

## 2019 年度科学研究費シンポジウム

### 多様な高次元モデルにおける理論と方法論，及び，関連分野への応用

科学研究費補助金 基盤研究 (S) 18H05290 (2018 年度～2022 年度) 「広汎な観測に対する因果性の導入とその最適統計推測論の革新 (研究代表者: 谷口 正信)」によるシンポジウムを下記のように催しますので、ご案内申し上げます。

青嶋 誠 (筑波大学)  
矢田和善 (筑波大学)  
石井 晶 (東京理科大学)

#### 記

日時：2020 年 2 月 17 日 (月)～18 日 (火)

場所：イーアスホール

(イーアスつくば 2 階 <https://tsukuba.iias.jp/floor/floor2.html>)

〒305-0817 茨城県つくば市研究学園 5 丁目 19 番地

#### プログラム

2 月 17 日 (月)

13:30～13:40 開会

13:40～14:20 藤森 洸 (早稲田大・基幹理工学部)

高次元・スパースな設定における確率過程の統計モデルに対する Dantzig selector

14:30～15:10 蛭川 潤一 (新潟大・自然科学系)

Sangyeol Lee (Department of Statistics, Seoul National University)

Asymptotic properties of mildly explosive processes with locally stationary disturbance

15:25～16:05 柿沢 佳秀 (北海道大・経済学研究院)

多変量バーンバウムサンダース型分布: 非負データの密度推定への応用

16:15～16:55 本田 敏雄 (一橋大・経済学研究科)

The de-biased group Lasso estimation for varying coefficient models

17:40～ 懇親会

2月18日(火)

10:40 ~ 11:10 小西 啓介 (筑波大・数理物質科学研究科)  
矢田 和善 (筑波大・数理物質系)  
青嶋 誠 (筑波大・数理物質系)

High-dimensional covariance matrix estimation under the strongly spiked eigenvalue model

11:15 ~ 11:45 江頭 健斗 (筑波大・数理物質科学研究科)  
矢田 和善 (筑波大・数理物質系)  
青嶋 誠 (筑波大・数理物質系)

Asymptotic properties of distance weighted discrimination and its bias correction  
in HDLSS settings

11:50 ~ 12:20 藤澤 健吾 (東京理科大・理工学研究科)  
田畑 耕治 (東京理科大・理工学部)

順序カテゴリをもつ多元分割表における非同等性のモデリング

12:20 ~ 13:45 昼食

13:45 ~ 14:25 石井 晶 (東京理科大・理工学部)  
矢田 和善 (筑波大・数理物質系)  
青嶋 誠 (筑波大・数理物質系)

データ変換を用いた高次元次判別分析について

14:35 ~ 15:15 菅澤 翔之助 (東京大・空間情報科学研究センター)

Robust bayesian modeling with synthetic posterior

15:30 ~ 16:00 桃崎 智隆, 齋藤 健 (東京理科大・理工学研究科)  
三枝 祐輔 (横浜市立大・医学部)  
生亀 清貴 (日本大・経済学部)  
中川 智之, 石井 晶, 富澤 貞男 (東京理科大・理工学部)

分割表における種々の対称モデルに関する尺度

16:05 ~ 16:45 島谷 健一郎 (統計数理研究所)  
荒木 希和子 (立命館大・生命科学部)

地上部と地下部のデータを統合させた推移行列モデル

16:45 ~ 16:50 閉会

# 高次元・スパースな設定における確率過程の統計モデルに対する Dantzig selector

藤森 洸

早稲田大学 基幹理工学部数学科

次の線型回帰モデルを考える。

$$Y = Z\beta + \epsilon,$$

ただし、 $Y \in \mathbb{R}^n$  は目的変数、 $Z = (Z_1^\top, \dots, Z_n^\top)^\top \in \mathbb{R}^{n \times p}$  は計画行列、 $\epsilon = (\epsilon_1, \dots, \epsilon_n)^\top$  は平均 0 の独立同分布な確率変数からなる、 $Z$  とは独立な確率ベクトル、 $\beta \in \mathbb{R}^p$  は未知の回帰係数である。従来の回帰分析においては、データ  $(Y, Z)$  に基づいて、パラメータ  $\beta$  の推定を最小二乗法、あるいは最尤法に基づいて行ってきた。特に、誤差項が正規分布に従うとき、適当な正則条件の下で上記の手法により得られた推定量が漸近的に良い性質を持つことはよく知られた事実である。

一方で、本講演ではパラメータ  $\beta$ 、あるいは説明変数  $Z_i$ ,  $i = 1, \dots, n$  の次元  $p$  がサンプル数  $n$  に比べて大きく、パラメータの非零成分の個数が小さいことを意味する、「高次元・スパース」と呼ばれる設定の下で回帰モデルの推定問題を扱う。高次元の設定は、目的変数に回帰しうる説明変数の成分がサンプル数に比べても多いという状況を想定しており、スパース性の設定は目的変数に真に影響を及ぼす説明変数の個数が少ない、すなわち説明変数の成分の中に無駄な情報が含まれているという状況を想定している。高次元の設定における最小二乗推定量は計算可能性や、一致性をはじめとする漸近的な良さが保証されないケースが多い。また、スパースの設定においてはパラメータの推定量に求められるのはノルムの意味での一致性だけではなく、目的変数に真に回帰する説明変数を正しく選択していることを意味する、変数選択の一致性である。このような背景から、一致性と変数選択の一致性を同時に満たす推定量の構成に関する研究が 1990 年代から盛んに行われてきた。特によく知られているのは Lasso 推定量に代表される、罰則項付き推定関数により構成される推定量である。例えば、線型回帰モデルに対する Lasso 推定量は次の凸最適化問題の解として定義される。

**Definition 1.** 次で  $\beta$  の推定量を定義する。

$$\hat{\beta}_L = \arg \min_{\beta} \left\{ \frac{1}{n} \|Y - Z\beta\|_2^2 + \lambda \|\beta\|_1 \right\},$$

ただし、 $\lambda \geq 0$  は解析者が設定する tuning parameter である。

ここで推定に用いている推定関数は、従来用いられてきた最小二乗誤差関数に、パラメータの  $l_1$  ノルムで定義された罰則関数を加えたものである。もし、 $\lambda = 0$  であれば、得られる推定量は最小二乗推定量に一致し、 $\lambda$  を大きく取れば罰則関数  $\lambda \|\beta\|_1$  の影響が大きくなって、推定量にスパース性が誘導される。適当な正則条件の下で推定関数を決定する tuning parameter の値を適切に設定することにより、得られる推定量は従来の意味での漸近的な良さと変数選択の一致性を両立する性質を持つことが知られている。さらに、上記の手

法は「最小二乗誤差」の部分を一一般のモデルにおける従来の M-推定に用いる推定関数に、罰則関数の部分を  $l_1$  ノルムそのものではなく、ノルムに関する多様な罰則関数に拡張することができるため、多様な推定手法が提案され現在も盛んに研究されている。

一方、本講演で扱うのは罰則項付き推定問題とは異なるアプローチで構成される、Dantzig selector と呼ばれる推定量である。例えば、線型回帰モデルに対する Dantzig selector は次の不等式制約付き最適化問題の解として定義される：

**Definition 2.** 次で  $\beta$  の推定量を定義する。

$$\hat{\beta}_D := \arg \min_{\beta \in \mathcal{C}} \|\beta\|_1, \quad \mathcal{C} := \{\beta \in \mathbb{R}^p : \|\psi_n(\beta)\|_\infty \leq \lambda\},$$

ただし、 $\psi_n$  は次で定まるスコア関数；

$$\psi_n(\beta) := \frac{1}{n} Z^\top (Y - Z\beta)$$

であり、 $\lambda \geq 0$  は tuning parameter である。

$\lambda = 0$  の場合には不等式制約の部分が等式になり、得られる推定量は従来の最小二乗推定量と一致する。また、 $\lambda > 0$  の場合、不等式制約により最小二乗推定量の近傍が定まり、 $l_1$  ノルムによりスパース性が誘導された推定量が得られる。適当な正則条件下で、 $\lambda$  を適切に選ぶことにより、 $\hat{\beta}_D$  は Lasso 推定量と同様、漸近的な良い性質と変数選択の一致性を満たすことが知られている。さらにこの手法は、不等式制約を課す関数を Z-推定で用いるスコア関数に置き換えることにより、一般のモデルに対する拡張を考えることは容易である。また、スコアがパラメータに関して線型関数になるモデルに関しては、推定量を定義する最適化問題が線型計画法に帰着できるということが知られており、実装の意味で利点を持つことが知られている。

Dantzig selector 型の推定量は Lasso 型の推定量とは別のアプローチで構成されたものであるが、得られる推定量の漸近的な性質は互いに似通っている。しかし、Lasso 型の推定量を様々なモデルに拡張する研究が盛んに行われているのに対して、Dantzig selector 型の推定量を拡張する研究は未開拓の部分が多い。そこで本講演では上記の Dantzig selector を確率過程、特に拡散過程の統計モデルに適用し、得られる推定量の漸近的性質に関する議論を行った。なお、本講演の内容は下記の Fujimori (2019) に基づくものである。

## 参考文献

Fujimori, K. The Dantzig selector for a linear model of diffusion processes. *Stat. Inference Stoch. Process.* **22**, no.3, p.475-498. (2019).

# Asymptotic Properties of Mildly Explosive Processes with Locally Stationary Disturbance

Junichi Hirukawa and Sangyeol Lee  
Niigata University and Seoul National University

**1. Introduction** Autoregressive processes of the form  $y_t = \rho y_{t-1} + \varepsilon_t$  with an explosive root  $|\rho| > 1$ , where  $\varepsilon_t$  are iid standard normal random variables, were first studied in White (1958) and Anderson (1959). By assuming a zero initial value for  $y_t$ , a Cauchy limit theory is derived for the least squares estimate  $\widehat{\rho}_n = (\sum_{t=1}^n y_{t-1} y_t) (\sum_{t=1}^n y_{t-1}^2)^{-1}$ :

$$(1) \quad \frac{\rho^n}{\rho^2 - 1} (\widehat{\rho}_n - \rho) \Rightarrow C,$$

where  $C$  denotes a Cauchy random variable and  $\Rightarrow$  denotes convergence in distribution when  $n$  tends to infinity. The Gaussian assumption imposed on the innovation sequence  $\{\varepsilon_t\}$  plays an important role and cannot be relaxed to obtain the same asymptotic distribution as in (1): see Anderson (1959) who demonstrate that the limit distribution of the least squares estimate depends upon the distributional assumptions imposed on the error sequence.

However, this difficulty can be avoided when the explosive root approaches unity as the sample size tends to infinity. Phillips and Magdalinos (2007a) consider autoregressive processes with root  $\rho_n = 1 + c/k_n$ , where  $k_n$  is a positive real sequence with  $k_n = o(n)$ . When  $c > 0$ , such roots are explosive in finite samples and approach unity at the rate slower than  $O(n^{-1})$ . It is well known that the asymptotic behavior of such mildly explosive autoregressions is more uniform than their purely explosive counterparts. Under the second moment condition on the iid innovations, Phillips and Magdalinos (2007a) establish limit theorems for sample moments generated by mildly explosive processes and obtain the following Cauchy limit result:

$$(2) \quad \frac{1}{2c} k_n \rho_n^n (\widehat{\rho}_n - \rho_n) \Rightarrow C.$$

This limit result is unaffected by both the distribution of the initial condition  $y_0$  as far as  $y_0 = o(k_n)$ . The result was extended by Phillips and Magdalinos (2007b) to a class of weakly dependent innovations. Aue and Horváth (2007) relax the moment conditions on the innovations by considering an iid innovation sequence that belongs to the domain of attraction of a stable law. Magdalinos and Phillips (2008) give multivariate extensions and Magdalinos (2012) considers mildly explosive autoregressions generated by a linear process that may exhibit long-range dependence. Oh et al. (2017) recently study mildly explosive autoregressions with strong mixing innovation sequence, showing that the least squares estimate has the same limit distribution as the iid innovation case.

However, the time homogenous assumption on the residuals seems to be restrictive. The analysis of relatively long stretches of time series data that may contain either slow or rapid changes in the spectrum is of interest in a number of areas. Although the idea of having locally approximately a stationary process was also the starting point of Priestley's theory of processes with evolutionary spectra (Priestley (1965)), recently one of the most important classes of non-stationary processes has been formulated in a rigorous asymptotic framework by Dahlhaus (1996a, 1996b, 1996c, 1997), called locally stationary processes. Locally stationary processes have time varying spectral densities whose spectral structures smoothly change in time. Dahlhaus (2012) also gave the extensive review about locally stationary processes. In this study, we investigate the asymptotic distribution of the LSE for the autoregression with locally stationary error process. The limit behavior of the LSE from the unit root and near stationary autoregressions with locally stationary disturbance was considered by Hirukawa and Skdakata (2012). For the iid unit root process case, see Chan and Wei (1980) and Lee and Wei (1999).

Although we mainly focus on the limit behavior of the LSE, we develop a method for identifying the onset and the end of a bubble period of an econometric time series as an application, originally considered in Phillips and Yu (2009) and Phillips et al. (2011). For this task, we investigate the limiting distribution of the Dickey-Fuller tests when the underlying process is either a unit-root process or explosively mild process and finally demonstrate that they are consistent.

**2. Limit distribution of LSE** In this section, we consider the following mildly explosive process with locally stationary disturbance:

$$\begin{aligned} y_{t,n} &= \rho_{t,n} y_{t-1,n} + u_{t,n}, \quad t = 1, \dots, n, \\ &= \left( \prod_{k=1}^t \rho_{k,n} \right) y_{0,n} + \sum_{j=1}^t \left( \prod_{k=j+1}^t \rho_{k,n} \right) u_{j,n}, \end{aligned}$$

where  $\{u_{j,n}\}$  is generated from the time varying MA ( $\infty$ ) model:

$$u_{t,n} = \sum_{l=0}^{\infty} \alpha_l \left( \frac{t}{n} \right) \varepsilon_{t-l} = \sum_{l=0}^{\infty} \alpha_l \left( \frac{t}{n} \right) L^l \varepsilon_t := \alpha \left( \frac{t}{n}, L \right) \varepsilon_t,$$

where  $\varepsilon_t \sim i.i.d. (0, \sigma^2)$ ,  $\rho_{t,n} = 1 + \frac{1}{k_n} \beta \left( \frac{t}{n} \right)$  with  $\beta \in C[0, 1]$ , the class of continuous real-valued functions on  $[0, 1]$ , satisfying  $0 < \beta(u) < \infty$ , the MA coefficients  $\alpha \left( \frac{t}{n}, L \right) := \sum_{l=0}^{\infty} \alpha_l \left( \frac{t}{n} \right) L^l$  with lag operator  $L$ , satisfying

$$\sum_{l=0}^{\infty} l \sup_{0 \leq u \leq 1} |\alpha_l(u)| < \infty \quad \text{and} \quad \sum_{l=0}^{\infty} l \sup_{0 \leq u \leq 1} \left| \frac{\partial}{\partial u} \alpha_l(u) \right| < \infty,$$

and  $y_{0,n} = o_P(\sqrt{k_n})$  with  $k_n = o(n)$  being a sequence of positive real numbers.

We then consider the normalized serial correlation coefficient:

$$(3) \quad S_n := \frac{\frac{1}{k_n (\prod_{k=1}^n \rho_{k,n})} \sum_{t=1}^n y_{t-1,n} u_{t,n}}{\frac{1}{k_n^2 (\prod_{k=1}^n \rho_{k,n})^2} \sum_{t=1}^n y_{t-1,n}^2} := \frac{U_n}{V_n}.$$

Note that if  $\rho_{t,n} \equiv \rho_n = 1 + \frac{\beta}{k_n}$ ,  $t = 1, \dots, n$ , we have

$$(4) \quad S_n = k_n \left( \prod_{k=1}^n \rho_{k,n} \right) (\widehat{\rho}_n - \rho_n),$$

where  $\widehat{\rho}_n := \frac{\sum_{t=1}^n y_{t-1,n} y_{t,n}}{\sum_{t=1}^n y_{t-1,n}^2}$  is the least squares estimator (LSE) of the AR(1) process with constant coefficient:

$$y_{t,n} = \rho_n y_{t-1,n} + u_{t,n}, \quad t = 1, \dots, n, \quad \rho_n = 1 + \frac{\beta}{k_n}.$$

Therefore, the  $\widehat{\rho}_n$  has bias for the estimation of  $\rho_{t,n}$ . We then obtain the following theorem.

**Theorem 1.** Let  $\kappa = 2 \sqrt{\beta(0)\beta(1)} \left| \frac{\alpha(1,1)}{\alpha(0,1)} \right|$ . Then,

$$S_n \Rightarrow \kappa C.$$

Therefore, if  $\rho_{t,n} \equiv \rho_n = 1 + \frac{\beta}{k_n}$ ,  $t = 1, \dots, n$ ,

$$k_n \left( \prod_{k=1}^n \rho_{k,n} \right) (\widehat{\rho}_n - \rho_n) \Rightarrow \kappa C$$

from (4).

# 多変量バーンバウムサンダース型分布: 非負データの密度推定への応用

## 柿沢 佳秀 (北大経済)

1. はじめに カーネル密度推定量の漸近的性質はよく知られている (Silverman(1986), Wand and Jones(1995)) が, 推定される密度  $f$  の台が  $\mathbb{R}^d$  でない (典型的に  $[0, 1]^d$  や  $[0, \infty)^d$  の台) のときに, 境界付近で  $O(1)$  のバイアスがあり, 一致性をもたないという境界バイアス問題がある. 事実, カーネル密度推定の背後にある畳み込み近似  $\int_{\mathbb{R}^d} \frac{1}{h^d} k\left(\frac{\mathbf{x}-\mathbf{v}}{h}\right) f(\mathbf{v}) d\mathbf{v} = \int_{\mathbb{R}^d} k(\mathbf{u}) f(\mathbf{x}-h\mathbf{u}) d\mathbf{u} \approx f(\mathbf{x})$  as  $h \rightarrow 0$  (カーネル  $k$  に  $k(-\mathbf{v}) = k(\mathbf{v})$ ,  $\int_{\mathbb{R}^d} k(\mathbf{v}) d\mathbf{v} = 1$  を仮定する) が, 有界閉領域や正象限の境界付近では成立しない. この境界バイアスの対処として, 台の変換法 (Marron and Ruppert(1994)) の他, renormalization, reflection, 一般化ジャックナイフのような境界カーネルへと修正するアプローチ (Jones(1993)) が主に 1次元データで議論されてきた. 報告者は Chen(1999,2000) 以降の『非対称カーネル法 (AK 推定法)』に焦点を絞る.

1.1. 1次元の AK 推定 1次元の非負データ  $X_1, \dots, X_n \sim \text{iid } f$  に基づいた (点  $x \in [0, \infty) = \mathbb{R}_+$  での) AK 密度推定量は  $n^{-1} \sum_{i=1}^n k_{x,b}(X_i)$  と定義される<sup>1</sup>. ここに,  $b = b(n) > 0$  は  $n \rightarrow \infty$  のとき適当な速さで 0 に収束する平滑化パラメータ,  $k_{x,b}(\cdot)$  は  $f$  と同じ台  $\mathbb{R}_+$  を持つカーネルである<sup>2</sup>.

1.2. 多次元の (積型の) AK 推定  $d$ 次元の非負データ  $\mathbf{X}_1, \dots, \mathbf{X}_n \sim \text{iid } f$  (ただし,  $\mathbf{X}_i = (X_{i1}, \dots, X_{id})'$  とする) に基づき  $f(\mathbf{x})$ ,  $\mathbf{x} = (x_1, \dots, x_d)' \in \mathbb{R}_+^d$  を推定するなら, いわゆる “積カーネル法” が簡単で, カーネル  $k_{x,b}(\cdot)$  を 1.1 節のように選び, 積カーネルを作り, 積型の AK 密度推定量  $n^{-1} \sum_{i=1}^n \prod_{j=1}^d k_{x_j,b}(X_{ij})$  を定義する. Bouezmarni and Rombouts(2010) がベータ/ガンマカーネルによる積型の AK 推定を考察し, 他の文献もある (Funke and Kawka(2015), Zougab et al.(2018), Igarashi and Kakizawa(2020)).

2. BS 型分布 本報告では, 非積型の AK 推定を開発するために, BS 分布の周辺から分布論を再検討した.

2.1. 準備 (分布論) 1次元の確率変数  $\xi$  が密度関数  $p(s; \alpha, \beta) = \phi\left(\frac{a(s/\beta)}{\alpha}\right) \frac{A(s/\beta)}{\alpha\beta}$ ,  $s \in \mathbb{R}_+$  (Birnbaum and Saunders(1969)) を持つとき,  $\xi$  は BS 分布  $BS(\alpha, \beta)$  に従う. ここに,  $\alpha, \beta > 0$  であり,  $a(u) = u^{1/2} - \frac{1}{u^{1/2}}$ ,  $A(u) = (1/2)\left(\frac{1}{u^{1/2}} + \frac{1}{u^{3/2}}\right)$ ,  $\phi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$ . また, [a] 非心拡張, [b] 正規分布ベースから対称分布ベースへの拡張, [c] BS と LN とを繋ぐような拡張 (ジョンソンシステムの 1 例とみなせる), [d] 歪分布ベースへの拡張もできて, BS 型分布の一般形は以下のように与えられる.

q-BS 型密度 (柿沢 (2017)):  $h$  を  $\mathbb{R}$  上の分布として,

$$p_{(h)}^{(q)}(s; \alpha, \beta, \theta) = h\left(\frac{a^{(q)}(s/\beta)}{\alpha} - \theta\right) \frac{A^{(q)}(s/\beta)}{\alpha\beta} \quad (s \in \mathbb{R}_+)$$

<sup>1</sup>この形式は “重み付き推定量” (例えば, Silverman(1986, Subsection 2.9: General weight function estimators)) であるが, Chen(1999,2000) 以降は “非対称カーネル推定量” がキーワードである.

<sup>2</sup>先行研究 (Jin and Kawczak(2003), Scaillet(2004), Koul and Song(2013), Marchant et al.(2013), Saulo et al.(2013) など) のオリジナルの推定量において

(i)  $\hat{f}(0) = 0$  の制約 (明らかに,  $f(0) > 0$  の場合には適さない)

(ii)  $O(1)$  の境界バイアス (Rosenblatt-Parzen カーネル法と同様, これを回避する提案をすべきである)

(iii) 積分分散の主要項の存在性に関わる問題 (その存在を仮定してもよいが,  $f(0) > 0$  なる  $[0, \infty)$  上の有界な密度関数を排除してしまう; この仮定を外すべきだろう)

などがあり, いずれかの問題があるものについては, それらが回避された “再定義” を用いるべきである.

(この  $q$ -BS 型密度は,  $q = 0$  として対数対称密度や対数 skew 密度を含む). ここに,

$$a^{(q)}(u) = \begin{cases} \frac{1}{2q} (u^q - u^{-q}), & q \neq 0, \\ \log u, & q = 0 \end{cases}, \quad A^{(q)}(u) = (d/du)a^{(q)}(u).$$

[例]  $h(u) = \phi(u)$  の場合が, 正規分布ベースの  $q$ -BS,  $f_{S_g}(u) = C_g g(u^2)$  の場合が, 対称分布ベースの  $q$ -BS である (なお, Díaz-García and Leiva(2005) は  $\theta = 0$  の中心ケースで  $q = 1/2$  を扱い, “GBS” と呼んだが, Kakizawa(2018) は一般化の意味を強調した別名から “対称分布ベースの  $q$ -BS” を採用). 柿沢 (2017) は,  $h$  として Azzalini 型 skew/epsilon-skew 密度  $h_{AS_{g,j,\lambda}}(u) = 2f_{S_g}(u)F_{S_g}(\lambda u)$ ,  $h_{ES_{g,\epsilon}}(u) = f_{S_g}\left(\frac{u}{1-\text{sgn}(u)\epsilon}\right)$  ( $u \in \mathbb{R}$ ) を 2 例としている. ここに,  $F_{S_g}(t) = \int_{-\infty}^t f_{S_g}(u) du$  ( $t \in \mathbb{R}$ ), 及び,  $\text{sgn}(u) = u/|u|$  if  $u \neq 0$  and  $\text{sgn}(0) = 0$  とする.

2.2. 多次元拡張 (2次元の場合のみを掲載)  $h$  を  $\mathbb{R}^2$  上の分布として,

$$p_{(h)}^{(q)}(\mathbf{s}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) = h\left(\frac{a^{(q_1)}(s_1/\beta_1)}{\alpha_1} - \theta_1, \frac{a^{(q_2)}(s_2/\beta_2)}{\alpha_2} - \theta_2\right) \prod_{j=1}^2 \frac{A^{(q_j)}(s_j/\beta_j)}{\alpha_j \beta_j} \quad (\mathbf{s} \in \mathbb{R}_+^2).$$

本報告では,  $h$  として楕円密度

$$h_{E_{g,\rho}}(u_1, u_2) = \frac{C_{g,2}}{(1-\rho^2)^{1/2}} g\left\{\frac{u_1^2 + u_2^2 - 2\rho u_1 u_2}{2(1-\rho^2)}\right\} \quad (\mathbf{u} \in \mathbb{R}^2)$$

を例とし, それをベースにした  $(q_1, q_2)$ -BS 分布を扱った.

3. 密度推定への応用 想定したカーネル (報告で, 楕円分布ベースの  $q$ -BS 型に限定) に付随するパラメータを推定位置に依存させ, ノンパラメトリックな (可変的)AK 密度推定量を構築できる. パラメータ  $(\alpha, \beta, \theta)$  に対し,  $\alpha_{b,c}(x) = \left(\frac{b}{x+bc}\right)^{1/2}$ ,  $\beta_{b,c}(x) = x + bc$ ,  $\theta_{b,c}(x) = \theta_0 \alpha_{b,c}(x)$  (ただし,  $c > 0, \theta_0 \in \mathbb{R}$ ) を採用し, 1次元の対称分布ベースの  $q$ -BS カーネル密度推定量  $\hat{f}_{b,g}^{(q)}(x) = \frac{1}{n} \sum_{i=1}^n p_g^{(q)}(X_i; \alpha_{b,c}(x), \beta_{b,c}(x), \theta_0 \alpha_{b,c}(x))$ ,  $x \in \mathbb{R}_+$  (中心ケース  $\theta_0 = 0$  は Kakizawa(2018), 非心ケース  $\theta_0 \neq 0$  は柿沢 (2017) の Azzalini 型 skew/epsilon-skew  $q$ -BS カーネルで “歪みパラメータ” を 0 にした特別な場合) を定義したのと同様にして, 積型/非積型<sup>3</sup>の楕円分布ベースの  $(q_1, q_2)$ -BS カーネル密度推定量 ( $d$ 次元の推定量へ拡張可能)

$$\begin{aligned} \hat{f}_{b,g}^{(q_1, q_2)} \Pi(\mathbf{x}) &= \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^2 p_g^{(q_j)}(X_{ij}; \alpha_{b,c}(x_j), \beta_{b,c}(x_j), \theta_0 \alpha_{b,c}(x_j)), \\ \hat{f}_{b,g,\rho}^{(q_1, q_2)}(\mathbf{x}) &= \frac{1}{n} \sum_{i=1}^n p_{g,2}^{(q_1, q_2)}(\mathbf{X}_i; \boldsymbol{\alpha}_{b,c}(\mathbf{x}), \boldsymbol{\beta}_{b,c}(\mathbf{x}), \rho, \theta_0 \boldsymbol{\alpha}_{b,c}(\mathbf{x})), \end{aligned} \quad \mathbf{x} = (x_1, x_2)' \in \mathbb{R}_+^2 \quad (1)$$

を提案した. ここに,  $\boldsymbol{\alpha}_{b,c}(\mathbf{x}) = (\alpha_{b,c}(x_1), \alpha_{b,c}(x_2))'$ ,  $\boldsymbol{\beta}_{b,c}(\mathbf{x}) = (\beta_{b,c}(x_1), \beta_{b,c}(x_2))'$  とする.

主結果として, (1) の漸近バイアス/分散, MSE, MISE, 強一貫性, 漸近正規性を示した (いくつかの主要項は  $\mathbf{q} = (q_1, q_2)'$  に依存せず, これは初期の正規分布ベースの  $q = 1/2$  の BS カーネル密度推定量と  $q = 0$  の LN カーネル密度推定量が, 多くの同じ漸近性質を持つ根拠であった).

<sup>3</sup>楕円分布においては, 多変量正規分布以外で, 相関行列を仮に単位行列 (無相関) としても, 各要素は独立にならないことが知られており, 従って, 一般には,  $\hat{f}_{b,g}^{(q_1, q_2)} \Pi \neq \hat{f}_{b,g,0}^{(q_1, q_2)}$  となる. そのため, “積型の AK 密度推定量” も個別に扱う.



# The de-biased group Lasso estimation for varying coefficient models

一橋大学大学院経済学研究科 本田敏雄

近年データ収集技術の進歩により、様々な分野において非常に多くの高次元データが利用可能になり、それらのデータ解析の必要性が高まってきた。Lasso は Tibshirani(1996) で提案された、高次元線形回帰モデルに対する推定法であり、高次元データ解析の必要性、重要性の高まりとともに、実用面および理論面での研究が進んできた。理論面では、Bickel et al.(2009) が最も重要な論文である。

研究の進展に伴い Lasso の問題点：1. 特殊な条件下を除けばモデル選択に関する一致性はない、2. 比較的バイアスの大きな推定値を与える、が明らかになってきた。後者の問題を解決するために Javanmard and Montanari(2014), van de Geer et al.(2014), Zhang and Zhang(2014) により提案されたのが de-biased Lasso である。厳しい仮定の下ではあるが、Lasso のバイアスを取り除き、高次元の設定を維持したままでの統計的推測を可能にした。

Honda(2020) では、de-biased Lasso の手法を高次元変動係数モデルに適用し、理論的な結果およびシミュレーションの結果等を与えている。本講演では Honda(2020) に基づき、変動係数モデルの場合の de-biased Lasso 推定の理論的な結果とシミュレーションの結果を報告した。

まず以下のように de-biased Lasso の理論を簡単に紹介した。

高次元線形モデル： $\mathbf{X} = (X_{ij})$  を  $n \times p$  のデザイン行列とし、 $p > n$  とする。

$$Y = \mathbf{X}\beta_0 + \epsilon, \text{ ここで } Y \in \mathbb{R}^n \text{ および } \beta_0 \in \mathbb{R}^p.$$

Lasso 推定量：上のモデルでは、0 でない係数は  $p$  に比べて非常に少ないとする。

Lasso 推定量は以下で定義される。

$$\widehat{\beta} := \operatorname{argmin}_{\beta \in \mathbb{R}^p} \left\{ n^{-1} \|Y - \mathbf{X}\beta\|^2 + 2\lambda \sum_{j=1}^p |\beta_j| \right\},$$

ここで  $\beta = (\beta_1, \dots, \beta_p)^T$  とし、 $S := \{j | \beta_j \neq 0\}$  かつ  $s := |S|$  である。

この  $\widehat{\beta}$  は以下の式を満たす.

$$n^{-1} \mathbf{X}^T Y = n^{-1} \mathbf{X}^T \mathbf{X} \widehat{\beta} + \lambda \boldsymbol{\kappa},$$

ここで  $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_p)^T \in \mathbb{R}^p$  および  $|\kappa_j| \leq 1$  である.

上記の設定では,  $\widehat{\Sigma} = n^{-1} \mathbf{X}^T \mathbf{X}$  の逆行列が存在しないので, その代わりとなる  $\widehat{\Theta}$  を Lasso などを用いて構成する. そして以下の式により, de-biased Lasso 推定量は定義される.

$$\begin{aligned} \widehat{\mathbf{b}} &:= \widehat{\beta} + \widehat{\Theta} \lambda \boldsymbol{\kappa} = \widehat{\beta} + \frac{1}{n} \widehat{\Theta} \mathbf{X}^T (Y - \mathbf{X} \widehat{\beta}) \\ &= \beta_0 + \frac{1}{n} \widehat{\Theta} \mathbf{X}^T \boldsymbol{\epsilon} + \Delta, \end{aligned}$$

ここで  $\Delta$  は無視可能な項である.

ついで変動線形モデルを紹介した上で, 以上の方法に倣って変動線形モデルに対する de-biased Lasso 推定量を定義し, Honda(2020) で与えられている結果を紹介した.

#### 参考文献

- P. J. Bickel, Y. Ritov, and A. B. Tsybakov (2009). Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.*, 37:1705–1732.
- A. Javanmard and A. Montanari (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *J. Machine Learning Research*, 15:2869–2909.
- T. Honda (2020). The de-biased group Lasso estimation for varying coefficient models. Forthcoming in *Ann. Inst. Statist. Math.*
- R. Tibshirani (1996). Regression shrinkage and selection via the lasso. *J. Royal Statist. Soc. B*, 58:267–288.
- S. van de Geer, P. Bühlmann, Y. Ritov, and R. Dezeure (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist.*, 42:1166–1202.
- C.-H. Zhang and S. S. Zhang (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. Royal Statist. Soc. B*, 76:217–242.

# High-dimensional covariance matrix estimation under the strongly spiked eigenvalue model

Keisuke Konishi<sup>a</sup>, Kazuyoshi Yata<sup>b</sup>, Makoto Aoshima<sup>b</sup>

<sup>a</sup>Graduate School of Pure and Applied Sciences, University of Tsukuba

<sup>b</sup>Institute of Mathematics, University of Tsukuba

## 1 Introduction

One of the features of high-dimensional data is that the data dimension  $d$  is high, however, the sample size  $n$  is low. This is the so-called ‘‘HDLSS’’ or ‘‘large  $d$ , small  $n$ ’’ data. For HDLSS data, the sample covariance matrix does not have its inverse matrix. The estimation for the inverse matrix of a covariance matrix is a crucial issue for high-dimensional data analyses, especially for pathway analysis and graphical modeling.

Bickel and Levina (2008, AS) gave a thresholding estimator for the inverse matrix of a covariance matrix when the covariance matrix is sparse and its eigenvalues are bounded. However, such sparsity conditions are severe for actual data and often out of touch with reality. In fact, Aoshima and Yata (2018, Sinica; 2019, MCAP) showed that the first several eigenvalues diverge as  $d$  grows and the bounded-eigenvalues condition is quite strict for microarray data sets. Fan et al. (2013, JRSS-B) proposed a different thresholding estimator called the principal orthogonal complement thresholding (POET) under the assumption that the first several eigenvalues diverge rapidly at the rate of  $d$ . Unfortunately, the assumption required in the POET cannot express the structure of actual data.

Suppose we have a  $d \times n$  data matrix  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ , where  $\mathbf{x}_i$ ,  $i = 1, \dots, n$  ( $< d$ ), are independent and identically distributed (i.i.d.) as a  $d$ -dimensional distribution with mean zero and covariance matrix  $\mathbf{\Sigma}$ . We denote the eigen-decomposition of  $\mathbf{\Sigma}$  by  $\mathbf{\Sigma} = \mathbf{H}\mathbf{\Lambda}\mathbf{H}^T$ , where  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_d)$  with eigenvalues  $\lambda_1 \geq \dots \geq \lambda_d \geq 0$  and  $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_d]$  is an orthogonal matrix with eigenvectors  $\mathbf{h}_1, \dots, \mathbf{h}_d$  corresponding to the  $\lambda_1, \dots, \lambda_d$ . Let  $\mathbf{x}_j = \mathbf{H}\mathbf{\Lambda}^{1/2}\mathbf{z}_j$ , where  $\mathbf{z}_j = (z_{1j}, \dots, z_{dj})^T$  is considered as a sphered data vector having the zero mean vector and identity covariance matrix. We assume that

$$\begin{aligned} \text{(C-i)} \quad & \limsup_{d \rightarrow \infty} E(z_{ri}^4) < \infty \text{ for all } r, \text{ and} \\ & E(z_{ri}^2 z_{si}^2) = E(z_{ri}^2)E(z_{si}^2) = 1 \text{ and } E(z_{ri} z_{si} z_{ti} z_{ui}) = 0 \text{ for all } r \neq s, t, u. \end{aligned}$$

When  $\mathbf{x}_j$ s are Gaussian, (C-i) naturally holds.

The sample covariance matrix is given by  $\mathbf{S} = n^{-1}\mathbf{X}\mathbf{X}^T$ . Let  $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_d \geq 0$  be the eigenvalues of  $\mathbf{S}$ . Then, we denote the eigen-decomposition of  $\mathbf{S}$  by  $\mathbf{S} = \sum_{i=1}^d \hat{\lambda}_i \hat{\mathbf{h}}_i \hat{\mathbf{h}}_i^T$ , where  $\hat{\mathbf{h}}_i$  is a unit eigenvector corresponding to the  $\hat{\lambda}_i$ . The dual sample covariance matrix is given by  $\mathbf{S}_D = n^{-1}\mathbf{X}^T\mathbf{X}$ . We have the eigen-decomposition of  $\mathbf{S}_D$  by  $\mathbf{S}_D = \sum_{i=1}^n \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^T$ , where  $\hat{\mathbf{u}}_i$  is a unit eigenvector corresponding to the  $\hat{\lambda}_i$ . We assume the following spiked model for the eigenvalues of  $\mathbf{\Sigma}$ :

(C-ii)  $\frac{\lambda_i}{d^{1/2}} \rightarrow \infty$  as  $d \rightarrow \infty$  for  $i = 1, \dots, m$ , and  $\lambda_i \in (0, \infty)$  as  $d \rightarrow \infty$  for all  $i \geq m + 1$ . Here,  $m$  is a positive and fixed integer.

Note that (C-ii) is one of the strongly spiked eigenvalue models given by Aoshima and Yata (2018, Sinica). We divide  $\Sigma$  into  $\Sigma_1 = \sum_{j=1}^m \lambda_j \mathbf{h}_j \mathbf{h}_j^T$  and  $\Sigma_2 = \sum_{j=m+1}^d \lambda_j \mathbf{h}_j \mathbf{h}_j^T$ , so that  $\Sigma = \Sigma_1 + \Sigma_2$ .

## 2 NR-POET and its asymptotic properties

In this section, we propose a new estimation of  $\Sigma^{-1}$ . Yata and Aoshima (2012, JMVA) proposed an eigenvalue estimation called the noise-reduction (NR) method as followings.

$$\tilde{\lambda}_j = \hat{\lambda}_j - \frac{\text{tr}(\mathbf{S}_D) - \sum_{l=1}^j \hat{\lambda}_l}{n - j}, \quad \tilde{\mathbf{h}}_j = (n\tilde{\lambda}_j)^{-1/2} \mathbf{X} \hat{\mathbf{u}}_j \quad (j = 1, \dots, n - 1). \quad (1)$$

For estimating  $\mathbf{x}_l^T \mathbf{h}_j$ , Aoshima and Yata (2018, Sinica) showed that  $\mathbf{x}_l^T \hat{\mathbf{h}}_j$  and even  $\mathbf{x}_l^T \tilde{\mathbf{h}}_j$  involve a huge bias. In order to overcome the inconvenience, they gave the modified NR method. We modify  $\mathbf{x}_l^T \tilde{\mathbf{h}}_j$  as  $\mathbf{x}_l^T \tilde{\mathbf{h}}_{jl}$  by

$$\tilde{\mathbf{h}}_{jl} = \left( \frac{n}{n-1} \right) \frac{\mathbf{X} \hat{\mathbf{u}}_{jl}}{(n\tilde{\lambda}_j)^{1/2}} = \frac{n^{1/2} \mathbf{X} \hat{\mathbf{u}}_{jl}}{(n-1)\tilde{\lambda}_j^{1/2}}, \quad \text{where} \quad \hat{\mathbf{u}}_{jl} = (\hat{u}_{j1}, \dots, \hat{u}_{jl-1}, 0, \hat{u}_{jl+1}, \dots, \hat{u}_{jn})^T.$$

We propose a new estimation of  $\Sigma^{-1}$  by applying the noise-reduction (NR) method to the POET. By using the modified NR method, we estimate  $\mathbf{w}_l$  by  $\tilde{\mathbf{w}}_l = (\mathbf{I}_d - \sum_{j=1}^m \hat{\mathbf{h}}_j \hat{\mathbf{h}}_{jl}^T) \mathbf{x}_l$ . We consider estimating  $\Sigma_2$  by  $T(\tilde{\Sigma}_2)$ , where  $\tilde{\Sigma}_2 = \sum_{l=1}^n \tilde{\mathbf{w}}_l \tilde{\mathbf{w}}_l^T / n$  and  $T(\cdot)$  is sparse operator proposed by Bickel and Levina (2008, AS). We denote the eigen-decomposition of  $T(\tilde{\Sigma}_2)$  by  $T(\tilde{\Sigma}_2) = \sum_{j=1}^d \acute{\lambda}_j \acute{\mathbf{h}}_j \acute{\mathbf{h}}_j^T$ , where  $\acute{\lambda}_j$ s are eigenvalues of  $T(\tilde{\Sigma}_2)$  having  $\acute{\lambda}_1 \geq \dots \geq \acute{\lambda}_d \geq 0$  and  $\acute{\mathbf{h}}_j$  is a unit eigenvector corresponding to the  $\acute{\lambda}_j$ . Note that

$$\Sigma^{-1} = \sum_{j=1}^m \lambda_j^{-1} \mathbf{h}_j \mathbf{h}_j^T + \sum_{j=m+1}^d \lambda_j^{-1} \mathbf{h}_j \mathbf{h}_j^T.$$

Finally, by applying the NR method to the intrinsic part and the noise part, we propose to estimate  $\Sigma^{-1}$  by

$$\tilde{\Sigma}^{-1} = \sum_{j=1}^m \tilde{\lambda}_j^{-1} \tilde{\mathbf{h}}_j \tilde{\mathbf{h}}_j^T + \left( \mathbf{I}_d - \sum_{j=1}^m \tilde{\mathbf{h}}_j \tilde{\mathbf{h}}_j^T \right) \left( \sum_{j=1}^{d-m} \acute{\lambda}_j^{-1} \acute{\mathbf{h}}_j \acute{\mathbf{h}}_j^T \right) \left( \mathbf{I}_d - \sum_{j=1}^m \tilde{\mathbf{h}}_j \tilde{\mathbf{h}}_j^T \right).$$

Note that the intrinsic and noise parts are orthogonal. We call this new estimation method the "NR-POET". We have the following result.

**Theorem 2.1.** *Assume (C-i) and (C-ii). Then, under some regularity conditions, it holds that as  $d \rightarrow \infty$  and  $n \rightarrow \infty$*

$$\|\Sigma^{1/2} \tilde{\Sigma}^{-1} \Sigma^{1/2} - \mathbf{I}_d\|_F / \sqrt{d} = o_P(1).$$

# Asymptotic Properties of Distance Weighted Discrimination and Its Bias Correction in HDLSS Settings

Kento Egashira<sup>a</sup>, Kazuyoshi Yata<sup>b</sup>, Makoto Aoshima<sup>b</sup>

<sup>a</sup>Graduate School of Pure and Applied Sciences, University of Tsukuba

<sup>b</sup>Institute of Mathematics, University of Tsukuba

## 1 Introduction

Suppose we have two independent and  $d$ -variate populations,  $\Pi_i$ ,  $i = 1, 2$ , having an unknown mean vector  $\boldsymbol{\mu}_i$  and unknown covariance matrix  $\boldsymbol{\Sigma}_i (> \mathbf{O})$  for each  $i = 1, 2$ . Let us have a training data set,  $\boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{in_i}$ , from  $\Pi_i$  for  $i = 1, 2$ . Let  $N = n_1 + n_2$ . We denote the class labels of  $y_i$  by  $-1$  for  $i = 1, \dots, n_1$ , and by  $+1$  for  $i = n_1 + 1, \dots, N$ . We simply write that  $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N) = (\boldsymbol{x}_{11}, \dots, \boldsymbol{x}_{1n_1}, \boldsymbol{x}_{21}, \dots, \boldsymbol{x}_{2n_2})$ . Let  $\boldsymbol{x}_0$  be an observation vector of an individual belonging to one of the  $\Pi_i$ s. We assume that  $\boldsymbol{x}_0$  and  $\boldsymbol{x}_{ij}$ s are independent.

The support vector machine (SVM) developed by Vapnick (2000) is well known as a representative tool of binary linear discriminant analysis. However, Marron et al. (2007) pointed out that the SVM causes data piling in the HDLSS context. Data piling is a phenomenon that the projection of a training data to the normal direction vector of a separating hyperplane is same for each class. In order to avoid the data piling problem of the SVM, Marron et al. (2007) proposed distance weighted discrimination (DWD). Whereas the SVM finds the optimal hyperplane by maximizing the minimum distances from each class to the hyperplane, the DWD finds a proper hyperplane by minimizing the sum of reciprocals of the distance from each data point to the hyperplane. The DWD cares all the data vectors that are not always used in the SVM. Unfortunately, the DWD is designed for balanced training data sets. For imbalanced training data sets, Qiao et al. (2010) developed weighted DWD (WDWD) that imposes different weights on two classes. However, the WDWD is sensitive for a choice of weight for each class and it is quite difficult to select a proper weight in actual data analyses.

In this talk, we investigated DWD theoretically in the HDLSS context where  $d \rightarrow \infty$  while  $n$  is fixed. We showed that the DWD includes a huge bias caused by heterogeneity of covariance matrices as well as sample imbalance. We proposed a bias corrected-DWD (BC-DWD) and showed that the BC-DWD can enjoy consistency properties about misclassification rates.

## 2 DWD and its asymptotic properties

We define the classifier for the DWD as  $\hat{y}(\mathbf{x}_0) = \hat{\mathbf{w}}^T \mathbf{x}_0 + \hat{b}$ . One classifies  $\mathbf{x}_0$  into  $\Pi_1$  if  $\hat{y}(\mathbf{x}_0) < 0$  and into  $\Pi_2$  otherwise. Let  $\Delta = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2$ . We assume the following assumptions:

$$\text{(A-i)} \quad \frac{\text{Var}(\|\mathbf{x}_{ij} - \boldsymbol{\mu}_i\|^2)}{\Delta^2} \rightarrow 0 \text{ as } d \rightarrow \infty \text{ for } i = 1, 2;$$

$$\text{(A-ii)} \quad \frac{\text{tr}(\boldsymbol{\Sigma}_i^2)}{\Delta^2} \rightarrow 0 \text{ as } d \rightarrow \infty \text{ for } i = 1, 2.$$

Let  $\Delta_* = \Delta + \text{tr}(\boldsymbol{\Sigma}_1)/n_1 + \text{tr}(\boldsymbol{\Sigma}_2)/n_2$  and

$$\delta = -\frac{(n_1^{3/2} - n_2^{3/2})\Delta_*}{N(\sqrt{n_1} + \sqrt{n_2})\Delta} + \frac{\text{tr}(\boldsymbol{\Sigma}_1) - \text{tr}(\boldsymbol{\Sigma}_2)}{N\Delta}.$$

We consider the following assumption:

$$\text{(A-iii)} \quad \limsup_{d \rightarrow \infty} |\delta| < \frac{\min_{i=1,2} n_i}{N}.$$

Let  $e(i)$  denote the error rate of misclassifying an individual from  $\Pi_i$  into the other class for  $i = 1, 2$ .

**Theorem 1.** *Under (A-i) to (A-iii), the DWD holds:*

$$e(1) \rightarrow 0 \text{ and } e(2) \rightarrow 0 \text{ as } d \rightarrow \infty.$$

**Corollary 1.** *Under (A-i) and (A-ii), the DWD holds:*

$$e(1) \rightarrow 1 \text{ and } e(2) \rightarrow 0 \text{ as } d \rightarrow \infty \text{ if } \liminf_{d \rightarrow \infty} \delta > \frac{n_2}{N}; \text{ and}$$

$$e(1) \rightarrow 0 \text{ and } e(2) \rightarrow 1 \text{ as } d \rightarrow \infty \text{ if } \limsup_{d \rightarrow \infty} \delta < -\frac{n_1}{N}.$$

From Corollary 1, the DWD brings the strong inconsistency because of a huge bias caused by heterogeneity of covariance matrices as well as sample imbalance. In order to overcome such difficulties, we proposed a bias-corrected DWD.

## References

- [1] Marron, J.S., Todd, M.J., Ahn, J. (2007). Distance-weighted discrimination. *Journal of the American Statistical Association*, 102, 1267–1271.
- [2] Qiao, X., Zhang, H. H., Liu, Y., Todd, M.J., Marron, J.S. (2010). Weighted distance weighted discrimination and its asymptotic properties. *Journal of the American Statistical Association*, 105, 401–414.
- [3] Vapnik, V. N. (2000). *The Nature of Statistical Learning Theory (second ed.)*. New York: Springer.

# 順序カテゴリをもつ多元分割表における非同等性のモデリング

東京理科大学大学院 藤澤 健吾  
東京理科大学 田畑 耕治

## 1. 正方分割表における一般化周辺非同等モデル

行変数  $X$  と列変数  $Y$  が順序のある同じ分類からなる  $r \times r$  正方分割表において,  $(i, j)$  セル確率を  $\pi_{ij}$  ( $i = 1, \dots, r; j = 1, \dots, r$ ) とする. 周辺同等 (MH) モデルは次のように表される (Stuart, 1955):

$$\pi_{i+} = \pi_{+i} \quad (i = 1, \dots, r),$$

ただし,  $\pi_{i+} = \sum_{t=1}^r \pi_{it}$ ,  $\pi_{+i} = \sum_{s=1}^r \pi_{si}$  である. MH モデルは累積確率を用いて次のように表すこともできる:

$$F_i^X = F_i^Y \quad (i = 1, \dots, r-1),$$

ただし,  $F_i^X = \sum_{s=1}^i \pi_{s+}$ ,  $F_i^Y = \sum_{s=1}^i \pi_{+s}$  である. MH モデルがデータに当てはまらないとき, MH モデルの拡張モデルを当てはめることに関心がある. 任意の  $k$  ( $k = 1, \dots, r-1$ ) に対して,  $k$  次一般化周辺非同等 ( $GM_k$ ) モデルを次のように提案した:

$$G^{-1}(F_i^X) = \sum_{l=0}^{k-1} i^l \Delta_l + G^{-1}(F_i^Y) \quad (i = 1, \dots, r-1),$$

ここで  $G$  は  $G(-\infty) = 0$ ,  $G(\infty) = 1$  を満たす連続な単調増加関数である. 特に  $\Delta_0 = \Delta_1 = \dots = \Delta_{k-1} = 0$  のときは MH モデルである. また,  $k = 1$ ,  $G^{-1}(x) = \log x / (1 - x)$  のときは周辺累積ロジットモデルである (McCullagh, 1977). さらに,  $k = 1$ ,  $G^{-1}(x) = \log(-\log(1 - x))$  のときは周辺累積補対数対数モデルである (Saigusa *et al.*, 2018). その他にも  $k = 1$ ,  $G^{-1}(x) = \Phi^{-1}(x)$  を考えることで周辺累積プロビットモデル等を構成することが可能である. ただし  $\Phi(x)$  は標準正規分布の分布関数である.

任意の  $k$  ( $k = 1, \dots, r-1$ ) に対して, 周辺  $k$  次積率一致 ( $ME_k$ ) モデルを次のように考える:

$$\sum_{i=1}^r i^l \pi_{i+} = \sum_{i=1}^r i^l \pi_{+i} \quad (l = 1, \dots, k).$$

このとき次の定理を得た:

**定理 1** 与えられた  $k$  ( $k = 1, \dots, r-1$ ) に対して, MH モデルが成立するための必要十分条件は  $GM_k$  モデルと  $ME_k$  モデルの両方が成立することである.

この定理は Saigusa *et al.* (2018) 等の結果を含む.

## 2. 多元分割表における一般化周辺非同等モデル

$X_t$  ( $t = 1, \dots, T$ ) が順序のある同じ分類からなる  $r^T$  分割表において,  $(i_1, \dots, i_T)$  セル確率を  $P(X_1 = i_1, \dots, X_T = i_T) = \pi_{i_1 \dots i_T}$  ( $i_t = 1, \dots, r$ ) とする. このとき  $r^T$  分割表における  $T$  元周辺同等 (MH[T]) モデルは次のように表される (Bhapkar and Darroch, 1990):

$$\pi_i^{(1)} = \dots = \pi_i^{(T)} \quad (i = 1, \dots, r),$$

ただし  $\pi_i^{(t)} = P(X_t = i)$  である. また累積確率を用いて次のように表すこともできる:

$$F_i^{X_1} = \dots = F_i^{X_T} \quad (i = 1, \dots, r-1),$$

ただし  $F_i^{X_t} = \sum_{s=1}^i \pi_s^{(t)}$  である. MH[T] モデルの当てはまりが悪い場合, MH[T] モデルより制約の弱いモデルに関心がある. 任意の  $k$  ( $k = 1, \dots, r-1$ ) に対して,  $T$  元  $k$  次一般化周辺非同等 (GM<sub>k</sub>[T]) モデルを次のように提案した:

$$G^{-1}(F_i^{X_t}) = \sum_{l=0}^{k-1} i^l \Delta_i^{(t)} + G^{-1}(F_i^{X_1}) \quad (i = 1, \dots, r-1; t = 2, \dots, T).$$

このモデルは Agresti (2002) 等のモデルを含む.

$k$  ( $k = 1, \dots, r-1$ ) に対して,  $T$  元周辺  $k$  次積率一致 (ME<sub>k</sub>[T]) モデルを次のように考える:

$$\sum_{i=1}^r i^l \pi_i^{(1)} = \dots = \sum_{i=1}^r i^l \pi_i^{(T)} \quad (l = 1, \dots, k)$$

このとき次の定理が成立する:

**定理 2** MH[T] モデルが成立するための必要十分条件は GM<sub>k</sub>[T] モデルと ME<sub>k</sub>[T] モデルの両方が成立することである.

### 参考文献

- [1] Agresti, A. (2002). *Categorical Data Analysis*, 2nd edition. Wiley, New York.
- [2] Bhapkar, V. P. and Darroch, J. N. (1990). Marginal symmetry and quasi symmetry of general order. *Journal of Multivariate Analysis*, **34**, 173-184.
- [3] McCullagh, P. (1977). A logistic model for paired comparisons with ordered categorical data. *Biometrika*, **64**, 449-453.
- [4] Saigusa, Y., Maruyama, T., Tahata, K., and Tomizawa, S. (2018). Extended marginal homogeneity model based on complementary log-log transform for square tables. *International Journal of Statistics and Probability*, **7**, 27-31.
- [5] Stuart, A. (1955). A test for homogeneity of the marginal distributions in a two-way classification. *Biometrika*, **42**, 412-416.



# High-dimensional classification by data transformation (データ変換を用いた高次元判別分析について)

Aki Ishii<sup>a</sup>, Kazuyoshi Yata<sup>b</sup>, and Makoto Aoshima<sup>b</sup>

<sup>a</sup> Department of Information Sciences, Tokyo University of Science

<sup>b</sup> Institute of Mathematics, University of Tsukuba

Suppose we have two classes  $\pi_i$ ,  $i = 1, 2$ , and define independent  $p \times n_i$  data matrices,  $\mathbf{X}_i = [\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i}]$ ,  $i = 1, 2$ , from  $\pi_i$ ,  $i = 1, 2$ , where  $\mathbf{x}_{ij}$ ,  $j = 1, \dots, n_i$ , are independent and identically distributed (i.i.d.) as a  $p$ -dimensional distribution with a mean vector  $\boldsymbol{\mu}_i$  and covariance matrix  $\boldsymbol{\Sigma}_i$  ( $\geq \mathbf{O}$ ). We assume that  $\limsup_{p \rightarrow \infty} \|\boldsymbol{\mu}_i\|^2/p < \infty$  for  $i = 1, 2$ , where  $\|\cdot\|$  denotes the Euclidean norm. Also, we assume that  $\text{tr}(\boldsymbol{\Sigma}_i)/p \in (0, \infty)$  as  $p \rightarrow \infty$  for  $i = 1, 2$ . Here, for a function,  $f(\cdot)$ , “ $f(p) \in (0, \infty)$  as  $p \rightarrow \infty$ ” implies  $\liminf_{p \rightarrow \infty} f(p) > 0$  and  $\limsup_{p \rightarrow \infty} f(p) < \infty$ . We assume  $n_i \geq 3$ ,  $i = 1, 2$ . The eigen-decomposition of  $\boldsymbol{\Sigma}_i$  is given by

$$\boldsymbol{\Sigma}_i = \mathbf{H}_i \boldsymbol{\Lambda}_i \mathbf{H}_i^T = \sum_{s=1}^p \lambda_{s(i)} \mathbf{h}_{s(i)} \mathbf{h}_{s(i)}^T,$$

where  $\boldsymbol{\Lambda}_i = \text{diag}(\lambda_{1(i)}, \dots, \lambda_{p(i)})$  having  $\lambda_{1(i)} \geq \dots \geq \lambda_{p(i)} (\geq 0)$  and  $\mathbf{H}_i = [\mathbf{h}_{1(i)}, \dots, \mathbf{h}_{p(i)}]$  is an orthogonal matrix of the corresponding eigenvectors. Let  $\mathbf{X}_i - [\boldsymbol{\mu}_i, \dots, \boldsymbol{\mu}_i] = \mathbf{H}_i \boldsymbol{\Lambda}_i^{1/2} \mathbf{Z}_i$  for  $i = 1, 2$ . Then,  $\mathbf{Z}_i$  is a  $p \times n_i$  sphered data matrix from a distribution with the zero mean and identity covariance matrix. Let  $\mathbf{Z}_i = [\mathbf{z}_{1(i)}, \dots, \mathbf{z}_{p(i)}]^T$  and  $\mathbf{z}_{j(i)} = (z_{j1(i)}, \dots, z_{jn_i(i)})^T$ ,  $j = 1, \dots, p$ , for  $i = 1, 2$ . Note that  $E(z_{jk(i)} z_{j'k(i)}) = 0$  ( $j \neq j'$ ) and  $\text{Var}(z_{j(i)}) = \mathbf{I}_{n_i}$ , where  $\mathbf{I}_{n_i}$  denotes the  $n_i$ -dimensional identity matrix. Also, note that if  $\mathbf{X}_i$  is Gaussian,  $z_{jk(i)}$ 's are i.i.d. as the standard normal distribution,  $N(0, 1)$ . We assume that the fourth moments of each variable in  $\mathbf{Z}_i$  are uniformly bounded for  $i = 1, 2$ . Let  $\mathbf{z}_{oj(i)} = \mathbf{z}_{j(i)} - (\bar{z}_{j(i)}, \dots, \bar{z}_{j(i)})^T$ ,  $j = 1, \dots, p$ ;  $i = 1, 2$ , where  $\bar{z}_{j(i)} = n_i^{-1} \sum_{k=1}^{n_i} z_{jk(i)}$ . We also assume that

$$P\left(\liminf_{p \rightarrow \infty} \|\mathbf{z}_{o1(i)}\| \neq 0\right) = 1 \text{ for } i = 1, 2.$$

Also, we consider the following assumption as necessary:

$$\mathbf{(A-i)} \quad E(z_{qk(i)}^2 z_{sk(i)}^2) = 1, \quad E(z_{qk(i)} z_{sk(i)} z_{tk(i)}) = 0 \text{ and } E(z_{qk(i)} z_{sk(i)} z_{tk(i)} z_{uk(i)}) = 0 \text{ for all } q \neq s, t, u.$$

We note that (A-i) naturally holds when  $\mathbf{X}_i$  is Gaussian.

Let  $\mathbf{x}_0$  be an observation vector of an individual belonging to  $\pi_i$  ( $i = 1, 2$ ). We assume  $\mathbf{x}_0$  and  $\mathbf{x}_{ijs}$  are independent. We estimate  $\boldsymbol{\mu}_i$  and  $\boldsymbol{\Sigma}_i$  by  $\bar{\mathbf{x}}_{in_i} = \sum_{j=1}^{n_i} \mathbf{x}_{ij}/n_i$  and  $\mathbf{S}_{in_i} = \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_{in_i})(\mathbf{x}_{ij} - \bar{\mathbf{x}}_{in_i})^T / (n_i - 1)$ . A typical classification rule is that one classifies an individual into  $\pi_1$  if

$$(\mathbf{x}_0 - \bar{\mathbf{x}}_{1n_1})^T \mathbf{S}_{1n_1}^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_{1n_1}) - \log \left\{ \frac{\det(\mathbf{S}_{2n_2})}{\det(\mathbf{S}_{1n_1})} \right\} < (\mathbf{x}_0 - \bar{\mathbf{x}}_{2n_2})^T \mathbf{S}_{2n_2}^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_{2n_2}), \quad (1.1)$$

and into  $\pi_2$  otherwise. However, the inverse matrix of  $\mathbf{S}_{in_i}$  does not exist in the HDLSS context ( $p > n_i$ ). When  $\Sigma_1 = \Sigma_2$ , Bickel and Levina (2004) considered the inverse matrix defined by only diagonal elements of the pooled sample covariance matrix. Yata and Aoshima (2012) considered using a ridge-type inverse covariance matrix derived by the *noise reduction (NR) methodology*. When  $\Sigma_1 \neq \Sigma_2$ , Dudoit et al. (2002) considered using the inverse matrix defined by only diagonal elements of  $\mathbf{S}_{in_i}$ . Aoshima and Yata (2011,2014) considered substituting  $\{\text{tr}(\mathbf{S}_{in_i})/p\}\mathbf{I}_p$  for  $\mathbf{S}_{in_i}$  in (1.1) by using the difference of a geometric representation of HDLSS data from each  $\pi_i$ . Aoshima and Yata (2019) considered quadratic classifiers in general and discussed asymptotic properties and optimality of the classifiers under high-dimensional settings. They showed that misclassification rates tend to zero as the dimension goes to infinity. On the other hand, Chan and Hall (2009) and Aoshima and Yata (2014) considered distance-based classifiers and Aoshima and Yata (2014) gave the misclassification rate adjusted classifier for multiclass, high-dimensional data whose misclassification rates are no more than specified thresholds under the following condition for eigenvalues:

$$\frac{\lambda_{1(i)}^2}{\text{tr}(\Sigma_i^2)} \rightarrow 0 \text{ as } p \rightarrow \infty \text{ for } i = 1, 2. \quad (1.2)$$

Recently, Aoshima and Yata (2018) considered the ‘‘strongly spiked eigenvalue (SSE) model’’ as follows:

$$\liminf_{p \rightarrow \infty} \left\{ \frac{\lambda_{1(i)}^2}{\text{tr}(\Sigma_i^2)} \right\} > 0 \text{ for } i = 1 \text{ or } 2. \quad (1.3)$$

On the other hand, Aoshima and Yata (2018) called (1.2) the ‘‘non-strongly spiked eigenvalue (NSSE) model’’. For example, we consider a spiked model such as

$$\lambda_{s(i)} = a_{s(i)} p^{\alpha_{s(i)}} \text{ (} s = 1, \dots, t_i \text{)} \quad \text{and} \quad \lambda_{s(i)} = c_{s(i)} \text{ (} s = t_i + 1, \dots, p \text{)} \quad (1.4)$$

with positive and fixed constants,  $a_{s(i)}$ ’s,  $c_{s(i)}$ ’s and  $\alpha_{s(i)}$ ’s, and a positive and fixed integer  $t_i$ . Note that (1.2) holds when  $\alpha_{1(i)} < 1/2$  for  $i = 1, 2$ . On the other hand, (1.3) holds for the spiked model in (1.4) with  $\alpha_{1(i)} \geq 1/2$ .

As for the SSE model, Aoshima and Yata (2018) considered a classifier by using a data transformation from the SSE model to the NSSE model. They gave a consistency property of the classifier and discussed the asymptotic normality when  $p \rightarrow \infty$  and  $n_i \rightarrow \infty$  ( $i = 1, 2$ ).

In this talk, we created a new quadratic classifier under one of the SSE models. We emphasize that one should construct a classification procedure by considering the eigenstructure of high-dimensional data. We handled the divergence condition for  $p$  and  $n_i$ ’s such as  $p \rightarrow \infty$  while  $n_i$ ’s are fixed. We introduced a quadratic classification procedure for high-dimensional data. We proposed a new quadratic classifier under the SSE model. We showed that our classification procedure has a consistency property for misclassification rates.

# Robust Bayesian Modeling with Synthetic Posterior\*

Shonosuke Sugasawa

Center for Spatial Information Science, The University of Tokyo

## 1 Introduction

The standard Bayesian method in which the posterior distribution is obtained as the product of likelihood and prior would be able to work reasonably well when the assumed model is approximately correct. However, it is often the case that real data contains outliers that the assumed model cannot explain. In this case, the standard Bayesian method would break down in the sense that the posterior distribution does not give reasonable Bayesian inference. In this project, we solve the difficulty of the standard Bayesian approaches by adopting synthetic posterior based on robust divergence. In particular, we focus on Bayesian linear regression and dynamic linear models and propose their robust alternatives.

## 2 Robust Bayesian Regression

Suppose we have independent observation  $(y_i, x_i)$  for  $i = 1, \dots, n$ , where  $y_i$  is a continuous response and  $x_i = (x_{i1}, \dots, x_{ip})^t$  is a  $p$ -dimensional vector of covariates. We consider fitting a regression model  $y_i = x_i^t \beta + \varepsilon_i$  with  $\varepsilon \sim N(0, \sigma^2)$ . Let  $\pi(\beta, \sigma^2)$  be prior distribution for the model parameters  $\beta$  and  $\sigma^2$ . Then the standard posterior distribution of the model parameters is given by the product of the prior and the likelihood function, which is known to be sensitive to outliers, and it could produce biased or inefficient posterior inference.

---

\*Joint work with Shintaro Hashimoto (Hiroshima University)

To overcome the problem, we propose replacing the log-likelihood function with robust alternatives. Specifically, we employ  $\gamma$ -divergence to define synthetic posterior for  $(\beta, \sigma^2)$ . We develop an efficient sampling strategy for the posterior distribution via Bayesian bootstrap or nonparametric Bayesian updating, which generate posterior samples by minimizing randomized loss function defined by the synthetic posterior. The minimization problem can be efficiently carried out via MM-algorithm. In order to address the case where the number of covariates is relatively large, we extended the propose method to incorporated shrinkage priors on regression coefficients.

### 3 Robust Dynamic Linear Modeling

Let  $(y_1, \dots, y_t)$  be a time series and we consider a simple dynamic linear model,  $y_t = \theta_t + \varepsilon_t$  and  $\theta_t = \theta_{t-1} + u_t$  for  $t = 1, \dots, T$ , where  $\varepsilon_t \sim N(0, \sigma^2)$ ,  $u_t \sim N(0, \tau^2)$ , and  $\theta_t$  is a latent true signal. This model is useful for smoothing the observed time series  $y_t$  to obtain latent trend (signal)  $\theta_t$ . It is noted that  $\theta_t$  is assumed to follow random-walk process, in which the variance parameter  $\tau^2$  determines the smoothness of  $\theta_t$ . The unknown static parameters are  $\sigma^2$  and  $\tau^2$ , which will be estimated by assigning prior distributions.

When  $\varepsilon_t$  contains outliers, the resulting signal estimator via the standard Bayesian approach would be biased or inefficient. To overcome the difficulty, we restrict the information from observation  $y_t$  by replacing the likelihood function  $\phi(y_t; \theta_t, \sigma^2)$  with robust alternatives. Specifically, we here employ density power divergence to define synthetic posterior distribution of latent signals. We developed a Gibbs sampling algorithm for posterior computation.

# 分割表における種々の対称モデルに関する尺度

東京理科大学大学院理工学研究科 桃崎 智隆・齋藤 健  
横浜市立大学医学部 三枝 祐輔  
日本大学経済学部 生亀 清貴  
東京理科大学理工学部 中川 智之・石井 晶・富澤 貞男

本講演では、分割表における種々の対称モデルに関する尺度について2部構成で述べる。

## 第1部 正方分割表における部分対称と部分非対称からの隔たりを測る2次元尺度

正方  $r \times r$  分割表を考え、 $(i, j)$  セル確率を  $p_{ij}$  ( $i = 1, \dots, r; j = 1, \dots, r$ ) とする。  $\{p_{ij} + p_{ji} > 0\}$  を仮定する。 Saigusa et al. (2016) は部分対称 (PS) モデルを提案した: 少なくとも1組の  $i \neq j$  に対して、  $p_{ij} = p_{ji}$ 。 いま  $p_{ij}^c = p_{ij}/(p_{ij} + p_{ji})$ ,  $p_{ij}^* = p_{ij}/\delta$ ,  $\delta = \sum \sum_{i \neq j} p_{ij}$  とおく。 Saigusa et al. (2016) は PS モデルからの隔たりを測る尺度を提案した:

$$\Phi^{(\lambda)} = \prod_{i=1}^{r-1} \prod_{j=i+1}^r \left[ \phi_{ij}^{(\lambda)} \right]^{(p_{ij}^* + p_{ji}^*)} \quad (\lambda > -1),$$

ただし,

$$\phi_{ij}^{(\lambda)} = 1 - \frac{\lambda 2^\lambda}{2^\lambda - 1} H_{ij}^{(\lambda)}, \quad H_{ij}^{(\lambda)} = \frac{1}{\lambda} \left[ 1 - (p_{ij}^c)^{\lambda+1} - (p_{ji}^c)^{\lambda+1} \right].$$

本講演では、新しいモデルと2つの尺度を提案する:

(1) 部分完全非対称 (PAS) モデル:

$$p_{ij} = 0 \ (p_{ji} \neq 0) \text{ または } p_{ji} = 0 \ (p_{ij} \neq 0) \text{ (少なくとも1組の } i \neq j)$$

(2) PAS モデルからの隔たりを測る尺度:

$$\tau^{(\lambda)} = \prod_{i=1}^{r-1} \prod_{j=i+1}^r \left[ 1 - \phi_{ij}^{(\lambda)} \right]^{(p_{ij}^* + p_{ji}^*)} \quad (\lambda > -1)$$

(3) 2次元尺度:

$$\mathbf{\Lambda}^{(\lambda)} = \left( \Phi^{(\lambda)}, \tau^{(\lambda)} \right)^T \quad (\lambda > -1)$$

$\mathbf{\Lambda}^{(\lambda)}$  は次の性質を満たす:

- (i)  $\mathbf{\Lambda}^{(\lambda)} = (0, 0)^T \Leftrightarrow$  PS モデルと PAS モデルが同時に成立
- (ii)  $\mathbf{\Lambda}^{(\lambda)} = (0, 1)^T \Leftrightarrow$  対称モデル (すべての  $i \neq j$  に対して  $p_{ij} = p_{ji}$ ) が成立
- (iii)  $\mathbf{\Lambda}^{(\lambda)} = (1, 0)^T \Leftrightarrow$  完全非対称モデル (すべての  $i \neq j$  に対して  $p_{ij} = 0$  ( $p_{ji} \neq 0$ ) または  $p_{ji} = 0$  ( $p_{ij} \neq 0$ )) が成立

推定尺度  $\widehat{\mathbf{\Lambda}}^{(\lambda)}$  は近似的に  $N(\mathbf{\Lambda}^{(\lambda)}, n^{-1}\mathbf{\Sigma})$  に従う。  $\mathbf{\Lambda}^{(\lambda)}$  に関するシミュレーション結果や、信頼楕円を用いた実データ解析については当日示す。

## 第2部 周辺点対称性からの隔たりを測る尺度

本節では、順序付き  $R \times C$  分割表における点対称性に着目する。ここで、 $R$  は行変数のカテゴリ数、 $C$  は列変数のカテゴリ数を表し、各カテゴリは、行と列で順序付けがされているとする。第1部と同様に、 $(i, j)$  セル確率を  $p_{ij}$  ( $i = 1, \dots, R; j = 1, \dots, C$ ) とし、 $i^* = R+1-i$ ,  $j^{**} = C+1-j$  とする。各  $i, j$  に対し、行周辺確率を  $p_{i\cdot} = \sum_{j=1}^C p_{ij}$ , 列周辺確率を  $p_{\cdot j} = \sum_{i=1}^R p_{ij}$  とする。Tomizawa (1985) は、次の周辺点対称モデルを提案した：

$$p_{i\cdot} = p_{i^*\cdot} \quad (i = 1, \dots, R) \quad \text{かつ} \quad p_{\cdot j} = p_{\cdot j^{**}} \quad (j = 1, \dots, C) \quad (1)$$

本講演では、周辺点対称モデル (1) からの隔たりを測る尺度を新たに提案する。 $\{p_{i\cdot} + p_{i^*\cdot} \neq 0\}$  と  $\{p_{\cdot j} + p_{\cdot j^{**}} \neq 0\}$  を仮定する。

$$\delta_1 = \sum_{i=1}^{\lfloor \frac{R}{2} \rfloor} (p_{i\cdot} + p_{i^*\cdot}), \quad \delta_2 = \sum_{j=1}^{\lfloor \frac{C}{2} \rfloor} (p_{\cdot j} + p_{\cdot j^{**}})$$

とする。ここで、 $\lfloor x \rfloor$  は  $x$  を超えない最大の整数を表す。提案尺度は次のように定義される。

$$\gamma_{\text{MPS}} = \frac{\delta_1 \gamma_1 + \delta_2 \gamma_2}{\delta_1 + \delta_2}$$

ここで、

$$\gamma_1 = \frac{4}{\pi} \sum_{i=1}^{\lfloor \frac{R}{2} \rfloor} (q_{i\cdot} + q_{i^*\cdot}) \left( \theta_{1(i)} - \frac{\pi}{4} \right); \quad \theta_{1(i)} = \arccos \left( \frac{p_{i\cdot}}{\sqrt{p_{i\cdot}^2 + p_{i^*\cdot}^2}} \right),$$

$$\gamma_2 = \frac{4}{\pi} \sum_{j=1}^{\lfloor \frac{C}{2} \rfloor} (q_{\cdot j} + q_{\cdot j^{**}}) \left( \theta_{2(j)} - \frac{\pi}{4} \right); \quad \theta_{2(j)} = \arccos \left( \frac{p_{\cdot j}}{\sqrt{p_{\cdot j}^2 + p_{\cdot j^{**}}^2}} \right)$$

であり、

$$q_{i\cdot} = \frac{p_{i\cdot}}{\delta_1}, \quad q_{i^*\cdot} = \frac{p_{i^*\cdot}}{\delta_1} \quad \left( i = 1, \dots, \left\lfloor \frac{R}{2} \right\rfloor \right);$$

$$q_{\cdot j} = \frac{p_{\cdot j}}{\delta_2}, \quad q_{\cdot j^{**}} = \frac{p_{\cdot j^{**}}}{\delta_2} \quad \left( j = 1, \dots, \left\lfloor \frac{C}{2} \right\rfloor \right)$$

とする。提案尺度  $\gamma_{\text{MPS}}$  は、以下の性質をもつ。

- (i)  $-1 \leq \gamma_{\text{MPS}} \leq 1$
- (ii)  $\gamma_{\text{MPS}} = 1 \Leftrightarrow \gamma_1 = \gamma_2 = 1$  かつ  $\gamma_{\text{MPS}} = -1 \Leftrightarrow \gamma_1 = \gamma_2 = -1$
- (iii) 周辺点対称モデル (1) が成立するならば、 $\gamma_{\text{MPS}} = 0$  となる。

当日は、提案尺度の推定量における漸近正規性を述べ、近似信頼区間を構成し、実データへの適用例を紹介する。また、多元分割表に対する周辺点対称性について考え、新たな尺度を提案する。

## 参考文献

Saigusa, Y., Tahata, K., and Tomizawa, S. (2016). Measure of departure from partial symmetry for square contingency tables. *Journal of Mathematics and Statistics*, 12:152–156.

Tomizawa, S. (1985). The decompositions for point symmetry models in two-way contingency tables. *Biometrical Journal*, 27:895–905.

# 地上部と地下部のデータを統合させた推移行列モデル

島谷健一郎(統計数理研究所)、荒木希和子(立命館大生命科学部)

マルコフ推移行列モデルは、生物種の生活史のモデル化として、個体群生態学で積極的に活用されている。特に、行列要素に繁殖が加わることで、生物学固有の行列モデルが展開する。行列の行の和は1ではなく、固有値は生物個体群の増殖率を表すなど、推移行列モデルの数学と生物の間に、様々な対応が付けられる。

さらに、実データから行列要素を推定する統計モデルも数多く提唱されている。データが多ければ単純な割り算しか必要ないが、野外の生物で大量のデータを得ることは困難な場合が多い。限られたデータから行列成分を推定する過程で統計モデルは必須となる。

植物は、大きく種子繁殖とクローン繁殖で拡大する。それぞれ行列モデルの中でモデル化されてきているが、実際の繁殖過程を大幅に簡略したモデルが主流である。とりわけクローン繁殖過程は、単なる年間平均クローン数程度に単純化されている。実際は、クローン繁殖する親株としない親株など、株の間に変異があるはずだが、モデル化以前に、どういう株がクローン繁殖するかといった基本的な野外データすら不十分な状況にある。

さらに、多年生の草本種には、地下茎でクローン繁殖する種も多い。その場合、どの地上部(株)がどの株の親株なのか、地上の観察ではわからない。遺伝子を調べることでクローン識別はできるが、遺伝子情報だけでは親子関係の判別はできない。

地下を掘り起こし株と株のつながりを直接観察することで、親子関係は確かめられる。ただし、1度掘り起こせばその株は枯死する(または大きなダメージを被る)。1度の破壊的調査で複雑なクローン繁殖の仕組みを定量的に知るには、掘り起こされた限られたデータから統計モデルによる推定が不可欠となる。

本研究では、クローナル繁殖動態について、(1) 数理モデルとしての定式化、(2) 実データによる検証法、を提唱することを主題に置いた。発表は、

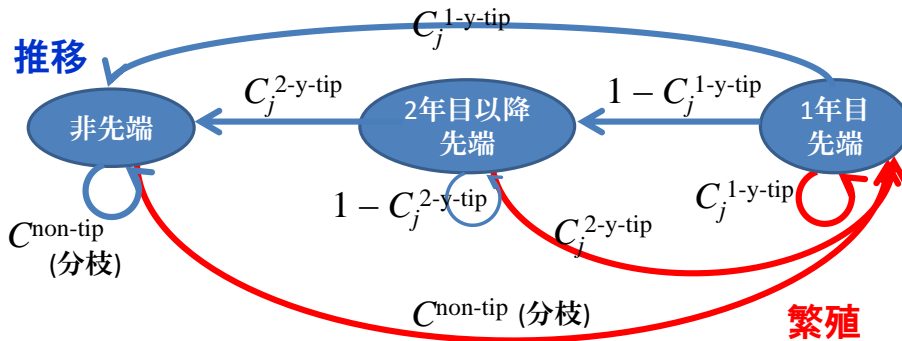
- 1 地上部と地下部のフィールドデータ
- 2 クローナル繁殖動態のマルコフ推移行列モデルによる定式化
- 3 地上部と地下部のデータからの行列成分推定法
- 4 スズランのクローナル繁殖動態: クローン間の違いとその要因の順に行った。

2 では、以下のような構造の行列モデルを提唱した。

$C_j^h$ : age-position  $h$ , size class  $j$  の株のクローナル繁殖率

Stage: age-position (1-y-tip, 2-y-tip, non-tip,  $h$ ) と size class (1 枚葉、2 枚葉 (<10cm,

10-15cm, 15-20cm, 20cm<,  $i, j$ )



さらに、以下の記法を加えて、行列成分は以下のようになる。

$N_i$ : サイズクラス  $i$  の新規クローン割合

$S_j$ : サイズクラス  $j$  の生残率

$T_{ij}$ : サイズクラス  $j$  が生残したときクラス  $j$  に推移する確率

$E$ : 平均娘数 (高々1シーズンに3クローン)

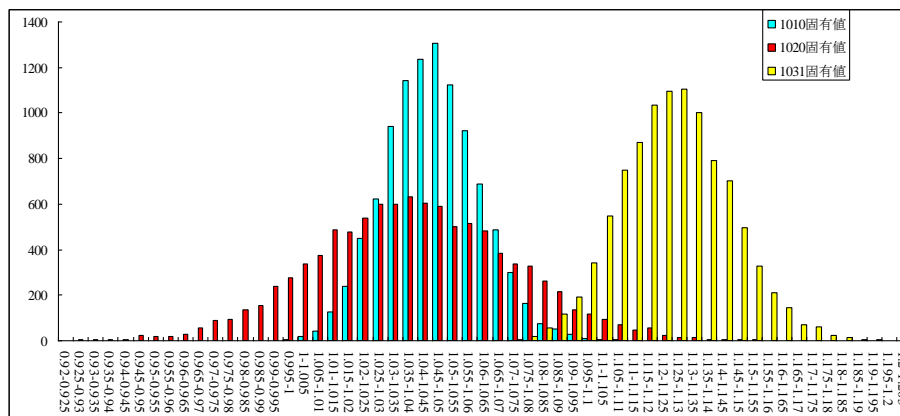
繁殖によるage-position変化 ( $C$ ), 生残率 ( $S$ ), 推移確率 ( $T$ )

Year $t$ (当年) size class $i$	Year $t - 1$ (前年), size class $j$		
	1年目先端 1-y-tip	2年目先端 2-y-tip	非先端 non-tip
1年目先端 1-y-tip	0	0	0
2年目先端 2-y-tip	$T_{ij}S_j(1 - C_j^{1-y-tip})$	$T_{ij}S_j(1 - C_j^{2-y-tip})$	0
非先端 non-tip	$T_{ij}S_jC_j^{1-y-tip}$	$T_{ij}S_jC_j^{2-y-tip}$	$T_{ij}S_j$

繁殖による増殖確率 ( $C$ ), 新規サイズ分布 ( $N_j$ ), 新規娘数 ( $E$ )

Year $t$ (当年) size class $i$	Year $t - 1$ (前年), size class $j$		
	1年目先端 1-y-tip	2年目先端 2-y-tip	非先端 non-tip
1年目先端 1-y-tip	$EN_iC_j^{1-y-tip}$	$EN_iC_j^{2-y-tip}$	$EN_iC_j^{non-tip}$
2年目先端 2-y-tip	0	0	0
非先端 non-tip	0	0	0

実データから各成分をベイズ推定し、得られた行列の固有値の分布を求めたところ、以下のように3つのクローン間で異なる分布を示す結果となった。



また、いくつかのモデルを比較・対照させることにより、クローン間の違いは、生残・推移 (成長) よりクローナル繁殖率に起因すると思われた。