2015 年度科学研究費シンポジウム

大規模複雑データの理論と方法論：最前線の動向

科学研究費補助金 基盤研究 (A) 15H01678「大規模複雑データの理論と方法論の総合的研究（研究代表者：青嶋誠)」, 学術研究助成基金助成金 挑戦的萌芽研究 26540010「ビッグデータの統計学: 理論の開拓と 3V への挑戦（研究代表者：青嶋誠)」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)
日野英逸　(筑波大学)

記

日時：2015 年 11 月 16 日 (月)〜18 日 (水)
場所：筑波大学自然系学系棟Ｄ棟 D509 (筑波キャンパス内)
〒305-8571 茨城県つくば市天王台 1-1-1

プログラム

11 月 16 日 (月)

14:00〜14:05　開会

14:05〜14:35　園田　翔　（早稲田大・先進理工学研究科）
村田　昇　（早稲田大・先進理工学研究科）
ReLU ネットワークの積分表現理論

14:45〜15:25　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 two-sample tests in general settings

15:40〜16:10　江口　翔一　（九州大・数理学府）
従属型一般化線形モデルにおけるモデル評価

16:20〜17:00　福地　純一郎（学習院大・経済学部）
Nonparametric bootstrap for Gupta’s selection procedure

11 月 17 日 (火)

9:20〜9:50　中川　智之　（広島大・理学研究科）
高次元の場合での誤判別確率の推定について

10:00〜10:30　岡　紘之　（慶應義塾大・理工学研究科）
白石　博　（慶應義塾大・理工学部）
高次元の下での有効フロンティアの統計的推定

10:45〜11:25　蛭川　潤一　（新潟大・自然科学系）
坂井　俊介　（新潟大・自然科学研究科）
Rank tests for an ARMA model against other tv-ARMA models

11:35〜12:15　星野　伸明　（金沢大・経済学類）
疎な分割表上の分布

12:15〜13:45　昼食

13:45〜14:25　柿沢　佳秀　（北海道大・経済学研究科）
Some multivariate asymmetric kernel estimators for density functions

14:35〜15:15　柳本　武美　（統計数理研究所）
事前密度のベイズ尤度とその含意

15:30〜16:10　若木　宏文　（広島大・理学研究科）
ランダム係数を持つ GMANOVA モデルの変数選択規準

16:20〜17:00　筑瀬　靖子　（香川大）
Peter E. Jupp (University of St. Andrews)
Distribution theory and its asymptotics in shape analysis

18:00〜　懇親会

11 月 18 日 (水)

9:10〜9:40　茶之原　直人（広島大・理学研究科）
中川　智之　（広島大・理学研究科）
若木　宏文　（広島大・理学研究科）
修正コレスキー分解を用いた分散共分散行列の縮小推定について

9:50〜10:20　渋谷　明　（東京理科大・理工学研究科）
生亀　清貴　（東京理科大・理工学部）
富澤　貞男　（東京理科大・理工学部）
順序カテゴリ正方分割表における対角指数対称モデルについて

10:35〜11:05　稲津　佑　（広島大・理学研究科）
パラレルプロファイルモデルにおけるランダムエフェクト共分散構造に対する
尤度比統計量の高次元漸近展開

11:15〜11:55　鎌谷　研吾　（大阪大・基礎工学研究科, JST CREST）
マルコフ連鎖モンテカルロ法の高次元漸近論

11:55〜12:00　閉会

報告書：ReLU ネットワークの積分表現理論 [*†]

園田　翔 [‡]　村田　昇
早稲田大学　先進理工学研究科
Sonoda Sho　Noboru Murata
Faculty of Science and Engineering
Waseda University

本発表では，まずディープラーニングの現状について簡単に説明したのち，ReLU ネットワークを具体例として積分表現理論について紹介した。

ディープラーニングでは，従来よりも多くの隠れ層を備えたニューラルネットを学習させる。隠れ層を重ねることで，情報表現は階層化され，情報の流れは組合せ的に複雑になる。この原理によって，浅いニューラルネットよりも効率的に学習ができることは，以前から予見されていた。しかし現実には，隠れ層が深くなるにつれて誤差信号が指数的に減衰するので，バックプロパゲーションは実行不可能であった。ディープニューラルネットでは，モデルパラメータの数は数百万から数億のオーダーにのぼることもあるのに対し，データサイズは多くとも数千から数万のオーダーに留まる。それにもかかわらず，画像認識を中心としとする応用分野で好成績を収めていることは，一種の驚きである。

ニューラルネットの積分表現とは，ニューラルネットを連続化した関数

$$g(\mathbf{x}) = \int_{\mathbb{R}^m \times \mathbb{R}} \mathrm{T}(\mathbf{a}, b)\, \eta(\mathbf{a} \cdot \mathbf{x} - b) \mathrm{d}\mathbf{a}\mathrm{d}b, \quad \mathbf{x} \in \mathbb{R}^m \tag{1}$$

である。適当な方法で積分表現を離散化することで，通常のニューラルネットを得る。積分表現 (1) は，$\mathrm{T}(\mathbf{a}, b)$ の η による双対リッジレット変換として知られる。T として適当な関数 f のリッジレット変換

$$\mathscr{R}_\psi f(\mathbf{a}, b) := \int_{\mathbb{R}^m} f(\mathbf{x})\, \overline{\psi(\mathbf{a} \cdot \mathbf{x} - b)} \mathrm{d}\mathbf{x}, \quad (\mathbf{a}, b) \in \mathbb{R}^m \times \mathbb{R} \tag{2}$$

---

[*] 「大規模複雑データの理論と方法論：最前線の動向」

[†] 本研究は JSPS 科研費 15J07517 の助成を受けたものです。

[‡] `s.sonoda0110@toki.waseda.jp`

をとると，f, ψ, η に対する適当な条件のもとで再生公式

$$\mathscr{R}_\eta^\dagger \mathscr{R}_\psi f(\mathbf{x}) = f(\mathbf{x}), \tag{3}$$

が成り立つ。再生公式は，ニューラルネットが万能であることを表している。

積分表現を用いて，活性化関数が ReLU のように非有界な場合にもニューラルネットの万能性が保たれることを説明した。形式的には，積分表現が b に関して畳込み積分になっていることを利用して，畳込みに対する微分の性質 (部分積分)

$$\int_{\mathbb{R}^m \times \mathbb{R}} \mathrm{T}(\mathbf{a}, b) \eta'(\mathbf{a} \cdot \mathbf{x} - b) \mathrm{d}\mathbf{a}\mathrm{d}b = \int_{\mathbb{R}^m \times \mathbb{R}} \partial_b \mathrm{T}(\mathbf{a}, b) \eta(\mathbf{a} \cdot \mathbf{x} - b) \mathrm{d}\mathbf{a}\mathrm{d}b, \tag{4}$$

から従う。

再生公式から，ニューラルネットが学習の結果として獲得する情報表現の正体がリッジレット変換であることを説明した。隠れ層における情報表現 $\mathrm{T} = \mathscr{R}_\psi f$ は，本質的に一意性がない。なぜならば，ひとつの与えられた η に対して許容条件を満たす ψ を複数構成できるためである。つまり隠れ層における情報表現 $\mathrm{T} = \mathscr{R}_\psi f$ には，ψ の不定性がある。

可能な（許容条件を満たす）ψ からさらに「よい」ψ を選択する方法として，積分表現 $\mathscr{R}_\eta^\dagger \mathscr{R}_\psi f$ において $\mathscr{R}_\psi f$ が係数に相当することに基いて，$\|\mathscr{R}_\psi f\|_1$ を最小化することでスパーシティの高い係数を選ぶ方法を紹介した。

$$\min_{\psi \in \mathcal{A}_\eta} \|\mathscr{R}_\psi f\|_1 \,. \tag{5}$$

通常のニューラルネットとして考えると，この定式化は L^1 正則化付きの最小二乗法に相当する。ただし積分表現の式では，常に許容条件を満たしていることが前提なので，二乗誤差は適当な近似誤差 $\varepsilon > 0$ で押さえられたうえでの最適化として考えてよい。

$$\min_{\mathbf{a}_j, b_j, c_j} \|f - \sum_j c_j \cdot \eta(\mathbf{a}_j \cdot \mathbf{x} - b_j)\|_2 + \lambda \|\mathbf{c}\|. \tag{6}$$

積分表現 $\mathscr{R}_\eta^\dagger \mathscr{R}_\psi(\mathbf{a}, b)$ を直接数値計算することで，バックプロパゲーションを経ずに学習する方法を紹介した。本発表では，ランダムに離散化を行うオラクル分布という方法と，数値積分を計算する方法を説明した。

ディープニューラルネットに対する積分表現はまだない。本発表では，例えば隠れ層が 2 層の場合でも，以下のように入れ子構造になるために難しいということを説明した。

$$g(\mathbf{x}) = \int_{\mathbb{R}^{m'} \times \mathbb{R}} \mathrm{T}'(\mathbf{a}', b) \eta(\mathbf{a}' \cdot \mathbf{h}(\mathbf{x}) - b) \mathrm{d}\mathbf{a}'\mathrm{d}b, \tag{7}$$

$$\mathbf{h}(\mathbf{x}) = \int_{\mathbb{R}^m \times \mathbb{R}} \mathbf{T}(\mathbf{a}, b) \eta(\mathbf{a} \cdot \mathbf{x} - b) \mathrm{d}\mathbf{a}\mathrm{d}b. \tag{8}$$

High-dimensional two-sample tests in general settings

矢田　和善　(筑波大数理物質)
青嶋　誠　　(筑波大数理物質)

1. はじめに

ゲノム科学等に見られる高次元データの一つの特徴は，データがもつ次元数の膨大さにある．次元数は標本数を遥かに超える．いま一つの特徴は，高次元データは豊富な情報を有するものの，それが巨大なノイズに埋もれて見つけ難いことである．これらの理由から，通常の多変量解析法では高次元データの推測に精度を保証することができず，間違った解析結果を導くことさえある．

高次元二標本問題について，Bai and Saranadasa (1996, Statist. Sinica)，Chen and Qin (2010, Ann. Statist.) や Aoshima and Yata [1] 等は，母集団分布が正規性よりも緩い仮定のもとで，ユークリッド距離に基づく統計量を与え，その漸近正規性を示し，それに基づく二標本検定法を考えた．一方で，マハラノビス距離に基づく統計量については，Cai et al. (2014, JRSS-B) が高次元の枠組みで扱っている．

本講演では，二つの固有値モデル，strongly spiked eigenvalue (SSE) モデルと non-SSE (NSSE) モデルに基づいて高次元二標本問題を考えた．まず，固有値の発散速度が遅い NSSE モデルのもと，高次元検定統計量の漸近一致性と漸近正規性を与えた．それらに基づき，ある種の最適性について議論し，ホテリングの T^2 検定等のマハラノビス距離に基づく統計量が最適性を持たないことを示した．さらに，検定方式を選ぶための理論的な指標も与えた．次に，固有値の発散速度が速い SSE モデルのもと，新たな高次元検定統計量を提案し，その漸近正規性を示した．それらに基づく検定法が，SSE モデルのもと他の検定法より優れていることを理論的かつ数値的に示した．

2. 高次元検定統計量に関する漸近的性質

母集団が 2 個あると想定し，各母集団 (π_i) は平均に p 次のベクトル $\boldsymbol{\mu}_i$，共分散行列に p 次の正定値対称行列 $\boldsymbol{\Sigma}_i$ $(> \boldsymbol{O})$ をもつと仮定する．いま，各母集団 π_i から $n_i(\geq 4)$ 個の p 次元データ $\boldsymbol{x}_{i1},...,\boldsymbol{x}_{in_i}$ を無作為に抽出する．母集団 $\pi_i,\ i=1,2$ の高次元データに，$\boldsymbol{x}_{ij}=\boldsymbol{\Gamma}_i\boldsymbol{w}_{ij}+\boldsymbol{\mu}_i$ $(j=1,...,n_i)$ なるモデルを考える．ここで，$\boldsymbol{\Gamma}_i$ は $\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_i^T=\boldsymbol{\Sigma}_i$ なる $p\times r_i$ 行列，$\boldsymbol{w}_{ij}$ は $E(\boldsymbol{w}_{ij})=\boldsymbol{0}$, $\mathrm{Var}(\boldsymbol{w}_{ij})=\boldsymbol{I}_{r_i}$ とし，$\boldsymbol{w}_{ij}=(w_{i1j},...,w_{ir_ij})^T$ の各成分は 4 次モーメントが一様有界と仮定する．ただし，$\boldsymbol{I}_{r_i}$ は r_i 次元の単位行列である．母集団 π_i の分布には，必要な箇所でその都度，次を仮定する．

(A-i)　$E(w_{iqj}^2w_{isj}^2)=1,\ E(w_{iqj}w_{isj}w_{itj}w_{iuj})=0,\ q\neq s,t,u;\ i=1,2.$

(A-i) は正規分布を緩めた仮定になっている．高次元のもと，

$$H_0:\boldsymbol{\mu}_1=\boldsymbol{\mu}_2 \quad \text{vs.} \quad H_1:\boldsymbol{\mu}_1\neq\boldsymbol{\mu}_2 \tag{1}$$

なる高次元平均ベクトルの検定を考え，漸近的な最適性を議論する．

いま，非負定値対称行列 $\boldsymbol{A}$ を考える．各 i で，標本平均を $\overline{\boldsymbol{x}}_{in_i}$，標本共分散行列を $\boldsymbol{S}_{in_i}$ とし，(1) の検定統計量を

$$T(\boldsymbol{A})=(\overline{\boldsymbol{x}}_{1n_1}-\overline{\boldsymbol{x}}_{2n_2})^T\boldsymbol{A}(\overline{\boldsymbol{x}}_{1n_1}-\overline{\boldsymbol{x}}_{2n_2})-\sum_{i=1}^{2}\mathrm{tr}(\boldsymbol{S}_{in_i}\boldsymbol{A})/n_i$$

とおく．$T(\boldsymbol{I}_p)$ は，Aoshima and Yata [1] 等で与えた，ユークリッド距離に基づく統計量と一致することに注意する．ここで，$\boldsymbol{\mu}_A = \boldsymbol{A}^{1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$，$\boldsymbol{\Sigma}_{i,A} = \boldsymbol{A}^{1/2}\boldsymbol{\Sigma}_i\boldsymbol{A}^{1/2}$ とおく．そのとき，$E(T(\boldsymbol{A})) = ||\boldsymbol{\mu}_A||^2$ ($= \Delta(\boldsymbol{A})$ とおく)，$\mathrm{Var}(T(\boldsymbol{A})) = K(\boldsymbol{A}) + 4\sum_{i=1}^{2}\boldsymbol{\mu}_A^T\boldsymbol{\Sigma}_{i,A}\boldsymbol{\mu}_A/n_i$ となる．ただし，

$$K(\boldsymbol{A}) = 2\sum_{i=1}^{2}\frac{\mathrm{tr}(\boldsymbol{\Sigma}_{i,A}^2)}{n_i(n_i-1)} + 4\frac{\mathrm{tr}(\boldsymbol{\Sigma}_{1,A}\boldsymbol{\Sigma}_{2,A})}{n_1n_2}$$

である．いま，任意の正定値対称行列 $\boldsymbol{M}$ の最大固有値を，$\lambda_{\max}(\boldsymbol{M})$ で表すとし，$\boldsymbol{\Sigma}_{i,A}$ には次の NSSE モデルを仮定する．

(A-ii) $\quad \dfrac{\lambda_{\max}(\boldsymbol{\Sigma}_{i,A})}{\mathrm{tr}(\boldsymbol{\Sigma}_{i,A}^2)^{1/2}} \to 0,\ p \to \infty;\ i = 1, 2.$

さらに，$m = \min\{p, n_1, n_2\}$ とおき，次のどちらかを仮定する．

(A-iii) $\quad \dfrac{K(\boldsymbol{A})}{\{\Delta(\boldsymbol{A})\}^2} \to 0,\ m \to \infty;$ $\quad$ **(A-iv)** $\quad \limsup_{m\to\infty}\dfrac{\{\Delta(\boldsymbol{A})\}^2}{K(\boldsymbol{A})} < \infty.$

H_0 のもと，(A-iv) を満たすことに注意する．このとき，次の定理を得た．

定理 1．(A-iii) を仮定する．$m \to \infty$ のとき次が成り立つ．

$$\frac{T(\boldsymbol{A})}{\Delta(\boldsymbol{A})} = 1 + o_P(1).$$

定理 2．(A-i) を仮定する．さらに，(A-ii) と (A-iv) のもとで，$m \to \infty$ のとき次が成り立つ．

$$\frac{T(\boldsymbol{A}) - \Delta(\boldsymbol{A})}{\sqrt{K(\boldsymbol{A})}} \Rightarrow N(0, 1),$$

ここで，$\Rightarrow$ は分布収束を表す．

定理 1 と 2 に基づき，各種 $\boldsymbol{A}$ による検定法の高次元における特徴と性質を報告し，NSSE モデルのもと漸近的な最適性を議論した．それにより，$\boldsymbol{A}$ が共分散行列の逆行列であるようなマハラノビス距離に基づく統計量が最適性を持たないことを示し，検定方式を選ぶための理論的な指標を与えた．さらに，固有値の発散速度が速い SSE モデルを与え，新たな高次元検定統計量を提案した．SSE モデルのもと，その漸近正規性を与えることで検定法を構築し，それが他の検定法より優れていることを理論的かつ数値的に示した．

参考文献

[1] Aoshima, M. and Yata, K. (2015a). Asymptotic normality for inference on multisample, high-dimensional mean vectors under mild conditions. *Methodol. Comput. Appl. Probab.* **17**, 419-439.

[2] Aoshima, M. and Yata, K. (2015b). Two-sample tests for high-dimension, strongly spiked eigenvalue models, submitted.

従属型一般化線形モデルにおけるモデル評価

九州大学大学院数理学府　江口 翔一

Bayes 因子に基づいた相対モデル記述評価においては, データの対数周辺尤度の漸近挙動に基づいて最適モデルを決定する．すなわち, 対数周辺尤度の主要項を使いやすい形で近似することができれば記述的モデル評価規準として用いることができる．このとき, 適当な正則条件および Taylor 展開により古典的な Bayes 情報量規準 BIC が導出されるが, 従属データモデルなどでは, BIC 導出の数学的正当性が厳密に議論がされない場合がしばしばある．そこで, 一般の疑似対数尤度関数を対象として, 漸近混合正規性やモデル誤特定の場合まで視野に入れたモデル設定で BIC の拡張を行うことは理論上重要である．本シンポジウムでは特に, 一種の従属型一般化線形モデルに対して, Schwarz 型統計量によるモデル評価の正当性を与えた．本結果により, Lv and Liu (2014, Section 3.1) の結果が大幅に拡張された.

$\mathbf{Y}_n = (Y_1,\ldots,Y_n)'$ を n 次元確率変数, $X_j = (x_{j,1},\ldots,x_{j,p})'$ $(j=1,\ldots,n)$, $\mathbf{X}_n = (X_1,\ldots,X_n)'$ を $n \times p$ random time series とする．ここで, $(Y_1,...,Y_n)$ が $(X_1,...,X_n)$-条件付き独立であり, 各 Y_j の $(X_1,...,X_n)$-条件付き分布は X_j にのみ依存し, $\mathbf{X}_n$ の分布については未知でよいが θ には依らないものとする．このとき, モデル選択の対象は $(\mathbf{X}_n,\mathbf{Y}_n)$ の同時分布であるが, 候補モデルにおける確率 (密度) 関数は

$$f(\mathbf{x}_n,\mathbf{y}_n;\theta) = f(\mathbf{x}_n)f(\mathbf{y}_n|\mathbf{x}_n;\theta) = f(\mathbf{x}_n)\prod_{j=1}^{n} f(y_j|x_j;\theta)$$

と表すことができ, 条件付き分布 $\mathbf{Y}_n|\mathbf{X}_n$ についてのみ考察すればよい．この条件付き分布に一般化線形モデルを仮定すると, 疑似対数尤度関数は

$$\mathbb{H}_n(\theta) = \sum_{j=1}^{n}\left(Y_j X_j'\theta - b(X_j'\theta)\right)$$

を考えればよい．ここで, $b(\cdot)$ は凸関数である (e.g. 線形 Gauss 回帰 : $b(\theta)=|\theta|^2/2$, 線形 Logistic 回帰 : $b(\theta)=\log(1+e^{\theta})$)．この疑似対数尤度関数に対して対数周辺疑似尤度の確率展開を導出するための条件を与え, 特に, $\mathbf{X}_n$ については広範な時系列モデルを扱える一般的な条件を与えた．このときパラメータ $\theta \in \Theta$ $(\subset \mathbb{R}^p)$ の事前分布

を π とし, 適当な仮定の下, 対数周辺疑似尤度の確率展開が導出できた:

$$\log\left(\int_\Theta \exp\{\mathbb{H}_n(\theta)\}\pi(\theta)d\theta\right) = \mathbb{H}_n(\hat{\theta}_n) - \frac{p}{2}\log n + \frac{p}{2}\log 2\pi + \log\pi(\hat{\theta}_n) - \frac{1}{2}\log\det\left(\frac{1}{n}\sum_{j=1}^{n}\partial^2 b(X_j'\hat{\theta}_n)X_jX_j'\right) + o_p(1)$$

ただし, $\hat{\theta}_n$ は疑似最尤推定量, つまり $\hat{\theta}_n \in \mathrm{argmax}_{\theta\in\Theta}\,\mathbb{H}_n(\theta)$ である. この発散項に着目することにより, Bayes モデル $(\pi, \mathbb{H}_n)$ に対する疑似 Bayes 型情報量規準 (Quasi-BIC; QBIC) が

$$\begin{aligned}\mathrm{QBIC} &:= -2\sum_{j=1}^{n}\left(Y_jX_j'\hat{\theta}_n - b(X_j'\hat{\theta}_n)\right) + \log\det\left(\sum_{j=1}^{n}\partial^2 b(X_j'\hat{\theta}_n)X_jX_j'\right)\\ &= -2\sum_{j=1}^{n}\left(Y_jX_j'\hat{\theta}_n - b(X_j'\hat{\theta}_n)\right) + p\log n + O_p(1)\end{aligned}$$

という形で得られた. これにより, Schwarz 型統計量 QBIC (BIC) の使用を正当化できた. 候補モデルごとに QBIC を計算し, それが最小となるモデルを最適なモデルとして選択する.

この QBIC と BIC, 赤池情報量規準 AIC を用いてモデル選択の実験を行い, その結果を比較した. すると, QBIC は BIC と AIC の中間となるような振る舞いをみせ, データの増加とともに最適なモデルを選択する回数は増加する傾向がみられた.

参考文献

[1] Lv, J. and Liu, J. S. (2014), Model selection principles in misspecified models. *J. R. Stat. Soc. Ser B. Stat. Methodol.*, **76**, 141–167.

Nonparametric Bootstrap for Gupta's Selection Procedure

福地純一郎　学習院大学

2015 年 11 月

1　母集団選択問題

$\Pi_1, \dots, \Pi_k$ は k 個の母集団で, 平均がそれぞれ $\mu_1, \dots, \mu_k$, 共通の分散 σ^2 を持ち, 分布関数はそれぞれ $F(x-\mu_1), \dots, F(x-\mu_k)$ であるとする. ただし F は未知であるとする. 母集団 $\Pi_i (i = 1, \dots, k)$ からの標本を $\{X_{ij} : j = 1, \dots, n\}$ で表すとする. $\overline{X}_i$ を標本 $\{X_{ij}, j = 1, 2, \dots, n\}$ の標本平均, s_n^2 をプールされた標本分散とする. 母平均のパラメータ空間を $\Omega = \{(\mu_1, \dots, \mu_k) : \mu_i \in \mathbb{R}\}$ として, 順序付けた母平均を $\mu_{[1]} \leq \mu_{[2]} \leq \cdots \leq \mu_{[k]}$ と表す. また, $\overline{X}_{(i)}$ は母平均が $\mu_{[i]}$ である母集団からの標本の標本平均を表すとする. また順序付けた標本平均を $\overline{X}_{[1]} \leq \overline{X}_{[2]} \leq \cdots \leq \overline{X}_{[k]}$ と表す. Gupta (1965) が正規母集団の場合に提案した母集団選択ルールは以下で定義される.
母集団選択ルール R :

$$\overline{X}_i \geq \overline{X}_{[k]} - \frac{ds_n}{\sqrt{n}} \tag{1}$$

となる母集団 Π_i を選択する.

　記号 CS で正しい選択, つまり最良の母集団が選択される部分集合に含まれる事象を意味するものとする. このとき正しい選択の確率 $P(CS|R)$ の Ω 上の下限が $\mu_1 = \cdots = \mu_k$ のときに与えられ,

$$\inf_{\Omega} P(CS|R) = P\Big(\sqrt{n}(\overline{X}_{(j)} - \mu_{[j]}) \leq \sqrt{n}(\overline{X}_{(k)} - \mu_{[k]}) + ds_n,\ j = 1, \dots, k-1\Big) \tag{2}$$

であることが示される. Gupta の部分集合選択の方法では, 定数 P^*, $(1/k < P^* < 1)$ に対して $\inf_{\Omega} P(CS|R) = P^*$ となる d を用いる. 定数 $d = d(P^*, n, k, F)$ は確率変数

$$W_n = s_n'^{-1} \left(\max_{1 \leq i \leq k-1} \sqrt{n}\ \overline{X}'_i - \sqrt{n}\ \overline{X}'_k \right).$$

の P^* 分位点であることがわかる. ただし $X'_{ij} \overset{iid}{\sim} F(x)$, $\overline{X}'_i = n^{-1} \sum_{j=1}^n X'_{ij}$, $(i = 1, \dots, k)$, $s_n'^2 = ((n-1)k)^{-1} \sum_{i=1}^k \sum_{j=1}^n \left(X'_{ij} - \overline{X}'_i \right)^2$. である. 分布関数 F は未知であるから d を推定する必要がある.

2　ブートストラップ法

ブートストラップ法を用いて d を推定する以下のような方法の性質を調べた. 1. F_n を中心化標本 $\mathcal{X}_n = \{X_{ij} - \overline{X}_i : i = 1, \dots, k,\ j = 1, \dots, n\}$ の経験分布関数とする. 2. $\mathcal{X}_n$ を与えたときの条件付き F_n からの無作為標本 $\{X^*_{ij} : i = 1, \dots, k,\ j = 1, \dots, n\}$ をブートストラップ標本とする. 3. ブートストラップ平均, ブートストラップ分散, W_n のブートストラップ・バージョン W_n^* を求める. 4. 2, 3 の手続きを多数 (B) 回繰り返すことにより, W_n^* の分布の P^* 分位点を求める.

　この方法について, 以下の結果が成り立つ.

Theorem 1 $X_{ij} \overset{iid}{\sim} F\left(x-\mu_i\right)$, $(i=1,\ldots,k,\ j=1,\ldots,n)$ であり, $\mathrm{E}(X_{ij})=\mu_i$, $\mathrm{V}(X_{ij})=\sigma^2$ を仮定する．このとき $\sup_{x\in\mathbb{R}}|P(W_n^*\le x|\mathcal{X}_n)-P(W_n\le x)|\to 0,\ w.p.1.$ が成り立つ.

次に W_n の分母を σ で置き換えた $\tilde{W}_n$ の分布に関して以下が成り立つ.

Theorem 2 確率変数列 $X_1,\ldots,X_n$ は *i.i.d.* 確率変数列で非格子点分布を持つとする．また, $\mathrm{E}(X_1)=\mu$, $\mathrm{V}(X_1)=\sigma^2$, $\mathrm{E}(|X_1|^3)<\infty$ を仮定する．G_n は $\sigma^{-1}\sqrt{n}(\overline{X}-\mu)$ の確率分布関数, $\gamma=\mathrm{E}(X_1-\mu)^3/\sigma^3$, k は正の整数,

$$Q_k(x)=6^{-1}\gamma k\Phi^{k-1}(x)(1-x^2)\phi(x),\quad x\in\mathbb{R}.$$

とする．このとき x について一様に $G_n^k(x)=\Phi^k(x)+n^{-1/2}Q_k(x)+o(n^{-1/2})$ が成り立つ.

Theorem 3 確率変数列 $X_{i1},\ldots,X_{in}$, $(i=1,\ldots,k)$ は *i.i.d.* 確率変数列で非格子点分布を持つとする．また, $\mathrm{E}(X_{11})=\mu$, $\mathrm{V}(X_{11})=\sigma^2$, $\mathrm{E}(|X_{11}|^3)<\infty$ を仮定する．また $\mathrm{E}(X_{11})=0$, $\mathrm{V}(X_{11})=\sigma^2$, $\mathrm{E}(|X_{11}|^3)<\infty$. $\tilde{W}_n$ を

$$\tilde{W}_n=\frac{1}{\sigma}\left(\max_{1\le i\le k-1}\sqrt{n}\ \overline{X}_i-\sqrt{n}\ \overline{X}_k\right).$$

と定義する．このとき $P\left(\tilde{W}_n\le x\right)=\int_{-\infty}^{\infty}\Phi^{k-1}\left(y+x\right)d\Phi(y)+n^{-1/2}J_n(x;\gamma)+o(n^{-1/2})$ が x について一様に成り立つ．ただし

$$J_n(x;\gamma)=6^{-1}\gamma\int_{-\infty}^{\infty}\left\{(k-1)\left(1-(y+x)^2\right)\phi(y+x)+(y^3-3y)\Phi\left(y+x\right)\right\}\Phi^{k-2}(y+x)\phi(y)dy.$$

である.

Theorem 4 確率変数列 $X_{i1},\ldots,X_{in}$, $(i=1,\ldots,k)$ は *i.i.d.* 確率変数列で非格子点分布を持つとする．また, $\mathrm{E}(X_{11})=\mu$, $\mathrm{V}(X_{11})=\sigma^2$, $\mathrm{E}(|X_{11}|^3)<\infty$ を仮定する．また $\mathrm{E}(X_{11})=0$, $\mathrm{V}(X_{11})=\sigma^2$, $\mathrm{E}(|X_{11}|^3)<\infty$. F の特性関数 $\hat{F}$ は *Cramér* 条件

$$\overline{\lim_{\|t\|\to\infty}}\left|\hat{F}(t)\right|<1.$$

満たすと仮定する．このとき

$$P\left(\tilde{W}_n^*\le x\middle|\mathcal{X}_n\right)=\int_{-\infty}^{\infty}\Phi^{k-1}\left(y+x\right)d\Phi(y)+\ n^{-1/2}J_n(x;\hat{\gamma}_n)+o(n^{-1/2}),\tag{3}$$

が確率 1 で x について一様に成り立つ．ただし $\hat{\gamma}_n=(nk)^{-1}\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij}-\overline{X}_i)^3/\hat{\sigma}_n^3.$ である. *(3)* 式から

$$\sup_{x\in\mathbb{R}}\left|P\left(\tilde{W}_n^*\le x\middle|\mathcal{X}_n\right)-P\left(\tilde{W}_n\le x\right)\right|=o(n^{-1/2}),\quad w.p.1.$$

が成り立つことがわかる.

References

[1] Babu, G.J. and Singh, K (1984). On one-term Edgeworth correction by Efron's boostrap Sankya Ser. A, 46, 219-232

[2] Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* 9, 1196-1217.

[3] Gupta, S.S. (1965). On Some Multiple Decision (Selection and Ranking) Rules. *Technometrics*, 7, No. 2, 225-245

[4] Singh, K. (1981). On asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* 9, 1187-1195.

高次元の場合での誤判別確率の推定について

広島大学大学院　理学研究科　中川 智之

本報告は誤判別確率の推定について行った．誤判別確率の推定方法としては Cross Validation(CV, Lachenbruch and Mickey(1968), Stone(1974)) や漸近展開 (Okamoto(1963, 1968), McLachlan(1974), Fujikoshi and Seo(1998), etc.) などがある．

本研究は 2 母集団 Π_1, Π_2 の判別分析を考える． Π_i の個体を誤って Π_j に判別してしまう確率 (誤判別確率) を $P(i|j)$ と表す． p 次元母集団 Π_k からのトレーニングデータ $\boldsymbol{X}_{k1}, ..., \boldsymbol{X}_{kN_k} (k = 1, 2)$ とする．このとき, 線形判別関数を

$$L(\boldsymbol{X}) = L(\boldsymbol{X}; \bar{\boldsymbol{X}}_1, \bar{\boldsymbol{X}}_2, \boldsymbol{S}) = \left(\bar{\boldsymbol{X}}_1 - \bar{\boldsymbol{X}}_2\right)' \boldsymbol{S}^{-1} \left\{\boldsymbol{X} - \left(\bar{\boldsymbol{X}}_1 + \bar{\boldsymbol{X}}_2\right)/2\right\},$$

とする．ここで, $\bar{\boldsymbol{X}}_k$, $\boldsymbol{S}$ は次で定義される不偏推定量である．線形判別関数に関する漸近展開としては, Okamoto(1963, 1968) で大標本漸近理論 ($N_k \to \infty$, p:fix) の場合での正規母集団に対する漸近展開が次の形で与えられている．

$$P(2|1) = P(L(\boldsymbol{X}) \le 0|\boldsymbol{X} \in \Pi_1) = \Phi\left(-\Delta/2\right) + \frac{1}{N_1}P_1 + \frac{1}{N_2}P_2 + \frac{1}{N-2}P_3 + O_2 \qquad (1)$$

ここで, O_m は $N_1^{-1}, N_2^{-1}, p^{-1}$ に関するオーダー, $\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, $N = N_1 + N_2$ である．また, McLachlan(1974) では漸近展開式に推定量を入れたときのバイアスの補正を行っている．Deev(1970) では $N_1 = N_2$ の場合の漸近近似式, Fujikoshi and Seo(1998) では $N_1 \neq N_2$ の場合を含む漸近近似式がそれぞれ高次元大標本漸近理論 ($N_k, p \to \infty, p/N \to c_0 \in (0, 1)$) の場合で与えられている．また, Tonda and Wakaki(2003, TR) では漸近展開が次の形で与えられている．

$$\begin{aligned} P(2|1) &= \Phi\left(\nu\right) + \phi(\nu) f_1\left(\Delta\right) + O_{3/2}, \qquad (2)\\ \nu &= \nu\left(\Delta, \frac{p}{N_1}, \frac{p}{N_2}\right)\\ &= -\frac{1}{2}\left(\frac{N-p}{N-1}\right)^{1/2}\left\{\Delta^2 + \frac{p}{N_1N_2}\left(N_1 - N_2\right)\right\}\left\{\Delta^2 + \frac{pN}{N_1N_2}\right\}^{-1/2} \end{aligned}$$

また, 推定量を入れたときのバイアス補正もしている．一方で, CV での推定に関しては大標本漸近理論の場合で, バイアスが O_2 であることが知られている．

本報告では高次元大標本漸近理論の場合での CV に関する漸近的評価とバイアス補正の提案を行った．この章では線形判別関数での判別の誤判別確率を leave-one-out CV で推定したときのバイアスと MSE の評価を行う．まず,

$$\bar{\boldsymbol{X}}_1^{(j)}, \boldsymbol{S}^{(j)} : \boldsymbol{X}_{1j} \text{ を除いた推定量},$$
$$L_j(\boldsymbol{X}_{1j}) = L(\boldsymbol{X}_{1j}, \bar{\boldsymbol{X}}_1^{(j)}, \bar{\boldsymbol{X}}_2, \boldsymbol{S}^{(j)}),$$

とおくと, leave-one-out CV の推定量は,

$$\hat{P}^{CV}(2|1) = \frac{1}{N_1}\sum_{j=1}^{N_1} 1(L_j(\boldsymbol{X}_{1j}) \le 0)$$

と書ける. ここで, $1(\cdot)$ は定義関数である. バイアスは (2) より, 次が成り立つ.

$$\mathrm{E}\left[\hat{P}^{CV}(2|1)\right] - P(2|1) = O_1,$$

大標本漸近理論の下ではバイアスは O_2 であったが, 高次元大標本漸近理論の場合ではオーダーが 1 つ上がることがわかる. 次に MSE の評価を行う. MSE は分散とバイアスの 2 乗に分解できる. すなわち,

$$\mathrm{E}\left[\left\{\hat{P}^{CV}(2|1) - P(2|1)\right\}^2\right] = \mathrm{Var}\left(\hat{P}^{CV}(2|1)\right) + \mathrm{E}\left[\hat{P}^{CV}(2|1)\right]^2,$$

なので, $\hat{P}^{CV}(2|1)$ は漸近不偏推定量であり, 正規分布の仮定から

$$\mathrm{Var}\left(\hat{P}^{CV}(2|1)\right) \to 0$$

が成り立つことが示せる. したがって, 以下が成り立つ.

$$\mathrm{E}\left[\left\{\hat{P}^{CV}(2|1) - P(2|1)\right\}^2\right] \to 0, \quad (N_i \to \infty, \quad p \to \infty, \quad p/N \to c_0 \in (0,1)).$$

また, 本報告では k-hold CV で推定した推定量 $\hat{P}^{CV_k}$ のバイアスは $O(k^{-1})$ であるので, Yanagihara and Fujisawa(2012) での考え方を応用して, k-hold CV に関してバイアス補正の方法を提案した. まず, $N_1 = km$ となると仮定する. $\hat{P}_{new}(2|1)$ 以下のように定義する.

$$\hat{P}_{new}(2|1) = \left\{\hat{P}^{CV_{k_1}}(2|1) - \frac{k_1 - k_2}{k_1 - 1}\left(\hat{P}^{CV_{k_2}}(2|1) - \hat{P}^{CV_{k_1}}(2|1)\right)\right\}$$

そのとき, 次のオーダーまでバイアス補正ができる.

$$\mathrm{E}\left[\hat{P}_{new}(2|1) - P(2|1)\right] = O(k_1^{-2})$$

また, 最後に数値比較を行った.

参考文献

[1] Fujikoshi, Y. , Seo, T. (1998), Asymptotic approximations of EPMC's of the linear and the quadratic discriminant functions when the sample sizes and the dimension are large, *Random Oper. Stochastic Equations*, **6**, 269-280.

[2] Lachenbruch, P. A. and Mickey, M. R. (1968), Estimation of Error Rates in Discriminant analysis. *Technometrics*, **10**, 1-11.

[3] Stone, M. (1974), Cross-Validatory choice and assessment of statistical predictions. *J. R. Statist. Soc.* , **B36**, 111-147.

[4] Tonda, T. and Wakaki, H. (2003), EPMC Estimation in Discriminant Analysis When the Dimension and Sample Sizes are Large. *Hiroshima Statistical Research Group Technical Report*, TR, 03-08.

[5] Yanagihara, H. and Fujisawa, H. (2012), Iterative bias correction of the cross-validation criterion. *Scand. J. Stat.* , Vol. 39, 116-130.

高次元の下での有効フロンティアの統計的推定

慶應義塾大学理工学研究科　岡 紘之
慶應義塾大学理工学部　白石 博

1　研究背景

ポートフォリオ理論は，リスク回避的な投資家が分散投資を行い，自身のポートフォリオにおいてポートフォリオ収益率を高めるためには，どのように最適化すればよいかを決定するための理論である．実際には，資産のリターンはランダムな変数であり，どのような平均と分散をもつ分布に従っているかは未知であるため，サンプルから推定することを考える．近年の株式市場などでは，市場の規模が大きくなっており，分散投資する資産の数が膨大となっている．一般に，d が大きくなるにつれて大標本漸近理論での近似が悪くなってしまうことが知られており，このような状況を鑑みて，高次元データにおける解析を考える必要性が主張されている．

本研究では，ポートフォリオによって実現化される点の領域（実現可能領域）の左側境界を表す有効フロンティア（efficient frontier）を次元 d，標本サイズ n のデータ行列 $\mathbb{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)$ から推定し，

$$d \to \infty, \quad n \to \infty, \quad \frac{d}{n} \to \rho \in (0, 1) \tag{1.1}$$

という設定のもとでの漸近的挙動について結果を報告した．

2　推定問題

$\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in \mathbb{R}^d$ を平均ベクトル $\boldsymbol{\mu}$，分散共分散行列 $\boldsymbol{\Sigma}$ をもつある分布からの無作為標本とする．ただし，$\boldsymbol{\Sigma}$ は正定値行列であるとし，$\boldsymbol{X}_j$ は 4 次までのモーメントを持つと仮定する．また，標本平均 $\bar{\boldsymbol{X}}_n$ と標本分散共分散行列を $\boldsymbol{S}$ を以下のように表す．

$$\bar{\boldsymbol{X}}_n = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{X}_j, \ \boldsymbol{S} = \frac{1}{n-1}\sum_{j=1}^{n}(\boldsymbol{X}_j - \bar{\boldsymbol{X}}_n)(\boldsymbol{X}_j - \bar{\boldsymbol{X}}_n)^\top \tag{2.1}$$

パラメータ $\boldsymbol{\theta} \in \Theta$ を

$$\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\ \mathbf{1}_d^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\ \mathbf{1}_d^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}_d \end{pmatrix} \tag{2.2}$$

と定義する．ここで，以下の仮定を考える．

仮定 2.1.
(A1) $n, d \to \infty$ かつ $d/n \to \rho \in (0, 1)$ とする．以後，これを (n, d)-asymptotic と呼ぶ．

(A2) 各 n, d を固定した下での $\boldsymbol{\theta}$ を $\boldsymbol{\theta}^{(n,d)} \in \Theta$ とする．このとき，ある $\boldsymbol{\theta}^{(\infty)} = (\alpha_1, \alpha_2, \alpha_3) \in \Theta \subset \mathbb{R}^3$ が存在して，(n, d)-asymptotic 下で以下が成り立つ．

$$\boldsymbol{\theta}^{(n,d)} \to \boldsymbol{\theta}^{(\infty)} \tag{2.3}$$

このような条件の下で，次の統計量 $\tilde{\boldsymbol{\theta}}$ を定義する．

$$\tilde{\boldsymbol{\theta}} = \begin{pmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \\ \tilde{\theta}_3 \end{pmatrix} = \begin{pmatrix} \bar{\boldsymbol{X}}^\top \boldsymbol{S}^{-1} \bar{\boldsymbol{X}}_n \\ \mathbf{1}_d{}^\top \boldsymbol{S}^{-1} \bar{\boldsymbol{X}}_n \\ \mathbf{1}_d{}^\top \boldsymbol{S}^{-1} \mathbf{1}_d \end{pmatrix} \tag{2.4}$$

このとき，次の結果が得られる．

定理 2.2. 仮定 2.1 下で次が成立する．

$$\tilde{\theta}_1 \overset{a.s.}{\to} \frac{1}{1-\rho}\alpha_1 + \frac{\rho}{1-\rho}, \quad \tilde{\theta}_2 \overset{a.s.}{\to} \frac{1}{1-\rho}\alpha_2, \quad \tilde{\theta}_3 \overset{a.s.}{\to} \frac{1}{1-\rho}\alpha_3 \tag{2.5}$$

系 2.3. $\hat{\boldsymbol{\theta}} = \frac{n-d}{n}(\frac{n-d}{n}\tilde{\theta}_1 + \frac{d}{n}, \tilde{\theta}_2, \tilde{\theta}_3)^\top$ とする．このとき，仮定 2.1 の下で，次が成立する．

$$\hat{\boldsymbol{\theta}} \overset{a.s.}{\to} \boldsymbol{\theta}^{(\infty)} \tag{2.6}$$

この $\boldsymbol{\theta}$ の一致推定量 $\hat{\boldsymbol{\theta}}$ を用いて，有効フロンティアの一致推定量が得られる．

定理 2.4. 仮定 2.1 の下で

$$\sup_{\gamma \in A} ||\ell_{\text{eff}}(\gamma; \hat{\boldsymbol{\theta}}) - \ell_{\text{eff}}(\gamma; \boldsymbol{\theta}^{(\infty)})|| \overset{a.s.}{\to} 0 \tag{2.7}$$

が成り立つ．

定理 2.5. 推定量 $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)^\top$ に対して仮定 2.1 の下で次が成立する．

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \overset{\mathbb{D}}{\to} N_3(\mathbf{0}, \boldsymbol{\Omega}) \tag{2.8}$$

ここで，

$$\boldsymbol{\Omega} = \frac{1}{1-\rho} \begin{pmatrix} 2\alpha_1^2 + 4\alpha_1 + 2\rho & * & * \\ 2\alpha_1\alpha_2 & \alpha_2^2 + \alpha_1\alpha_3 + \alpha_3 & * \\ 2\alpha_1\alpha_3 & 2\alpha_2\alpha_3 & 2\alpha_3^2 \end{pmatrix} \tag{2.9}$$

である．

定理 2.6. $\ell_{\text{eff}}(\gamma; \boldsymbol{\theta}) = (\sigma_{\text{opt}}(\gamma; \boldsymbol{\theta}), \mu_{\text{opt}}(\gamma; \boldsymbol{\theta}))$ とし，ℓ_{eff} は $\boldsymbol{\theta} \in \mathbb{R}^3$ まわりで微分可能であるとする．このとき，

$$\sqrt{n}\left(\ell_{\text{eff}}(\cdot\,; \hat{\boldsymbol{\theta}}) - \ell_{\text{eff}}(\cdot\,; \boldsymbol{\theta})\right) \overset{\mathbb{D}}{\to} N_2(\mathbf{0}, \dot{\ell}_{\text{eff}}(\cdot\,; \boldsymbol{\alpha})\boldsymbol{\Omega}\dot{\ell}_{\text{eff}}(\cdot\,; \boldsymbol{\alpha})^\top) \tag{2.10}$$

が成立する．ここで，

$$\dot{\ell}_{\text{eff}}(\cdot\,; \theta) = \left(\frac{\partial \ell_{\text{eff}}}{\partial \boldsymbol{\theta}}\right)^\top = \begin{pmatrix} \gamma & \frac{-2\gamma\theta_2 + 1}{\theta_3} & \frac{\gamma\theta_2^2 - \theta_2}{\theta_3^2} \\ \gamma^2 & \frac{-2\gamma^2\theta_2}{\theta_3} & \frac{\gamma^2\theta_2^2 - 1}{\theta_3^2} \end{pmatrix} \tag{2.11}$$

である．

Rank tests for an ARMA model against other tv-ARMA models

Shunsuke Sakai and *Junichi Hirukawa

Abstract For a class of locally stationary process introduced by Dahlhaus, we apply the idea of the problem of testing ARMA model against other non-stationary ARMA model. When testing the problem, we use linear serial rank statistics and contiguity of LeCam's notion. And then, if null hypothesis is white noise, under null and alternative, the asymptotic normality of the proposed statistics is established by using the locally asymptotic normal (LAN) property. We incorporate the locally stationary phenomena in the testing problem.

Nonparametric methods have been developed for analysis of univariate and multivariate observations without the distributional assumption (e.g., Gaussian assumption). The nonparametric procedures are even desired in the area of time series analysis. Hallin et al. (1985) considered the systematic time series oriented study of testing for randomness against ARMA alternatives. They proposed the statistic of the form

$$S_T = \frac{1}{T-p} \sum_{t=p+1}^{T} c_T(R_t^{(T)}, R_{t-1}^{(T)}, \ldots, R_{t-p}^{(T)}),$$

where $R_t^{(T)}$ denotes the rank of the observation $X_t^{(T)}$ in an observed series $\boldsymbol{X}^{(T)} = (X_1^{(T)}, \ldots, X_T^{(T)})$ of length T, and $c_T(\cdots)$ is some given *score function*. These statistics are so-called *linear serial rank statistics* (of order p).

1 Locally asymptotic normality of time varying ARMA models

We consider the sequence of locally stationary ARMA(p_1, p_2) (tv-ARMA) models

$$X_{t,T}^{(T)} + \sum_{k=1}^{p_1} a_{\theta^{(p_2+k)}}(\frac{t}{T}) X_{t-k,T}^{(T)} = b_{\theta^{(0)}}(\frac{t}{T})\varepsilon_t + \sum_{k=1}^{p_2} b_{\theta^{(k)}}(\frac{t}{T})\varepsilon_{t-k},\ t \in \mathbb{Z},$$

where $\{\varepsilon_t; t \in \mathbb{Z}\}$ is a discrete-time stationary white noise, i.e., a family of i.i.d. random variables with mean $E[\varepsilon_t] = 0$, variance $E[\varepsilon_t^2] = 1,\ t \in \mathbb{Z}$ and density function f. Let $b_{\theta^{(k)}}(u), a_{\theta^{(p_2+k)}}(u)$ be

$$\begin{aligned} b_{\theta^{(k)}}(u) &= \theta^{(k)} + (\theta^{(k)} - \theta_k)(\tilde{b}_{\theta_k}(u) - 1),\ k = 0, \ldots, p_2, \\ a_{\theta^{(p_2+k)}}(u) &= \theta^{(p_2+k)} + (\theta^{(p_2+k)} - \theta_{p_2+k})(\tilde{a}_{\theta_{p_2+k}}(u) - 1),\ k = 1, \ldots, p_1, \end{aligned}$$

where the functions $\tilde{b}_{\theta_k}(u)$, $\tilde{a}_{\theta_{p_2+k}}(u)$ are sufficiently smooth bounded functions.

Consider the null hypothesis $H(f;\theta)$ and the alternative hypothesis $K(f;\theta_T)$ as follows;

$$H(f;\theta) : \theta^{(k)} = \theta_k\ (k = 0, \ldots, p_1 + p_2), \quad K(f;\theta_T) : \theta^{(k)} = \theta_T^{(k)} = \theta_k + \frac{h_k}{\sqrt{T}}(k = 0, \ldots, p_1 + p_2).$$

where $\theta = (\theta_0, \ldots, \theta_{p_1+p_2})' \in \Theta \subset \mathbb{R}^{p_1+p_2+1},\ h = (h_0, \ldots, h_{p_1+p_2})' \in \mathscr{H} \subset \mathbb{R}^{p_1+p_2+1}$. Under the two hypotheses, the coefficient functions $b_{\theta^{(k)}}(u)$ and $a_{\theta^{(p_2+k)}}(u)$ are given by

$$b_{\theta^{(k)}}(u) = \begin{cases} \theta_k & (H(f;\theta)) \\ \theta_k + \frac{h_k}{\sqrt{T}}\tilde{b}_{\theta_k}(u) & (K(f;\theta_T)), \end{cases} \qquad a_{\theta^{(p_2+k)}}(u) = \begin{cases} \theta_k & (H(f;\theta)) \\ \theta_k + \frac{h_k}{\sqrt{T}}\tilde{a}_{\theta_{p_2+k}}(u) & (K(f;\theta_T)). \end{cases}$$

Under null hypotheses $H(f;\theta)$, for the Fisher information matrix $\Gamma(\theta)$ of Lemma 5, $-\frac{1}{2}h'\Gamma(\theta)h$ can be written as follows;

$$-\frac{1}{2}h'\Gamma(\theta)h = -\frac{I(f)}{2}\int_0^1 \sum_{k=0}^{p} \big(A_k(u) + B_k(u) + C_k(u) + D_k(u)\big)du.$$

For the definitions of $A_k(u)$, $B_k(u)$, $C_k(u)$ and $D_k(u)$, refer to the master thesis.

In the special case of the white noise null hypothesis $H_0^{(T)} : X_{t,T} = \theta_0\varepsilon_t$, we can obtain

$$-\frac{1}{2}h'\Gamma(\theta)h = -\frac{I(f)}{2}\int_0^1 \left(\sum_{k=0}^{p} d_k^2(u)\right) du,$$

where

$$d_k(u) = \begin{cases} \sqrt{\left|\frac{1}{I(f)}\left(\frac{\tilde{b}_{\theta_0}(u)h_0}{\theta_0}\right)^2 (E[\varepsilon_t^2\phi^2(\varepsilon_t)] - 1)\right|} & k = 0 \\ \tilde{a}_{\theta_{p_2+k}}(u)h_{p_2+k} - \frac{1}{\theta_0}\tilde{b}_{\theta_k}(u)h_k & 1 \le k \le \min(p_1, p_2) \\ \tilde{a}_{\theta_{p_2+k}}(u)h_{p_2+k} & p_2 < k \le p_1 \text{ if } p_2 < p_1 \\ -\frac{1}{\theta_0}\tilde{b}_{\theta_k}(u)h_k & p_1 < k \le p_2 \text{ if } p_1 < p_2. \end{cases}$$

Under null hypothesis $H(f;\theta)$, we derive the central sequence $\Delta_T(\theta)$ of Lemma 5 for locally stationary ARMA model

$$h'\Delta_T(\theta) = \frac{1}{\sqrt{T}\theta_0}\sum_{t=p+1}^{T}\Big\{\phi(\varepsilon_t)\sum_{k=1}^{t-1}D'_k\big(\frac{t}{T}\big)X_{t-k,T} - \tilde{b}_{\theta_0}\big(\frac{t}{T}\big)(1+\phi(\varepsilon_t)\varepsilon_t)h_0\Big\} + o_p(1).$$

For the definition of $D'_k\big(\frac{t}{T}\big)$, refer to the master thesis.

In the special case of the white noise null hypothesis $H_0^{(T)} : X_{t,T} = \theta_0\varepsilon_t$, we have

$$h^{'}\Delta_T(\theta) = \frac{1}{\sqrt{T}\theta_0}\sum_{t=p+1}^{T}\Big\{\phi(\varepsilon_t)\sum_{k=1}^{p}d_k\big(\frac{t}{T}\big)X_{t-k,T} - \tilde{b}_{\theta_0}\big(\frac{t}{T}\big)(1+\phi(\varepsilon_t)\varepsilon_t)h_0\Big\} + o_p(1),$$

where

$$d_k\big(\frac{t}{T}\big) = \begin{cases} \tilde{a}_{\theta_{p_2+k}}\big(\frac{t}{T}\big)h_{p_2+k} - \frac{1}{\theta_0}\tilde{b}_{\theta_k}\big(\frac{t}{T}\big)h_k & 1 \le k \le \min(p_1, p_2) \\ \tilde{a}_{\theta_{p_2+k}}\big(\frac{t}{T}\big)h_{p_2+k} & p_2 < k \le p_1 \text{ if } p_2 < p_1 \\ -\frac{1}{\theta_0}\tilde{b}_{\theta_k}\big(\frac{t}{T}\big)h_k & p_1 < k \le p_2 \text{ if } p_1 < p_2. \end{cases}$$

2 Main results

Now, we attempt to derive the asymptotic distribution of the linear serial rank statistics $S^{(T)}$ under $K(f;\theta_T)$. If we can derive the joint asymptotic normality of $\begin{pmatrix} T^{1/2}(S^{(T)} - m^{(T)}) \\ \Lambda_T(\theta,\theta_T) \end{pmatrix}$ under $H_0^{(T)}$, then it will lead to the asymptotic normality of linear serial rank statistics $S^{(T)}$ under $K(f;\theta_T)$, from LeCam's third lemma (see, e.g. Hajek et al. (1999)). In what follows, we assume that the null hypothesis is white noise hypothesis $H_0^{(T)} : X_{t,T} = \theta_0\varepsilon_t$.

From Hallin and Puri (1985), we can establish the asymptotic equivalence of $(T-p)^{1/2}(S^{(T)} - m^{(T)})$ with $\mathscr{S}^{(T)} - \mathscr{M}^{(T)}$. (For the definitions $\mathscr{S}^{(T)}$ and $\mathscr{M}^{(T)}$, refer to the master thesis). Furthermore, we can show that the U-statistics $\mathscr{U}_{\mathscr{S}}^{(T)} - \mathscr{U}_{\mathscr{M}}^{(T)}$ which are asymptotically equivalent to $T^{-1/2}(\mathscr{S}^{(T)} - \mathscr{M}^{(T)})$.

We define the U-statistics $\mathscr{U}_{\mathscr{L}}^{(T)}$ associated with the central sequence $\mathscr{L}^{(T)} = h^{'}\Delta_T(\theta)$ as follows. Then, we show that this U-statistics is asymptotically equivalent to the central sequence $T^{-1/2}\mathscr{L}^{(T)}$.

Consider the function $G_{t,T}(\boldsymbol{y})$ (of $(p+1)$ arguments $\boldsymbol{y} = (y_1,\dots,y_{p+1}) \in [0,1]^{p+1}$) where

$$G_{t,T}(\boldsymbol{y}) = \phi(F^{-1}(y_1))\sum_{i=1}^{p}d_i\big(\frac{t}{T}\big)F^{-1}(y_{i+1}) - \frac{\tilde{b}_{\theta_0}(\frac{t}{T})}{\theta_0}\{1+\phi(F^{-1}(y_1))F^{-1}(y_1)\}h_0.$$

Additionally, put

$$\Phi^{\mathscr{L}}_{(t_1,\dots,t_{p+1}),T}(\boldsymbol{Y}_{t_1}^{(T)},\dots,\boldsymbol{Y}_{t_{p+1}}^{(T)}) = \frac{1}{p+1}\sum_{j=1}^{p+1}G_{t_j,T}(\boldsymbol{Y}_{t_j}^{(T)}),$$

where $\boldsymbol{Y}_t^{(T)} = (Y_{t,1}^{(T)},\dots,Y_{t,p+1}^{(T)})' = (F(X_{t,T}/\theta_0),\dots,F(X_{t-p,T}/\theta_0))'$, $\quad p+1 \le t \le T$. Then, the corresponding U-statistic is

$$\mathscr{U}_{\mathscr{L}}^{(T)} = \begin{pmatrix} T-p \\ p+1 \end{pmatrix}^{-1}\sum_{p+1\le t_1<\cdots<t_{p+1}\le T}\Phi^{\mathscr{L}}_{(t_1,\dots,t_{p+1}),T}(\boldsymbol{Y}_{t_1}^{(T)},\dots,\boldsymbol{Y}_{t_{p+1}}^{(T)}) = T^{-1/2}\mathscr{L}^{(T)} + o_p(T^{-1/2}).$$

From the above, we can use the theory of asymptotic normality about U-statistics.

Our purpose is showing that the following joint asymptotic normality;

Under $H_0^{(T)}$, $\begin{pmatrix} T^{1/2}(S^{(T)} - m^{(T)}) \\ \Lambda_T(\theta,\theta_T) \end{pmatrix}$ is joint asymptotically normal, with mean $\begin{pmatrix} 0 \\ -\frac{1}{2}h^{'}\Gamma(\theta)h \end{pmatrix}$ covariance $\begin{pmatrix} V^2 & Q \\ Q & h^{'}\Gamma(\theta)h \end{pmatrix}$, where V^2 is given by Hallin et al. (1985), Proposition 4.2.

If we can derive this joint asymptotic normality, from LeCam's third lemma, we can derive the following desire results;

Under $K(f;\theta_T)$, $T^{1/2}(S^{(T)} - m^{(T)})$ is asymptotically normal, with mean Q and variance V^2.

Therefore, we need to prove this joint asymptotic normality and so that calculate the covariance Q.

References

[1] Hallin, M., Ingenbleek, J-Fr., and Puri, M. L. (1985). Linear Serial Rank Tests for Randomness Against Arma Alternatives. *Ann Statist.* **13** 1156-1181.

[2] Hirukawa, J. and Taniguchi, M. (2006). LAN theorem for non-Gaussian locally stationary processes and its applications. *Journal of Statistical Planning and Inference* **136** 640-688.

疎な分割表上の分布

星野伸明

2015 年 11 月 17 日

1 高次元小標本（疎）な分割表

公的統計や社会調査では、個体属性の調査結果が離散変数とみなせる。このような状況は、変数が p 個として、標本の n 個体が p 次元の分割表上で分布していると考えられる。公的統計の場合、典型的な p のオーダーは 10^2 程度で、n のオーダーは十万 (10^6) なら大きい方である。この場合は $n \gg p$ なので、一見、高次元小標本とは言えない。しかし変数がほぼカテゴリカルなので、実質的なデータの次元は p ではない。一つのカテゴリカル変数が c 分類されていれば、実質的に $c-1$ 個の変数が存在すると考えられる。故にカテゴリカルな分割表の実質的な次元 d のオーダーは、c^p である。典型的な c のオーダーは 10 なので、$d \asymp 10^{100} \gg 10^6 \asymp n$ となり高次元小標本と言えよう。

通常の漸近論は、中心極限定理を使いたいので $n \to \infty$ とする。しかし統計的開示制限では n が所与での度数分布を用いるため、そのような極限操作は不都合である。一方、個体属性を詳細に表す場合に興味があり、それは c を増やすことに他ならない。その場合実質的な次元 d は増加する。従って本報告では n を固定して $d \to \infty$ という漸近論を考える。中心極限定理の代わりに小数法則（ポアソン分布:Po）を用いれば、極限分布の取り扱いは解析的に容易となる。

2 疎な分割表の漸近論

分割表の実質的な次元 d のオーダーはセル総数 J と等しいので、以下では $J \to \infty$ とする。また第 j セルの度数を $F_{j,J}$、寸法指標を $S_{i,J} := \sum_{j=1}^{J} I(F_{j,J} = i), i \in \mathbb{N},$ で表す。そのベクトルは $\boldsymbol{S}_J := (S_{1,J}, S_{2,J}, \dots)$ である。総度数は N_J で表す。

本報告の漸近論は次の命題 1 を用いる。なお複合ポアソン分布 $\mathtt{CP}(\mu, \boldsymbol{q})$ を確率母関数 $G(z) = \exp(\lambda(g(z)-1)), \lambda > 0,$ で定義する。ただし $g(z) = \sum_{i=1}^{\infty} q_i z^i$ は正の整数上の分布の確率母関数である（q_i を i の確率と見る）。

命題 1 確率変数 $F_{j,J}, j \in [J],$ が独立に非負整数上の分布に従うとする。この時 $J \to \infty$ の極限で

$$(\boldsymbol{S}_J, N_J) \xrightarrow{d} (\mathop{\perp\!\!\!\perp}_{i=1}^{\infty} \mathtt{Po}(\mu q_i), \mathtt{CP}(\mu, \boldsymbol{q})) \tag{1}$$

が成立することは以下の条件 (2) と (3) が成立することと同値である。

$$\lim_{J\to\infty} \max_j \mathbb{P}(F_{j,J} = i) = 0, \quad i \in \mathbb{N}, \tag{2}$$

$$\lim_{J\to\infty} \mathrm{E}(S_{i,J}) = \mu q_i, \quad i \in \mathbb{N}. \tag{3}$$

命題 1 は Hoshino (2012) の結果を書き直した。寸法指標 $S_{i,J}$ の周辺分布はポアソンの二項分布をする。条件 (2) と (3) は、ポアソンの二項分布で各成功確率は小さいが成功回数の期待値が（正値に）収束することを意味している。つまり極限 (1) で寸法指標がポアソン分布に従う理屈は、小数法則に他ならない。

命題 1 において N_J の極限分布が $\mathtt{CP}(\mu, \boldsymbol{q})$ となることは、以下のように説明出来る。確率変数ベクトル

$$\boldsymbol{S} := (S_1, S_2, \ldots) \sim \mathop{\perp\!\!\!\perp}_{i=1}^{\infty} \mathtt{Po}(\mu q_i)$$

としよう。この時 S_i の確率母関数は $\exp(\mu q_i(z-1))$ であり、iS_i の確率母関数は $\exp(\mu q_i(z^i-1))$ となる。各 S_i は独立なので $N := \sum_{i=1}^{\infty} iS_i$ の確率母関数は $\prod_i \exp(\mu q_i(z^i-1)) = \exp(\mu(\sum_i z^i q_i - 1))$ となる。これは $\mathtt{CP}(\mu, \boldsymbol{q})$ の確率母関数であり、$N_J = \sum_i iS_{i,J}$ の極限分布を与えている。この独立なポアソン分布に従う確率変数の整数倍の和としての複合ポアソン分布の表現は "composed Poisson" と呼ばれる

ここで $\boldsymbol{S}$ の N 所与での条件付き分布を考えよう。全ベル多項式を用いれば $x_i = i!\, q_i$ の時

$$\mathbb{P}(N = n) = \frac{\exp(-\mu)}{n!} B_n(\mu x_1, \ldots, \mu x_n), \quad n \in \mathbb{N}_0,$$

と書ける。$N \sim \mathtt{CP}(\mu, \boldsymbol{q})$ だったので、複合ポアソン分布の確率関数はベル多項式を用いて表現可能ということになる。結局

$$\mathbb{P}(\boldsymbol{S} = \boldsymbol{s} | N = n) = \frac{n! \mu^k \prod_{i=1}^{n} {q_i}^{s_i} \frac{1}{s_i!}}{B_n(\mu x_1, \ldots, \mu x_n)}, \tag{4}$$

ただし $x_i = i!\, q_i, k = \sum_{i=1}^{n} s_i$ ということになる。

分布 (4) は CCP 分布の $J \to \infty$ における極限分布である。従って分布 (4) を母数 $(\mu, \boldsymbol{q})$ の Limiting Conditional Compound Poisson (LCCP) 分布と呼ぶ。値域を見れば、LCCP 分布は自然数の順序の着かない確率分割族であることが分かる。

LCCP 分布族は Ewens 分布など解析的に取り扱いが容易な例を含み、族としての性質も議論できる。これまでに報告者は周辺分布やモメントを明らかにしている。

参考文献

[1] Hoshino, N. (2012). Random partitioning over a sparse contingency table. *Annals of the Institute of Statistical Mathematics*, **64**, 457–474.

Some multivariate asymmetric kernel estimators for density functions

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 $\mathbf{R}^d$ でない (典型的に $[0,1]^d$ や $[0,\infty)^d$ の台) のときに, 境界バイアス問題がある. カーネル密度推定の漸近論は, その背後にある積分近似 $\int_{\mathrm{R}^d} \frac{1}{h^d} K\left(\frac{\mathbf{x}-\mathbf{v}}{h}\right) f(\mathbf{v})\, d\mathbf{v} = \int_{\mathrm{R}^d} K(\mathbf{u}) f(\mathbf{x}-h\mathbf{u})\, d\mathbf{u} \approx f(\mathbf{x})$ $(h \to 0)$ が本質的 (ただし, $K(-\mathbf{v}) = K(\mathbf{v})$, $\int_{\mathrm{R}^d} K(\mathbf{v})\, d\mathbf{v} = 1$ を仮定) で, 有界閉領域や正象限の境界付近ではこの近似が成り立たない. このような境界バイアス問題への対処として, 境界カーネル法が主に１次元データに対し議論されてきた (Jones (1993)). 一方, 推定密度 f が区間 $[0,1]$/半無限区間 $[0,\infty)$ の台を持つ場合に, Chen (1999,2000) がベータ/ガンマカーネル密度推定量を提案して以来, 最近１０年間において『非対称カーネル法 (AK 推定法)』が脚光を浴びてきた. 以下, 非負データを扱う.

2. １次元の AK 推定　１次元の非負データ $X_1,\ldots,X_n \sim \mathrm{iid}\ f$ に基づいた (点 $x \in [0,\infty) = \mathrm{R}_+$ 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[1]. ここで, $k_{x,b}(\cdot)$ は f と同じ台 R_+ を持つカーネルとし, $b = b(n) > 0$ は $n \to \infty$ のとき適当な速さで 0 に収束する平滑化パラメータである. このような先行研究として, ガンマカーネル (Chen (2000)), Birnbaum–Saunders (BS)/対数正規カーネル (Jin and Kawczak (2003)), 逆ガウス/相反逆ガウスカーネル (Scaillet (2004)), 逆ガンマカーネル (Koul and Song (2013)), 一般化 BS カーネル族 (Marchant et al. (2013)), 歪 BS カーネル族 (Saulo et al. (2013)), 一般化逆ガウスカーネル族 (Igarashi and Kakizawa (2014)), 一般化ガンマカーネル族 (Hirukawa and Sakudo (2015)), 重み付き対数正規カーネル族 (Igarashi (2015; to appear)) がある.

3. 多次元の AK 推定　d 次元の非負データ $\mathbf{X}_1,\ldots,\mathbf{X}_n \sim \mathrm{iid}\ f$ (ただし, $\mathbf{X}_i = (X_{i1},\ldots,X_{id})'$ とする) に基づいて, $f(\mathbf{x})$, $\mathbf{x} = (x_1,\ldots,x_d)' \in \mathrm{R}_+^d$ を推定したいとする.

3.1. (積型の)AK 推定　カーネル $k_{x,b}(\cdot)$ を 2 節のように選び, その積として d 次元のカーネルを作れば, このときの (積型の)AK 密度推定量は $n^{-1}\sum_{i=1}^n \prod_{j=1}^d k_{x_j,b_j}(X_{ij})$ と定義される. ここで, $b_j = b_j(n) \propto b(n)$ とする (Bouezmarni and Rombouts (2010) が, Chen (1999,2000) のベータ/ガンマカーネルによる (積型の)AK 密度推定量を考察).

3.2. (非積型の)AK 推定　Koul and Song (2013) に述べられた注意点も一理があり, 多変量ガンマ分布 (Kotz et al. (2000) のテキストからその定義自体も複数存在, かつ, 無限級数を伴う定義になっている) を直接, 密度推定に応用するのは無理があろうことは致し方ない (勿論, 理論的な関心は残されている).

一方, 要素毎に適当な変換を施して多変量分布が構成される場合, 計算機の面からも解析的にも扱いが容易になることが期待できる. 典型例として, (重み付き) 多変量対数正規分布があり, これに基づいた (非積型の)AK 密度推定量を五十嵐 (2015; 統計関連

[1] この形式は “重み付き推定量” (例えば, Siliverman (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.

学会連合大会) が報告している．本報告では, 多変量 BS 分布を (非積型の) カーネルとして用いることに焦点をおき, その拡張についても議論した (なお, 対数正規分布と BS 分布との関係も触れた).

4. 多変量 BS カーネル密度推定量とその拡張

4.1. 準備 (分布論)　1 次元の確率変数 ξ が密度関数

$$k(s;\alpha,\beta)=\phi\Big(\frac{a(s/\beta)}{\alpha}\Big)\frac{A(s/\beta)}{\alpha\beta}\quad (s\in \mathrm{R}_+) \tag{BS}$$

(Birnbaum and Saunders (1969)) を持つとき, ξ は BS 分布 $BS(\alpha,\beta)$ に従うといい, $\xi\sim BS(\alpha,\beta)$ と書く．ここに, $\alpha,\beta>0$ であり, $a(t)=t^{1/2}-\frac{1}{t^{1/2}}$, $A(t)=\frac{1}{2}\Big(\frac{1}{t^{1/2}}+\frac{1}{t^{3/2}}\Big)$, $\phi(u)=\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{u^2}{2}\Big)$. Kundu et al. (2010) は自然な拡張として 2 次元 BS(BBS) 分布 $BBS(\alpha_1,\alpha_2,\beta_1,\beta_2,\rho)$ を定義した (この d 次元化も同様に定義できるが, ここでは省略):

$$k(\mathbf{s};\alpha_1,\alpha_2,\beta_1,\beta_2,\rho)=\phi_\rho\Big(\frac{a(s_1/\beta_1)}{\alpha_1},\frac{a(s_2/\beta_2)}{\alpha_2}\Big)\prod_{j=1}^{2}\frac{A(s_j/\beta_j)}{\alpha_j\beta_j}\quad (\mathbf{s}=(s_1,s_2)'\in \mathrm{R}_+^2)\,.$$

ここに, $\alpha_1,\alpha_2,\beta_1,\beta_2>0$, $|\rho|<1$, $\phi_\rho(u_1,u_2)=\frac{1}{2\pi(1-\rho^2)^{1/2}}\exp\Big\{-\frac{(u_1^2+u_2^2-2\rho u_1u_2)}{2(1-\rho^2)}\Big\}$.

4.2. 密度推定への応用　想定したカーネル (ここでは多変量 BS, 及びその拡張) に付随するパラメータを推定位置に依存させることから, ノンパラメトリックな (可変的な)AK 密度推定量を構築できる．すなわち, 上記のパラメータ (α,β) に対し,

$$\alpha_{b,c}(x)=\Big(\frac{b}{x+bc}\Big)^{1/2},\quad \beta_{b,c}(x)=x+bc\quad (\text{ただし}, c>0)$$

を採用し, (非積型の) 2 次元 BS カーネル密度推定量 (d 次元化も同様)

$$\widehat{f}^{(BBS_\rho)}_{(b_1,b_2),(c_1,c_2)}(\mathbf{x})=\frac{1}{n}\sum_{i=1}^{n}k^{(BBS_\rho)}_{(b_1,b_2),(c_1,c_2)}(\mathbf{X}_i;\mathbf{x})\,,\quad \mathbf{x}=(x_1,x_2)'\in \mathrm{R}_+^2$$

を提案した ($\rho=0$ のとき (積型の) 密度推定量に他ならない). さらに，漸近バイアス/分散/平均積分 2 乗誤差/漸近正規性/強一致性の結果, 及び数値例を報告した.

4.3. 拡張　d 次元でも同様の議論があるので, ここでは 1 次元に限定する.

(I) BS 密度は, 標準正規密度 ϕ の変換分布である．その ϕ を他の対称密度にするとか, 歪みを加えることもできる (BS 分布よりも柔軟なモデル化を扱った文献は膨大).

(II) 分布生成としての変換法の観点からは, おそらく BS 密度よりも対数正規密度

$$f_{\mathrm{LN}}(s)=\frac{1}{\sqrt{2\pi}\sigma s}\exp\Big\{-\frac{1}{2\sigma^2}(\log s-\mu)^2\Big\}\quad (s\in \mathrm{R}_+) \tag{LN}$$

がよく知られている．変換による特徴付け『$Z=a(\xi^{(BS)}/\beta)/\alpha\sim N(0,1)$』の関数 a を

$$a^{[q]}(t)=\begin{cases}\dfrac{1}{2q}(t^q-t^{-q})\,, & q\neq 0\,,\\ \log t\,, & q=0\end{cases}$$

(Yang (2006)) へと拡張することで両者を含む分布族になる.

これらの方向からの密度推定の研究は現在進行中である (進展があれば, 別の機会に報告したい).

事前密度のベイズ尤度とその含意

柳 本 武 美: 統計数理研究所

1. 問題の設定

確率モデルを仮定して得られたデータ x を評価する頻度論は、今日の統計学に確固たる基盤をなしている。一方、標本密度 $p(x|\,\theta)$ を母数 θ の関数と見て尤度と見なすことがある。尤度に基づいた 尤度推論が優れていることが広く受け入れられている。対応するベイズモデルにおける尤度は事前密度の確からしさである、との仮定の下に尤度の定義を試みている。尤度が導入できると様々な適用が期待できる。実際の適用では標本密度より事前密度の評価を行う場合が殆どだから、この報告では確率密度の評価には言及しない。

2. ベイズ尤度の定義

ベイズ尤度を予測子応じて定義する。その意味で複数のベイズ尤度を定義する。標本密度を $p(x|\,\theta)$ 事前密度を $\pi(\theta)$ とする。また、y を x と同じ標本空間を上の将来 (未観測) の標本として予測子は一般に $p(y|x)$ と書く。

2.1. 予測子

ベイズモデルにおける予測子 $p_m(y|\,x)$ は Aitchison (1975) により与えられ、$\mathrm{E}\{p(y|\,\theta);\,\pi(\theta|\,x)\}$ で定義される。ここで y は x と同じ標本空間の変数であり、未来の観測値 (あるいは未知観測値) とする。事前密度ではなく事後密度の下での期待値から導かれている。この予測子は、密度関数の空間の微分幾何学的接続の概念を用いて一般化できることを Corcuera and Giummole (1999) が示した。彼らは一般的な接続係数 α に対して様々な予測子を議論した。その両端である m-混合と e-混合が多くの場合に重要である。Yanagimoto and Ohnishi (2009) 後者の予測子が実際的あることを論じている。特に Kullback-Leibler sparator との関連が深くなる。ベイズ尤度の関連では予測子を定義するためには、事後密度だけではなくて事前密度による混合も用いられる。結局以下のように四つの予測が議論の対象とする。

Definition 1. 四つの予測子は次のように定義される:

1) $p_e(y;\pi) \;=\; \exp\{\mathrm{E}[\log p(y|\theta);\,\pi(\theta)]\}k_1,$

2) $p_m(y;\pi) \;=\; \mathrm{E}\{p(y|\theta);\,\pi(\theta)\}.$

3) $p_e(y|x;\pi) \;=\; \exp\{\mathrm{E}[\log p(y|\theta);\,\pi(\theta|x)]\}k_2$ and

4) $p_m(y|x;\pi) \;=\; \mathrm{E}\{p(y|\theta);\,\pi(\theta|x)\}$

ここで $k_2 = \exp\big[\mathrm{E}\{\mathrm{D}\big(p_e(y|\,x;\pi), p(y|\theta)\big);\,\pi(\theta|\,x)\}\big(\;=\;\exp\{p_D\}\big).$

2.2. 四つの尤度

予測子 $p(y|\,x)$ からベイズ尤度を導くために $\log p(y|\,x)$ の不変な量を x と $\pi(\theta)$ を用いて構成する。考え方は情報量規準の導出と同じである。その上で、改めて指数変換を施してベイズ尤度を導

出する。この指針は多様なベイズ尤度の定義を導いてしまう意味で一意には定められない。

Definition 2. Definition 1 で与えた予測しに対応して次の四つのベイズ尤度が定義される.

1) $\mathrm{L}_e(\pi;\, x) \;=\; p_e(x;\pi)$,

2) $\mathrm{L}_m(\pi;\, x) \;=\; p_m(x;\pi)$,

3) $\mathrm{L}_e(\pi|\, x) \;=\; p_e(x|x;\pi) \cdot \exp(-2p_D)$ and

4) $\mathrm{L}_m(\pi|\, x) \;=\; p_m(x;\, \pi) \cdot \exp\{\mathrm{D}(p_m(y|x;\pi),\, p_m(y;\pi))\}$.

2.3. 定義に見る尤度の特徴

尤度 $\mathrm{L}_m(\pi;\, x)$ は通常は周辺尤度と呼ばれる。問題は対応する予測子が事後密度ではなくて事前密度を用いて導出されている点である。尤度 $\mathrm{L}_e(\pi|\, x)$ を対数変換して -2 を乗じると Spiegelhalter ら (2002) が導入した DIC の特定化版になる。特定化では予測子を前面に出して plug-in 予測子を排除した点にあるからごく真っ当である。その結果、調整項 p_D が常に正になる。DIC は BUGS に搭載されるなど広く用いられているが、様々な議論がある (例えば Spiegelhalter ら, 2014)。

多くの研究者は予測子 $p_m(y|x;\pi)$ に基づいた尤度の導入を期待しているようである (例えば Gelman ら. 2014; Spiegelhalter ら, 2014)。この予測子は対数変換すると不偏な量を見いだすのは容易ではない。その理由は、標本分布族にしばしば見られる積構造と m-混合と折り合いの悪さに起因する。これ以外のベイズ尤度は、小修正を除けばほぼ確定できる思われる。

3. 含意

この節では、ベイズ尤度が定義できたときの含意を議論する。

3.1. 経験ベイズモデルにおける超母数の推定

尤度が定義できると最初の適用は経験ベイスモデルへの適用である。事前密度の評価は、超母数の評価と読み換えられるから、ごく自然に最大ベイズ尤度推定量により超母数を推定することになる。周辺尤度最大化と最小 DIC の比較についての研究は、二つのベイズ尤度を用いたときの性能の比較と見なされる。ベイズ尤度として一つの枠組みとして理解することにより高い視点からの比較が可能である。

3.2. 情報量規準

Spiegelhalter ら (2002, p168) による DIC は AIC のベイズ版として調整項 (ペナルティ項) を変更したものである。既に述べたように、$-0.5\times$ 特定版 DIC は $\log \mathrm{L}_e(\pi|\, x)$ に一致する。そして、特に無情報事前密度が仮定された下では DIC が AIC と一致することがあることを指摘されている。情報量規準 BIC はサイズ 1 の標本相当の情報量を含む事前情報を仮定したときの周辺尤度で近似されることが知られている 。この事実は次のように見ることができる。情報量規準はベイズ尤度の特殊な場合である。

3.3. Lindley's paradox

Libdkey paradox はベイズ尤度が低い二つのベイズモデルの比較から生じている。より柔軟な二つのベイズモデルと繋ぐモデルの族を考える。その分布族に関してベイズ尤度を評価することにより現実的な推論が構成される。 推定量が意味あるものだとすれば、尤度比検定統計量も役立つと期待される。

ランダム係数を持つGMANOVAモデルの変数選択規準

広島大　理　　若木　宏文

$Y_{ij}^{(l)}$ $(l=1,\ldots,g; i=1,\ldots,n_l; j=1,\ldots,p)$ を第 l 群に属する個体 i の t_j 時点での観測値とし, 次の成長曲線モデルを考える.

$$Y_{ij}^{(l)} = b_i^{(l)} + \sum_{k=1}^{q} \beta_k^{(l)} x_k(t_j) + \varepsilon_{ij}^{(l)} \tag{1}$$

ここで, $x_k(t)$ $(k=1,\ldots,q)$ は基底関数, $b_i^{(l)}$ は個体変動を表わすランダム係数, $\varepsilon_{ij}^{(l)}$ は観測誤差で,

$$\begin{aligned} &b_i^{(l)}\ (i=1,\ldots,n_l; l=1,\ldots,g) \overset{i.i.d.}{\sim} N(\beta_0^{(l)}, \tau^2), \\ &\varepsilon_{ij}^{(l)}\ (l=1,\ldots,g; i=1,\ldots,n_l; j=1,\ldots,p) \overset{i.i.d.}{\sim} N(0, \sigma^2) \end{aligned}$$

とする.

$\boldsymbol{\theta}$ をすべての未知母数を並べたベクトル, $\mathbf{Y}$ をすべての観測変量を並べた確率行列とし, $\hat{\boldsymbol{\theta}}(\mathbf{Y})$ を $\mathbf{Y}$ による最尤推定量とすると, カルバック-ライブラーダイヴァージェンスに基づくモデルの良さは, $\mathbf{Z}$ を $\mathbf{Y}$ と独立に同じ分布に従う確率行列として

$$\mathrm{Risk}(f) = -2\mathrm{E}_{\mathbf{Y}}\mathrm{E}_{\mathbf{Z}}[\log f(\mathbf{Z}; \hat{\boldsymbol{\theta}}(\mathbf{Y}))] \tag{2}$$

で与えられる. ここで, $f(\mathbf{Y}, \boldsymbol{\theta})$ は, $\mathbf{Y}$ の同時確率密度関数である.

$$\mathbf{Y}_i^{(l)} = \begin{pmatrix} Y_{i1}^{(l)} \\ \vdots \\ Y_{ip}^{(l)} \end{pmatrix},\ \boldsymbol{\varepsilon}_i^{(l)} = \begin{pmatrix} \varepsilon_{i1}^{(l)} \\ \vdots \\ \varepsilon_{ip}^{(l)} \end{pmatrix},\ \mathbf{X} = \begin{pmatrix} x_1(t_1) & \cdots & x_q(t_1) \\ \vdots & & \vdots \\ x_1(t_p) & \cdots & x_q(t_p) \end{pmatrix}$$

$$\mathbf{B} = \begin{pmatrix} \beta_1^{(1)} & \cdots & \beta_1^{(g)} \\ \vdots & & \vdots \\ \beta_q^{(1)} & \ldots & \beta_q^{(g)} \end{pmatrix},\quad \begin{aligned} &\boldsymbol{b}_i = (b_i^{(1)}, \ldots, b_i^{(g)}), \\ &\boldsymbol{e}_l = (0, \ldots, 0, \underset{l}{1}, 0, \ldots, 0)^{\mathrm{T}} \end{aligned}$$

とおき, 説明変数行列について rank $\mathbf{X} = q$, $\mathbf{1}_p^{\mathrm{T}}\mathbf{X} = \mathbf{0}_p^{\mathrm{T}}$ と仮定する.

$$\lambda^2 = \sigma^2 + p\tau^2, \quad \mathbf{B} = (\boldsymbol{\beta}^{(1)}, \ldots, \boldsymbol{\beta}^{(l)}, \ldots, \boldsymbol{\beta}^{(g)})$$

とおくと, $\sigma^2, \lambda^2, \beta_0^{(l)}, \boldsymbol{\beta}^{(l)}$ $(l=1,\ldots,g)$ の最尤推定量は

$$\widehat{\beta}_0^{(l)} = \frac{1}{n_l p}\mathbf{1}_p^{\mathrm{T}}\mathbf{Y}^{(l)}\mathbf{1}_{n_l}, \quad \widehat{\boldsymbol{\beta}}^{(l)} = \frac{1}{n_l}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}^{(l)}\mathbf{1}_{n_l},$$

$$(\widehat{\sigma}^2, \widehat{\lambda}^2) = \begin{cases} \left(\dfrac{S_1}{n(p-1)}, \dfrac{S_2}{n}\right) & \text{if } S_1 \le (p-1)S_2 \\ \left(\dfrac{S_1+S_2}{np}, \dfrac{S_1+S_2}{np}\right) & \text{if } S_1 > (p-1)S_2 \end{cases}$$

で与えられる．ここで,

$$
\begin{aligned}
S_1 &= \sum_{l=1}^{g}\Big\{\mathrm{tr}[(\mathbf{I}_p-\mathbf{G}_p)\mathbf{Y}^{(l)}(\mathbf{I}_{n_l}-\mathbf{G}_{n_l})(\mathbf{Y}^{(l)})^{\mathrm{T}}]+\frac{1}{n_l}\mathbf{1}_{n_l}^{\mathrm{T}}(\mathbf{Y}^{(l)})^{\mathrm{T}}(\mathbf{I}_p-\mathbf{G}_p-\mathbf{\Pi}_X)\mathbf{Y}^{(l)}\mathbf{1}_{n_l}\Big\}\\
S_2 &= \frac{1}{p}\sum_{l=1}^{g}\mathbf{1}_p^{\mathrm{T}}\mathbf{Y}^{(l)}(\mathbf{I}_{n_l}-\mathbf{G}_{n_l})(\mathbf{Y}^{(l)})^{\mathrm{T}}\mathbf{1}_p,\\
&\quad \mathbf{Y}^{(l)}=[\boldsymbol{Y}_1^{(l)},\ldots,\boldsymbol{Y}_{n_l}^{(l)}],\ \mathbf{G}_{n_l}=\frac{1}{n_l}\mathbf{1}_{n_l}\mathbf{1}_{n_l}^{\mathrm{T}},\ \mathbf{G}_p=\frac{1}{p}\mathbf{1}_p\mathbf{1}_p^{\mathrm{T}},\ \mathbf{\Pi}_X=\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}
\end{aligned}
$$

である．(2) を $(-2)\times$ 最大対数尤度 $-2\log f(\mathbf{Y};\hat{\boldsymbol{\theta}}(\mathbf{Y}))$ によって推定するときのバイアスは

$$
\begin{aligned}
&\mathrm{E}_{S_1,S_2}\left[\frac{S_1-n_1\sigma^2}{S_1\{n(p-1)\}^{-1}}+\frac{S_2-n_2\lambda^2}{S_2n^{-1}}\right]\\
&\quad+\mathrm{E}_{S_1,S_2}\left[1_{\{S_1>(p-1)S_2\}}\left\{\frac{n_1\sigma^2}{S_1\{n(p-1)\}^{-1}}+\frac{n_2\lambda^2}{S_2n^{-1}}-\frac{n_1\sigma^2+n_2\lambda^2}{(S_1+S_2)(np)^{-1}}\right\}\right],\\
&n_1=n(p-1)+gq,\ n_2=n-g
\end{aligned}
$$

と表される．ラプラス展開を用いてバイアスを評価することにより次の結果を得た.

定理　$(0,1)$ 上の C^3-級関数 $H(x)$ の $x=p_0=(p-1)/p$ における微係数が

$$
H(p_0)=1,\ H'(p_0)=\frac{gp(p-q-1)}{p-1},\ H''(p_0)=-\frac{2p^2}{p-1}
$$

であるとする.

$$
\begin{aligned}
d_1&=2\left(-p+2-\sqrt{p^2-2p+2}\right),\ d_2=2\left(-p+2+\sqrt{p^2-2p+2}\right),\\
c&=\frac{2(p^2-p+1)+g(3p^2-(7+q)p+4(q+1))}{p(p-1)}
\end{aligned}
$$

とし,

$$
\begin{aligned}
\mathrm{AIC}_2&=-2\log f(\mathbf{Y};\widehat{\boldsymbol{\theta}}(\mathbf{Y}))+2\left[\{g(q+1)+2\}+\frac{(qg+1)(qg+2)}{n(p-1)-qg-2}+\frac{(g+1)(g+2)}{n-g-2}\right]\\
&\quad-\frac{np}{np-(q+1)g-2}\Big\{1_{\{\frac{S_1}{S_2}>p-1+\frac{d_1}{n}\}}+1_{\{\frac{S_1}{S_2}>p-1+\frac{d_2}{n}\}}\Big\}\Big\{H\Big(\frac{S_1}{S_1+S_2}\Big)+\frac{c}{n}\Big\},
\end{aligned}
$$

と定義すると

$$
\mathrm{E}[\mathrm{AIC}_2]-\mathrm{Risk}(f)=O(n^{-3/2})
$$

を満たす.

$$
\mathrm{AIC}_2=-2\log f(\mathbf{Y};\hat{\boldsymbol{\theta}}(\mathbf{Y}))+2\big[\{g(q+1)+2\}-1_{\{S_1>(p-1)S_2\}}\big]+O_p(n^{-1/2})
$$

であり, σ^2,λ^2 の最尤推定値がパラメータ空間の境界 $\{(\sigma^2,\lambda^2);\ \sigma^2=\lambda^2>0\}$ 上にあるときは, 通常の AIC の罰則項よりパラメータ数をほぼ一つ分 (罰則項としては 2) 減らすべきことがわかった．また, 数値実験により $O(n^{-1})$ までバイアスを補正した AIC_2 の有効性が確認された.

Distribution Theory and Its Asymptotics in Shape Analysis

(Joint Work with Professor Peter E. Jupp, University of St. Andrews, U. K.)

Yasuko Chikuse

Professor Emeritus, Kagawa University

1 INTRODUCTION

This paper is concerned with developing some families of distributions on the shape spaces. The "shape" of an object is the geometrical information that remains when allowance is made for changes in location, scale and orientation. The shape of a set X $(m \times k)$ of k (not totally coincident) labelled points in the m-dimensional real space is transformed by the Helmert matrix (centred and scaled) to give a "pre-shape" Z $(m \times (k-1))$. The "(similarity) shape space" is then the coset space of the pre-shape space obtained by the special orthogonal transformations UZ of Z for U in the special orthogonal group $SO(m)$.

It is useful to identify distributions on the shape space with distributions of Z that are invariant under the special orthogonal transformations UZ for U in $SO(m)$. Here we can construct the "reflection shape space" obtained by the orthogonal transformations UZ for U in the orthogonal group $O(m)$, removing the effect of both rotations and reflections. The distributions on this reflection shape space can be identified with distributions of Z that are invariant under the orthogonal transformations UZ for U in $O(m)$.

The main parametric families of distributions which are concerned in this paper are

1. Shape Bingham distributions having densities proportional to $\exp[\mathrm{tr}(AZ'Z)]$, and

2. Determinantal shape angular central Gaussian distributions having densities proportional to some power of the determinant of ZAZ'.

We consider the derivations of the normalizing constants of these distributions.

Also we can consider

3. the distributions without reflective symmetry, the volume Fisher-Bingham distributions, and the cardioid-type distributions.

Asymptotic behaviours of the distributions are considered for the large concentrations and for the large labels.

REFERENCES

[1] Chikuse, Yasuko and Jupp, Peter (2015). Some families of distributions on higher shape spaces, In "Geometry Driven Statistics", Dryden and Kent, eds, Wiley, pp.206-217.

[2] Chikuse, Yasuko and Jupp, Peter (2004). A test of uniformity on shape spaces, J. Multivariate Analysis 88, 163-176.

[3] Chikuse, Yasuko (2003). "Statistics on Special Manoifolds", Lecture Notes in Statistics, Vol. 174, Springer.

[4] Mardia, K. and Jupp, P. (2000). "Directional Statistics", Wiley, New York.

[5] Mardia, K. and Dryden, I. (1998). "Statistical Shape Analysis", Wiley.

[6] Goodall, C. and Mardia, K. (1993). Multivariate aspects od shape theory, Annals of Statistics, 21, 848-566.

[7] Davis, A. W, (1980). Invariant polynomials with two matrix arguments, extending the zonal polynomials, In "Multivariate Analysis, vol. V", Krishnaiah, ed, pp. 287-299.

[8] Andrews, G., Askey, R. and Roy, R. (1999). "Special Functions", Encyclopedia of Mathematics and Its Applications, Vol. 71, Cambridge University Press.

[9] Azzalini, A. (2013). "The Skew-Normal and Related Families", IMS Monographs Series, Cambridge University Press.

修正コレスキー分解を用いた分散共分散行列の縮小推定について

広島大学　茶之原直人　中川智之　若木宏文

1　導入

本研究は共分散行列の推定に関する研究である. 従来の共分散行列の推定方法としては標本共分散行列が用いられ, 大標本下においては不偏性・正定値性が保証される面で良い推定が可能である. しかし, 次元数が標本数と比べて小さくない場合 (以下高次元と呼ぶ) においては, 最大 (最小) 固有値のバイアスが増加する (Johnstone(2001)) ことや, 正定値性が保証されないことが問題視されている. したがって, 高次元下における共分散行列のより良い推定量が望まれており, 数多くの手法が提案されている (詳しくは Pourahmadi(2011) で多く紹介されている). 中でも Pourahmadi(1999) による修正コレスキー分解を用いた回帰問題への帰着は, 推定量が必ず正定値性が成立すること, LASSO(Tibushirani(1996)) などの線形回帰モデルのテクニックが応用できることから, Huang et al(2006), Chang and Tsay(2010), Levina, Rothman and Zhu(2008) など修正コレスキー分解を用いた罰則付き最尤推定法が提案されている. しかし多くの手法では共分散構造 (モデル) を限定した上での理論的性質しか分かっておらず, 構造が未知の場合の良さは保証されていない. そこで本研究では, 共分散構造を正定値性のみの仮定の下で, 修正コレスキー分解を用いた回帰係数の縮小型推定量の提案, 大標本下におけるリスクの漸近近似から理論的性質の導出を行った. 漸近近似の結果から, 罰則パラメータの選択法を提案するが, 罰則パラメータに真の共分散行列が含まれているため, その推定量について数値実験を行った.

2　提案手法

仮定 1. $\boldsymbol{Y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n)'$, $\boldsymbol{y}_i \overset{i.i.d.}{\sim} N_p(\boldsymbol{0}, \Sigma)$, $\Sigma = (\sigma_{ij})$: 正定値

仮定 1 の下で, Σの推定を行いたいとする. 本研究では修正コレスキー分解 Pourahmadi(1999) を用いた推定量として, 回帰係数 $\boldsymbol{\phi}_j$に対する縮小推定を用いた推定量$\hat{\Sigma}_{PLSE}$ を提案する.

$$
\left.
\begin{aligned}
\hat{\boldsymbol{\phi}}_{j\ PLSE} &= \left(\boldsymbol{Y}'_{j-1}\boldsymbol{Y}_{j-1} + \lambda_j I_{j-1}\right)^{-1}\boldsymbol{Y}'_{j-1}\boldsymbol{y}_{(j)} \\
\hat{\sigma}^2_{j\ PLSE} &= \tfrac{1}{n}(\boldsymbol{y}_{(j)} - \boldsymbol{Y}_{j-1}\hat{\boldsymbol{\phi}}_{j\ PLSE})'(\boldsymbol{y}_{(j)} - \boldsymbol{Y}_{j-1}\hat{\boldsymbol{\phi}}_{j\ PLSE})
\end{aligned}
\right\} \quad j = 2, \cdots, p \qquad (1)
$$

ここで, $\boldsymbol{y}_{(k)} = (y_{1,k},\ \cdots,\ y_{n,k})$, $\boldsymbol{Y}_k = (\boldsymbol{y}_{(1)},\ \cdots,\ \boldsymbol{y}_{(k)})$, $\hat{\sigma}^2_{1\ PLSE} = \boldsymbol{y}'_{(1)}\boldsymbol{y}_{(1)}/n$, $\lambda_j \geq 0$ であり, $\lambda_j = O(1)$ とする. 回帰残差平方和に Ridge 型罰則 (Hoerl and Kennard(1974)) を課したもので, 高次元下 ($n < p$) においても, $\lambda_j > 0$ であれば残差平方和は 0 にならないため, どのような状況下においても正定値な推定量が構成できる.

3　リスクの漸近近似

KL ロスに基づくリスクの漸近近似を用いて罰則パラメータの選択を考える. 罰則パラメータと

して, 別々の罰則 $\boldsymbol{\lambda} = (\lambda_2, \cdots, \lambda_p)'$と共通の罰則 $(\lambda_2, \cdots, \lambda_p)' = (\lambda, \cdots, \lambda)'$の 2 通りを (1) へ代入したときの KL リスクの最小化を行った.

系 1. $n > p$, n は十分大きい, p : 固定, $(\sigma_{1k}, \cdots, \sigma_{k-1,k})' \neq \mathbf{0}(^\forall k = 2, \cdots, p)$(※) の下で, $O(n^{-2})$ までのリスクを最小とするパラメータ $\boldsymbol{\lambda}_*$, λ_*とそのリスクは次である.

$$E\left[KL(\Sigma, \hat{\Sigma}_{PLSE}(\boldsymbol{\lambda}_*))\right] = \frac{p}{2}(p+1)n^{-1} + \left\{\frac{p}{12}\left(10p^2 + 21p + 13\right) - f(\Sigma)\right\} n^{-2} + O(n^{-\frac{5}{2}})$$

$$f(\Sigma) = \sum_{k=2}^{p}\left\{\left(tr\Sigma_{k-1}^{-1}\right)^2 \frac{\sigma_{kk\cdot k-1}}{\boldsymbol{\phi}_k'\Sigma_{k-1}^{-1}\boldsymbol{\phi}_k}\right\}, \ \ \lambda_{k*} = tr\Sigma_{k-1}^{-1}\frac{\sigma_{kk\cdot k-1}}{\boldsymbol{\phi}_k'\Sigma_{k-1}^{-1}\boldsymbol{\phi}_k}, \ \ k = 2, \cdots, p$$

$$E\left[KL(\Sigma, \hat{\Sigma}_{PLSE}(\lambda_*))\right] = \frac{p}{2}(p+1)n^{-1} + \left\{\frac{p}{12}\left(10p^2 + 21p + 13\right) - g(\Sigma)\right\} + O(n^{-\frac{5}{2}})$$

$$g(\Sigma) = \frac{\left(\sum_{k=2}^{p} tr\Sigma_{k-1}^{-1}\right)^2}{\sum_{k=2}^{p}\frac{\boldsymbol{\phi}_k'\Sigma_{k-1}^{-1}\boldsymbol{\phi}_k}{\sigma_{kk\cdot k-1}}}, \ \ \lambda_* = \frac{\sum_{k=2}^{p} tr\Sigma_{k-1}^{-1}}{\sum_{k=2}^{p}\frac{\boldsymbol{\phi}_k'\Sigma_{k-1}^{-1}\boldsymbol{\phi}_k}{\sigma_{kk\cdot k-1}}}$$

ここで, $\Sigma_j = (\sigma_{kl}), \ 1 \leq k, \ l \leq j, \ \sigma_{jj\cdot j-1} = \sigma_{jj} - \boldsymbol{\phi}_j'\Sigma_{j-1}\boldsymbol{\phi}_j, \ j = 2, \cdots, p$ である.

(※) 例えば $(\sigma_{1k}, \cdots, \sigma_{k-1,k})' = \mathbf{0}$ とする. このときλ_{k*}は, $\boldsymbol{\phi}_k = \mathbf{0}$ であるため, 分母が 0 となり$\lambda_{k*} = \infty$ となる. これは$\lambda_k = O(1)$ であることを満たしていないので, このような共分散構造下では漸近近似の精度は保証されない. しかしながら, $\lambda_{k*} = \infty$ である場合$\hat{\boldsymbol{\phi}}_{k\ PLSE} = \mathbf{0}$ となるため, 正しい推定 (罰則) を行っていると言える.

参考文献

[1] Chang, C. and Tsay, R. S. (2010). Estimation of covariance matrix via the sparse Cholesky factor with lasso. *J. Statist. Plann. Inference* 140 3858-3873.

[2] Hoerl, A. E. and Kennard, R. W. (1970). Ridge Regression. Applications to nonorthogonal problems. *Technometrics*. 12.

[3] Huang, J. Z., Liu, N., Pourahmadi, M. and Liu, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika* 93 85-98.

[4] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* 29 295-327.

[5] Levina, E., Rothman, A. and Zhu, J. (2008). Sparse estimation of large covariance matrices via a nested Lasso penalty. *Ann. Appl. Statist.* 2 245-263.

[6] Pourahmadi, M. (2011). Covariance estimation: The GLM and Regularization perspectives. *Statist. Sci.* 26 369-387.

[7] Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* 86 677-690.

[8] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B* 58 267-288.

順序カテゴリ正方分割表における
対角指数対称モデルについて

渋谷　明　(東京理科大学大学院 理工学研究科)
生亀　清貴　(東京理科大学 理工学部)
富澤　貞男　(東京理科大学 理工学部)

1. はじめに

$R\times R$ 順序カテゴリ正方分割表において, (i,j) セル確率を p_{ij} とする $(i=1,\dots,R;\ j=1,\dots,R)$. 対角指数対称 (DES) モデルは次のように定義される (Tomizawa, 1992):

$$p_{ij}=\begin{cases}\delta^{i+j}d_{|j-i|} & (i\neq j),\\ \psi_{ii} & (i=j).\end{cases}$$

DES モデルは, セル確率が主対角線に対して対称であり, さらに主対角線に沿って指数的に変化する構造をもつ. 準対角指数対称 (QDES) モデルは次のように定義される (Iki, Yamamoto and Tomizawa, 2014):

$$p_{ij}=\begin{cases}\alpha^{i}\beta^{j}d_{|j-i|} & (i\neq j),\\ \psi_{ii} & (i=j).\end{cases}$$

特に $\alpha=\beta$ とおいたこのモデルは DES モデルである. 行変数を X, 列変数を Y とし, 平均一致 (ME) モデルを $\mathrm{E}(X)=\mathrm{E}(Y)$ として定義する. Iki et al. (2014) は定理「DES モデルが成り立つための必要十分条件は, QDES モデルと ME モデルの両方が成り立つことである」を与えた.

QDES モデルは, 潜在分布として周辺分散が等しい 2 変量正規分布が想定される場合においてよく適合する (Iki et al., 2014). 本講演では, 潜在分布として周辺分散の等分散性を仮定しない 2 変量正規分布が想定される場合において, よく適合するモデルを提案する. さらに, DES モデルと QDES モデルの一般化を考える.

2. モデルの提案

モデルを次のように提案する:

$$p_{ij}=\begin{cases}\delta_1^{i+j}\delta_2^{i^2+j^2}d_{|j-i|} & (i\neq j),\\ \psi_{ii} & (i=j).\end{cases}$$

特に $\delta_2=1$ とおいたこのモデルは DES モデルである. よって, このモデルを拡張対角指数対称 (EDES) モデルと呼ぶことにする. さらに, モデルを次のように提案する:

$$p_{ij}=\begin{cases}\alpha_1^{i}\alpha_2^{i^2}\beta_1^{j}\beta_2^{j^2}d_{|j-i|} & (i\neq j),\\ \psi_{ii} & (i=j).\end{cases}$$

特に $\alpha_2=\beta_2=1$ とおいたこのモデルは QDES モデルである. よって, このモデルを拡張準対角指数対称 (EQDES) モデルと呼ぶことにする. また, $\alpha_1=\beta_1$ かつ $\alpha_2=\beta_2=1$ とおいたこのモデルは DES モデル, $\alpha_1=\beta_1$ かつ $\alpha_2=\beta_2$ とおいたこのモデルは EDES モデルである.

3. モデルの分解

平均分散一致 (MVE) モデルを $\mathrm{E}(X) = \mathrm{E}(Y)$ かつ $\mathrm{Var}(X) = \mathrm{Var}(Y)$ として定義する. 次の定理を得る:

定理 1. EDES モデルが成り立つための必要十分条件は, EQDES モデルと MVE モデルの両方が成り立つことである.

モデルの適合度を検定するための尤度比カイ二乗統計量を G^2 とする.

定理 2. EDES モデルに対する G^2 は, EQDES モデルに対する G^2 と MVE モデルに対する G^2 との和に漸近的に同等である.

4. モデルの一般化

DES モデルと QDES モデルの一般化を考える. 任意に与えられた k $(k = 1, \ldots, R-1)$ に対して, モデルを次のように提案する:

$$p_{ij} = \begin{cases} \prod_{t=1}^{k} \delta_t^{i^t + j^t} d_{|j-i|} & (i \neq j), \\ \psi_{ii} & (i = j). \end{cases}$$

$k = 1$ のときは DES モデルであり, $k = 2$ のときは EDES モデルである. すなわち, このモデルは DES モデルを一般化させたモデルである. このモデルを k 次対角指数対称 (DES(k)) モデルと呼ぶことにする. さらに, 任意に与えられた k $(k = 1, \ldots, R-1)$ に対して, モデルを次のように提案する:

$$p_{ij} = \begin{cases} \prod_{t=1}^{k} \alpha_t^{i^t} \beta_t^{j^t} d_{|j-i|} & (i \neq j), \\ \psi_{ii} & (i = j). \end{cases}$$

$k = 1$ のときは QDES モデルであり, $k = 2$ のときは EQDES モデルである. すなわち, このモデルは QDES モデルを一般化させたモデルである. このモデルを k 次準対角指数対称 (QDES(k)) モデルと呼ぶことにする. また, $\alpha_t = \beta_t$ $(t = 1, \ldots, k)$ とおいた QDES(k) モデルは DES(k) モデルである.

任意に与えられた k $(k = 1, \ldots, R-1)$ に対して, k 次周辺積率一致 (MME(k)) モデルを $\mathrm{E}(X^t) = \mathrm{E}(Y^t)$ $(t = 1, \ldots, k)$ として定義する. 次の定理を得る:

定理 3. 任意に与えられた k $(k = 1, \ldots, R-1)$ に対して, DES(k) モデルが成り立つための必要十分条件は, QDES(k) モデルと MME(k) モデルの両方が成り立つことである.

定理 4. 任意に与えられた k $(k = 1, \ldots, R-1)$ に対して, DES(k) モデルに対する G^2 は, QDES(k) モデルに対する G^2 と MME(k) モデルに対する G^2 との和に漸近的に同等である.

参考文献

[1] Iki, K., Yamamoto, K. and Tomizawa, S. (2014). *Statistics and Probability Letters*, **92**, 33-38.

[2] Tomizawa, S. (1992). *Journal of Applied Statistics*, **19**, 509-512.

パラレルプロファイルモデルにおけるランダムエフェクト共分散構造に対する尤度比統計量の高次元漸近展開

広島大学大学院 理学研究科　稲津佑

パラレルプロファイルモデルは多変量解析の場においてよく用いられている重要なモデルである. $\boldsymbol{x}_j^{(g)}$ を第 g グループの j 番目の個体から得られた p 次元目的変数ベクトルとする. ただし, $j=1,\ldots,N_g$, $g=1,\ldots,k$ であり, N を $N=N_1+\cdots+N_k$ とする. また, $N-p-k-1>0$ は常に成り立っているものとし, k は 2 以上の固定された自然数とする. ここで, $\boldsymbol{x}_1^{(1)},\ldots,\boldsymbol{x}_{N_k}^{(k)}$ は互いに独立に以下の分布,

$$\boldsymbol{x}_j^{(g)} \sim N_p(\boldsymbol{\mu}^{(g)},\boldsymbol{\Sigma}),\ \boldsymbol{\mu}^{(g)}=\delta_g\mathbf{1}_p+\boldsymbol{\mu},\ \boldsymbol{\mu}=(\mu_1,\ldots,\mu_p)',\ \delta_k=0, \tag{1}$$

に従っていると仮定する. このとき, モデル (1) をパラレルプロファイルモデルと呼ぶ. このモデルに対して, Yokoyama & Fujikoshi (1993) は $\boldsymbol{\Sigma}$ に以下のランダムエフェクト共分散構造を仮定した.

$$\boldsymbol{\Sigma}=\lambda^2\mathbf{1}_p\mathbf{1}_p'+\sigma^2\boldsymbol{I}_p,\ \lambda^2\geq 0,\ \sigma^2>0. \tag{2}$$

この $\boldsymbol{\Sigma}$ に対する強い仮定により, 彼らは平均パラメータ $\delta_g,\ \boldsymbol{\mu}$ に対するより有効な推定量, もしくはより強力な検定を導出することに成功している. とはいえ, これらの結果は $\boldsymbol{\Sigma}$ が (2) の構造をしているときに成り立つ結果であるから, $\boldsymbol{\Sigma}$ がランダムエフェクト共分散構造をしているかどうかを検定することが重要となる.

Yokoyama (1995) は (2) の妥当性を評価するための次の仮説

$$\mathrm{H}_0:\boldsymbol{\Sigma}=\lambda^2\mathbf{1}_p\mathbf{1}_p'+\sigma^2\boldsymbol{I}_p,\ \lambda^2\geq 0,\ \sigma^2>0 \text{ v.s. } \mathrm{H}_1:\text{not } \mathrm{H}_0, \tag{3}$$

について考察しており, (3) を検定するための尤度比検定統計量 $-2\log\Lambda$ を導出している. Yokoyama (1995) は $-2\log\Lambda$ の帰無分布の漸近分布を, Inatsu & Wakaki (2015) は $-2\log\Lambda$ の帰無分布の漸近展開公式をそれぞれ導出しており, これらを用いることで (3) を検定することが可能となる.

しかしながら, Yokoyama (1995) あるいは Inatsu & Wakaki (2015) の結果はいずれも, 次元数 p を固定した下で N を無限大に飛ばす大標本漸近理論に基づいている. そのため, 次元数 p も N と同じくらい大きい高次元の枠組みでは, 近似精度は極めて悪い. そこで, 本研究では以下に仮定する高次元の枠組み,

(HD1)　$N\to\infty$ かつ $p\to\infty$.
(HD2)　$(N-p)\to\infty$.

の下, $-2\log\Lambda$ を基準化した統計量 V の高次元漸近展開公式を導出した.

パラメータ $\lambda^2>0$ の場合と $\lambda^2=0$ の 2 通りの場合において, 高次元漸近展開公式を導出した. まず $\lambda^2>0$ の場合, V の帰無分布の高次元漸近展開は以下で与えられる.

$$\mathrm{P}(V\leq c)=Q_2(c)+O(\vartheta^{-3}),\quad \vartheta=\left\{\begin{array}{ll} p & \text{if}\quad p/N\to\varrho\in[0,1) \\ N-p & \text{if}\quad p/N\to 1,\ N-p\to\infty \end{array}\right. .$$

ただし, $Q_2(c)$ は

$$Q_2(c)=\Phi(c)-\phi(c)\left\{\frac{\tilde{\kappa}^{(3)}}{6}h_2(c)+\left(\frac{\tilde{\kappa}^{(4)}}{24}h_3(c)+\frac{(\tilde{\kappa}^{(3)})^2}{72}h_5(c)\right)\right\},$$

で定義され, $\Phi(c)$, $\phi(c)$ はそれぞれ標準正規分布の分布関数と密度関数である. また, $h_r(x)$ は

$$h_2(x) = x^2 - 1,\ h_3(x) = x^3 - 3x,\ h_5(x) = x^5 - 10x^3 + 15x,$$

である. また, $\tilde{\kappa}^{(s)} = \kappa^{(s)}\{\kappa^{(2)}\}^{-s/2}$ $(s \geq 3)$ であり,

$$\begin{aligned}\kappa^{(s)} &= (-1)^s \left\{\psi^{(s-1)}\left(\frac{N-(p-1)-k}{2}\right) - \psi^{(s-1)}\left(\frac{N-k}{2}\right)\right\} \\ &\quad + (-1)^s \sum_{l=1}^{p-2}\left\{\psi^{(s-1)}\left(\frac{N-l-1}{2}\right) - \psi^{(s-1)}\left(\frac{N-1}{2}\right)\right\} \\ &\quad + (-1)^s \left\{(p-1)\psi^{(s-1)}\left(\frac{N-1}{2}\right) - (p-1)^s\psi^{(s-1)}\left(\frac{(N-1)(p-1)}{2}\right)\right\} \quad (s = 1, 2, \ldots),\end{aligned}$$

となる. $\psi^{(s)}(\alpha)$ は

$$\psi^{(s)}(\alpha) = \left(\frac{d}{d\alpha}\right)^{s+1} \log\Gamma(\alpha) = \begin{cases} -C_{\mathrm{E}} + \displaystyle\sum_{k=0}^{\infty}\left(\frac{1}{1+k} - \frac{1}{k+\alpha}\right), & (s = 0), \\ \displaystyle\sum_{k=0}^{\infty}\frac{(-1)^{s+1}s!}{(k+\alpha)^{s+1}}, & (s = 1, 2, \ldots), \end{cases},$$

で表され, C_{E} はオイラー定数である.

一方, $\lambda^2 = 0$ の場合, V の帰無分布の高次元漸近展開は以下で与えられる.

$$\mathrm{P}(V \leq c) = Q_2(c) + \sum_{j=1}^{2}\frac{Q_2^{(j)}(c)}{j!}(-L^{-1})^j H_j + O(\vartheta^*), \quad \vartheta^* = \max\{p^{-1}N^{-3/2}, p^{-3}, \vartheta^{-3}\}.$$

ただし,

$$Q_2^{(j)}(c) = \left(\frac{d}{dc}\right)^j Q_2(c), \quad L = N\sqrt{\kappa^{(2)}},$$

であり, H_j は以下で与えられる.

$$H_j = 2^j\left(J_0\frac{\Gamma(1/2+j)}{\Gamma(1/2)} + \frac{J_1}{\sqrt{N}}\frac{\Gamma(2/2+j)}{\Gamma(2/2)} + \frac{J_2}{N}\frac{\Gamma(3/2+j)}{\Gamma(3/2)}\right).$$

係数 J_0, J_1, J_2 はそれぞれ以下で与えられる.

$$\begin{aligned}J_0 &= \frac{1}{2}\left\{1 - \frac{3k^2(p-1)^2 + 5p^2 - 2p - 1}{12Np(p-1)}\right\},\ J_1 = \frac{3k(p-1) - p - 1}{6\sqrt{\pi p(p-1)}}, \\ J_2 &= \frac{1}{2}\frac{3k^2(p-1)^2 + 5p^2 - 2p - 1}{12p(p-1)}.\end{aligned}$$

References

Inatsu, Y. and Wakaki, H. (2015). Asymptotic expansions of the null distribution of the LR test statistic for random-effects covariance structure in a parallel profile model. *TR 15–14, Statistical Research Group, Hiroshima University*, Hiroshima.

Yokoyama, T. (1995). LR test for random-effects covariance structure in a parallel profile model. *Ann. Inst. Statist. Math.*, **47**, 309–320.

Yokoyama, T. and Fujikoshi, Y. (1993). A parallel profile model with random-effects covariance structure. *J. Japan Statist. Soc.*, **23**, 83–89.

マルコフ連鎖モンテカルロ法の高次元漸近論

大阪大学大学院基礎工学研究科，金融・保険教育研究センター，JST CREST　　鎌谷研吾

マルコフ連鎖モンテカルロ (MCMC; Metropolis et al. [1953]) 法は終戦後のロスアラモス研究所で開発された積分の近似計算法の一つである．$P_d(\mathrm{d}x) = p_d(x)\mathrm{d}x$ は $\mathbb{R}^d$ 上の確率分布とする．最も基本的な MCMC 法であるランダムウォーク型メトロポリス (RWM) 法は次の手続きで定義された．ある $x \in \mathbb{R}^d$ を出発点として以下を繰り返す：

Random walk Metropolis algorithm

- $x^* \leftarrow x + w,\ w \sim N_d(0, \sigma^2 I_d)$.
- $x \leftarrow x^*$ with probability $\alpha(x, x^*)$, ただし,

$$\alpha(x, x^*) = \min\left\{1, \frac{p_d(x^*)}{p_d(x)}\right\}.$$

上記のステップにより d 次元の確率変数列 $\{X^d_m\}_m$ が定義される．このマルコフ連鎖は $P_d(\mathrm{d}x)$ を不変分布として持つように出来ていて，例えば $p_d(x) > 0$ $(x \in \mathbb{R}^d)$ であればエルゴード的になり，下記右辺の積分が存在する限り大数の法則

$$\frac{1}{M}\sum_{m=1}^{M} f(X^d_m) \to \int f(x)P_d(\mathrm{d}x) \text{ a.s. } (M \to \infty) \tag{1}$$

が成り立つ．したがって右辺の積分の近似値として左辺の有限和を用いることが正当化されるわけである．

MCMC 法は Gelfand and Smith [1990] を契機としてベイズ統計学で利用されるようになり，特に高次元での積分計算には欠かせなくなった．状態空間の次元の高さはしかし，MCMC 法のパフォーマンスを低下させる主要因でもある．高次元であればあるほど $p_d(x)$ の計算負荷が増加する上，$\{X^d_m\}_m$ の動きも制限されるからである．MCMC 法の高次元漸近論はしばしばこれらの次元の影響を見積もる際に有用な情報を提供する．

MCMC 法の高次元漸近論はもともと量子力学の分野で研究されていた (Creutz [1988])．しかしパラメータ次元 d の増加に対する採択率 $\alpha(x, x^*)$ の安定性の解析であり，MCMC 法のパフォーマンスとのつながりは明瞭とは言えなかった．転機は Roberts et al. [1997] らによるベイズ統計学的な解釈で，彼らの研究により，高次元漸近論が MCMC 法のチューニングパラメータの調整の Criterion として使われるようになる，実用面でのインパクトをもたらした．ここで彼らの結果を端的に述べる．

興味のある分布は次の形であるとする[*1]．

$$P_d(\mathrm{d}x) = \prod_{i=1}^{d} f(x_i)\mathrm{d}x_i\ (x = (x_1, \ldots, x_d)).$$

RWM 法により d 次元の系列 $\{X^d_m\}_m$ が定義され，各 d, m で

$$X^d_m = (X^d_{m,1}, \ldots, X^d_{m,m}).$$

[*1] この仮定は一般的のように見えるかもしれないが，ベイズ統計学での事後分布に対応する量であるから，現実から乖離したものであり，またおおよそ正規分布に近似されることも意味する．

と書けるとする．ここで時間スケーリング：$t \mapsto [dt]$ ($dt \in \mathbb{R}$ の整数部分) を定義し，$X^d_{[dt]}$ とすることにより連続時間の確率過程として $\{X^d_m\}_m$ を捉える．また任意の集合 $E \subset \{1, \ldots, d\}$ に対し射影 $\pi_E(x) = (x_i)_{i \in E}$ ($x = (x_1, \ldots, x_d)$) を定める．その上で $Y^d_t := \pi_{\{1\}}(X^d_{[dt]})$. と置く．提案分散のスケーリングとして，$\sigma^2 = l^2/d$ と定める．

Theorem 1 (Roberts et al. [1997])
正則条件かつ定常性のもと，$Y^d \Rightarrow Y$ *where*

$$\mathrm{d}Y_t = a(Y_t)\mathrm{d}t + b(Y_t)\mathrm{d}W_t,$$

where

$$\begin{aligned} a(y) =& \frac{(\log f)'(y)}{2} h(l),\ b(y)^2 = h(l) \\ h(l) =& 2l^2 \Phi\left(-\frac{l\sqrt{I}}{2}\right),\ I = \int \left(\frac{f'(x)}{f(x)}\right)^2 f(x)\mathrm{d}x < \infty, \end{aligned}$$

Roberts et al. [1997] らはこの定理に実用的解釈を与え，それが大きなインパクトを与えた．本発表と関連する点は，上述の定理において $t \mapsto [dt]$ とした事実である．これにより，RWM 法は次元が高くなるつれ．d のオーダーで繰り返し回数を増やしていかなければいけないことがわかる．

本発表では様々な MCMC 法の高次元漸近論を考え，手法に比較や条件の比較を行うことで，高次元における有用な MCMC 法のデザインを考えることが目的である．主に次のアルゴリズム，MpCN 法を解析する．

- $r \sim \mathrm{Gamma}(d/2, \|x - \mu\|^2/2)$,
- $x^* \leftarrow \mu + \rho^{1/2}(x - \mu) + (1 - \rho)^{1/2} r^{-1/2} w, w \sim N_d(0, \sigma^2 I_d)$,
- $x \leftarrow x^*$ with probability $\alpha(x, x^*)$, ただし，

$$\alpha(x, y) = \min\left\{1, \frac{p(y)\|x - \mu\|^{-d}}{p(x)\|y - \mu\|^{-d}}\right\}.$$

本発表は Kamatani [2014a,b] に基づいている．また，本研究は JSPS 科研費 24740062 の助成を受けている．

参考文献

Michael Creutz. Global monte carlo algorithms for many-fermion systems. *Phys. Rev. D*, 38:1228–1238, Aug 1988. doi: 10.1103/PhysRevD.38.1228.

Alan E. Gelfand and Adrian F. M. Smith. Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, 85(410):398, 1990.

Kengo Kamatani. Rate optimality of Random walk Metropolis algorithm in high-dimension with heavy-tailed target distribution. *Arxiv*, 2014a. URL `http://arxiv.org/abs/1406.5392`.

Kengo Kamatani. Efficient strategy for the Markov chain Monte Carlo in high-dimension with heavy-tailed target probability distribution. *Arxiv*, 2014b. URL `http://arxiv.org/abs/1412.6231`.

N. Metropolis, W. A. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

Gareth O. Roberts, Andrew Gelman, and Walter R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Probab.*, 7(1):110–120, 1997. ISSN 1050-5164. doi: 10.1214/aoap/1034625254.