情報処理学会研究報告
IPSJ SIG Technical Report

Vol.2013-MPS-93 No.11
2013/5/23

ランダム行列と自由キュムラントを用いた時系列データ解析

長谷川 彩子[†1]　吉田 裕亮[†1]

概要：本研究では, MA モデリングされる時系列データについて, ランダム行列と整合する自由確率論に現れる自由キュムラントを用いて, データから構成される行列と MA モデルでのパラメータから定まる行列との関係性を, データ行列からの標本自由キュムラントとパラメータから定まる行列の標本モーメントで示し, 数値実験による検証を行った. また, その関係性を用いた MA モデルのパラメータ推定の評価基準を提案し, その有効性を数値実験により検証した. 結果として, 理論値からのずれの大きさによって, MA モデルのパラメータ推定の良さが測れることが示唆された. さらに, 無限次 MA 表現される定常 AR および ARMA モデルについても同様の数値実験による検証を行った.

キーワード：ランダム行列, 自由キュムラント, 時系列データ解析, MA モデル

Time Series Data Analysis by Using Random Matricies and Free Cumulants

Ayako Hasegawa[†1]　Hiroaki Yoshida[†1]

Abstract: In this study, for Moving Average(MA) modeled time series data, using the free cumulants appear in free probability theory which is a tool for the theory of random matricies, we have investigated the relevance between the matricies constructed from given time series data and the matricies determined by the parameters in MA models, by using the sample free cumulants of the data matricies and the sample moments of the parameters matricies. We also have performed numerical experiments. Then we have proposed a criteria for the estimation of MA parameters by this relevance, and have performed numerical experiments. Conseqently, we have obtained suggestion that we will be able to measure the goodness of parameters by the displacement from the theoretical values. Moreover, we also have performed numerical experiments on the stationary AutoRegressive(AR) models and ARMA models which are expressed as MA models of infinite degree.

Keywords: random matrix, free cumulant, time series data analysis, Moving Average model

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

†1 現在, お茶の水女子大学大学院 人間文化創成科学研究科理学専攻
Presently with Graduate School of Humanities and Sciences, Ochanomizu University

Vol.2013-MPS-93 No.11
2013/5/23

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, 過去の観測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 は一般にホワイトガウスノイズを考える.

$$\begin{aligned} v_t \ &\sim \ N(0,\sigma^2) \ \ i.i.d. \\ \mathrm{Cov}(v_t, v_s) &= 0 \ (t \neq s) \end{aligned} \tag{2}$$

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. 一般に, AR 部分のパラメータ推定にはユールウォーカー法, burg 法, 最小二乗法および最尤推定法が用いられ, MA 部分のパラメータ推定には最尤推定法が用いられる.

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 $p \times n$ ランダム行列を C とすると,

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

で与えられる $p \times p$ 対称ランダム行列 S を Wishart 行列という. $n/p = \lambda$ 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 分布の密度関数は

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

で与えられる.

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 $R_k = R_k(\mu)$ が定義される.

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

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. k 次モーメントおよび k 次自由キュムラントは, 各々 k 次以下の他方の値から導くことができる.

例えば, $p \times p$ 行列 X の k 次モーメント m_k は

$$m_k \ = \ \frac{1}{p} \mathrm{Tr}\left(X^k\right) \tag{9}$$

で定められ, さらに行列 X の k 次自由キュムラント r_k は k 次以下のモーメントを用いて

$$\begin{aligned} r_1 \ &= \ m_1 \\ r_2 \ &= \ m_2 - m_1^2 \\ r_3 \ &= \ m_3 - 3m_1m_2 + 2m_1^3 \\ r_4 \ &= \ m_4 - 4m_1m_3 - 2m_2^2 + 10m_1^2m_2 - 5m_1^4 \\ r_5 \ &= \ m_5 - 5m_1m_4 - 5m_2m_3 + 15m_1^2m_3 \\ &\quad + 15m_1m_2^2 - 35m_1^3m_2 + 14m_1^5 \\ &\ \vdots \end{aligned} \tag{10}$$

と表される.

5. 提案手法

5.1 データから構成される行列と MA モデルでのパラメータから定まる行列との関係性

MA モデル $y_t \ = \sum_{i=0}^{l} b_i v_{t-i}$ $(v_t \sim N(0,\sigma^2)\ i.i.d.)$ で表される時系列データについて,

データから構成される $p \times n$ 行列

$$X \ = \ \begin{bmatrix} y_1 & \cdots & y_n \\ y_{n+1} & \cdots & y_{2n} \\ \vdots & \ddots & \vdots \\ y_{(p-1)n+1} & \cdots & y_{pn} \end{bmatrix} \tag{11}$$

から求まる

$$\hat{X} \ = \ \frac{1}{p} X\, {}^tX \tag{12}$$

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 b_i (ただし $i > l$ のとき $b_i = 0$) を用いて表わされる

$$\gamma(h) = \sum_{j=0}^{\infty} b_j b_{j+|h|}, \quad h \in \mathbb{Z} \tag{13}$$

を要素に持つ $m \times m$ 行列

$$\Gamma(i,j) = \gamma(i-j) \tag{14}$$

の固有値経験分布の k 次モーメント m_k の間には,

$$r_k = \theta \sigma^{2k} m_k \tag{15}$$

の関係が成り立つ. ただし θ は定数, m は $m \geq p$ を満たす任意の自然数である.

5.2 数値実験

前小節で示した関係性の精度を数値計算で検証するために, ある MA モデルから次数, 係数, ホワイトノイズの分散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 (15) における $(k+1)$ 次/k 次,

$$\frac{r_{k+1}\, m_k}{\sigma^2\, m_{k+1}\, r_k} \tag{16}$$

を計算し, その理論値を 1.0 として検証を行った.

実験結果の一例

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, 上の一例のように 1.0 に十分近い値が観測された. このことから, 次数や係数およびホワイトノイズの分散に大きく影響されることなく式 (15) の関係性が成り立つことが検証された.

5.3 補足

前々小節の, データから構成される行列と MA モデルでのパラメータから定まる行列との関係性において, 式 (14) の行列 Γ は, $2m \times m$ 行列

$$H = \begin{bmatrix} b_m & b_{m-1} & \cdots & b_1 & b_0 & 0 & \cdots & 0 \\ 0 & b_m & \cdots & b_2 & b_1 & b_0 & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & b_m & b_{m-1} & \cdots & \cdots & b_0 \end{bmatrix}^t \tag{17}$$

を用いて求まる

$$\Gamma = H\,{}^tH \tag{18}$$

と考える.

6. MA モデルのパラメータ推定の評価基準としての応用

前節で示した関係性を, MA モデルのパラメータ推定の評価基準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, 適切なパラメータ推定が行われた場合にのみ式 (15) が成り立つと考えられる. そこで, 次のような実験を行った.

ある MA モデルに従う時系列データを生成し, 以下の条件, (i) 正しい係数, (ii) 係数の一つを全く違う値に, (iii) 次数を 1 次増やし適当な係数を加える, (iv) 係数の一つに誤差 0.1 を加える, (v) 係数の一つに誤差 0.3 を加える, (vi) 係数の一つに誤差 0.5 を加える, (vii) 係数すべてに誤差 0.1 を加える, (viii) 係数すべてに誤差 0.3 を加える, (ix) 係数すべてに誤差 0.5 を加える, により行列 Γ を作り, それぞれの場合について式 (16) の値を観測した.

実験結果

	k=1	k=2	k=3	k=4	k=5
i	**0.9953**	**1.0020**	**1.0020**	**0.9996**	**1.0024**
ii	1.3922	1.4053	1.3895	1.3725	1.3661
iii	0.7743	0.7399	0.7337	0.7306	0.7321
iv	0.8160	0.7974	0.7861	0.7797	0.7705
v	0.6626	0.6407	0.6310	0.6257	0.6183
vi	0.5525	0.5268	0.5174	0.5127	0.5066
vii	0.8080	0.7870	0.7748	0.7681	0.7589
viii	0.6095	0.5992	0.5889	0.5824	0.5746
ix	0.4381	0.4415	0.4394	0.4365	0.4309

(i) と (ii)〜(ix) との比較により, パラメータ推定が適切に行われた場合には式 (16) の値が理論値 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, 提案手法のパラメータ推定の評価基準としての有効性が示唆される.

また, (iv),(v),(vi) および (vii),(viii),(ix) より, 係数に加えた誤差が大きくなるにつれ理論値 1.0 からのずれが大きくなっていることから, 理論値からのずれの大きさでパラメータ推定の良さを測れることが示唆される.

7. AR および ARMA モデルへの拡張

7.1 インパルス応答関数を用いた ARMA モデルの無限次 MA 表現

定常な有限次 AR および ARMA モデルは無限次 MA 表現が可能である. $Ly_t \equiv y_{t-1}$ によって定義されるラグオペレータ L を用いると, 式 (1) の ARMA モデルは

$$\left(1 - \sum_{i=1}^{m} a_i L^i\right) y_t = \left(1 - \sum_{i=1}^{l} b_i L^i\right) v_t \tag{19}$$

と表現できる. ここで, さらに AR オペレータと MA オペレータをそれぞれ

$$a(L) \equiv \left(1 - \sum_{i=1}^{m} a_i L^i\right) \quad (20)$$

$$b(L) \equiv \left(1 - \sum_{i=1}^{l} b_i L^i\right) \quad (21)$$

によって定義すると, ARMA モデルは

$$a(L)y_t = b(L)v_t \quad (22)$$

と簡潔に表すことができる. 式 (22) の両辺を $a(L)$ で割ると, ARMA モデルは

$$y_t = a(L)^{-1}b(L)v_t \quad (23)$$

と表現できる. したがって, 形式的に得られる無限級数

$$g(L) \equiv a(L)^{-1}b(L) = \sum_{i=0}^{\infty} g_i L^i \quad (24)$$

によって $g(L)$ を定義すると, ARMA モデルに従う時系列 y_t は, 現在までに加わったホワイトノイズ v_t だけを用いて

$$y_t = g(L)v_t = \sum_{i=0}^{\infty} g_i v_{t-i} \quad (25)$$

と無限次の MA モデルで表現できる. $\{g_i; i = 0, 1, \cdots\}$ は ARMA モデルのインパルス応答関数と呼ばれる. g_i は

$$\begin{aligned} g_0 &= 1 \\ g_i &= \sum_{j=1}^{i} a_j g_{i-j} - b_i, \quad i = 1, 2 \cdots \end{aligned} \quad (26)$$

によって計算できる. ただし, $j > m$ のとき $a_j = 0, j > l$ のとき $b_j = 0$ とおくものとする.

7.2 数値実験

定常 AR および ARMA モデルで表されるデータの場合, 式 (25) における AR および ARMA モデルの無限次 MA 表現に用いる, 式 (26) の無限次 MA 係数 g_i のうち, 十分大きな m 個を用いて Γ 行列を作ると, 式 (15) の関係が成り立つことが数値計算により検証された.

そこで, 前節同様, 提案手法の定常 AR および ARMA モデルのパラメータ推定の評価基準としての有効性を検証する実験を行った. ある定常 ARMA モデルに従う時系列データを生成し, 以下の条件, (i) 正しい係数, (ii) 係数の一つを全く違う値に, (iii) 次数を 1 次増やし適当な係数を加える, (iv) 係数の一つに誤差 0.1 を加える, (v) 係数の一つに誤差 0.3 を加える, (vi) 係数の一つに誤差 0.5 を加える, (vii) 係数すべてに誤差 0.1 を加える, (viii) 係数すべてに誤差 0.3 を加える, (ix) 係数すべてに誤差 0.5 を加える, により定まる ARMA 係数を用いて表される無限次 MA 係数のうち, 十分大きな m 個を用いて行列 Γ を作り, それぞれの場合の式 (16) の値を観測した.

実験結果

	k=1	k=2	k=3	k=4	k=5
i	**1.0035**	**1.0179**	**1.0334**	**1.0579**	**1.1007**
ii	0.4774	0.5091	0.5201	0.5174	0.5070
iii	0.2752	0.2392	0.2305	0.2245	0.2182
iv	1.2622	1.2954	1.2829	1.2578	1.2253
v	1.7756	1.9737	2.0192	2.0033	1.9593
vi	1.9929	2.2494	2.3330	2.3406	2.3087
vii	1.1619	1.1845	1.1757	1.1552	1.1267
viii	1.4709	1.5736	1.5924	1.5778	1.5454
ix	1.7096	1.8937	1.9483	1.9464	1.9151

前節同様に, 正しい係数を用いた場合にのみ理論値に十分近くなり, 加えた誤差が大きくなるにつれ理論値からのずれも大きくなることから, 提案手法は定常 AR および ARMA モデルのパラメータ推定の評価基準としても有効であることが示唆される.

8. 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解析する手法を提案した. データから構成される行列と MA モデルでのパラメータから定まる行列との関係性を基に, その精度と, パラメータ推定の評価基準としての有効性の検証を数値実験により行った. さらに, 無限次 MA 表現される定常 AR および ARMA モデルについても同様の検証を行った. 今後の課題として ランダム行列理論を用いた時系列データ解析のさらなる可能性の検討を行う.

参考文献

[1] O.Pfaffel and E.Schlemm, Eigenvalue distribution of large sample covariance matrices of linear processes, Probab. Math. Statist. 31(2)(2011), pp.313-329.
[2] D.V.Voiculescu, Lectures on free probability theory, In; Lectures on probability theory and statistics, Lecture Notes in Math., 1738, Springer, Berlin, 2000, pp.279-349.
[3] D.V.Voiculescu, K.J.Dykema, and A.Nica, Free random variables, CRM Monograph Series, Vol.1, Amer. Math. Soc., Providence RI,1992.
[4] Ayako Hasegawa, Noriyoshi Sakuma and Hiroaki Yoshida, Random matrices by MA models and compound free poisson laws, preprint, (2013).
[5] 洞彰人, 自由確率論と対称群の表現の漸近理論, 数理解析研究所講究録 1418(2005), 10-40.
[6] 北川源四郎, 時系列解析プログラミング, 岩波書店, (1993).