ランダム行列理論を用いた時系列モデルの評価

理学専攻・情報科学コース　郷治 雅　(指導教員：吉田　裕亮)

1　はじめに

ランダム行列理論とは, 20 世紀初めの Wishart らによる数理統計学の研究に起源を持ち, その後, 物理学の研究対象として注目され, 近年では数学分野において発展を続けている理論である．時系列モデルとは, 時系列データを表現するモデルであり, 主に AR モデル, MA モデル, ARMA モデルが用いられているが, 本研究では最も一般の ARMA モデルによるシミュレーションを行う．

時系列データモデルのパラメータ推定に関する研究はこれまで多くされており, 複数の手法が提案されている．本研究では, 定常条件を満たす時系列データを折りたたんだ行列を Compound Wishart 行列の標本とみなし, その行列の標本モーメント列と ARMA パラメータのみに依存する値から求まるモーメントとその分散の理論値を調べ, 推定されたモーメント値が理論値からどの程度離れているかを検証することで ARMA モデルの評価を行う手法の提案を目的としている．

2　ランダム行列理論

一般に，ランダム行列とは確率変数を要素に持つ行列であり，その代表例として Wishart 行列が挙げられる．$n \times n$ 対称ランダム行列 S で $p/n = \lambda$ を保ちながら，$n \to \infty$, $p \to \infty$ の極限を取ると, Wishart 行列 S の固有値の経験分布は, $\lambda_{min} \leq t \leq \lambda_{max}$ のときに以下の確率密度関数 $p(t)$ に収束する．この分布は Marchenko-Pastur 分布と呼ばれている．

$$p(t) = \frac{1}{2\pi}\frac{\sqrt{-(t-\lambda_{max})(t-\lambda_{min})}}{\lambda t}$$

$$\lambda_{min}^{max} = (1 \pm \sqrt{\lambda})^2$$

3　ARMA モデル

時系列データを表すモデルはに主に AR モデル (自己回帰モデル) と MA モデル (移動平均モデル) が使われており, 更に, 2 つのモデルを一般化した ARMA モデルがある．

AR モデルは AR(p) と表現され, p 次の自己回帰モデルを意味し, 以下の式で書ける．

$$X_t = c + \sum_{i=1}^{p} \phi_p X_{t-p} + Z_t$$

ϕ_1, ..., ϕ_p はモデルのパラメータ, c は定数, Z_t は, 誤差項である．

MA モデルは MA(q) と表現され, q 次の移動平均モデルを意味し, 以下の式で表せる．

$$X_t = Z_t + \sum_{i=1}^{q} \sigma_i Z_{t-i}$$

σ_1, ..., σ_q はモデルパラメータ, Z_t, Z_{t-1}, ... は標準正規分布に従っている．

ARMA(p, q) は p 次の自己回帰と q 次の移動平均を組み合わせたモデルのことであり,

$$X_t = Z_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{q} \sigma_i Z_{t-i}$$

のように AR モデルと MA モデルの組み合わせで表現される．

4　ランダム行列と時系列データ

ある時系列データ $X_1, ..., X_n$ が定常である時, 自己分散は

$$\gamma(h) = \mathrm{Cov}(X_{n+|h|}, X)$$

で与えられ, 自己分散関数を用いて行列

$$\Gamma = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-1) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n) & \gamma(n-1) & \cdots & \gamma(0) \end{pmatrix} \tag{1}$$

を考える．ここで, 時系列データ X_n を以下の様に $n \times m$ に折りたたんで．

$$H = \begin{pmatrix} X_1 & X_2 & \cdots & X_m \\ X_{m+1} & X_{m+2} & \cdots & X_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ X_{(n-1)(m+1)} & \cdots & \cdots & X_{mn} \end{pmatrix} \tag{2}$$

を構成する．ただし, 行列の縦横比 $\lambda = \frac{m}{n}$ とする．

$$D = \frac{1}{n} H^t H$$

とすると, D は

$$\frac{1}{n} G \Gamma^t G$$

と相似であることが知られている. ただし G はガウスランダム行列である．

$$W = \frac{1}{n} G \Gamma^t G$$

は compund Wishart 行列と呼ばれる．また行列 Γ のモーメント列は

$$m_k = \frac{1}{2\pi} \int_0^{2\pi} f(\omega)^k dt \tag{3}$$

と求めることができる. ここで, $f(\omega)$ は $\gamma(h)$ のフーリエ変換を用いて

$$f(\omega) = \mathcal{F}(\gamma(h))(\omega) = \sum_{h=0}^{\infty} e^{\sqrt{-1}\gamma(h)\omega} \tag{4}$$

と与えられる．さらに, W のモーメント列は $m_1,...,m_k$ と行列の比 λ で表せることが知られている.
実際に, W のモーメント列を M_k とおくと, M_1, M_2 は以下の様に与えられる.

$$M_1 = \lambda m_1,$$
$$M_2 = \lambda^2 m_1^2 + \lambda m_2.$$

また, Compound-Wishart 行列の一般論より M_1, M_2 の分散 σ_1^2, σ_2^2 も知られている.

$$\sigma_1^2 = \frac{1}{n}\lambda m_1^2,$$

$$\sigma_2^2 = \frac{2}{n^2}(4\lambda^3 m_1^2 m_2 + 2\lambda^2 m_2^2 + 8\lambda^2 m_1 m_3 + 4\lambda m_4).$$

5 提案手法

本研究では以下の手順で実験を進める.

1. 時系列データを $n \times m$ に折りたたんで行列 X を作成.
2. $D = \frac{1}{n}X^t X$ のモーメント列を調べる.
3. データから推定された ARMA モデルのパラメータと行列の縦横比を用いてモーメントの理論値 M_k, σ_k を求める.
4. 2 で得られたモーメント値と 3 で得られた M_k とどの程度離れているかを σ_k を用いて評価する.

6 実験

6.1 実験設定

1. 7 年分の日経平均株価の一階差分
データ数：2140 $\lambda = \frac{5}{4}, \frac{12}{7}, \frac{7}{3}, \frac{75}{28}$

2. 15 年間の為替レート (USD $\to$ JPY) の二階差分
データ数:4310 $\lambda = 1, \frac{7}{6}, \frac{15}{11}, \frac{8}{5}$

6.2 実験結果

以下に推定した 2 次のモーメントと, 対応するモーメントとその標準偏差の理論値を示す.
平均値の推定値を α_2, 平均値の理論値を M_2, 標準偏差の理論値を σ_2 とし, 理論値と推定値がどの程度離れているかを $\frac{|\alpha_2 - M_2|}{\sigma_2}$ を用いて表現した.
1.
ARMA(2,3)

λ	α_2	σ_2	M_2	$\frac{\lvert m_2-M_2\rvert}{\sigma_2}$
$\frac{5}{4}$	19.0168	14.3677	2.13717	2.180935
$\frac{12}{7}$	24.0028	29.6787	2.13717	2.6626167
$\frac{7}{3}$	40.3583	46.9265	3.73177	1.76941773
$\frac{75}{28}$	37.1634	58.0506	4.27174	4.8896232

ARMA(1,3)

λ	α_2	M_2	σ_2	$\frac{\lvert m_2-M_2\rvert}{\sigma_2}$
$\frac{5}{4}$	6.31206	14.3633	0.647265	12.43886198
$\frac{12}{7}$	10.044	23.9953	0.980967	14.22198
$\frac{7}{3}$	16.1891	40.3456	1.54132	15.672605
$\frac{75}{28}$	20.1964	37.1518	1.990052	8.921453

2.
ARMA(2,1)

λ	α_2	M_2	σ_2	$\frac{\lvert m_2-M_2\rvert}{\sigma_2}$
1	10.8065	11.8461	1.08393	0.9591025
$\frac{7}{6}$	13.4331	16.4841	1.042029	2.9331605
$\frac{15}{11}$	16.8413	28.2094	1.191311	9.54251241
$\frac{8}{5}$	21.366	31.1529	2.38851	4.0975089

ARMA(1,1)

λ	α_2	M_2	σ_2	$\frac{\lvert m_2-M_2\rvert}{\sigma_2}$
1	7.71605	9.6927	0.754427	2.620068
$\frac{7}{6}$	9.60685	13.4876	0.990213	3.9181974
$\frac{15}{11}$	12.061	23.0814	1.33613	8.24799982
$\frac{8}{5}$	15.03233	25.4899	1.67115	15.5815689

6.3 考察

2 種類のデータを用いた実験から, 推定された ARMA モデルのうち, パラメータから求まるモーメントの理論値とデータから求められる推定値が近いものと離れているものがあることがわかった．このことから, 本研究での提案手法を用いて ARMA モデルの評価を行うことは可能であると考えられる．しかし, データによって精度のばらつきが見られるため, トレンドを除去する方法や定常化の方法に改善の余地があると考えられる．また, モデルの推定方法として 2 種類の手法を用いて実験したが, それぞれ値が異なっていたことについても検証する必要があると考える.

7 おわりに

本研究では，時系列パラメータから求められるモーメントの理論値とデータから求められる推定値を比較することで時系列モデルの評価手法を提案することを目的とした. 2 つのデータを用いた実験から AR が 2 次以上の ARMA モデルに対してこの手法が有用であると考えられる．また, モデルの推定方法としてモーメント法を用いた場合により精度の高い結果を得られることがわかった．今後更に多くのデータに対してこの手法を適用するためには、トレンド除去や定常化手法を改善することで精度が向上する可能性が考えられる.

参考文献

[1] 茨木志織, モーメント法によるノイズ推定を用いたスペクトラルクラスタリグ, お茶の水女子大学大学院理学専攻情報科学コース修士論文, 2011.

[2] 長谷川彩子, ランダム行列理論を用いた時系列データ解析, お茶の水女子大学大学院理学専攻情報科学コース修士論文, 2014

[3] 長谷川彩子, 吉田裕亮, ランダム行列と自由キュムラント行列を用いた時系列データ解析, 研究報告数理モデル化と問題解決, 2013-MPS-93