

(様式 2)

## 学位論文の概要及び要旨

氏 名 楊 欣 印

題 目 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

### 学位論文の概要及び要旨

近年、気象、医療、金融、国勢調査など様々な分野でデジタル化が進んだ結果、膨大なデータの取得・保存が可能になり、そのような大容量データから知識抽出を有効に行える手法の必要性が増している。その候補として、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のに適した、ビッグデータ解析向きの主成分解析手法として注目を集めている。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への応用は、高頻度株価データを用いてトレンド株を算出できることでその有効性が注目され、気象データ等にも応用が進みつつある。

本論文の主眼である RMT テストは、この RMT-PCA と同様の原理を応用した乱数度測定法である。RMT-PCA では、異なる時系列間の相関行列の固有値分布を対応する理論式と比較することにより、主成分とノイズ成分を分離する。RMT テストでは、データとして 1 次元数列を用い、これを長さ  $L$  ずつに切り分けて、 $N$  個の時系列とみなし、各時系列を平均 0 分散 1 に規格化した上で自己相関行列を作ってその固有値分布を RMT 理論式と比較する。RMT テストの『定性評価』は、実データの固有値分布と RMT 理論式が一致すれば RMT テストに合格したとし、不一致ならば RMT テストに不合格と判断する。『定性評価』だけでは微小な差異を見分けにくいので、『定量評価』を用いて数値化する。これは固有値分布の実測値と理論式の差異を数値化して比較する方法で、両者の  $k$  次モーメントを計算し、そ

の比と 1 との誤差( $E$ )で乱数度を評価する. すなわち, RMT テストの定量評価に於いては「この誤差が小さいほど乱数度が高い」という基準を用いることになる. RMT 理論式は, 次元  $N$  と時系列長  $L$  が無限大の極限で成立するため, 実際の応用に於いて有効な  $N$  と  $L$  の範囲を定めた上で, モーメントの次数  $k$  と, 「良い乱数」として要求されるモーメント誤差  $E$  の最大値を, 乱数の定量評価基準として数値実験に基づいて定める必要があった. これらの数値実験を, 擬似乱数や物理乱数データ, すなわち, 一般に利用される線形合同法を用いた rand 乱数やメルセンヌ・ツイスタ (Mersenne Twister) 等の擬似乱数と, スーパーコンピュータに組み込まれた物理乱数発生器による物理乱数等を用いて行い, これら全てが「良い乱数」として合格する条件として,

「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」

という定量評価基準を定めたのが本論文の前半の結果である. 後半は RMT テストの応用として, ハッシュ関数の比較と株価選定に応用した結果を述べる.

ここで提案する RMT テストは, 従来の乱数度検定法と比較して以下の三利点を持つ. 第一に可視化手法であり, 乱数度が非常に低い数列に対して, RMT テストの『定性評価』を用いれば, 直感的にすぐ判断できる簡便な検定法として使用できる. 第二に, NIST 等の従来の乱数度検定アルゴリズムにあったような, 対象とするデータの長さやデータ型に対する強い制限が本手法にはないため, この RMT テストは, データ型を問わず, 実数・整数・二進数等と同じアルゴリズムが適用できる. 第三に, 乱数度の低い実データに対してもランダム性を RMT テストの『定量評価』を用いて数値化し比較できる. このような 3 大利点を生かし, ハッシュ関数の安全性判定と高収益株の抽出のために RMT テストを適用した. 暗号学的ハッシュ関数の安全性の為には, 出力値の無規則性と予測不可等が必要であるが, 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 SHA1 の乱数度を比較検定した結果, SHA1 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が収益性が高い」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.