ランダム行列理論を用いた時系列データ解析

長谷川彩子　(指導教員：吉田裕亮)

1　はじめに

近年, データベース技術の発達などにより, 膨大な時系列データが蓄積可能となり, それらのより精度の高い解析が必要とされている. また, 近年, ランダム行列理論を用いた研究が様々な分野で注目されており, データ解析の分野においてもその応用が期待されている.

しかし, 既存のランダム行列理論を用いた時系列データ解析は, ランダム行列理論に基づき, i 種 t 時の it 個のデータから作られたデータ行列の相関行列を解析し, その中の相関関係を調べるというものであり, ランダム行列理論を用いて 1 種 t 時の個別時系列データの特徴を見出すことは困難とされていた. そこで本研究では, 1 種 t 時の個別時系列データから構成された行列を作り, ランダム行列理論を用いて解析する手法を考察する.

2　時系列モデル

時系列モデルは多数存在するが, 最も多く使われているのが以下の ARMA モデル (AutoRegressive Moving Average Model) である.

ARMA(m,l) モデル

時点 t の観測値 y_t を, 自分自身の過去の値と誤差項の現在および過去の値の線形和で表わしたモデル.

$$y_t \;=\; \sum_{i=1}^{m} a_i y_{t-i} + v_t - \sum_{i=1}^{l} b_i v_{t-i} \tag{1}$$

誤差項 v_t は一般にホワイトノイズを考える.

$$\begin{aligned} &\mathrm{E}(v_t) = 0\ ,\ \ \mathrm{Var}(v_t) = \sigma^2\ , \\ &\mathrm{Cov}(v_t, v_s) = 0\ (t \neq s) \end{aligned} \tag{2}$$

特に, $l = 0$ としたものを AR(m) モデル, $m = 0$ としたものを MA(l) モデルといい, これらは最も基本的な時系列モデルである. 有限次の定常 AR および ARMA モデルは無限次の MA モデルで表現可能である. 一般に, ARMA モデリングには最尤推定法がよく用いられる.

3　ランダム行列理論

ランダム行列とは, 確率変数を要素にもつ行列である. 各成分が独立に標準正規分布に従う変数をもつ $n \times p$ ランダム行列を C とすると,

$$S = \frac{1}{n} C\,{}^tC \tag{3}$$

で求められる $n \times n$ 対称ランダム行列 S を Wishart 行列という. $p/n = \lambda$ を保ちながら $n \to \infty, p \to \infty$ の極限をとると, Wishart 行列 S の固有値経験分布は Marcenko-Pastur 分布に収束することが知られている.

Marcenko-Pastur 分布の密度関数

$$\begin{aligned} p(t) \;&=\; \frac{1}{2\pi} \frac{\sqrt{-(t-\lambda_{\max})(t-\lambda_{\min})}}{\lambda t}\ , \\ \lambda_{\max} &= (1+\sqrt{\lambda}\,)^2\ ,\ \ \lambda_{\min} = (1-\sqrt{\lambda}\,)^2 \end{aligned} \tag{4}$$

4　モーメント・自由キュムラント

$\mathbb{R}$ 上の台がコンパクトな確率測度 μ の k 次キュムラント $\kappa_k = \kappa_k(\mu)$ とは, μ の Laplace 変換の対数の展開係数をいい,

$$\log \int_{\mathbb{R}} e^{zx} \mu(dx) \;=\; \sum_{k=1}^{\infty} \frac{\kappa_k}{k!} z^k \ \ (z \in \mathbb{C}) \tag{5}$$

である. 式 (5) の指数関数をとると, 左辺は μ の k 次モーメント

$$M_n \;=\; M_n(\mu) \;=\; \int_{\mathbb{R}} x^n \mu(dx) \tag{6}$$

の指数型の母関数であるから, 両辺を展開することにより, モーメント-キュムラント公式

$$M_n \;=\; \sum_{\pi \in P(n)} \kappa_\pi \;=\; \sum_{\rho \in \mathbb{Y}_n} |P(\rho)|\, \kappa_1^{m_1(\rho)} \cdots \kappa_n^{m_n(\rho)} \tag{7}$$

を得る. 式 (7) の分割を非交差のものに制限することにより, μ の k 次自由キュムラント $R_k = R_k(\mu)$ を定義する.

$$M_n \;=\; \sum_{\pi \in NC(n)} R_\pi \;=\; \sum_{\rho \in \mathbb{Y}_n} |NC(\rho)|\, R_1^{m_1(\rho)} \cdots R_n^{m_n(\rho)} \tag{8}$$

自由キュムラントは自由独立な確率変数の和 (自由合成積) を線形化する. n 次モーメントおよび自由キュムラントは, n 次以下の他方の値から導くことができる.

5　CONJECTURE

MA モデル $y_t \;=\; \sum_{i=0}^{N} b_i v_{t-i}$ で表される時系列データについて,

データから作られる $p \times n$ 行列

$$K \;=\; \begin{bmatrix} y_1 & \cdots & y_n \\ y_{n+1} & \cdots & y_{2n} \\ \vdots & \ddots & \vdots \\ y_{(p-1)n+1} & \cdots & y_{pn} \end{bmatrix} \tag{9}$$

から求まる

$$\hat{K} \;=\; K\,{}^tK \tag{10}$$

の固有値経験分布の k 次自由キュムラント r_k と,

MA 係数からなる $n \times 2n$ 行列

$$H = \begin{bmatrix} b_n & b_{n-1} & \cdots & b_1 & b_0 & 0 & \cdots & 0 \\ 0 & b_n & \cdots & b_2 & b_1 & b_0 & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & b_n & b_{n-1} & \cdots & \cdots & b_0 \end{bmatrix} \tag{11}$$

から求まる

$$\hat{H} = {}^{t}HH \tag{12}$$

の固有値経験分布の k 次モーメント m_k の間には,

$$r_k = \theta p^k m_k \ (\theta : \text{定数}) \tag{13}$$

の関係が成り立つ.

6 数値実験による検証

CONJECTURE の数値計算的な検証を行うために, ある MA モデルから次数, 係数, ホワイトノイズを各々変えたモデルで対照実験を行った. ここで, 実験結果には, 式 (13) における $(k+1)$ 次/k 次,

$$\frac{r_{k+1}\, m_k}{p\, m_{k+1}\, r_k} \tag{14}$$

を計算し, その値が 1.0 に十分近ければ CONJECTURE が数値計算的に検証されたと判断する.

実験結果の 1 例

k=1	k=2	k=3	k=4	k=5	k=6
0.9936	1.0023	1.0066	1.0129	1.0237	1.0375

全ての対照実験において, 上の 1 例のように 1.0 に十分近い値となったことから, 次数や係数およびホワイトノイズの値に関係なく CONJECTURE が成り立つことが検証された.

7 モデリング評価法としての応用

前節で検証した CONJECTURE を, MA モデリングの評価法として応用することを考える. 与えられた時系列データに対して, 適切なモデリングが行われた場合にのみ CONJECTURE が成り立つことが推測される. そこで, 次のような実験を行った.

ある MA モデルから時系列データを生成し, 以下の条件, ①正しい係数, ②係数の 1 つを全く違う値に, ③次数を 1 次増やし適当な係数を加える, ④正しい係数に誤差 ±0.1 を加える, ⑤正しい係数に誤差 ±0.3 加える, ⑥正しい係数に誤差 ±0.5 を加える, により行列 H を作り, それぞれの場合の式 (14) の値を観測した.

実験結果

	k=1	k=2	k=3	k=4	k=5
①	**1.0137**	**1.0143**	**1.0175**	**1.0304**	**1.0569**
②	0.4490	0.4439	0.4448	0.4489	0.4582
③	0.6688	0.6137	0.5936	0.5916	0.6026
④	0.8294	0.8187	0.8154	0.8231	0.8430
⑤	0.5039	0.4914	0.4904	0.4962	0.5086
⑥	0.3267	0.3328	0.3334	0.3348	0.3405

①と②〜⑥との比較により, MA モデリングが適切に行われた場合には式 (14) の値が 1.0 に十分近くなり, 不適切であった場合には 1.0 からずれることから, CONJECTURE の MA モデリング評価法としての有効性が示唆される.

また, ④,⑤,⑥より, 係数に加えた誤差が大きくなるにつれ 1.0 からのずれが大きくなっていることから, 1.0 からのずれの大きさで MA モデリングの適切さを測れることが示唆される.

8 AR および ARMA モデルへの拡張

無限次 MA 表現された有限次の定常 AR および ARMA モデルに関しても CONJECTURE の数値実験による検証を行った. すると, n を十分大きくとると CONJECTURE が成り立つことが検証された. そこで, AR および ARMA モデリングについても前節同様に実験を行った. ある AR および ARMA モデルから時系列データを生成し, ①正しい②誤差 ±0.1③誤差 ±0.3④誤差 ±0.5 の AR および ARMA 係数を用いて表現した無限次の MA 係数から十分大きな n 個を取り出して行列 H を作り, それぞれの場合の式 (14) の値を観測した.

実験結果

	k=1	k=2	k=3	k=4
AR①	**1.0030**	**0.9892**	**0.9728**	**0.9647**
AR②	0.8035	0.7373	0.7112	0.7044
AR③	0.3986	0.3259	0.3129	0.3115
AR④	0.1052	0.0809	0.0782	0.0781
ARMA①	**0.9882**	**0.9953**	**1.0043**	**1.0100**
ARMA②	1.4715	1.5384	1.5618	1.5735
ARMA③	2.3933	2.6651	2.7448	2.7788
ARMA④	3.1888	3.7255	3.8936	3.9657

実験結果から, 無限次 MA 表現可能な定常 AR および ARMA モデルに関しても, 式 (14) の 1.0 からのずれの大きさでそのモデリングの適切さを測れることが示唆される.

9 まとめ

MA モデリングされる個別時系列データについて, ランダム行列理論を用いて CONJECTURE を提案し, 数値実験による検証を行った. また, CONJECTURE のモデリング評価法としての有効性を検証した. そこで, CONJECTURE の理論値からのずれの大きさによって, MA モデリングの適切さを測れることが示唆された. さらに, 無限次 MA 表現された定常 AR および ARMA モデルについても同様に CONJECTURE の検証やモデリング評価法としての有効性の検証を行った.

参考文献

[1] O. Pfaffel and E. Schlemm, Eigenvalue distribution of large sample covariance matrices of linear processes, Probab. Math. Statist. **31** (2011), 313–329.

[2] Hora, Akihito, Free probability and asymptotic representation theory of sysmetric groups, Surikaisekikenkyusho Kokyuroku **1418** (2005), 10–40.