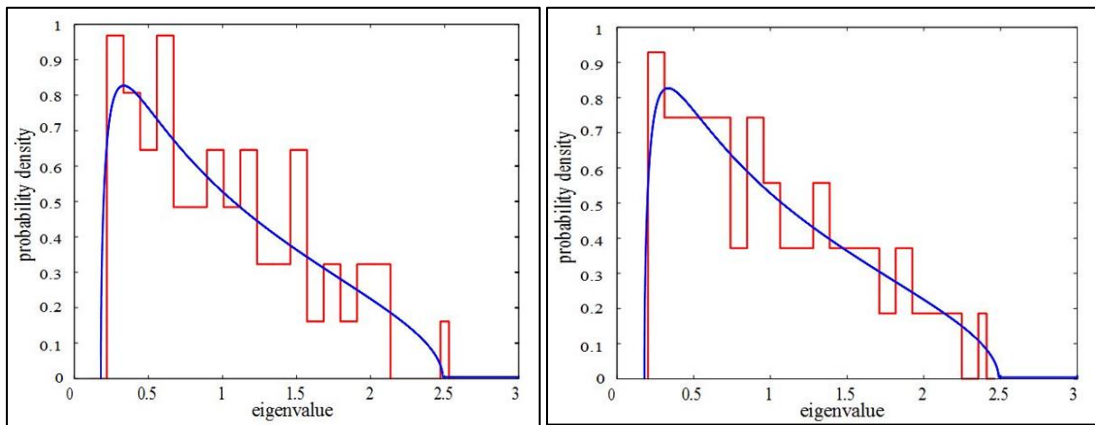


図 5.2 MD5 (左) と SHA-1 (右) の定性評価結果

表 5.2 MD5 と SHA-1 の出力データのランダム性

K	MD5(Q=3)	SHA-1(Q=3)	MD5(Q=6)	SHA-1(Q=6)
2	-0.0012	-0.0019	-0.0008	-0.0013
3	-0.0051	-0.0032	-0.0031	-0.0033
4	-0.0115	-0.0034	-0.0065	-0.0049
5	-0.0194	-0.0029	-0.0108	-0.0058
6	-0.0282	-0.0020	-0.0159	-0.0057

## 5.2 RMT テストにより株投資の研究

ランダム行列理論を利用した主成分分析法として、特にビッグ・データ解析への応用効果が期待される RMT-PCA を、解析対象となるデータ時系列の乱数度測定法として用いた「RMT テスト」を株価 tick 価格時系列に対して適用し、測定された乱数度とその株式の収益率の推移との間に一定の関連性が見られることを見出した。その結果の導出過程を詳説する。

RMT テストは、ランダム性が十分高く良い乱数と見做せる時系列間の乱数度を比較出来るだけでなく、乱数とは見做せないほど乱数度の低い時系列に対しても同様に適用できてその乱雑さの度合いを可視化・数値化できる点において、多方面の分野への応用を期待できるため、実用性が高いと期待できる[4-8].

節3.5で説明したが最大値付近の誤差が大きく反映するk=6の場合が最も比較しやすいので、本稿では、各データの6次モーメントの誤差絶対値をRMTテストの定量評価結果とし、それにより株価変動のランダム性を数値化することにより、「6次モーメントの誤差絶対値が小さい方がランダム性が良い」と定義する。

### 5.2.1 株価変動ランダム性

企業コード4452（花王(株)）の2009年のティックデータにより作成した株価変動データのRMTテストの定性評価結果を図5.3の右図に例として示す。データから求めた固有値分布が、理論曲線の最大値2.5より大きい範囲にはみ出していることがはっきり見える。比較のために、株4452のデータと同じパラメータ $Q$ を持つ、線形合同法(Linear Congruential Generator, LCGと略)[9]により発生させた擬似乱数列から作成したデータの定性評価結果を図5.3の左図に示す。こちらの方はRMTの理論曲線とほぼ一致し、乱数度が高いことが見て取れる。このLCG乱数列をRMTテストで定量評価した結果の誤差絶対値は1.6%である。一方、株4452のRMTテストの定量評価結果は19.9%であることから、株4452の乱数度はLCG乱数列の10倍以上であり、株価変動は擬似乱数のような高い乱数度を持ってはいないということが解る。とは言え、理論曲線との際は小さく、ヒストグラムは理論曲線の範囲にほぼ収まっていることから、この株価はある程度ランダム性が高いデータであると言える。

本稿では、株価時系列の持つランダム性とその収益率との関係に注目してこれを定量的に調べることで明らかにして行く。

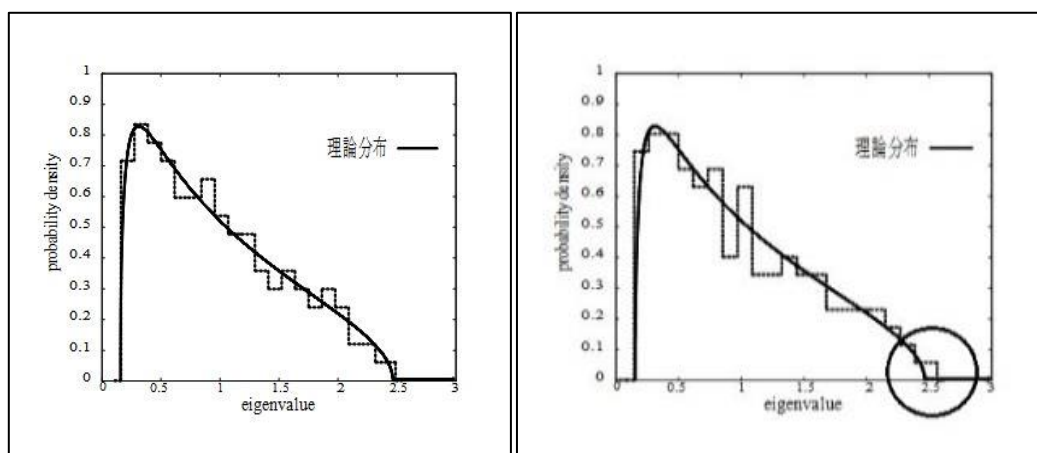


図 5.3 LCG の乱数度 (左), 株価ティックデータの乱数度 (右)

### 5.2.2 株価変動のランダム性と株投資

#### 1) データ処理

本稿では、2007年から2009年の東証 TOPIX500の3年間1分毎のティックデータを対象として分析する。ティックデータとは売値買値と取引オーダーを売買システムに入力され、分単位あるいは秒単位のタイムスタンプを付けた時系列である。よって、データは取引が

発生した時点に対してのみ記録される。そこで同時刻相関行列を作るため下記のような 3 段階のデータ処理を行う。

①データ成形（等間隔ブロック化と補完による同時刻化）

全ての銘柄に対して全ての時刻にデータが記録されるように時間間隔を決めてブロック化し、中間点に最も近い時点のデータを代表値として採用すると共に、データが空白の時点に対しては直前の値で補完を行い、データを成形する。具体的には、図 5.4 のように、データがない時点の一つ前のデータをコピーし、各銘柄のティックデータを同じ長さに補正する。このとき、補正分が全データ長の 2 割以上となる銘柄は採用しない(表 5.2)。このため、分析対象となる銘柄は元データより少なくなる。結果として、TOPIX 500 の全 33 業種の三年毎の分析対象は、2007 年は 211 銘柄、2008 年は 240 銘柄、2009 年は 229 銘柄となった(表 5.3)。

② 株価時系列から対数収益列への変換:

株価変動のデータに変換する為に株価  $S_{i,k}$  とその 1 時刻前の株価との変化率を対数収益

$$X_{i,k} = \ln(S_{i,k+1}/S_{i,k}) \quad (5.1)$$

として計算する。ここで  $i,k$  は  $i$  番目の銘柄( $i=1, \dots, N$ )の  $k$  時刻目の株価( $k=1, \dots, L-1$ ) である事を示す。

③ 次年度に直近のデータを用いて乱数度を測定:

時系列を RMT テストに掛けるために長さ  $L$  で切り分けて行列  $L \times N$  を作成し、長さ  $L$  未満の剰余部分は捨てるのだが、測定した乱数度と、次年度の株価収益との関係を調べるため、時間的に近いデータの最後の部分は残し、次年度に対して時間的に遠い、データの初期部分を捨てる。

以上の 3 ステップにより、同時刻相関を計算するための同長の対数収益時系列データ  $N$  個を作成した。

時間	株価			
	銘柄1	銘柄2	銘柄3	銘柄4
09:00	502	3810	1902	
09:01	502	3801	1906	
09:02	500	3804	1905	除外
...	一つ前の値を補正			
11:00	521	3788	1910	
12:30	515	3788		963
12:31	518	3792		953
...				
15:00	522	3820		853

図 5.4 ティックデータの補足

表 5.3 毎年の分析対象

year	Data length	Number of stocks
2007	66338	211
2008	66338	240
2009	65945	229

## 2) パラメータ $Q$ の選定

RMT 理論式(1.2)と(1.3)は 1 個のパラメータ  $Q=L/N$  のみに依存する。同じデータ時系列を用いた場合、長さ  $L$  の選び方次第で  $Q$  が決まるため、パラメータ  $Q$  としてどの値を選ぶべきかを実験的に決定することにした。

ランダム行列理論を適用できる条件である  $Q>1$ ,  $N\rightarrow\infty$  は、擬似乱数や物理乱数を用いた実験により [10,11],  $N=500$  程度で誤差が 1%以下に収まることがわかっているが、もう少し緩い条件を課して、誤差数%以下に抑えるための  $N$  の下限値としては少なくとも  $N\geq 100$  が必要となる。

前述のデータ処理ステップ①でティックデータの補足を行ったが、補足後のデータ長は 2007 年, 2008 年, 2009 年で順にそれぞれ, 66338, 66338, 65945 となった。最初の 2 年間は同長の 66338 であるが、これを長さ 100 で切り分けた場合、 $100\times 660$  の行列が作成でき、 $Q$  は  $1<Q<6.6$  の範囲内に設定できる。整数値を選ぶと  $Q=2,3,4,5,6$  が可能である。

このそれぞれの場合について、各年のデータから乱数度を計算し、翌年の収益と何らかの  
関係性が見られるのかを調査した。

まず、1年のデータ長が一定以上の長さの株価を日経33業種の各業種から4社ずつ選定  
し、その中で乱数度最高の株価と乱数度最低の株価を其々HとLとして、翌年の収益をグ  
ラフ化した。その結果、 $Q=4$ とした場合は、どの業種に於いても乱数度の高い株価が低い  
株価に比べて一貫して高い収益を示すが、 $Q<4$ の場合や $Q>4$ の場合は関係性が途中で変化  
したり、余り明確に見えないことがわかった。そこで $Q=4$ の場合にのみ明確に示される規  
則性ではあるが、『1年分の株価1分足から計算した乱数度が高い株を買えば翌年の収益は  
他の株より良好』という規則性を抽出できた。図5.5に $Q=2, \dots, 6$ を比較することによりこの  
様子を示す。

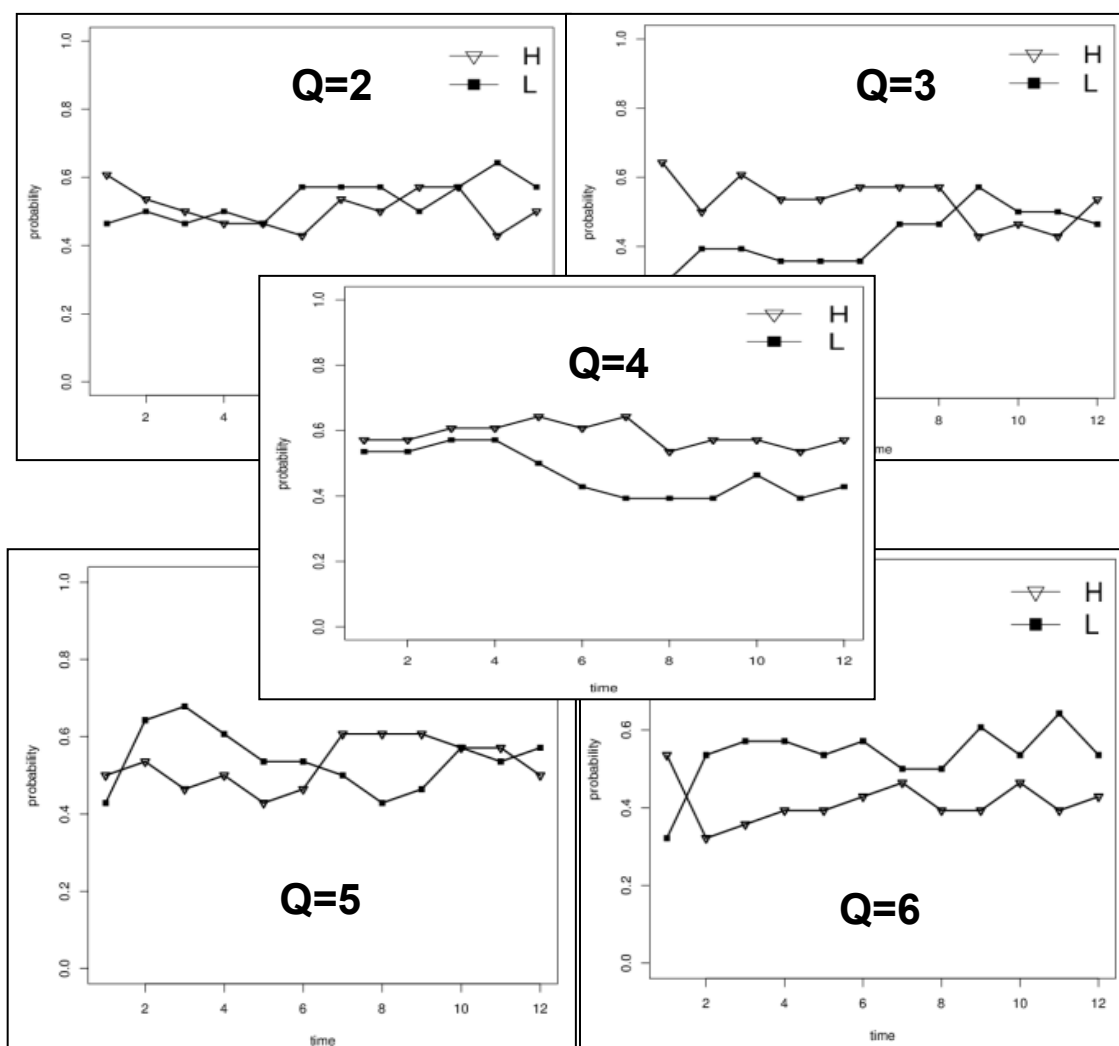


図 5.5 最適なパラメータは  $Q=4$



### 3) 実験

実験手順は下記の通りである。

- ①分析対象とする銘柄の各年の乱数度を計算する。
- ②乱数度最高と乱数度最低の株を抽出する (H, L)。
- ③抽出された銘柄の収益変化を比較する。
- ④翌年の株価変動を比較し,  $L < H$  となる期間を確認する。

### 4) 結果

#### (1)株のランダム性と安定性

東証 2007 年の 1 分足株価 1 年分をデータとして RMT テストにより乱数度を評価した結果を表 5.4 に示す。乱数度を式(3.19)により理論値との誤差で評価すると、誤差最小、すなわち乱数度最高の株は 9504, 乱数度最低の株は 7201 となる。それぞれについて次年度の収益を比較すると、図 5.6 に示すように、翌 2008 年の株価の動き (式(5.2.1)によって表す) は 2007 年度に於いて乱数度最高であった株 9504 が日経平均株価より勝っており、逆に乱数度最低であった株 7201 は日経平均株価より下回っている。つまり、前述の経験則

【1 年分の株価 1 分足から計算した乱数度が高い株を買えば翌年の収益は他の株より良好】が 2007 年から 2008 年にかけての東証株価に対して成立していると言える。ここで抽出された乱数度最高株 9504 は、市況からの影響を受けにくい電力・ガス業に属し、リーマンショックから受けた影響も少なく、安定株と言える。乱数度最低株 7201 は小輸送用機器業に属し、景気や外需変動の影響を強く受けたと考えられる。

表 5.4 2007 年の東証 1 分足から求めた乱数度順位

順位	業種	企業コード	6 次モーメントの誤差絶対値
1	電力・ガス	9504	26.4
2	機械	6460	37.6
3	電力・ガス	9506	38.2
4	電力・ガス	9508	43.3
5	情報・通信	4676	44.9
...			
207	電気機器	6506	740.9
208	非鉄金属	5802	797.3

209	化学	4043	799.8
210	鉄鋼	5541	1001.5
211	輸送用機器	7201	1209.6

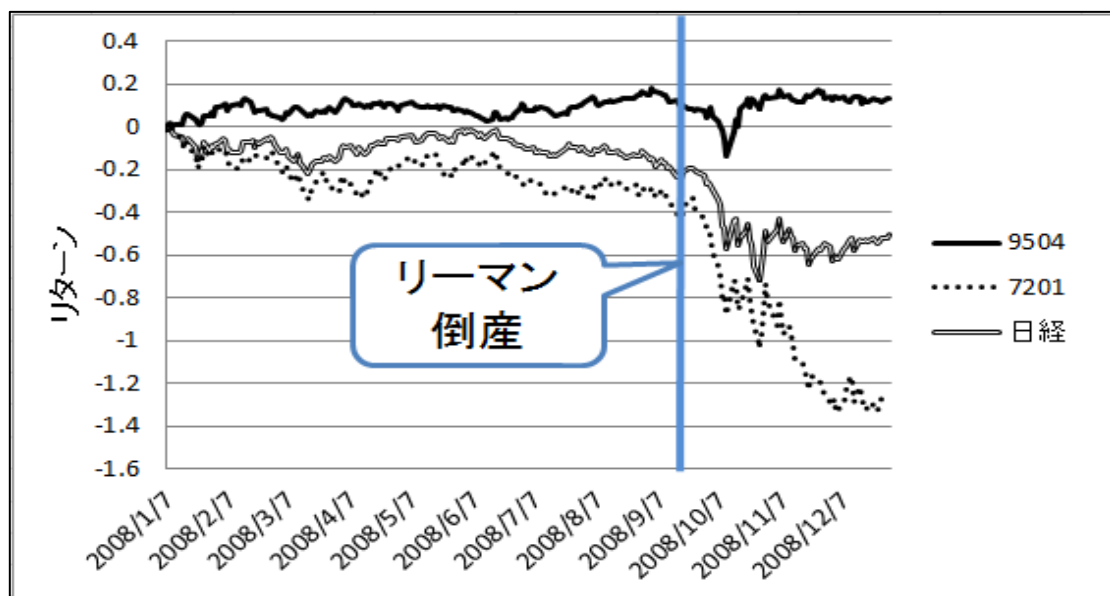


図 5.6 乱数度最高株(9504)の翌年収益は日経平均株価より安全であり、日経平均は乱数度最低株(7201)より安全である

また、乱数度最上位と最下位の株だけではなく、上位5位の株と最下位の7201を図5.6のように比較し、結果は図5.7に示す。図5.7を分析すると、乱数度上位5位の収益の上下関係は乱数度と完全一致ではないが、電気・ガス業種の三銘柄9504、9506と9508の収益変化線が他の銘柄の上に居り、つまり、安定であり安全株と言える。

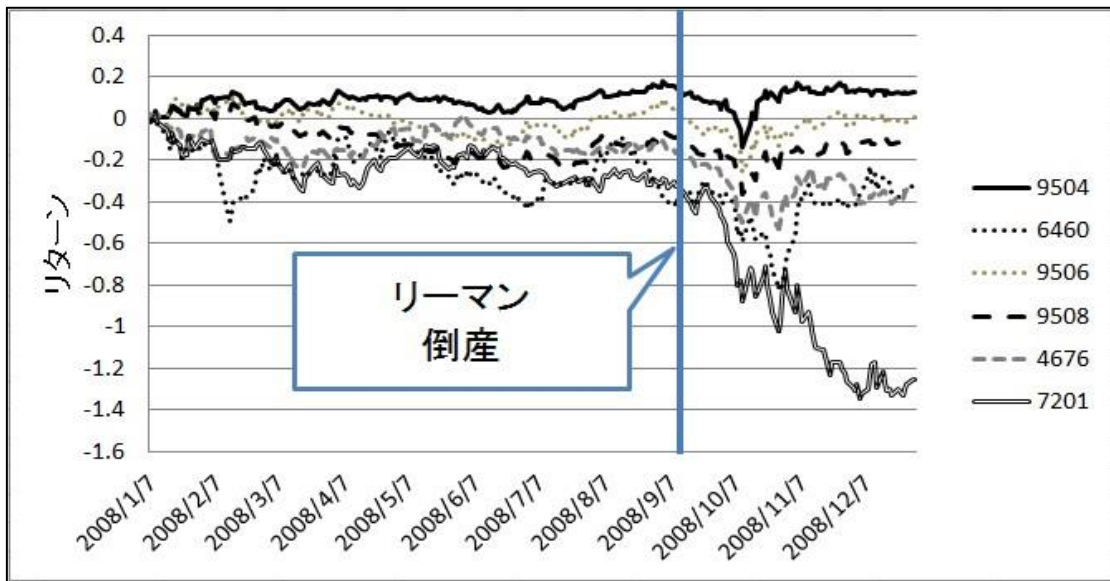


図 5.7 乱数度上位 5 位銘柄の翌年収益は乱数度最低株(7201)より安全である

(2) 高収益と高リスク

一方、2008 年の 1 分足から計算した乱数度は経験則とは異なる動きを示すことが表 5.5 からわかる。ここで乱数度最高の株として抽出されたのは銀行業の 8308 であり、これは翌年に於いて大幅に下落したため、安定株ではなかった。

この年が経験則に従わない理由としては、一応次のように考えられる。2008 年の 9 月に起きたリーマン・ブラザーズの倒産は、金融業に対する大幅な不信を世界規模で引き起こした。これは異常な事態であり、その結果、経験則が成立しないような株価変動を引き起こしたとも考えられる。

これを確かめるため、2008 年の株価変動の中から、リーマンショックが発生する以前、すなわち 2008 年 8 月の最終取引日迄のデータだけから測定した乱数度を使用した結果と比較する必要がある。

そこで東証 2008 年の 1 分足データを 2008 年初めから 8 月の最後の取引日まで用いて RMT 法により乱数度を測定した結果を表 5.6 に示す。リーマン・ショック直後の異常なデータを除外したことで、安定株の電力・ガス業の 9506 を抽出でき、経験則が成立していると言える。この 9506 と乱数度最低株であった 7201 の比較は図 5.8 に示す。図 5.8 の前半は経験則を支持するが後半は輸送用機器の 7201 の回復が大きく傾向がつかみにくい。しかし、株 7201 は 2007 年の乱数度順位が最下位で、その結果 2008 年は金融危機の影響もあって大幅下が

った事から、次の年に回復す可能性がもともと高かったのだと考えられる。そこで、株 7201 はリーマンショックの影響を受けた後で、2009 年の 2 月から上昇し始めたのだが、その時にたまたま購入すると、確かに乱数度が高い株 9506 より収益が高くなったのは当然とも考えられる、しかし、上昇時期の予測は困難であり、高リターンを狙えば高リスクを伴う。それよりは、企業配当も期待できる安定株を購入するのが安全だと考えられる。

表 5.5 2008 年の東証 1 分足から求めた乱数度順位

順位	業種	企業コード	6 次モーメントの 誤差絶対値
1	銀行業	8308	28.4
2	機械	7004	30.9
3	輸送用機器	7211	31.8
4	電力・ガス	9502	32.3
5	電力・ガス	9508	36.3
...			
236	電気機器	4902	1604.0
237	電気機器	6954	1611.2
238	電気機器	7752	1646.9
239	証券業	8604	2059.9
240	海運業	9104	2097.5

表 5.6 2008(Jan.-Aug.)のデータによる乱数度順位

順位	業種	企業コード	6 次モーメントの 誤差絶対値
1	電力・ガス	9506	11.6
2	電気機器	6728	13.1
3	食料品	2267	19.1
4	電力・ガス	9502	19.1
5	小売業	2685	22.2
...			
230	電気機器	6762	183.0
231	電気機器	6503	183.7
232	銀行業	8306	190.0
233	証券業	8604	222.8
234	輸送用機器	7201	257.9

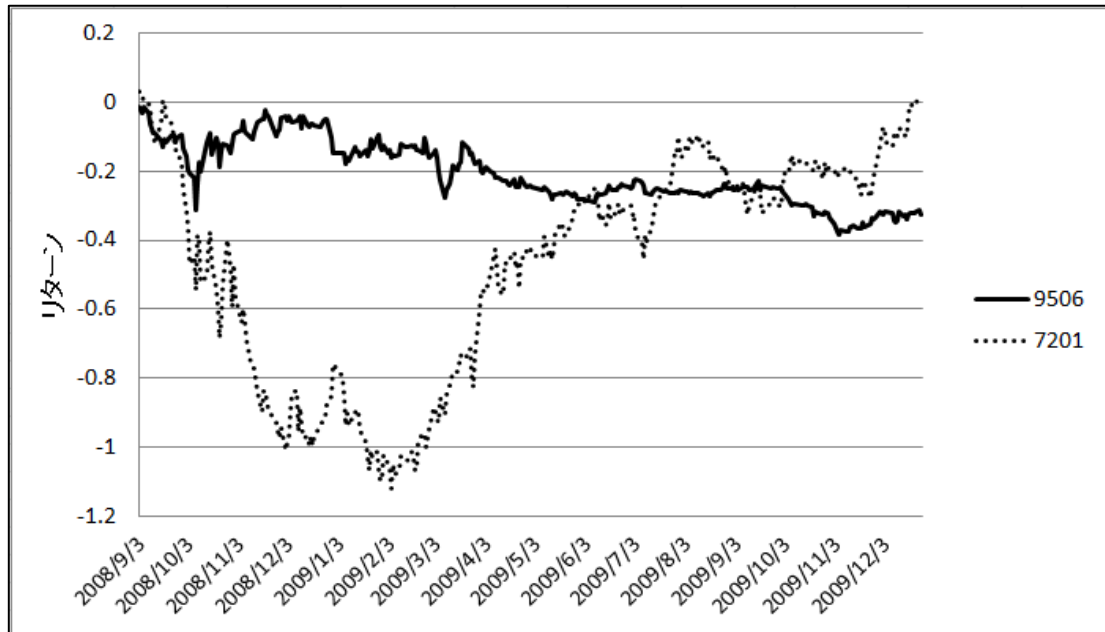


図 5.8 乱数度最高の9506が乱数度最低の8604より翌年の対数収益が高く安全であることを示す

また、図 5.7 のように、乱数度最上位と最下位の株だけではなく、上位 5 位の株と最下位の 7201 を比較し、結果は図 5.9 に示す。図 5.9 を分析すると、電気機器の 6728 はノイズのよう見え、乱数度が高いだが収益の変化が激しいという銘柄も存在した。一方、2008 の前 8 ヶ月間の乱数度により、乱数度上位 5 位のなかに収益一番良いのは小売業の 2685 であり、つまり、株式市場の外部環境が厳しいなると、投資者は景気の波により受注動向が動き、業績に影響する銘柄、電気・ガスや食品などの以外に、国内の消費関連株も注目しているということが分かった。

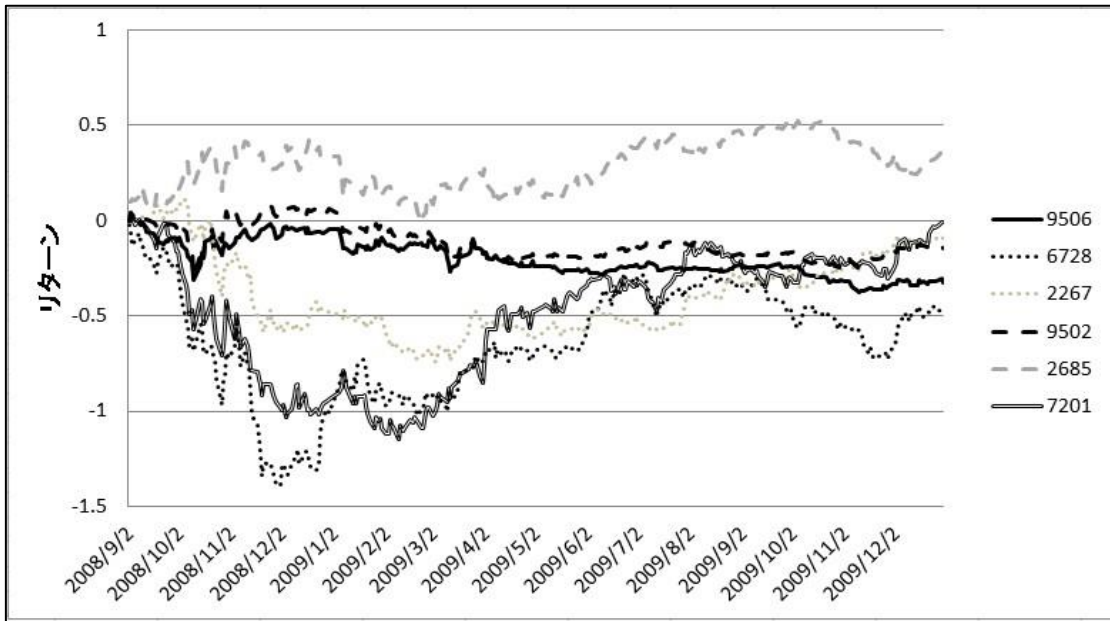


図 5.9 乱数度上位 5 位銘柄の内 6728 以外の銘柄の翌年収益は乱数度最低株(7201)より安全である

東証 2009 年の 1 分足から計算した乱数度を表 5.7 に示す。乱数度最高の株は電力・ガスの 9508 であり、乱数度最低の株は卸売業の 8058 であった。翌 2010 年の収益を図 5.10 に示す。9508 の収益は 8058 より上回り、経験則を満たしていると言える。また、乱数度上位 5 位の銘柄と乱数度最下位銘柄の収益比較結果の図 5.11 を分析してみると、第 2 位の 9509 のパフォーマンスが一番良い、上位 4 位の銘柄は同じ業種なので動きも似ている。一方、小売業の 2651 も同じ動きを見える。しかし、乱数度最下位の卸売業の銘柄 8058 はまったく異なる動きを確認できた。すなわち、図 5.6 から図 5.11 まではのいずれも乱数度最高の株の収益がそれ以外より上で、経験則はほぼ満たされている。

表 5.7 2009 年の東証 1 分足から求めた乱数度順位

順位	業種	企業コード	6 次モーメントの 誤差絶対値
1	電力・ガス	9508	25.5
2	電力・ガス	9509	28.1
3	電力・ガス	9506	31.6
4	電力・ガス	9502	33.2
5	小売業	2651	36.4
...			

225	非鉄金属	5713	1039.5
226	電気機器	4902	1072.4
227	機械	6301	1090.9
228	鉄鋼	5541	1128.2
229	卸売業	8058	1249.5

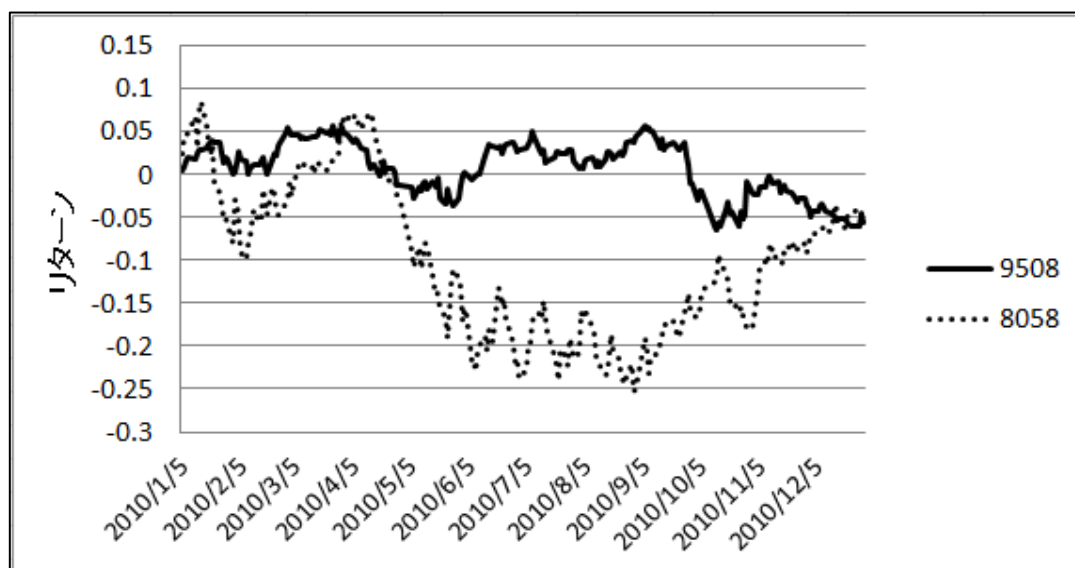


図 5.10 乱数度最高株 9508 が最低株 8058 より安全性が高い

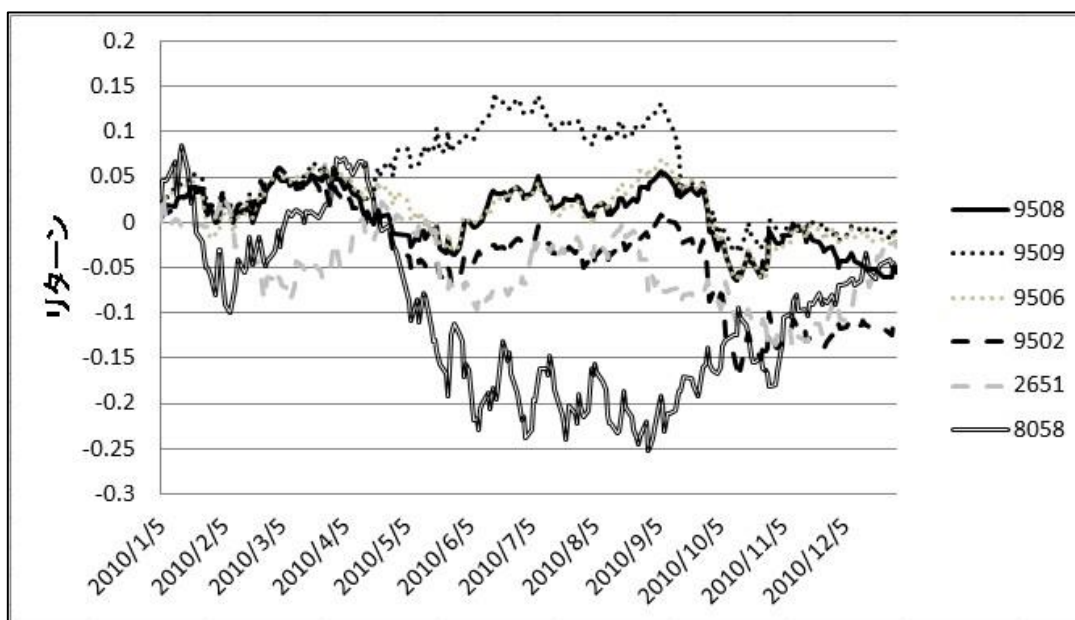


図 5.11 乱数度上位 5 位銘柄の翌年収益は乱数度最低株(8058)より良いである

### (3)ディフェンシブ株のランダム性が高い

東証 tick 株価の RMT に基づく主成分分析に於いて、田中、山本等[6]は表 5.8 に示す日本株式の新分類法を用いることで四半期ごとの主成分追跡に良い結果を出せることを示した。各分類の銘柄の特徴を表 5.9 に示す。その中に、業績が景気の影響を受けにくいと言われる銘柄、「食料品、医薬品、電力・ガス」で構成される「ディフェンシブ株」という分類がある。これに注目すると、実は表 5.4 にある、2007 年の乱数度上位 5 位のうちの三社はディフェンシブ株の電力・ガスに対応することがわかる。

また、表 5.6 に示す 2008 年の金融危機直後のデータを除外した乱数度測定の結果は、乱数度最高の株がディフェンシブ株に属する「電力・ガス」であり、上位 5 位のうちの三社がディフェンシブ株になっている。2009 年に入ると景気は回復傾向となり、それに従って株価変動のノイズが少なくなる。その結果は、表 5.7 に示すように乱数度の上位 4 位までがディフェンシブ株になっている。つまり、業績が景気の影響を受けにくいといわれる銘柄のランダム性が高いと言える。一方、乱数度下位 5 位は市況関連株と外需関連株が多く、これらは株価変動幅が大きく、高収益の可能性があると考えられる。図 5.12 に示すように、乱数度前 5 位はディフェンシブ株が主要セクタ(左図)、乱数度下 5 位は市況や外需に関連する銘柄の割合が高い(右図)。

表 5.8 関連株分類

外需関連株	65:電気・精密, 70:自動車・輸送機
内需関連株	17:建設・資材, 88:不動産
景気循環株	30:素材・化学, 60:機械
ディフェンシブ	20:食品, 45:医薬品, 95:電力・ガス
消費関連株	81:小売, 90:運送・物流, 94:情報通信・サービスその他
金利敏感株	83:銀行, 85:金融
市況関連株	50:エネルギー資源, 54:鉄鋼・非鉄, 80:商社・卸売



表 5.9 分類された株の特徴

外需関連株	国外での収益(輸出や現地の売上等)が影響する銘柄
内需関連株	事業基盤が国内にあり， 内需拡大にウエイトをおく銘柄
景気循環株	景気の波により受注動向が動き， 業績に影響する銘柄
ディフェンシブ	業績が景気の影響を受けにくいと言われる銘柄
消費関連株	消費者向けの事業を展開する銘柄
金利敏感株	金利の動向によって業績が大きく変わる銘柄
市況関連株	原油など商品価格に業績が左右される銘柄

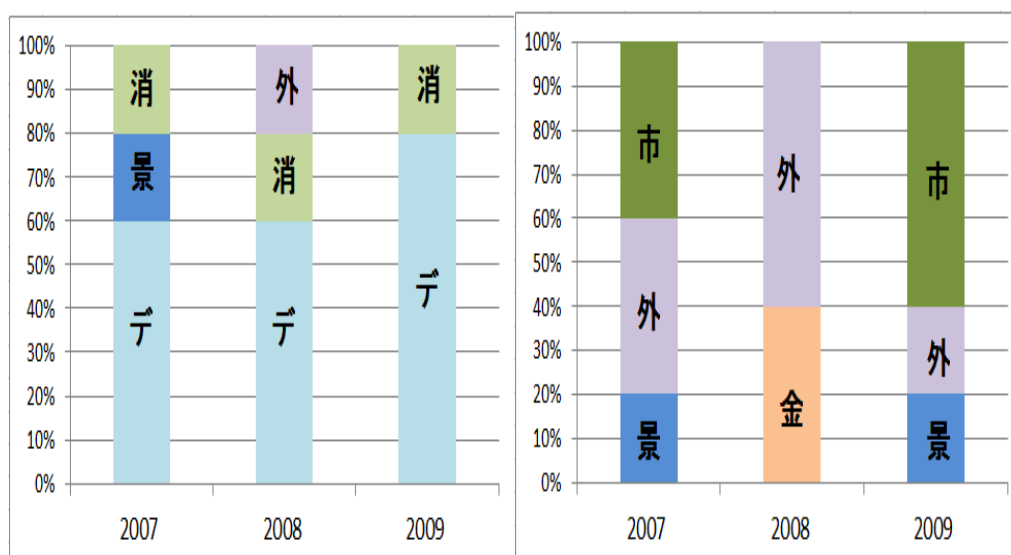


図 5.12 乱数度前 5 位 (左) と乱数度下 5 位 (右) の経年変化 (消：消費関連株，景：景気関連株，デ：ディフェンシブ株，外：外需関連株，市：市況関連株，金：金利敏感株)

### 5) 考察

RMT-PCA によって時系列の乱数度を直接測定する手法を提案し，それを tick 株価に応用した結果をまとめると共に，ここで得られた結果を 2007-2009 年にわたる 3 年間の経済状況との比較に於いて考察する。

この 3 年間は図 5.13 に示すように，2007 年にサブプライムローン問題が発生し，株式市場は下落相場に入った。しかし，サブプライムローン問題絡みで直接に破綻する日本企業

はなかった。よって、RMT テスト法により抽出された乱数度最高の株は株ディフェンシブ株の電力・ガスの 9504、乱数度最低の外需関連株 7201 より次年度の金融危機に受けた影響が少ない。一方、2008 年はリーマンショックがあったために、9 月以降、日経平均が大幅に下落した特異な年であったと考えることができる。この状況がどれほど以上であったかを示すため、2010 年末以前の TOPIX の対前日騰落率の下落記録の上位 10 位を示したのが表 5.10 である。この 10 位迄のうち 6 回が 2008 年に起きていることから、2008 年が特別に異常な年であったことが見て取れる。

この異常な株価変動を除外して、リーマンショック発生以前のデータのみを使用して乱数度を測った場合、表 5.6 に示すように、乱数度最高の株としてディフェンシブ株である電力・ガス業に属する 9506 を抽出できた。また、二年間連続して外需関連株の 7201 は乱数度が最低であったが、このために 7201 は次年度に乱数度とも収益とも上がる可能性が高かった。しかし、何時本当に上昇するかが特定積ない以上、高リスクを避けられない。2009 年に入ると、経済は少しずつ回復に向かい、株式や債券による新規資金調達も回復して来た。このときの乱数度結果は、上位 4 位がディフェンシブ株であり、下 5 位は市況や外需と景気関連株となって明確に分離した(表 5.7)。上位の安定株に投資した場合、次年度の収益は乱数度の低い株に比べて高かった。

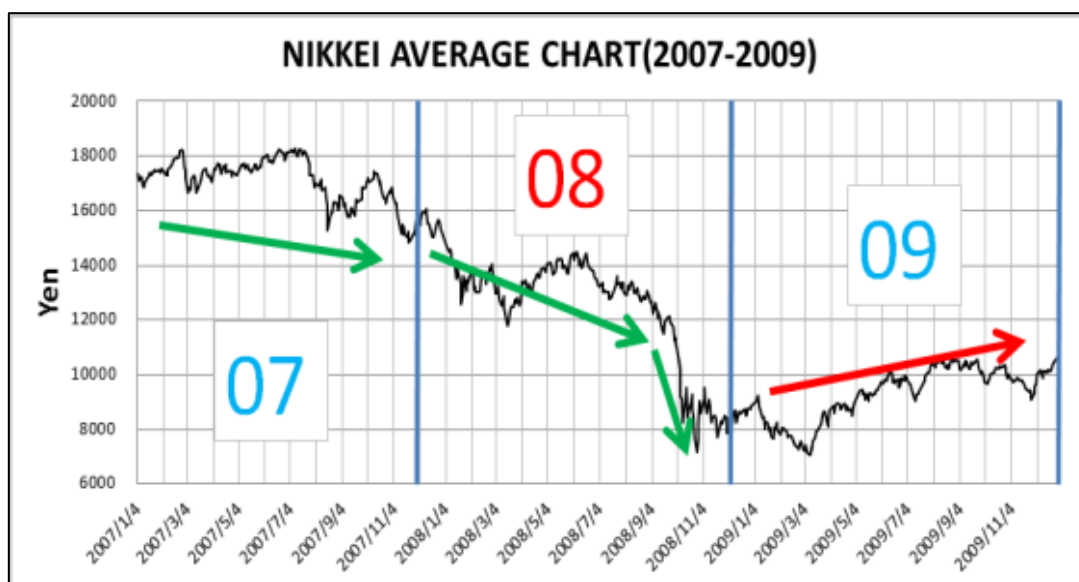


図 5.13 日経平均株価(2007-2009)

表 5.10 TOPIX の対前日騰落率の下落記録(-2010 年末)

順位	率(%)	TOPIX	年月日
1	-14.2	1793.9	1987/10/20
2	-9.52	864.5	<b>2008/10/16</b>
3	-8.75	32.3	1953/3/5
4	-8.04	899.0	<b>2008/10/8</b>
5	-7.52	806.1	<b>2008/10/24</b>
6	-7.47	159.3	1970/4/30
7	-7.40	746.5	<b>2008/10/27</b>
8	-7.10	2069.3	1990/4/2
9	-7.10	840.9	<b>2008/10/10</b>
10	-7.05	889.2	<b>2008/10/22</b>

<出所>東京証券取引所

### 5.3 株価時系列の特徴を分析するため使った時系列の考察

本稿では、株価変動のランダム性を調査するために、式(5.1)により株価時系列から変換した対数収益列を作成し、分析を行った。本節では、データを対数収益化の必要性などを考察する。

#### (1) 対数収益時系列は株価変動の特徴を反映する

節 5.2 では、株価の対数時系列のランダム性と株の特徴の関係を調査した。その結果、ランダム性高い株はディフュジブ株、ランダム性低い株は市況関連株と外需関連株ということが分かった。また、乱数度高い株は株式市場の下がり相場における悪い影響を受けにくいため、安全性が高いという結論を得た。

一方、株価というものは取引時間内に様々な要因の影響を受けて上がり下がりを行っている。その上がり下がりの変動のランダム性とも株価の特徴を反映するかを調査するため、実験を行った。ここで、株価の上がることを‘1’、株価の下がることを‘-1’、変化なしのことを‘0’で表し、収益時系列を代わりにアップ/ダウン時系列を作成した。まず、2007年の株価ティックデータ 6503(三菱電機(株))と 6841(横河電気(株))を例として出す。表 5.11 は 6503 と 6841 の元データを示す。

表 5.11 株価ティックデータの「株価時系列」の例(1分毎)

銘柄	株価時系列												
6503	1160	1159	1156	1159	1156	1161	1161	1158	1158	1155	1154	1156	1174
6841	1239	1238	1234	1228	1230	1231	1231	1238	1244	1245	1246	1245	1243

よって、作成したアップ/ダウン時系列は表 5.12 に示す。

表 5.12 ‘表 5.11’の例により作成した「アップ/ダウン時系列」の例

銘柄	アップ/ダウン時系列												
6503	-1	-1	1	-1	1	0	-1	0	-1	-1	1	1	
6841	-1	-1	-1	1	1	0	1	1	1	1	-1	-1	

次に、比較するために、二つ銘柄の対数収益時系列のランダム性評価結果、RMT テストの定性評価を図 5.14 に示し、乱数のようなランダム性が持っていないと見える。それに対して RMT テストの定量評価結果は  $Error_{6503} = 320.25\%$ ,  $Error_{6841} = 673.37\%$  である(付表 1)。従って、表 5.12 のアップ/ダウン時系列を評価すると、RMT テストの評価結果は図 5.15 に示す。図 5.14 と比べると両方とも理論曲線からはみ出す部分が見えなくなり; RMT テストの定量評価結果は  $Error'_{6503} = 0.41\%$ ,  $Error'_{6841} = 0.27\%$  乱数度は非常に高いという結果が出た。つまり、株価変動の幅を考えずに上がり下がりの変化だけを考えると、ランダム性が良く、株価の特徴を抽出できなくなり予測不可ということが分かった。

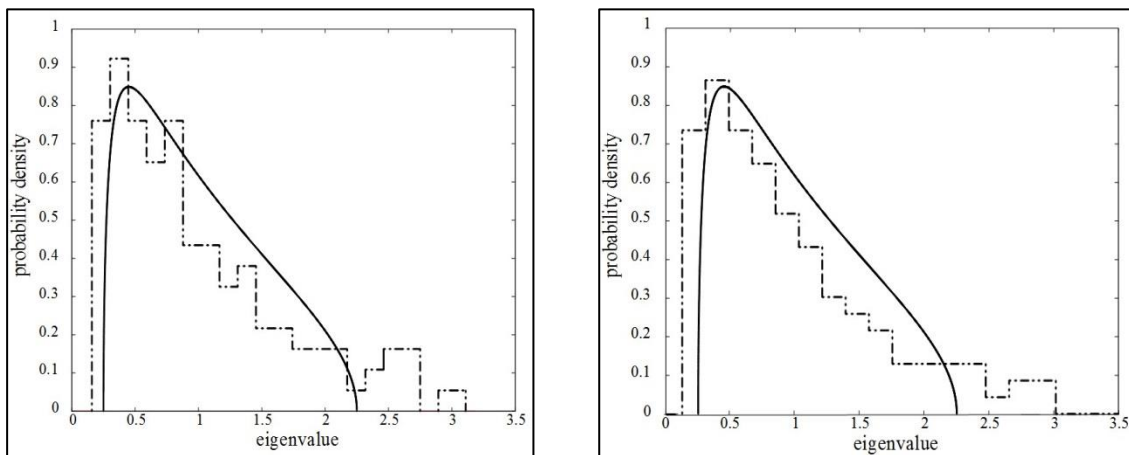


図 5.14 対数収益時系列の RMT テストの定性評価結果(6503: 左, 6841: 右)

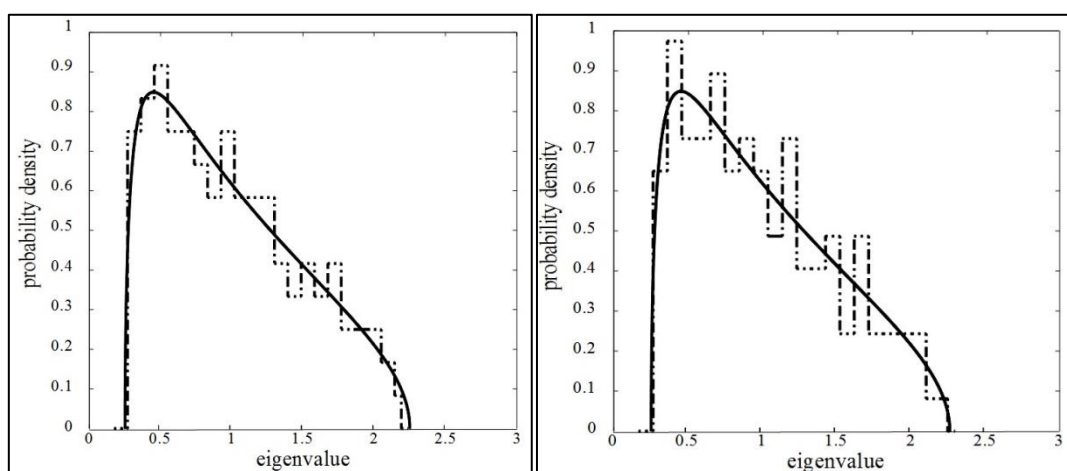


図 5.15 アップ/ダウン時系列の RMT テストの定性評価結果(6503:左, 6841:右)

(2)他の時系列について

株の変動幅を考えると、対数収益時系列ではなく隣接株価の差による作成する時系列は使えるかを調査した。株価を対数収益化する理由は株価の収益を比較するために、株価上昇と下降の割合を表すことが出来るということである。例えば、銘柄 1 の株価が 1000 円で 50 円上がり、銘柄 2 の株価が 10000 円で同じように 50 円上がりということに対し、対数収益は銘柄 1、銘柄 2 共に変動量は同じですが対数差を取ることで投資額を考慮した結果が得られる。しかし、提案した RMT テストでは、実データから作成した行列の自己相関行列を作る際、手順の一つは元行列の列毎を「正規化」した。よって、対数収益だけではなく、先ほど説明した銘柄 1 と銘柄 2 の変動量により構成する時系列も、即ち、表 5.11 の株価データから隣接値の差を取る時系列を表 5.13 に示す、同じような RMT テストの分析結果を得た。ここでは、表 5.14 に示すように 2007 年の乱数度計算結果の一部を例として出す。従って、株価のような実データのランダム性を分析する際、対数収益時系列と株価変動量時系列はどちらを利用しても構わない、RMT テストの定量評価結果に影響がないということが分かった。

表 5.13 ‘表 5.11’の例により作成した「株価の変動量時系列」の例

銘柄	株価の変動量時系列(隣接株価の差)											
6503	-1	-3	3	-3	5	0	-3	0	-3	-1	2	8
6841	-1	-4	-6	2	1	0	7	6	1	1	-1	-2

表 5.14 株価の対数収益時系列と変動量時系列の RMT テストの評価結果は一致である

順位	業種	銘柄	結果
1	電気・ガス	9504	26.37
2	機械	6460	37.62
3	電気・ガス	9506	38.16
4	電気・ガス	9508	43.29
5	情報・通信	4676	44.88
6	陸運業	9021	48.87
7	金属製品	5938	49.34
8	サービス	4324	49.73
9	電気・ガス	9502	52.70
10	機械	7013	54.87

順位	業種	銘柄	結果
1	電気・ガス	9504	26.33
2	機械	6460	37.49
3	電気・ガス	9506	38.07
4	電気・ガス	9508	43.30
5	情報・通信	4676	44.88
6	陸運業	9021	48.93
7	金属製品	5938	49.74
8	サービス	4324	49.95
9	電気・ガス	9502	52.65
10	機械	7013	55.51

## 第5章のまとめ

本章は乱数度検定ツール RMT テストを用い、暗号学的ハッシュ関数の安全性と東証株価ティックデータのランダム性を分析した。まず、セキュリティー分野でよく使われる、SHA-1 と MD5 の二つのハッシュ関数の出力データの乱数度を RMT テスト比較することにより、安全度の高い SHA-1 の方が乱数度の高いことを確認した。次に、2007 年から 2009 年までの 3 年間の乱数度順位から見ると、電力・ガスのような安定株の乱数度順位が高く、リーマンショックのような金融危機にもあまり影響されない。一方、『関連株分類』に基づけば、「電力・ガス」だけではなく景気の影響を受けにくいといわれる銘柄の集まりとしての、「ディフェンシブ株」の乱数度が高く、それに対して、乱数度が低い「外需関連株」と「市況関連株」などは株価変動幅が大きく、高収益可能性がある代わりに高リスクも伴う。

今後の研究としては、分析期間を広げると共に、株価変動のランダム性を用いた投資指標の可能性について考察を深めたいと考える。

注釈及び参考文献：

- [1] 安田 幹, 佐々木 悠：暗号学的ハッシュ関数—安全神話の崩壊と新たなる挑戦, 電子情報通信学会, 基礎・境界ソサイテイ, Fundamentals Review, Vol.4, No.1, pp.57-67 (2010)
- [2] J. Black, M. Cochran and T. Highland: A Study of the MD5 Attacks: Insights and Improvements (online), <http://www.cs.colorado.edu/~jrblack/papers/md5efull.pdf> (2008.07.27)
- [3] M. Locktyukhin and K. Farrel: Improving the Performance of the Secure Hash Algorithm (SHA-1) <http://software.intel.com/en-us/articles/improving-the-performance-of-the-secure-hash-algorithm-1/> (2010.03.29)
- [4] 楊 欣, 糸井 良太, 田中 美栄子：RMT 公式を用いた乱数度評価法の提案, 研究報告 数理モデル化と問題解決(MPS), 2011-MPS-83(2)
- [5] X. Yang, R. Itoi and M. Tanaka-Yamawaki: Testing randomness by means of RMT formula, Intelligent Decision Technologies, Smart Innovation, Systems and Technologies, Vol.10, pp. 589-596 (2011)
- [6] X. Yang and M. Tanaka-Yamawaki: Testing randomness by means of Random Matrix Theory, 2011 Kyoto Workshop on NOLTA pp.1-1 (2011)
- [7] 楊 欣, 糸井 良太, 田中 美栄子：ランダム行列理論を用いた乱数度評価法の提案, 統計数理研究所共同研究レポート, Vol.271, 経済物理とその周辺(8), pp.19-31, (2012)
- [8] 楊 欣, 糸井 良太, 田中 美栄子：乱数度計 RMT テストの実データへの応用—ハッシュ値と Tick 株価, 統計数理研究所共同研究レポート 292, 経済物理とその周辺(9), pp.70-83, (2013)
- [9] 松本 真：有限体の擬似乱数の応用(2004)  
<http://www.math.sci.hiroshima-u.ac.jp/m-mat/TEACH/0407-2.pdf>.
- [10] X. Yang, R. Itoi and M. Tanaka-Yamawaki: Testing randomness by means of Random Matrix Theory, Progress of Theoretical Physics Supplement, No.194, pp. 73-83 (2012)
- [11] M. Tanaka-Yamawaki, X. Yang and R. Itoi: Moment approach for quantitative evaluation of randomness based on RMT formula, Intelligent Decision Technologies, Smart Innovation, Systems and Technologies, Vol.16, pp. 423-432 (2012)



## 第6章 結 章

本稿では、現存の乱数検定法の使用難点を解決するために、ランダム行列理論を用いて新しい乱数度評価法 RMT テストを開発した。RMT テストは時系列の相関行列の固有値分布をランダム行列理論式と比較する方法である。前2章で RMT テストの理論基礎、ランダム行列理論(RMT)とランダム行列理論 (Random Matrix Theory: RMT と略) を用いた主成分分析法 (Principal Component Analysis: PCA と略)、RMT-PCA の理論知識や研究現状と今後の発展について述べた。次に、第3章では、乱数度検定法として RMT テストの研究背景を分析する上に、RMT テストの定性評価方法と定量評価方法のアルゴリズムを記述した。第4章の検証研究と第5章の応用研究を行い、以下は結論である。

- 1) 乱数度が高い擬似乱数と物理乱数は RMT テストに合格することがわかった。実験に使用した擬似乱数は代表的な機械乱数である線形合同法(Linear Congruential Generator, LCG と略)とメルセンヌ・ツイスター(Mersenne Twister, MT と略)の生成する乱数列である。定性評価により目視で両者の区別を判断できないが、定量評価の結果は MT により発生した乱数列の乱数度のほうが良いことが分かった。また、ランダムと言われる物理現象を利用して発生する物理乱数の乱数度と比較すると、長さ 75 万の数列に対し、物理乱数は発生源のノイズがあるため、乱数度のばらつきは擬似乱数より大きい。しかし、乱数列をその2倍である 150 万にすると、物理乱数の乱数度のばらつきが収まり、擬似乱数より乱数度が良いという結果となった。また、3種類の物理乱数発生器のうち最も新しい Tokyo ED 製が最も乱数度の高い乱数を発生することを検出した。従って、擬似乱数や物理乱数の乱数度を比較することにより RMT テストの有効性を実証した。
- 2) 乱数度の低いデータ列の乱数度を測定し、乱数度の差異を検出することができた。擬似乱数や物理乱数より乱数度が少し低い例がなかなかないため、人工的に例を作って実験を行った。それは擬似乱数として LCG と MT の2種を採用しその初期乱数の乱数度を測定した。本稿で採用した初期乱数というのは、異なる SEED に対して発生させた乱数の初期 500 個のみを繋げて用いる乱数列である。その結果、LCG の初期乱数の乱数度は MT より低いことが分かった。また、も一つの乱数度低い例は時系列の対数収益時系列であり、つまり、元々ランダム性が良い時系列の対数収益を取ると、特有の癖が時系列に付与され、乱数度が低下する。その乱数収益時系列の乱数度がどの程度低下するのかを RMT テストにより検出できた。特有の癖とは時系列を対数収益とる際に、重複使用される項があるということである。時系列を半分にしてこれを削除すると、時系列のランダム性が戻ることを発見した。
- 3) 実データには、乱数と言えないほど乱数度の低い数列が多いが、そういったデータも RMT テストは適用できる。この点にこそ RMT テストの利点があると考えられる。そのような応用として、ハッシュ関数の安全性判定と高収益株の抽出に対する

結果を報告する。まず、セキュリティー分野でよく使われる SHA-1 と、それより古い MD5 との二つのハッシュ関数の出力データの乱数度を RMT テストにより比較すると、確かに SHA-1 の方が乱数度が高く安全性の高いことが確認できた。次に、更にこれを用いた新たな応用として、東証株価ティックデータのランダム性を分析した。2007 年から 2009 年までの 3 年間の乱数度順位から見ると、電力・ガスのような安定株の乱数度順位が高く、リーマンショックのような金融危機にもあまり影響されない。一方、『関連株分類』に基づけば、「電力・ガス」だけではなく景気の影響を受けにくいといわれる銘柄の集まりとしての、「ディフェンシブ株」の乱数度が高く、それに対して、乱数度が低い「外需関連株」と「市況関連株」などは株価変動幅が大きく、高収益可能性がある代わりに高リスクも伴う。今後の研究としては、分析期間を広げると共に、株価変動のランダム性を用いた投資指標の可能性について考察を深めたいと考える。

## 謝 辞

最後になるが、私の指導教員である田中美栄子教授に心から誠に感謝しておく。博士の論文がスムーズにできあがり、また、早めに卒業できるのが田中美栄子教授のこころを尽くしていただいた指導のおかげだった。特に私の心の中に一番印象に残ったのが田中教授の厳しい研究態度、豊かな知識、新しいものを作り出す思考力、および上品な人格であった。私の生涯の成長にも大きな役割を果たしていたんじゃないのかと感じた。彼女から指導の上で、経済物理学領域において、私の視野が大きく広げようになった、また、十何回の国際や国内での学術シンポジウムに参加することによって、毎回の講演から自分を鍛えたり、進歩したりすることができた。将来、仕事に立ってもこれらのことが私にとって宝物のような大切な富になるだろう。そのほかには、年二回の一緒に統計数理研究所における研究集会に参加に来られる業界のベテランたちからの貴重な意見や提案をいただくことで、研究の上に、わたしがますます謹厳になれ、深く入り込めるようになった。ここでは、援助をくださった皆様にも私がこころから深く感謝している。次に、私の博士論文に手を入れていただいた知識A研究室の清水忠昭教授、機械宇宙工学専攻応用数理工学講座の石井晃教授、公共システム研究室の谷本圭志教授から貴重な提案や建設的な意見のおかげで、私の論文がいつそう整えられるようになり、その中でのまだ解決できていない問題にも研究の方向が示されていた。ここでは、再び無私に援助をくださった先生方に心をこめて謝意をお持ちしておく。わたしにとってはこの二年半の研究がただ一つのスタートである。お世話になった先生方の期待が答えられるように、これからの仕事にも前よりもっとがんばっていきたいと思う。また、私を育てた鳥取大学に恩返しをするために、立派な成果をあげたいと考えている。

次に、知識A研究室で一緒に勉強したり、生活したりし、がんばっていた同窓たちからの支持や激励にも謝意を持っている。留学生活がたいへんだったが、君たちがいることで、私の留学生活での楽しさと感動がたくさん増えることになった。今の私が卒業をするとすぐに職をつかないといけないのだ、君たちがぜひ知識Aという温かなファミリーを守ってほしい。

私の親、家族、友達にも感謝する。あなたたちから支持があるだけで、わたしが勇気を持って、人生の挑戦に直面できるようになった。最後に、私の旦那様にも「ありがとう」と言いたかった、彼に対する感激と愛を表すために一番ふさわしい言葉がどうしても見つけられなかったが、唯一の方法としては、私が心血を注ぎつくした極めて大切な博士論文を彼に差し上げたいと思っている。これからもたくさんの困難や挑戦に私たちが面に向かわなければならないかもしれないが、お互いに手を携えて肩を並べ、きっと一緒に美しい未来の人生を作り出すことができるだろうと私たちが信じている。

## 参考文献

- [1] 乱数について : <http://sehermitage.web.fc2.com/crypto/random.html>
- [2] 松本 真, 西村 拓士, 間違いだらけの疑似乱数選び:  
[http://www.soi.wide.ad.jp/class/20010000/slides/03/index\\_1.html](http://www.soi.wide.ad.jp/class/20010000/slides/03/index_1.html)
- [3] JIS:乱数発生及びランダム化の手順, JIS Z 9031(2010),JSA Web Store order No.911023999
- [4] 伏見正則, 乱数, 東京大学出版 (1989)
- [5] M.L. Mehta: Random Matrices,3rd edition, Academic Press (2004)
- [6] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral and H. E. Stanley: Universal and Non-Universal properties of cross-correlations in financial time series, Phys Rev Lett, Vol.83, 1471-1474 (1999)
- [7] A.M. Sengupta and P.P. Mitra: Distribution of singular values for some random matrices, Phys Rev E, Vol.60, pp.3389-3392 (1999)
- [8] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral and H.E. Stanley: Random matrix approach to cross correlations in \_nancial data, Phys Rev E, p.066126 (2002)
- [9] L. Laloux, P. Cizeaux, J.-P. Bouchaud and M. Potters: Phys Rev Lett, Vol.83, pp.1467-1470 (1999)
- [10] 山本 敦史 , 木戸 丈剛 , 田中 美栄子 , 高石 哲也: 株式市場の主成分追跡~RMT-PCAによる四半期間の主要業種抽出~, 研究報告数理モデル化と問題解決(MPS), 2012-MPS-88(17)
- [11] 村上和正, 安本典生, 家富洋: ランダム行列理論を用いた気象データ相関抽出とその有効性の検討, 素粒子論研究, Vol.117, No.5,pp.116-117 (2009)
- [12] 楊 欣, 糸井 良太, 田中 美栄子: RMT 公式を用いた乱数度評価法の提案, 研究報告数理モデル化と問題解決(MPS), 2011-MPS-83(2)
- [13] X. Yang, R. Itoi and M. Tanaka-Yamawaki: Testing randomness by means of RMT formula, Intelligent Decision Technologies, Smart Innovation, Systems and Technologies, Vol.10, pp. 589-596 (2011)
- [14] X. Yang and M. Tanaka-Yamawaki: Testing randomness by means of Random Matrix Theory, 2011 Kyoto Workshop on NOLTA pp.1-1 (2011)
- [15] X. Yang, R. Itoi and M. Tanaka-Yamawaki: Testing randomness by means of Random Matrix Theory, Progress of Theoretical Physics Supplement, No.194, pp. 73-83 (2012)
- [16] M. Tanaka-Yamawaki, X. Yang and R. Itoi: Moment approach for quantitative evaluation of randomness based on RMT formula, Intelligent Decision Thechnologies, Smart Innovation, Systems and Technologies, Vol.16, pp. 423-432 (2012)
- [17] 永尾 太郎: ランダム行列の基礎, 東京大学出版会(2005)
- [18] E.P. Wigner: Ann. Math., Vol. 67, pp. 325-327 (1958)
- [19] A. Edelman and N.R. Rao: Random Matrix Theory, Acta Numerica, pp. 1-65. Cambridge University Press, Cambridge (2005)
- [20] G. Akemann, G.M. Cicuta and L. Molinari: Compact support probability distributions in random matrix theory, Phys Rev E Stat Nonlinear Soft Matter Phys, Vol.59,No.2, pp.1489-1497(1999)
- [21] A. ZEE: Law of addition in random matrix theory, Nucl Phys B, Vol.474, No.3, pp.726-744(1996)

- [22] R.E. Prange: Quasiclassical Random Matrix Theory, Phys Rev Lett, Vol.77, No.12, pp.2447-2450 (1996)
- [23] A. Lakshminarayan, N.R. Cerruti and S. Tomsovic: Classical diffusion and quantum level velocities: Systematic deviations from random matrix theory, Phys. Rev. E Vol.60, No.4, pp. 3992–3999 (1999)
- [24] C. A. Tracy and H. Widom: Level-spacing distributions and the Airy kernel, Comm. Math.Phys. Vol.159, No.1, pp.151-174 (1994)
- [25] C. A. Tracy and H. Widom: On orthogonal and symplectic matrix ensembles, Comm. Math. Phys. Vol.177, pp.727-754 (1996)
- [26] V. Dotsenko: Universal randomness, Phys.-Usp, Vol.54, No.3, 259, doi:10.3367/UFNe.0181.201103b.0269 (2011)
- [27] 金澤拓也, クォーク物質とランダム行列理論, 原子核研究, Vol.55, No.2, pp.22-33 (2011)
- [28] E. Brezin and S. Hikami: Spectral form factor in a random matrix theory, Phys Rev E Stat Nonlinear Soft Matter Phys, Vol.55, No.4, pp.4067-4083 (1997)
- [29] B. Dietz and F. Haake: Random matrix theory as statistical mechanics, Euro phys Lett, Vol.9, No.1, pp.1-6 (1989)
- [30] F. Luo: Application of random matrix theory to biological networks, Phys Lett A, Vol.357, No.6, pp.420-423(2006)
- [31] P.L. Yang, R. Li and J. Chen: Application of random matrix theory to microarray data for discovering functional gene Modules of hepatocellular carcinoma, Shengwu Wuli Xuebao, Vol.25, No.3, pp.192-202 (2009)
- [32] R. Li, P.L. Yan and J. Chen: Application of random matrix theory to identification of lung cancer gene networks, Wuli Xuebao, Vol.58, No.10, pp.6703-6708 (2009)
- [33] 相馬亘: 経済物理学の実用化に向けて—ランダム行列理論からのアプローチ, 応用数理, Vol.15, No.3, pp.239-253 (2005)
- [34] 家富洋: “経済物理学”景気循環の物理: ランダム行列理論と揺動散逸定理, パリティ, Vol.27, No.6, pp.13-16 (2012)
- [35] 新井優太, 家富洋, 相馬亘: ランダム行列理論を用いた Hi-net 地震時系列データの解析, 日本物理学会講演概要集, 巻 66, 号 2 (2011)
- [36] J. Daly, M. Crane and H.J. Ruski: Random matrix theory filters in portfolio optimisation: A stability and risk assessment, Phys A Stat Mech Appl, Vol.387, No.16-17, pp.4248-4260 (2008)
- [37] Y. Nakayama and H. Iyetomi: Random matrix theory of dynamical cross correlations in financial data, Prog Theor Phys Suppl, Vol.179, 60-70 (2009)
- [38] Z.X. Xu, Y. Wang and H.B. Si: Analysis of urban human mobility behavior based on random matrix theory, Wuli xuebao, Vol.60, No.4 pp.040501.1-7 (2011)
- [39] J. Liu, D.P. Jin and W.Z. Zhang: Monitoring of link patterns based on random matrix theory, Qinghua Daxue Xuebao (Ziran kexue ban), Vol.50, No.1, pp.117-120 (2010)
- [40] S. Wold, K. Esbensen and P. Geladi: Principal component analysis, Chemometr Intell Lab Syst, Vol.2, No.1-3, pp.37-52 (1987)
- [41] 相馬亘: ランダム行列理論に基づいた新たな相関分析手法, 日本大学理工学部学術講演

- 会予稿集, Vol.56, O-23, pp.1287-1288 (2009)
- [42] V.A. Marcenko and L.A. Pastur: Distribution of eigenvalues for some sets of random matrices, *Mathematics of the USSR-Sbornik*, Vol.1-(4), pp. 457-483(1994)
- [43] M.O. Ulfarsson and V. Solo: Dimension estimation in noisy PCA with SURE and random matrix theory, *IEEE Trans Signal Process*, Vol.56, No.12, pp.5804-5816 (2008)
- [44] 青山秀明, 他 4 名 : 経済物理学, 共立出版, 第 5 章 (2008)
- [45] 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, 電子情報通信学会・情報処理学会, pp.153-156 (2010)
- [46] M. Tanaka-Yamawaki: Cross correlation of intra-day stock prices in comparison to Random Matrix Theory, *Intelligent Information Management*, Vol. 3, pp.65-70 (2011)
- [47] M. Tanaka-Yamawaki, T. Kido and R. Itoi: Trend extraction of stock prices in the american market by means of RMT-PCA, J. Watada et. al. (Eds.), *Intelligent Decision Technologies, SIST Vol. 10*, pp. 637-646, doi:10.1007/973-642-22194-1 (2011)
- [48] 木戸丈剛, 楊 欣, 田中美栄子, 高石哲弥:ランダム行列理論を用いた主成分抽出法による日本と米国の株式市場における主要セクタの変遷,情報処理学会論文誌数理モデル化と応用 4 巻 4 号, pp.104-110 (2011)
- [49] 脇本和昌 : 乱数の知識, 森北出版 (1970)
- [50] L. Hanne: PLAB a system for testing random numbers, *Proceedings of the International Workshop Parallel Numerics'94*, pp.89-99 (1994)
- [51] ENT: A pseudorandom number sequence test program(online), <http://www.fourmilab.ch/random/>
- [52] The Information Security Research Centre at Queensland University of Technology, "Crypt-X," (<http://www.isrc.qut.edu.au/resource/cryptx/>)
- [53] NIST: A statistical test suite (online), [csrc.nist.gov/groups/ST/toolkit/rng/documentation\\_software.html](http://csrc.nist.gov/groups/ST/toolkit/rng/documentation_software.html) (2010.08.13)
- [54] Dieharder: A random number test suite version 3.31.1 (online),<http://www.phy.duke.edu/rgb/General/dieharder.php>.
- [55] 木戸丈剛, 田中美栄子 : ランダム行列理論との比較による米国株価日次変動のトレンド抽出, FIT2010:第9回情報科学技術フォーラム講演論文集, pp.157-162 (2010)
- [56] 小国力 : Fortran 95, C&JAVAによる新数値計算法—数値計算とデータ分析, サイエンス社 (1997)
- [57] 尾山壮一, 伊藤保昌, 上原政之: 行列を分割した場合の固有値問題とその主成分分析における変数増減への適用について II 固有値回転行列理論の主成分分析への適用, 日本品質管理学会年次大会講演・研究発表要旨集, Vol.15, pp.46-49 (1985)
- [58] 楊欣, 糸井良太, 田中美栄子 : RMT 公式を用いた乱数度評価法の提案, 数理モデル化と問題解決研究報告, Vol.2012-MPS-83, No.2, 2011.5.10
- [59] 田中 美栄子, 糸井 良太, 楊 欣 : RMT 公式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)
- [60] 楊欣, 糸井良太, 田中美栄子 : モーメント分析法によるランダム行列理論を用いた乱数度評価法の改良, 数理モデル化と問題解決研究報告, Vol.2012-MPS-88, No.16, 2012.5.10

- [61] 楊欣, 三賀森悠太, 田中美栄子: 乱数度計 RMT テストの実データへの応用～ハッシュ値と Tick 株価～, 数理モデル化と問題解決研究報告, Vol.2012-MPS-91, No.33,2012.11.29
- [62] 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, No.1, pp.57-63 (2013)
- [63] W.G. Cochran: Some methods for strengthening the common  $\chi^2$  tests. Biometrics, Vol.10, 417-451 (1954)
- [64] 松本 真: 有限体の擬似乱数の応用(2004)  
<http://www.math.sci.hiroshima-u.ac.jp/m-mat/TEACH/0407-2.pdf>.
- [65] M. Matsumoto and T. Nishimura: Mersenne Twister: A Dimensionally Equidistributed Uniform Pseudorandom Number Generator, ACM Transactions on Modeling and Computer Simulation, Vol.8, pp.3–30 (1998)
- [66] Y. Tamura: Random Number Library (2010). <http://random.ism.ac.jp/random>.
- [67] 吉田等明, 村上武, 川村暁: NIST SP800-22rev.1a による疑似乱数の検定に関する考察, 電子通信学会技術研究報告, Vol.112, No.301, pp.13-18 (2012)
- [68] 安田 幹, 佐々木 悠: 暗号学的ハッシュ関数—安全神話の崩壊と新たなる挑戦, 電子情報通信学会, 基礎・境界ソサイテイ, Fundamentals Review, Vol.4, No.1, pp.57-67 (2010)
- [69] J. Black, M. Cochran and T. Highland: A Study of the MD5 Attacks: Insights and Improvements (online), <http://www.cs.colorado.edu/~jrblack/papers/md5efull.pdf> (2008.07.27)
- [70] M. Locktyukhin and K. Farrel: Improving the Performance of the Secure Hash Algorithm (SHA-1) <http://software.intel.com/en-us/articles/improvingthe-performance-of-the-secure-hash-algorithm-1/> (2010.03.29)
- [71] 楊 欣, 糸井 良太, 田中 美栄子: ランダム行列理論を用いた乱数度評価法の提案, 統計数理研究所共同研究レポート, Vol.271, 経済物理とその周辺(8), pp.19-31 (2012)
- [72] 楊 欣, 糸井 良太, 田中 美栄子: 乱数度計 RMT テストの実データへの応用—ハッシュ値と Tick 株価, 統計数理研究所共同研究レポート 292, 経済物理とその周辺(9), pp.70-83 (2013)
- [73] 谷口礼偉, 64 ビット整数の乗算とシフトによる非再帰的な擬似乱数の生成と検定, 統計数理研究所共同研究レポート 210, 乱数の応用指向性評価とその周辺 (2008)
- [74] 梅野 健, 決定論的カオスと「RanSure」による乱数の特性評価, 統計数理研究所共同研究レポート 210, 乱数の応用指向性評価とその周辺 (2008)
- [75] 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」(2003)

# 付 表

2007年の東証1分足から求めた乱数度順位(全)

順位	業種	企業コード	誤差	順位	業種	企業コード	誤差
1	電気・ガス業	9504	26.4	38	精密機器	4543	94.2
2	機械	6460	37.6	39	精密機器	7741	95.7
3	電気・ガス業	9506	38.2	40	輸送用機器	6902	97.7
4	電気・ガス業	9508	43.3	41	小売業	3382	99.0
5	情報・通信業	4676	44.9	42	医薬品	4523	100.3
6	陸運業	9021	48.9	43	電気・ガス業	9532	100.3
7	金属製品	5938	49.3	44	不動産業	8848	100.5
8	サービス業	4324	49.7	45	精密機器	7729	101.2
9	電気・ガス業	9502	52.7	46	保険業	8766	102.3
10	機械	7013	54.9	47	ガラス・土石製品	5233	104.5
11	空運業	9202	58.0	48	電気機器	6724	106.3
12	銀行業	8306	59.1	49	輸送用機器	7259	114.4
13	情報・通信業	9766	61.7	50	機械	6481	116.2
14	鉱業	1605	62.3	51	機械	6367	116.2
15	不動産業	8933	64.5	52	輸送用機器	7012	117.8
16	電気・ガス業	9503	64.9	53	電気機器	6967	123.7
17	輸送用機器	7211	65.6	54	情報・通信業	4684	124.8
18	輸送用機器	7269	67.2	55	卸売業	8058	125.7
19	サービス業	2432	67.8	56	繊維製品	3401	126.8
20	機械	7004	69.6	57	食料品	2811	127.7
21	建設業	1801	70.6	58	医薬品	4502	132.4
22	電気機器	6504	71.6	59	情報・通信業	9613	133.4
23	証券業	8698	72.8	60	輸送用機器	7267	136.2
24	その他金融業	8515	72.9	61	化学	4901	137.0
25	銀行業	8308	74.1	62	その他金融業	8253	137.4
26	化学	4208	74.8	63	水産・農林業	1332	137.5
27	化学	4004	76.0	64	食料品	2809	140.3
28	医薬品	4568	76.2	65	鉄鋼	5406	144.2
29	銀行業	8316	78.2	66	サービス業	4732	144.7
30	建設業	1808	78.8	67	繊維製品	3402	146.5
31	その他金融業	8572	79.8	68	医薬品	4519	146.5
32	不動産業	3231	80.2	69	情報・通信業	4307	150.4
33	情報・通信業	9984	82.4	70	化学	4217	152.0
34	卸売業	8015	84.4	71	情報・通信業	9433	153.8
35	情報・通信業	9437	87.4	72	ガラス・土石製品	5202	154.1
36	銀行業	8303	89.0	73	電気機器	6767	156.1
37	電気・ガス業	9501	90.3	74	食料品	2914	158.5



75	化学	4185	161.1	118	銀行業	8411	242.5
76	石油・石炭製品	5002	161.7	119	石油・石炭製品	5007	243.4
77	非鉄金属	5803	162.4	120	非鉄金属	5801	243.8
78	電気機器	6923	163.4	121	電気機器	6752	248.9
79	機械	6305	163.5	122	鉄鋼	5401	251.4
80	不動産業	8802	167.8	123	不動産業	8801	255.6
81	小売業	2730	168.0	124	輸送用機器	7003	264.0
82	情報・通信業	9432	169.6	125	医薬品	4503	267.3
83	輸送用機器	7272	171.8	126	鉄鋼	5407	271.2
84	建設業	1812	173.0	127	化学	4188	273.8
85	非鉄金属	5727	173.7	128	非鉄金属	5706	274.7
86	卸売業	2768	174.9	129	化学	3407	278.0
87	機械	6473	181.5	130	化学	3405	280.8
88	不動産業	8830	183.4	131	建設業	1803	281.5
89	陸運業	9020	184.1	132	輸送用機器	7202	282.8
90	非鉄金属	5711	194.7	133	情報・通信業	9404	283.3
91	その他製品	7984	195.6	134	電気機器	6857	283.7
92	ゴム製品	5108	201.0	135	小売業	2685	288.0
93	その他金融業	8593	201.3	136	銀行業	8309	288.5
94	機械	6301	202.8	137	卸売業	8053	289.0
95	化学	4063	204.6	138	化学	6988	289.2
96	サービス業	9735	205.7	139	電気機器	6502	296.3
97	精密機器	7762	205.8	140	鉄鋼	5411	298.4
98	情報・通信業	4689	206.9	141	鉄鋼	5471	309.6
99	輸送用機器	7240	207.4	142	卸売業	8002	311.5
100	小売業	8267	209.5	143	ゴム製品	5110	312.9
101	化学	4042	210.0	144	小売業	9983	313.7
102	化学	4005	211.8	145	証券業	8473	314.0
103	電気機器	6758	215.0	146	電気機器	6503	320.2
104	機械(32)	6361	217.7	147	建設業	1928	322.2
105	陸運業	9005	218.6	148	機械	6586	323.9
106	鉄鋼	5405	219.6	149	不動産業	4666	325.7
107	卸売業	7459	219.7	150	化学	4183	329.9
108	食料品	2502	224.1	151	情報・通信業	8056	331.9
109	その他金融業	8570	224.2	152	卸売業	8031	337.6
110	電気・ガス業	9531	224.9	153	電気機器	6773	340.1
111	機械	6370	227.3	154	電気機器	6753	341.2
112	機械	7011	232.0	155	その他製品	7832	344.0
113	銀行業	8332	232.8	156	海運業	9132	347.4
114	小売業	8252	237.1	157	電気機器	8035	354.1
115	電気機器	6952	240.5	158	金属製品	3436	361.9

116	電気機器	7751	242.0	159	機械	6136	370.9
117	輸送用機器	7203	242.4	160	その他金融業	8511	375.0
161	小売業	8233	377.3	199	鉄鋼	5423	619.4
162	情報・通信業	9697	379.6	200	ガラス・土石製品	5214	645.1
163	銀行業	8331	380.9	201	建設業	1925	647.2
164	電気機器	6971	382.2	202	海運業	9101	672.2
165	その他金融業	8591	383.0	203	電気機器	6841	673.4
166	その他製品	7912	386.7	204	電気機器	4062	673.7
167	ガラス・土石製品	5301	387.9	205	機械	6471	678.7
168	輸送用機器	7261	391.0	206	機械	6273	679.9
169	陸運業	9064	392.2	207	電気機器	6506	740.9
170	小売業	9831	394.9	208	非鉄金属	5802	797.3
171	保険業	8795	395.0	209	化学	4043	799.8
172	非鉄金属	5726	399.4	210	鉄鋼	5541	1001.5
173	電気機器	6501	401.2	211	輸送用機器	7201	1209.6
174	証券業	8604	423.3				
175	非鉄金属	5713	424.7				
176	電気機器	6770	427.5				
177	非鉄金属	5714	435.1				
178	電気機器	6702	435.6				
179	証券業	8628	436.2				
180	食料品	2503	439.9				
181	化学	4401	443.6				
182	食料品	2802	444.8				
183	電気機器	6701	455.3				
184	電気機器	4902	457.3				
185	機械	5631	460.1				
186	ガラス・土石製品	5201	483.8				
187	海運業	9107	486.0				
188	機械	6302	502.5				
189	卸売業	8001	507.2				
190	電気機器	6762	544.5				
191	機械	6103	545.7				
192	不動産業	8804	562.7				
193	証券業	8601	564.2				
194	機械	6326	570.9				
195	海運業	9104	573.4				
196	電気機器	6954	599.4				
197	不動産業	8815	609.6				
198	化学	4182	616.6				

2008(Jan.-Aug.)のデータによる乱数度順位(全)

順位	業種	企業 コード	誤差	順位	業種	企業 コード	誤差
1	電気・ガス業	9506	11.6	40	輸送用機器	7012	44.5
2	電気機器	6728	13.1	41	精密機器	4543	44.6
3	食料品	2267	19.1	42	食料品	2593	44.6
4	電気・ガス業	9502	19.1	43	不動産業	8933	45.0
5	小売業	2685	22.2	44	情報・通信業	4676	45.6
6	食料品	2810	25.2	45	陸運業	9005	45.6
7	非鉄金属	5711	25.5	46	情報・通信業	9697	46.2
8	情報・通信業	4307	26.8	47	電気・ガス業	9503	46.2
9	その他金融業	8593	28.2	48	電気機器	6674	46.3
10	陸運業	9020	29.4	49	サービス業	4732	46.3
11	電気・ガス業	9532	30.6	50	その他金融業	8586	46.4
12	電気・ガス業	9508	32.3	51	小売業	2730	46.6
13	電気・ガス業	9531	32.4	52	建設業	1808	47.5
14	小売業	8028	33.0	53	化学	4004	47.9
15	サービス業	4324	33.6	54	小売業	8273	48.4
16	卸売業	9962	34.6	55	鉱業	1605	48.5
17	機械	6460	35.0	56	情報・通信業	4689	48.6
18	電気機器	6770	36.2	57	その他製品	7951	48.7
19	情報・通信業	9432	36.3	58	食料品	2502	48.8
20	不動産業	3231	36.7	59	空運業	9202	49.3
21	水産・農林業	1332	37.3	60	輸送用機器	7269	50.4
22	小売業	2651	37.6	61	不動産業	4666	51.4
23	建設業	1925	37.6	62	化学	4042	51.5
24	機械	7004	38.5	63	ゴム製品	5110	52.2
25	鉄鋼	5406	38.8	64	その他金融業	8515	52.7
26	医薬品	4519	39.9	65	機械	7013	53.1
27	ガラス・土石製品	5233	41.2	66	卸売業	2768	53.7
28	食料品	2809	41.3	67	情報・通信業	9613	53.9
29	小売業	8282	41.8	68	医薬品	4523	54.0
30	その他金融業	8511	42.2	69	卸売業	8060	54.7
31	非鉄金属	5803	42.6	70	機械	6417	56.0
32	保険業	8766	42.9	71	電気機器	4062	56.1
33	化学	4401	43.0	72	医薬品	4503	56.3
34	小売業	9843	43.2	73	繊維製品	3401	56.5
35	化学	4208	43.3	74	建設業	1812	56.8
36	電気・ガス業	9501	43.4	75	ガラス・土石製品	5333	57.0
37	その他金融業	8572	43.4	76	食料品	2503	58.1
38	機械	6370	44.1	77	電気機器	6501	58.4
39	銀行業	8303	44.5	78	医薬品	4502	58.6

79	輸送用機器	7240	58.7	122	化学	6988	78.4
80	情報・通信業	8056	58.8	123	卸売業	8058	79.1
81	医薬品	4568	59.0	124	非鉄金属	5714	80.2
82	サービス業	2432	59.4	125	小売業	8252	80.5
83	情報・通信業	9404	59.5	126	電気機器	6752	81.0
84	小売業	3382	60.3	127	精密機器	7762	81.2
85	情報・通信業	9766	60.5	128	不動産業	8815	82.1
86	証券業	8698	60.6	129	非鉄金属	5801	82.4
87	輸送用機器	7003	60.7	130	輸送用機器	7259	83.0
88	輸送用機器	7230	60.9	131	電気機器	6724	83.5
89	陸運業	9021	61.4	132	不動産業	8802	83.6
90	その他製品	7966	61.6	133	機械	7011	83.6
91	情報・通信業	9984	61.8	134	機械	6305	83.8
92	小売業	7532	62.0	135	金属製品	3436	83.9
93	輸送用機器	6902	62.7	136	鉄鋼	5401	84.4
94	化学	4217	62.8	137	不動産業	8848	84.9
95	その他金融業	8253	63.1	138	精密機器	7729	85.2
96	陸運業	9022	64.7	139	化学	4185	86.4
97	電気機器	6923	65.4	140	その他製品	7832	87.5
98	機械	6136	65.7	141	情報・通信業	9433	88.0
99	金属製品	5901	66.2	142	電気機器	6448	88.1
100	鉄鋼	5405	66.6	143	陸運業	9064	88.3
101	小売業	8233	67.6	144	機械	6326	88.6
102	機械	6586	84.9	145	電気機器	6767	89.1
103	銀行業	8308	68.0	146	輸送用機器	7202	89.4
104	銀行業	8331	68.5	147	鉄鋼	5471	90.6
105	サービス業	9735	68.6	148	卸売業	8031	90.7
106	輸送用機器	7251	68.7	149	卸売業	7459	91.1
107	機械	6301	70.1	150	輸送用機器	6995	91.3
108	輸送用機器	7270	70.9	151	食料品	2802	92.9
109	金属製品	5938	72.7	152	不動産業	8801	93.4
110	銀行業	8411	73.6	153	機械	6103	93.6
111	機械	5631	73.8	154	小売業	8267	94.2
112	電気機器	6925	74.4	155	鉄鋼	5541	94.3
113	輸送用機器	7211	74.4	156	銀行業	8355	94.5
114	食料品	2914	75.5	157	不動産業	8830	94.6
115	化学	3407	76.2	158	輸送用機器	7203	94.7
116	輸送用機器	3116	76.3	159	電気機器	6702	96.2
117	その他金融業	8570	76.8	160	鉄鋼	5423	97.2
118	精密機器	7741	77.2	161	銀行業	8309	97.7
119	情報・通信業	9437	77.6	162	機械	6273	97.7
120	電気機器	6952	77.6	163	輸送用機器	7261	97.8
121	卸売業	8015	78.2	164	化学	4183	97.9

165	輸送用機器	7267	98.6	201	電気機器(54)	6841	123.2
166	化学	4188	99.5	202	電気機器(54)	7751	123.7
167	海運業	9132	101.2	203	電気機器(54)	4902	124.8
168	化学	4063	101.5	204	電気機器(54)	6773	126.0
169	海運業(4)	9107	101.5	205	精密機器(8)	7731	127.1
170	ガラス・土石製品	5202	101.5	206	不動産業(11)	8804	128.7
171	電気機器	6971	103.3	207	その他製品(9)	7911	130.6
172	証券業	8628	105.0	208	非鉄金属(11)	5726	134.2
173	海運業	9101	106.6	209	証券業(9)	8601	134.9
174	電気機器	6701	107.3	210	機械(32)	6471	135.2
175	情報・通信業	4684	108.0	211	電気機器(54)	6967	137.8
176	機械	6302	108.1	212	非鉄金属(11)	5802	137.9
177	小売業	9831	108.1	213	銀行業(39)	8316	138.4
178	その他金融業	8591	108.9	214	機械(32)	6581	140.3
179	輸送用機器	7272	109.4	215	ガラス・土石製品(9)	5201	142.8
180	化学	4182	110.9	216	化学(40)	3405	143.2
181	化学	4901	112.0	217	ガラス・土石製品(9)	5301	143.3
182	電気機器	6753	112.3	218	建設業(17)	1928	144.1
183	その他製品	7912	112.7	219	非鉄金属(11)	5713	148.8
184	繊維製品	3402	112.8	220	小売業(27)	9983	155.3
185	石油・石炭製品	5002	113.5	221	卸売業(17)	8036	160.1
186	鉄鋼	5411	114.1	222	電気機器(54)	8035	160.4
187	機械	6113	114.6	223	証券業(9)	8473	170.6
188	ガラス・土石製品	5214	115.4	224	卸売業(17)	8053	171.3
189	卸売業	8002	115.4	225	電気機器(54)	6954	173.3
190	電気機器	6758	116.1	226	電気機器(54)	7752	173.6
191	電気機器	6506	116.3	227	機械(32)	6367	177.2
192	化学	4043	116.6	228	海運業(4)	9104	178.5
193	保険業	8795	116.6	229	電気機器(54)	6502	180.4
194	ゴム製品	5108	118.3	230	電気機器(54)	6762	183.0
195	銀行業	8332	119.1	231	電気機器(54)	6503	183.7
196	機械	6481	119.9	232	銀行業(39)	8306	190.0
197	電気機器	6857	121.3	233	証券業(9)	8604	222.8
198	化学	4005	121.3	234	輸送用機器(26)	7201	257.9
199	機械	6473	121.7				
200	卸売業	8001	121.8				

2009年の東証1分足から求めた乱数度順位(全)

順位	業種	企業コード	誤差	順位	業種	企業コード	誤差
1	電気・ガス業	9508	25.5	40	卸売業	2768	91.6
2	電気・ガス業	9509	28.1	41	電気機器	6501	92.5
3	電気・ガス業	9506	31.6	42	化学	4183	93.6
4	電気・ガス業	9502	33.2	43	小売業	8028	98.8
5	小売業	2651	36.4	44	情報・通信業	9433	102.1
6	陸運業	9021	37.1	45	情報・通信業	3626	102.6
7	食料品	2809	44.6	46	鉄鋼	5401	104.1
8	電気・ガス業	9531	45.8	47	機械	6349	105.9
9	電気・ガス業	9503	47.0	48	その他金融業	8572	107.6
10	電気・ガス業	9501	48.8	49	電気機器	6723	108.5
11	機械	7013	52.5	50	非鉄金属	5711	108.9
12	電気・ガス業	9532	52.8	51	電気機器	6773	109.6
13	空運業	9202	54.4	52	サービス業	9735	111.5
14	その他金融業	8515	55.8	53	小売業	9843	115.9
15	輸送用機器	7211	56.5	54	電気機器	6674	116.8
16	化学	4004	56.6	55	情報・通信業	9984	118.1
17	輸送用機器	7012	58.1	56	小売業	2685	119.6
18	ガラス・土石製品	5233	58.3	57	パルプ・紙	3893	120.6
19	鉄鋼	5406	59.2	58	食料品	2267	125.8
20	医薬品	4502	62.1	59	電気機器	6845	127.2
21	銀行業	8303	62.8	60	小売業	2730	127.3
22	建設業	1808	63.4	61	情報・通信業	9437	130.9
23	情報・通信業	9432	65.2	62	食料品	2802	131.0
24	陸運業	9022	65.6	63	海運業	9101	132.5
25	銀行業	8316	72.7	64	化学	4188	132.7
26	機械	7011	74.4	65	その他金融業	8253	135.6
27	輸送用機器	7202	75.8	66	小売業	9831	137.8
28	鉄鋼	5405	76.0	67	電気機器	6849	138.2
29	医薬品	4523	76.0	68	情報・通信業	9684	138.9
30	食料品	2810	76.3	69	小売業	3382	144.3
31	電気機器	6967	80.1	70	輸送用機器	7203	146.3
32	保険業	8729	80.4	71	精密機器	4543	146.5
33	医薬品	4503	80.6	72	小売業	3099	147.1
34	陸運業	9020	82.2	73	非鉄金属	5801	148.8
35	銀行業	8411	83.2	74	電気機器	6925	151.4
36	化学	4208	83.8	75	医薬品	4506	158.8
37	輸送用機器	7261	85.4	76	金属製品	5938	161.8
38	化学	4042	85.7	77	情報・通信業	9602	165.1
39	化学	4061	86.9	78	ガラス・土石製品	5202	168.8

79	電気機器	6701	170.5	122	情報・通信業	9613	272.3
80	化学	3407	170.7	123	精密機器	7762	272.9
81	銀行業	8309	173.9	124	電気機器	6448	280.5
82	小売業	7532	176.1	125	輸送用機器	7251	281.0
83	繊維製品	3402	177.2	126	不動産業	8804	289.1
84	化学	4452	177.2	127	機械	6367	289.7
85	その他製品	7832	180.6	128	輸送用機器	7269	292.5
86	機械	6370	183.8	129	輸送用機器	7270	293.3
87	不動産業	8933	185.6	130	証券業	8601	296.7
88	保険業	8766	187.1	131	不動産業	4666	299.4
89	医薬品	4507	189.4	132	卸売業	8002	299.8
90	輸送用機器	7312	193.6	133	電気機器	6767	301.0
91	化学	7988	201.7	134	化学	4901	301.0
92	銀行業	8308	202.2	135	銀行業	8306	304.3
93	化学	4005	202.3	136	機械	6302	308.2
94	サービス業	2432	205.4	137	電気機器	6770	309.2
95	海運業	9107	205.9	138	電気機器	6923	313.9
96	保険業	8795	211.0	139	情報・通信業	4307	316.0
97	証券業	8604	213.3	140	ガラス・土石製品	5333	318.2
98	医薬品	4519	214.0	141	機械	6460	320.7
99	証券業	8628	217.4	142	精密機器	7733	323.4
100	鉄鋼	5411	227.8	143	その他製品	7966	323.6
101	その他製品	7951	229.2	144	化学	4043	325.1
102	化学	8113	233.5	145	保険業	8725	330.4
103	化学	4063	233.7	146	情報・通信業	7518	338.4
104	電気機器	6502	233.8	147	金属製品	5901	340.3
105	小売業	8267	234.7	148	電気機器	6702	342.2
106	小売業	8252	235.9	149	電気機器	6506	345.9
107	鉄鋼	5423	237.5	150	その他金融業	8591	346.1
108	その他金融業	8570	237.7	151	輸送用機器	7259	355.8
109	建設業	1928	239.3	152	輸送用機器	7282	362.0
110	情報・通信業	9404	239.4	153	不動産業	8905	363.5
111	不動産業	8815	239.4	154	海運業	9104	370.6
112	電気機器	7751	241.2	155	電気機器	6996	377.0
113	電気機器	6965	244.7	156	建設業	1925	377.2
114	精密機器	7741	246.1	157	食料品	2503	377.4
115	卸売業	8060	247.8	158	その他金融業	8593	378.0
116	情報・通信業	4689	252.5	159	電気機器	6952	380.5
117	情報・通信業	9697	254.9	160	情報・通信業	9766	380.6
118	サービス業	4324	265.0	161	銀行業	8332	380.9
119	卸売業	7459	265.5	162	食料品	2502	383.8
120	ゴム製品	5110	270.2	163	電気機器	8035	392.6
121	医薬品	4568	271.0	164	電気機器	6762	393.6

165	機械	6326	397.5	201	機械	6471	601.3
166	情報・通信業	4676	399.1	202	不動産業	8802	602.6
167	石油・石炭製品	5002	402.0	203	電気機器	6954	607.1
168	小売業	9983	404.5	204	機械	6305	637.8
169	電気機器	6841	405.3	205	電気機器	6753	643.7
170	電気機器	4062	407.4	206	輸送用機器	7272	658.5
171	化学	3405	410.2	207	電気機器	6976	673.3
172	不動産業	3231	411.5	208	証券業	8473	676.2
173	非鉄金属	5714	417.4	209	化学	4185	679.4
174	卸売業	8036	440.2	210	電気機器	6758	689.1
175	化学	4911	447.4	211	電気機器	6857	700.6
176	機械	6586	449.4	212	不動産業	8830	702.5
177	電気機器	6728	455.1	213	建設業	1963	726.4
178	機械	6481	456.1	214	不動産業	8801	730.1
179	輸送用機器	6995	457.0	215	電気機器	7752	786.1
180	化学	4217	462.9	216	非鉄金属	5802	806.2
181	電気機器	6724	467.0	217	鉱業	1605	836.0
182	不動産業	8848	467.8	218	ゴム製品	5108	836.4
183	機械	5631	468.9	219	金属製品	3436	863.4
184	輸送用機器	7240	474.4	220	陸運業	9064	891.3
185	機械	6146	476.9	221	卸売業	8031	913.2
186	輸送用機器	7267	488.7	222	ガラス・土石製品	5214	974.8
187	輸送用機器	7230	493.9	223	食料品	2914	1016.5
188	機械	6473	507.0	224	精密機器	7731	1017.3
189	ガラス・土石製品	5201	515.6	225	非鉄金属	5713	1039.5
190	卸売業	8001	521.3	226	電気機器	4902	1072.4
191	電気機器	6503	522.2	227	機械	6301	1090.9
192	証券業	8698	525.0	228	鉄鋼	5541	1128.2
193	輸送用機器	7201	535.3	229	卸売業	8058	1249.5
194	輸送用機器	6902	535.3				
195	電気機器	6971	536.9				
196	化学	6988	541.0				
197	卸売業	8015	556.0				
198	電気機器	6752	558.5				
199	卸売業	8053	595.2				
200	輸送用機器	3116	598.7				



## 分析対象とした銘柄

本論文の分析対象は 2009 年までの TOPIX500 の銘柄である。上場廃止等に伴い、TOPIX 500 の構成銘柄は毎年変動している。よって、2013 年現在上場廃止した 25 銘柄を除外し、残る 475 銘柄は分析対象となる。詳しくは下表に示す。

<p>上場廃止の 25 銘柄：2202(明治製菓 食料品),2261(明治乳業 食料品),3404(三菱レイヨン 繊維製品),4535(大正製薬 医薬品),5001(新日本石油 石油・石炭製品),5016,(新日鉱ホールディングス 石油・石炭製品),5855(アサヒプリテック 非鉄金属),6665(エルピーダメモリ 電気機器),6764(三洋電機 電気機器),6991(パナソニック 電工 電気機器),8112(東京スタイル 繊維製品),8354(ふくおかフィナンシャルグループ 銀行業),8403(住友信託銀行 銀行業),8404(みずほ信託銀行 銀行業),8564(武富士 その他金融業),8574(プロミス その他金融業),8606(みずほ証券 証券業),8607(みずほインベスターズ証券 証券業),8754(日本興亜損害保険 保険業),8755(損保ジャパン 保険業),8759(ニッセイ同和損害保険 保険業),8761(あいおい損害保険 保険業),8763(富士火災海上保険 保険業),9205(日本航空 空運業),9737(CSK 通信業)</p>	
<p>分析対象：</p> <p>1332 日本水産 水産・農林業</p> <p>1605 国際石油開発帝石 鉱業</p> <p>1662 石油資源開発 鉱業</p> <p>1721 コムシスホールディングス 建設業</p> <p>1801 大成建設 建設業</p> <p>1802 大林組 建設業</p> <p>1803 清水建設 建設業</p> <p>1808 長谷工コーポレーション 建設業</p> <p>1812 鹿島 建設業</p> <p>1820 西松建設 建設業</p> <p>1833 奥村組 建設業</p> <p>1860 戸田建設 建設業</p> <p>1878 大東建設 建設業</p> <p>1911 住友林業 建設業</p> <p>1925 大和ハウス工業 建設業</p> <p>1928 積水ハウス 建設業</p> <p>1944 きんでん 建設業</p> <p>1951 協和エクシオ 建設業</p> <p>1963 日揮 建設業</p> <p>2002 日清製粉グループ本社 食料品</p> <p>2206 江崎グリコ 食料品</p> <p>2212 山崎製パン 食料品</p> <p>2267 ヤクルト本社 食料品</p> <p>2282 日本ハム 食料品</p> <p>2331 総合警備保障 サービス業</p> <p>2432 ディー・エヌ・イー サービス業</p> <p>2433 博報堂 D Y ホールディングス サービス業</p>	<p>2501 サッポロホールディングス 食料品</p> <p>2502 アサヒビール 食料品</p> <p>2503 キリンホールディングス 食料品</p> <p>2531 宝ホールディングス 食料品</p> <p>2579 コカ・コーラウエスト 食料品</p> <p>2593 伊藤園 食料品</p> <p>2651 ローソン 小売業</p> <p>2685 ポイント 小売業</p> <p>2730 エディオン 小売業</p> <p>2768 双日 卸売業</p> <p>2784 アルフレッサ ホールディングス 卸売業</p> <p>2801 キッコーマン 食料品</p> <p>2802 味の素 食料品</p> <p>2809 キューピー 食料品</p> <p>2810 ハウス食品 食料品</p> <p>2811 カゴメ 食料品</p> <p>2871 ニチレイ 食料品</p> <p>2875 東洋水産 食料品</p> <p>2897 日清食品ホールディングス 食料品</p> <p>2914 J T 食料品</p> <p>3002 ゲンゼ 繊維製品</p> <p>3086 Jフロント リテイリング 小売業</p> <p>3088 マツモトキヨシホールディングス 小売業</p> <p>3099 (株)三越伊勢丹ホールディングス</p> <p>3101 東洋紡 繊維製品</p> <p>3105 日清紡ホールディングス 繊維製品</p> <p>3116 トヨタ紡織 輸送用機器</p>

3231 野村不動産ホールディングス 不動産業	4502 武田薬品工業 医薬品
3382 セブン&アイ 小売業	4503 アステラス製薬 医薬品
3401 帝人 繊維製品	4506 大日本住友製薬 医薬品
3402 東レ 繊維製品	4507 塩野義製薬 医薬品
3405 クラレ 化学	4508 田辺三菱製薬 医薬品
3407 旭化成 化学	4519 中外製薬 医薬品
3436 SUMCO 金属製品	4521 科研製薬 医薬品
3591 ワコールホールディングス 繊維製品	4523 エーザイ 医薬品
3626 ITホールディングス(株)	4527 ロート製薬 医薬品
3861 王子製紙 パルプ・紙	4528 小野薬品工業 医薬品
3880 大王製紙 パルプ・紙	4530 久光製薬 医薬品
3893 日本製紙グループ本社 パルプ・紙	4534 持田製薬 医薬品
3941 レンゴー パルプ・紙	4536 参天製薬 医薬品
4004 昭和電工 化学	4540 ツムラ 医薬品
4005 住友化学 化学	4543 テルモ 精密機器
4021 日産化学工業 化学	4544 みらかホールディングス サービス業
4042 東ソー 化学	4547 キッセイ薬品工業 医薬品
4043 トクヤマ 化学	4568 第一三共 医薬品
4044 セントラル硝子 化学	4612 日本ペイント 化学
4045 東亜合成 化学	4613 関西ペイント 化学
4061 電気化学工業 化学	4631 DIC 化学
4062 イビデン 電気機器	4661 オリエンタルランド サービス業
4063 信越化学工業(化学)	4665 ダスキン サービス業
4088 エア・ウォーター 化学	4666 パーク24 サービス業
4091 大陽日酸 化学	4676 フジ・メディア・ホールディングス 通信業
4114 日本触媒 化学	4684 オービック 通信業
4118 カネカ 化学	4689 ヤフー 通信業
4151 協和発酵キリン 医薬品	4704 トレンドマイクロ 通信業
4182 三菱ガス化学 化学	4716 日本オラクル 通信業
4183 三井化学 化学	4732 ユー・エス・エス サービス業
4185 JSR 化学	4739 伊藤忠テクノソリューションズ 通信業
4186 東京応化工業 化学	4768 大塚商会 通信業
4188 三菱ケミカルホールディングス 化学	4901 富士フイルムホールディングス 化学
4202 ダイセル化学工業 化学	4902 コニカミノルタホールディングス 電気機器
4203 住友ベークライト 化学	4911 資生堂 化学
4204 積水化学工業 化学	4912 ライオン 化学
4205 日本ゼオン 化学	4922 コーセイ 化学
4208 宇部興産 化学	4967 小林製薬 化学
4217 日立化成工業 化学	5002 昭和シェル石油 石油・石炭製品
4272 日本化薬 化学	5007 コスモ石油 石油・石炭製品
4307 野村総合研究所 通信業	5012 東燃ゼネラル石油 石油・石炭製品
4324 電通 サービス業	5019 出光興産 石油・石炭製品
4401 ADEKA 化学	
4452 花王 化学	

5101 横浜ゴム ゴム製品	6103 オークマ 機械
5108 ブリヂストン ゴム製品	6113 アマダ 機械
5110 住友ゴム工業 ゴム製品	6136 O S G 機械
5201 旭硝子 ガラス・土石製品	6141 森精機製作所 機械
5202 日本板硝子 ガラス・土石製品	6146 ディスコ 機械
5214 日本電気硝子 ガラス・土石製品	6201 豊田自動織機 輸送用機器
5232 住友大阪セメント ガラス・土石製品	6222 島精機製作所 機械
5233 太平洋セメント ガラス・土石製品	6268 ナブテスコ 機械
5301 東海カーボン ガラス・土石製品	6273 S M C 機械
5332 T O T O ガラス・土石製品	6301 コマツ 機械
5333 日本ガイシ ガラス・土石製品	6302 住友重機械工業 機械
5334 日本特殊陶業 ガラス・土石製品	6305 日立建機 機械
5401 新日本製鐵 鉄鋼	6326 クボタ 機械
5405 住友金属工業 鉄鋼	6349 小森コーポレーション 機械
5406 神戸製鋼所 鉄鋼	6361 荏原 機械
5407 日新製鋼 鉄鋼	6366 千代田化工建設 建設業
5411 ジェイ エフ イー ホールディングス 鉄鋼	6367 ダイキン工業 機械
5423 東京製鐵 鉄鋼	6370 栗田工業 機械
5444 大和工業 鉄鋼	6383 ダイフク 機械
5451 淀川製鋼所 鉄鋼	6395 タダノ 機械
5463 丸一鋼管 鉄鋼	6417 S A N K Y O 機械
5471 大同特殊鋼 鉄鋼	6448 ブラザー工業 電気機器
5481 山陽特殊製鋼 鉄鋼	6457 グローリー 機械
5482 愛知製鋼 鉄鋼	6460 セガサミーホールディングス 機械
5486 日立金属 鉄鋼	6471 日本精工 機械
5541 大平洋金属 鉄鋼	6472 N T N 機械
5631 日本製鋼所 機械	6473 ジェイテクト 機械
5701 日本軽金属 非鉄金属	6474 不二越 機械
5706 三井金属 非鉄金属	6479 ミネベア 電気機器
5711 三菱マテリアル 非鉄金属	6481 T H K 機械
5713 住友金属鉱山 非鉄金属	6501 日立製作所 電気機器
5714 D O W Aホールディングス 非鉄金属	6502 東芝 電気機器
5726 チタニウムテクノロジーズ 非鉄金属	6503 三菱電機 電気機器
5727 東邦チタニウム 非鉄金属	6504 富士電機 電気機器
5801 古河電気工業 非鉄金属	6506 安川電機 電気機器
5802 住友電気工業 非鉄金属	6581 日立工機 機械
5803 フジクラ 非鉄金属	6586 マキタ 機械
5812 日立電線 非鉄金属	6588 東芝テック 電気機器
5901 東洋製罐 金属製品	6592 マブチモーター 電気機器
5929 三和ホールディングス 金属製品	6594 日本電産 電気機器
5938 住生活グループ 金属製品	6645 オムロン 電気機器
5947 リンナイ 金属製品	6674 ジーエス・ユアサ コーポレーション 電 気機器
5991 ニッパツ 金属製品	6701 N E C 電気機器

6702 富士通 電気機器	7203 トヨタ自動車 輸送用機器
6703 O K I 電気機器	7205 日野自動車 輸送用機器
6707 サンケン電気 電気機器	7211 三菱自動車 輸送用機器
6723 ルネサスエレクトロニクス 電気機器	7221 トヨタ車体 輸送用機器
6724 セイコーエプソン 電気機器	7230 日信工業 輸送用機器
6728 アルバック 電気機器	7240 N O K 輸送用機器
6752 パナソニック 電気機器	7241 フタバ産業 輸送用機器
6753 シャープ 電気機器	7251 ケーヒン 輸送用機器
6758 ソニー 電気機器	7259 アイシン精機 輸送用機器
6762 T D K 電気機器	7261 マツダ 輸送用機器
6767 ミツミ電機 電気機器	7262 ダイハツ工業 輸送用機器 7
6770 アルプス電気 電気機器	7267 ホンダ 輸送用機器
6773 パイオニア 電気機器	7269 スズキ 輸送用機器
6804 ホシデン 電気機器	7270 富士重工業 輸送用機器
6806 ヒロセ電機 電気機器	7272 ヤマハ発動機 輸送用機器
6841 横河電機 電気機器	7276 小糸製作所 電気機器
6845 山武 電気機器	7282 豊田合成 輸送用機器
6849 日本光電 電気機器	7309 シマノ 輸送用機器
6856 堀場製作所 電気機器	7312 タカタ 輸送用機器
6857 アドバンテスト 電気機器	7453 良品計画 小売業
6861 キーエンス 電気機器	7459 メディパルホールディングス 卸売業
6869 シスメックス 電気機器	7518 ネットワンシステムズ 通信業
6902 デンソー 輸送用機器	7532 ドン・キホーテ 小売業
6923 スタンレー電気 電気機器	7649 スギホールディングス 小売業
6925 ウシオ電機 電気機器	7701 島津製作所 精密機器
6952 カシオ計算機 電気機器	7729 東京精密 精密機器
6954 ファナック 電気機器	7731 ニコン 精密機器
6963 ローム 電気機器	7733 オリンパス 精密機器
6965 浜松ホトニクス 電気機器	7735 大日本スクリーン製造 電気機器
6967 新光電気工業 電気機器	7741 H O Y A 精密機器
6971 京セラ 電気機器	7751 キヤノン 電気機器
6976 太陽誘電 電気機器	7752 リコー 電気機器
6981 村田製作所 電気機器	7762 シチズンホールディングス 精密機器
6986 双葉電子工業 電気機器	7832 バンダイナムコホールディングス その他製品
6988 日東電工 化学	7911 凸版印刷 その他製品
6995 東海理化 輸送用機器	7912 大日本印刷 その他製品
6996 ニチコン 電気機器	7915 日本写真印刷 その他製品
7003 三井造船 輸送用機器	7936 アシックス その他製品
7004 日立造船 機械	7951 ヤマハ その他製品
7011 三菱重工業 機械	7966 リンテック その他製品
7012 川崎重工業 輸送用機器	7974 任天堂 その他製品
7013 I H I 機械	7984 コクヨ その他製品
7201 日産自動車 輸送用機器	7988 ニフコ 化学
7202 いすゞ自動車 輸送用機器	

8001 伊藤忠商事 卸売業	8339 東京都民銀行 銀行業
8002 丸紅 卸売業	8341 七十七銀行 銀行業
8012 長瀬産業 卸売業	8355 静岡銀行 銀行業
8015 豊田通商 卸売業	8356 十六銀行 銀行業
8016 オンワードホールディングス 卸売業	8358 スルガ銀行 銀行業
8028 ファミリーマート 小売業	8359 八十二銀行 銀行業
8031 三井物産 卸売業	8361 大垣共立銀行 銀行業
8035 東京エレクトロン 卸売業	8363 北國銀行 銀行業
8036 日立ハイテクノロジーズ 卸売業	8366 滋賀銀行 銀行業
8053 住友商事 卸売業	8368 百五銀行 銀行業
8056 日本ユニシス 通信業	8369 京都銀行 銀行業
8058 三菱商事 卸売業	8377 ほくほくフィナンシャルグループ 銀行
8060 キヤノンマーケティングジャパン 卸売	8379 広島銀行 銀行業
8078 阪和興業 卸売業	8381 山陰合同銀行 銀行業
8086 ニプロ 精密機器	8382 中国銀行 銀行業
8113 ユニチャーム 化学	8385 伊予銀行 銀行業
8129 東邦ホールディングス 卸売業	8386 百十四銀行 銀行業
8184 島忠 小売業	8388 阿波銀行 銀行業
8218 コメリ 小売業	8390 鹿児島銀行 銀行業
8219 青山商事 小売業	8394 鹿児島銀行 銀行業
8227 しまむら 小売業	8411 みずほフィナンシャルグループ 銀行業
8233 高島屋 小売業	8415 紀陽ホールディングス 銀行業
8242 エイチ・ツー・オー リテイリング 小売	8418 山口フィナンシャルグループ 銀行業
8252 丸井グループ 小売業	8473 SBIホールディングス 証券業
8253 クレディセゾン その他金融業	8511 日本証券金融 その他金融業
8267 イオン 小売業	8515 アイフル その他金融業
8270 ユニー 小売業	8522 名古屋銀行 銀行業
8273 イズミ 小売業	8544 京葉銀行 銀行業
8282 ケーズホールディングス 小売業	8570 イオンクレジット その他金融業
8303 新生銀行 銀行業	8572 アコム その他金融業
8304 あおぞら銀行 銀行業	8586 日立キャピタル その他金融業
8306 三菱UFJフィナンシャル・グループ 銀行業	8591 オリックス その他金融業
8308 りそなホールディングス 銀行業	8593 三菱UFJリース その他金融業
8309 三井トラスト・ホールディングス 銀行業	8595 ジャフコ 証券業
8316 三井住友フィナンシャルグループ 銀行	8601 大和証券グループ本社 証券業
8324 第四銀行 銀行業	8604 野村ホールディングス 証券業
8327 西日本シティ銀行 銀行業	8609 岡三証券グループ 証券業
8328 札幌北洋ホールディングス 銀行業	8616 東海東京フィナンシャル・ホールディングス 証券業
8331 千葉銀行 銀行業	8628 松井証券 証券業
8332 横浜銀行 銀行業	8698 マネックスグループ 証券業
8333 常陽銀行 銀行業	8703 カブドットコム証券 証券業
8334 群馬銀行 銀行業	8725 MS&ADホールディングス
8336 武蔵野銀行 銀行業	8729 ソニーフィナンシャルホールディングス

保険業	9501 東京電力 電気・ガス業
8766 東京海上ホールディングス 保険業	9502 中部電力 電気・ガス業
8795 T&D ホールディングス 保険業	9503 関西電力 電気・ガス業
8801 三井不動産 不動産業	9504 中国電力 電気・ガス業
8802 三菱地所 不動産業	9505 北陸電力 電気・ガス業
8804 東京建物 不動産業	9506 東北電力 電気・ガス業
8815 東急不動産 不動産業	9507 四国電力 電気・ガス業
8830 住友不動産 不動産業	9508 九州電力 電気・ガス業
8848 レオパレス 2 1 不動産業	9509 北海道電力 電気・ガス業
8905 イオンモール 不動産業	9513 JPOWER 電気・ガス業
8933 NTT 都市開発 不動産業	9531 東京ガス 電気・ガス業
9001 東武鉄道 陸運業	9532 大阪ガス 電気・ガス業
9005 東京急行電鉄 陸運業	9533 東邦ガス 電気・ガス業
9006 京浜急行電鉄 陸運業	9602 東宝 通信業
9007 小田急電鉄 陸運業	9613 NTTデータ 通信業
9008 京王電鉄 陸運業	9684 スクウェア・エニックス・ホールディングス 通信業
9009 京成電鉄 陸運業	9697 カプコン 通信業
9020 東日本旅客鉄道 陸運業	9706 日本空港ビルデング 不動産業
9021 西日本旅客鉄道 陸運業	9735 セコム サービス業
9022 東海旅客鉄道 陸運業	9744 メイテック サービス業
9031 西日本鉄道 陸運業	9747 アサツーディ・ケイ サービス業
9041 近畿日本鉄道 陸運業	9766 コナミ 通信業
9042 阪急阪神ホールディングス 陸運業	9783 ベネッセホールディングス サービス業
9045 京阪電気鉄道 陸運業	9793 ダイセキ サービス業
9048 名古屋鉄道 陸運業	9831 ヤマダ電機 小売業
9062 日本通運 陸運業	9832 オートバックスセブン 卸売業
9064 ヤマトホールディングス 陸運業	9843 ニトリホールディングス 小売業
9065 山九 陸運業	9861 吉野家ホールディングス 小売業
9076 セイノーホールディングス 陸運業	9962 ミスミグループ本社 卸売業
9101 日本郵船 海運業	9983 ファーストリテイリング 小売業
9104 商船三井 海運業	9984 ソフトバンク 通信業
9107 川崎汽船 海運業	9987 スズケン 卸売業
9132 第一中央汽船 海運業	9989 サンドラッグ 小売業
9202 全日本空輸 空運業	
9301 三菱倉庫 倉庫・運輸関連業	
9303 住友倉庫 倉庫・運輸関連業	
9364 上組 倉庫・運輸関連業	
9401 TBSホールディングス 通信業	
9404 日本テレビ放送網 通信業	
9409 テレビ朝日 通信業	
9432 日本電信電話 通信業	
9433 KDDI 通信業	
9435 光通信 通信業	
9437 NTTドコモ 通信業	



## English Summary

A random number plays an important role by various scenes of the information security which protects our life, especially encryption communication. Moreover, the quality of random numbers used in scientific and engineering calculations also affect the accuracy and speed of the calculations.

In order to investigate the quality of a random number, using the statistical technique to analyze the character is common and various statistical inspection methods are proposed. For example, PLAB is used in the simulations field, ENT is used to evaluate the quality of random number generators and the compression algorithm, NIST-SP 800-22, DIEHARD which are used in the cryptography field, etc. However, each methods and tools adopted different type and number of test items, and it is difficult to use them.

Therefore, a tool is proposed to measure the randomness of long sequences based on RMT formula which is called RMT-test. And it is theoretically intelligible, judged visually, apply to lots kind of data format with a single valuation basis. The random matrix theory (RMT, hereafter) has a wide variety of application from the nuclear energy levels to the principal component analysis of stock markets. Especially, the latter example has attracted much attention of the community of econo-physics in which many researchers have been working to discover possible rules or structures behind the motion of stock prices. In this paper, firstly, the theoretical foundation of RMT-test is introduced which is the Random Matrix Theory (RMT) and in Chapter 1 and Principal Component Analysis by Random Matrix Theory (RMT-PCA) in Chapter 2.

In Chapter 3, the RMT-test is proposed as following, the aim is to test the randomness of a long 1-dimensional sequence of numerical data,  $S$ . The algorithm of the qualitative version of the RMT-test is show in Table 1. First, a string  $S$  is prepared to be tested and cut it into  $N$  pieces of length  $L$ ; Each piece,  $S_i = (S_{i,1}, S_{i,2}, \dots, S_{i,L})$ , is converted to a normalized vector  $x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,L})$  of zero mean and unit variance. Since the original sequence  $S$  is random, in general all those vectors are independent. A cross correlation matrix  $C$  is constructed by taking the inner product of two vectors  $x_i$  and  $x_j$  (Eq.1), which is symmetric under the interchange of  $i$  and  $j$  by definition. An eigenvalue problem of  $C$  is solved to obtain  $\lambda_k$  (Eq.2), the  $k$ -th eigenvalue, and its eigenvector. The distribution of the eigenvalue is compared to the theoretical formula (Eq.3) derived in the RMT at the limit of  $N \rightarrow \infty, L \rightarrow \infty, Q=L/N=\text{const}$ , where the upper bound and the lower bound are given by the formula Eq.4. At last, use all of the eigenvalues to draw a histogram, and compare the histogram with theoretical curve to judge the randomness of the data. If the two lines match, that data passes the RMT-test (Fig.1 left), and if they do not match, it fails the RMT-test (Fig.1 right), this is called the qualitative evaluation of RMT-test.

On the other hand, in order to compare a subtle difference of randomness between given random sequences, a quantitative formulation of the randomness-test is developed based on the random matrix theory, for this purpose the moment method is employed in order to compare the shape of the eigenvalue distribution of the cross correlation matrix obtained from the data to the theoretically derived eigenvalue distribution from the random matrix theory, which is so-called Marcenko-Pastur distribution. The algorithm of the qualitative version of the quantitative evaluation of RMT-test is show in Table 2. The first 3 step is the same to qualitative evaluation of RMT-test, the 4<sup>th</sup> step is compare the  $k$ -th moment of the obtained eigenvalues from Eq.5 to the corresponding theoretical formula obtained from Eq.6. Base on the result that have been obtained, a criterion is decided when the length of sequence when  $N \geq 200, Q \geq 3$ , that the sample sequence passes the quantitative RMT-test if the ratio



of the moment  $m_k^{EXP}$  over its theoretical value  $m_k^{RMT}$  is under 5% (Eq.7).

Table 1 Algorithm of the qualitative version of the RMT-test

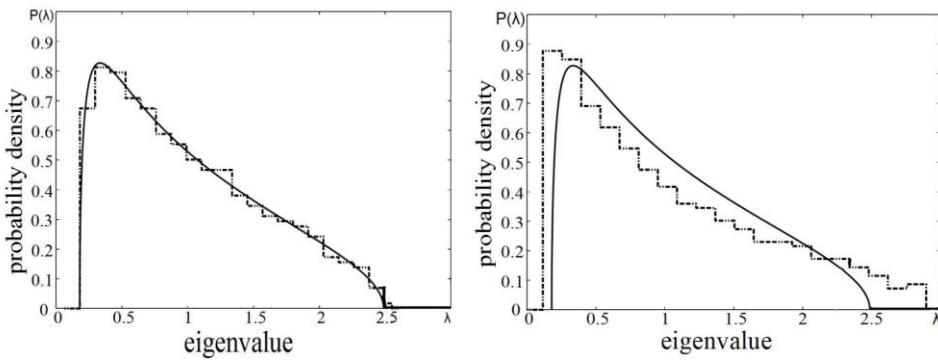
<p>1. Prepare a string to be tested and cut it into N pieces of length T.</p> <p>2. Each piece, <math>S_i</math> is converted to a normalized vector <math>x_i</math>, then make the correlation matrix C in</p> $C_{i,j} = \frac{1}{T} \sum_{t=1}^T x_{i,t} x_{j,t} \quad (1).$ <p>3. Solve an eigenvalue problem of C in <math>CV_k = \lambda_k V_k \quad (k = 1, \dots, N) \quad (2)</math></p> <p>4. Compare the eigenvalue distribution to the corresponding theoretical formula in <math>P_{RMT}(\lambda) = Q/2\pi\lambda \sqrt{(\lambda - \lambda_-)(\lambda_+ - \lambda)}</math> (3), where <math>\lambda_{\pm} = (1 \pm Q^{-\frac{1}{2}})^2</math> (4). If the two lines match, that data passes the RMT-test (fig.1 left), and if they do not match, it fails the RMT-test (fig.1 right).</p> <div style="display: flex; justify-content: space-around;">  </div> <p style="text-align: center;">Fig.1 Examples of Qualitative evaluation of RMT-test</p>
--

Table 2 Algorithm of the quantitative version of the RMT-test

<p>1. The first 3 step is the same to the qualitative version of the RMT-test. Calculate the k-th moments of the eigenvalues, obtained from the Eq. (2) for the data <math>m_k^{EXP} = 1/N \sum_{i=1}^N (\lambda_i)^k \quad (5)</math></p> <p>2. Calculate the theoretical moment in <math>m_k^{RMT} = \int_{\lambda_-}^{\lambda_+} \lambda^k P_{RMT}(\lambda) d\lambda \quad (6)</math></p> <p>3. Calculate the deviation of <math>m_k^{EXP}</math> from <math>m_k^{RMT}</math> : Error=<math>(m_k^{EXP}/m_k^{RMT} - 1) \times 100\% \quad (7)</math></p> <p>4. When <math>(N \geq 200, Q \geq 3)</math>, if the  Error  of moment up to 6<sup>th</sup> &lt; 5% that the sequence passes the quantitative evaluation of RMT-test, and if  Error  &gt; 5% it fails the test .</p>
---

In Chapter 4, in order to examine its effectiveness, the RMT-test is tested on two random number generators, LCG and MT, and three physical random numbers made by Toshiba, Hitachi and Tokyo-Electron. The result shows that all of them pass the RMT-test for a wide range of parameters. Although the physical random numbers by Tokyo-Electron are relatively better than the two other physical random numbers, the degrees of randomness of the three are indistinguishable both in qualitative test and in quantitative test. The author further tested the validity of RMT-test on the sequences of low randomness and showed that the RMT-test can detect off-randomness successfully which are the initial part of LCG and Log-return sequences of randomness numbers made by random number generators.

From the practical point of view, however, the advantage of this RMT-test resides in its applicability on real-world data, whose randomness level is far below the reach of the conventional randomness tests, such as NIST or JIS. In Chapter 5, the RMT-test is applied on two examples: choice of hash functions and the stock prices, assuming that the high randomness of the sequences indicates the high security level in both cases. In the first example, two popular hash functions are compared, one is the latest SHA-1 and the other one is the older MD5. The result of the RMT test shows that the randomness of the output sequences of SHA-1 is indeed higher than the output of MD5. In the second example, the author compare the performance of various stock prices in relation to their randomness and show that the stock prices of higher randomness perform better than the stocks of lower randomness, for the stocks in TOPIX500 in the year 2010 and 2011 based on their randomness in the previous three years, 2007-2009.

## 学会誌掲載論文リスト

- 題 目 Testing Randomness by Means of Random Matrix Theory  
著 者 名 Xin Yang, Ryota Itoi and Mieko Tanaka-Yamawaki  
学 術 雑 誌 名 Progress of Theoretical Physics Supplement  
( 卷, 号, 頁 ) ( 194 , , 73 ~ 83 )  
発 行 年 月 日 2012 年 6 月 14 日
- 題 目 Testing Randomness by Means of RMT Formula  
著 者 名 Xin Yang, Ryota Itoi and Mieko Tanaka-Yamawaki  
学 術 雑 誌 名 Intelligent Decision Technologies, SIST (Springer)  
( 卷, 号, 頁 ) ( 10 , , 589 ~ 596 )  
発 行 年 月 日 2011 年 7 月 20 日
- 題 目 Moment Approach for Quantitative Evaluation of Randomness  
basedon RMT Formula  
著 者 名 Mieko Tanaka-Yamawaki, Xin Yang, and Ryota Itoi  
学 術 雑 誌 名 Intelligent Decision Technologies, SIST (Springer)  
( 卷, 号, 頁 ) ( 16 , , 423 ~ 432 )  
発 行 年 月 日 2012 年 5 月 23 日
- 題 目 Randomness as the Security Levels of Investments  
著 者 名 Mieko Tanaka-Yamawaki, Xin Yang, Yuuta Mikamori  
学 術 雑 誌 名 Intelligent Decision Technologies, SIST (Springer)  
( 卷, 号, 頁 ) ( , , ~ )  
発 行 年 月 日 2013 年 6 月 26 日 (予定)
- 題 目 RMT 公式を用いた乱数度評価法の提案  
著 者 名 楊 欣, 糸井 良太, 田中 美栄子  
学 術 雑 誌 名 研究報告数理モデル化と問題解決(MPS)  
( 卷, 号, 頁 ) ( 83 , 2 , 1 ~ 5 )  
発 行 年 月 日 2011 年 5 月 10 日
- 題 目 モーメント分析法によるランダム行列理論を用いた乱数度  
評価法の改良  
著 者 名 楊 欣, 糸井 良太, 田中 美栄子  
学 術 雑 誌 名 研究報告数理モデル化と問題解決(MPS)  
( 卷, 号, 頁 ) ( 88 , 16 , 1 ~ 5 )  
発 行 年 月 日 2012 年 5 月 10 日

- 題 目** 乱数度計 RMT テストの実データへの応用～ハッシュ値と Tick 株価～
- 著 者 名** 楊 欣, 三賀森 悠太, 田中 美栄子
- 学 術 雑 誌 名** 研究報告数理モデル化と問題解決(MPS)
- ( 卷, 号, 頁 )** ( 91 , 33 , 1 ~ 6 )
- 発 行 年 月 日** 2012 年 11 月 29 日
- 
- 題 目** ランダム行列理論を用いた乱数度評価法の提案
- 著 者 名** 楊 欣, 糸井 良太, 田中 美栄子
- 学 術 雑 誌 名** 統計数理研究所共同研究リポート, 経済物理とその周辺(8)
- ( 卷, 号, 頁 )** ( 271 , , 19 ~ 31 )
- 発 行 年 月 日** 2012 年 3 月 日
- 
- 題 目** 乱数度計 RMT テストの実データへの応用～ハッシュ値と Tick 株価～
- 著 者 名** 楊 欣, 三賀森 悠太, 田中 美栄子
- 学 術 雑 誌 名** 統計数理研究所共同研究リポート, 経済物理とその周辺(9)
- ( 卷, 号, 頁 )** ( 292 , , 70 ~ 83 )
- 発 行 年 月 日** 2013 年 3 月 日
- 
- 題 目** ランダム行列理論を用いた主成分抽出法による日本と米国の株式市場における主要セクタの変遷
- 著 者 名** 木戸 丈剛, 楊 欣, 田中 美栄子, 高石 哲弥
- 学 術 雑 誌 名** 情報処理学会論文誌 数理モデル化と応用
- ( 卷, 号, 頁 )** ( 4 , 4 , 104 ~ 110 )
- 発 行 年 月 日** 2012 年 4 月 15 日
- 
- 題 目** RMTテストの性能検証—NIST乱数検定との比較
- 著 者 名** 三賀森悠大, 楊欣, 糸井良太, 田中美栄子
- 学 術 雑 誌 名** 情報処理学会論文誌 数理モデル化と応用
- ( 卷, 号, 頁 )** ( 6 , 1 , 57 ~ 63 )
- 発 行 年 月 日** 2013 年 3 月 12 日