2017 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，及び，関連分野への応用

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　挑戦的研究 (萌芽) 17K19956 「非スパースモデリングによるビッグデータの新展開（研究代表者：青嶋 誠)」，文部科学省委託事業「数学アドバンストイノベーションプラットフォーム (AIMaP)」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：2017 年 12 月 1 日 (金)〜3 日 (日)
場所：筑波大学自然系学系Ｄ棟 509 (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

12 月 1 日 (金)

14:00〜14:05　開会

14:05〜14:45　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）
Equality tests of covariance matrices based on eigenstructures in the high-dimensional context

14:55〜15:25　藤森　洸　（早稲田大・基幹理工学研究科）
西山　陽一　（早稲田大・国際教養学部）
The Dantzig selector for Cox's proportional hazards model

15:30〜16:00　牧草　夏実　（島根大・大学院総合理工学研究科)
内藤　貫太　（島根大・大学院総合理工学研究科)
再生核ヒルベルト空間における正規性の検定

16:05〜16:45　片山　翔太　（東京工業大・工学院 経営工学系）
Support recovery of adaptive generalized lasso under high-dimensionality

16:55〜17:35　筑瀬　靖子　（香川大）
High dimensional limit theorems on the stiefel manifold

12 月 2 日 (土)

9:40〜10:20　今倉　暁　（筑波大・システム情報系，人工知能科学センター）
複素モーメント型並列固有値解法とその応用

10:30〜11:10　園田　翔　（早稲田大・先進理工学部）
深層学習の Wasserstein 幾何学的解析にむけた取組み

11:20〜12:00　内藤　貫太　（島根大・大学院総合理工学研究科）
Spiridon Penev（University of New South Wales)
Locally robust methods and near-parametric asymptotics

12:00〜13:15　昼食

13:15〜13:55　柳本　武美　（統計数理研究所）
多項出現確率の推定法としての深層学習分類器

14:05〜14:45　塩川　浩昭　（筑波大・計算科学研究センター）
大規模グラフクラスタリングの高速化

14:55〜15:35　仲木　竜　（株式会社 Rhelixa・代表取締役 CEO）
全ゲノムシーケンシング時代を勝ち抜く計算科学技術

15:45〜16:25　大川　英希　（筑波大・数理物質系，宇宙史研究センター）
素粒子実験における多変量解析・機械学習・深層学習などのビッグデータ解析
− LHC-ATLAS 実験を例に

16:35〜17:15　小副川　健　（株式会社 富士通研究所）
ディープラーニングのビジネス適用に向けた取り組み

18:00〜　懇親会

12 月 3 日 (日)

9:20〜9:50　川村　健太　（島根大・大学院総合理工学研究科)
内藤　貫太　（島根大・大学院総合理工学研究科)
Divergence に基づく局所密度推定

9:55〜10:35　永井　勇　（中京大・国際教養学部）
分散共分散行列の逆行列における縮小推定法の提案

10:45〜11:25　種市　信裕　（北海道教育大・札幌校）
関谷　祐里　（北海道教育大・釧路校)
外山　淳　（数学利用研究所)
多次元分割表の完全独立性検定における変換検定統計量の構築について

11:35〜12:15　新村　秀一　（成蹊大学）
なぜ癌の遺伝子解析は 30 年以上成功しなかったのか？

12:15〜12:20　閉会

Equality tests of high-dimensional covariance matrices based on spiked eigenstructures

Aki Ishii[1], **Kazuyoshi Yata**[2], **and Makoto Aoshima**[2]
[1]Department of Information Sciences, Tokyo University of Science, Chiba, Japan
[2]Institute of Mathematics, University of Tsukuba, Ibaraki, Japan

1 Introduction

One of the features of modern data is that the data dimension d is high and the sample size n is relatively low. We call such data HDLSS data. In HDLSS situations such as $d/n \to \infty$, new theories and methodologies are required to develop for statistical inferences. Suppose we have two classes $\pi_i,\ i = 1, 2$, and define independent $d \times n_i$ data matrices, $\boldsymbol{X}_i = [\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{in_i}],\ i = 1, 2$, for $\pi_i,\ i = 1, 2$, where $\boldsymbol{x}_{ij},\ j = 1, ..., n_i$, are independent and identically distributed (i.i.d.) as a d-dimensional distribution with a mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i\ (\geq \boldsymbol{O})$. The eigen-decomposition of $\boldsymbol{\Sigma}_i$ is given by $\boldsymbol{\Sigma}_i = \boldsymbol{H}_i\boldsymbol{\Lambda}_i\boldsymbol{H}_i^T$, where $\boldsymbol{\Lambda}_i = \mathrm{diag}(\lambda_{1(i)}, ..., \lambda_{d(i)})$ having $\lambda_{1(i)} \geq \cdots \geq \lambda_{d(i)} (\geq 0)$ and $\boldsymbol{H}_i = [\boldsymbol{h}_{1(i)}, ..., \boldsymbol{h}_{d(i)}]$ is an orthogonal matrix of the corresponding eigenvectors. We considered the following test problem:

$$H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_1 \quad \text{vs.} \quad H_1 : \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_1. \tag{1.1}$$

Aoshima and Yata [1] gave a test procedure based on the quantity of $\mathrm{tr}(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2)$. They also discussed sample size determination so as to have prespecified size and power simultaneously. Li and Chen [6] considered the problem by using the quantity of $\mathrm{tr}\{(\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_2)^2\}$. The above literatures discussed asymptotic properties of the test procedures when $d \to \infty$ and $n_i \to \infty$ under the following eigenvalue condition:

$$\frac{\lambda_{1(i)}^2}{\mathrm{tr}(\boldsymbol{\Sigma}_i^2)} \to 0 \ \text{ as } d \to \infty \text{ for } i = 1, 2. \tag{1.2}$$

Aoshima and Yata [2] called (1.2) the "non-strongly spiked eigenvalue (NSSE) model". On the other hand, Ishii et al. (2016) investigated asymptotic properties of the first principal component and considered the test problem (1.1) when $d \to \infty$ while n_is are fixed under the following eigenvalue condition:

(A-i) $\sum_{s=2}^{d} \lambda_{s(i)}^2 / \lambda_{1(i)}^2 = o(1)$ as $d \to \infty$ for $i = 1, 2$.

The model (A-i) is generalized by

(A-ii) $\liminf_{d\to\infty} \left\{ \lambda_{1(i)}^2 / \mathrm{tr}(\boldsymbol{\Sigma}_i^2) \right\} > 0$ for $i = 1$ or 2.

Aoshima and Yata (2018) called (A-ii) the "strongly spiked eigenvalue (SSE) model" and showed that high-dimensional data often have the SSE model. Ishii (2017a, b) considered two-sample tests under (A-i) when $d \to \infty$ while n_is are fixed. The SSE model is very difficult to handle because of the influence of strongly spiked noise. Aoshima and Yata (2018) created a data-transformation technique for two-sample tests which transforms the SSE model to the NSSE model.

2 A new test procedure for the SSE model

We focused on the SSE model and gave new test procedures for (1.1). First, we considered modifying the test statistic given by Li and Chen [6] under (A-i). Li and Chen [6] proposed a test statistic as follows:

$$T_{n_1,n_2} = A_{n_1} + A_{n_2} - 2\mathrm{tr}\left(\boldsymbol{S}_{1n_1}\boldsymbol{S}_{2n_2}\right), \tag{2.1}$$

where $\boldsymbol{S}_{in_i}$ is the sample covariance matrix having $E(\boldsymbol{S}_{in_i}) = \boldsymbol{\Sigma}_i$ and

$$\begin{aligned} A_{n_i} =& \frac{1}{n_i(n_i-1)}\sum_{j\neq k}^{n_i}(\boldsymbol{x}_{ij}^T\boldsymbol{x}_{ik})^2 - \frac{2}{n_i(n_i-1)(n_i-2)}\sum_{j\neq k\neq l}^{n_i}\boldsymbol{x}_{ij}^T\boldsymbol{x}_{ik}\boldsymbol{x}_{ij}^T\boldsymbol{x}_{il} \\ &+ \frac{1}{n_i(n_i-1)(n_i-2)(n_i-3)}\sum_{j\neq k\neq l\neq l'}^{n_i}\boldsymbol{x}_{ij}^T\boldsymbol{x}_{ik}\boldsymbol{x}_{il}^T\boldsymbol{x}_{il'}. \end{aligned}$$

Let $m = \min(d, n_1, n_2)$. We estimated $\lambda_{1(i)}$s by using the *noise reduction (NR) methodology* given by Yata and Aoshima [7]. Let $\tilde{\lambda}_{1(i)}$ be the NR estimator for $i = 1, 2$. We proposed the following new test statistic:

$$\widetilde{T}_{n_1,n_2} = \frac{1}{2}\left(\frac{\tilde{\lambda}_{1(1)}^2}{n_1} + \frac{\tilde{\lambda}_{1(2)}^2}{n_2}\right)^{-1} T_{n_1,n_2} + 1.$$

Then, we gave the following result.

Theorem 2.1. *Under (A-i), H_0 in (1.1) and some regularity conditions, it holds that as $m \to \infty$*

$$\widetilde{T}_{n_1,n_2} \Rightarrow \chi_1^2.$$

Here, χ_1^2 denotes a random variable distributed as the χ^2 distribution with 1 degree of freedom.

We discussed the asymptotic power of $\widetilde{T}_{n_1,n_2}$. In addition, we considered extending eigenstructures from the model (A-i) to (A-ii) and gave a new test procedure. We find the difference of covariance matrices by dividing the high-dimensional eigenstructure into the first eigenspace and the others. We also discussed asymptotic properties of the new test procedure. we studied the performance of our test procedures in numerical simulations. Finally, we gave actual data analyses using a microarray data set.

References

[1] Aoshima, M. and Yata, K. (2011). Two-Stage Procedures for High-Dimensional Data. *Sequential Analysis* (*Editor's special invited paper*) 30: 356-399.

[2] Aoshima, M. and Yata, K. (2018). Two-sample tests for high-dimension, strongly spiked eigenvalue models. *Statistica Sinica* 28: 43-62.

[3] Ishii, A., Yata, K., and Aoshima, M. (2016). Asymptotic properties of the first principal component and equality tests of covariance matrices in high-dimension, low-sample-size context. *Journal of Statistical Planning and Inference* 170: 186-199.

[4] Ishii, A. (2017). A two-sample test for high-dimension, low-sample-size data under the strongly spiked eigenvalue model, *Hiroshima Mathematical Journal* 47: 273-288.

[5] Ishii, A. (2017). A high-dimensional two-sample test for non-Gaussian data under a strongly spiked eigenvalue model. *Journal of the Japan Statistical Society* 47: 273-291.

[6] Li, J. and Chen, S.X. (2012). Two-sample tests for high-dimensional covariance matrices. *Annals of Statistics* 40: 908-940.

[7] Yata, K. and Aoshima, M. (2012). Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. *Journal of Multivariate Analysis* 105: 193-215.

The Dantzig selector for Cox's proportional hazards model

藤森洸　　西山陽一
早稲田大学

固定された観測期間 $[0,1]$ の下で, $t \rightsquigarrow N_i(t),\ i=1,2,\ldots,n,$ は同時にジャンプしない計数過程とし, それぞれの強度が,

$$\lambda_i(t, Z_i) = Y_i(t)\lambda_0(t)\exp(Z_i^T\beta_0)$$

で与えられる, Cox 比例ハザードモデルを考える. ただし, $Y_i(t)$ は at risk 過程, $\lambda_0(t)$ は未知のベースラインハザード, $Z_i = (Z_i^1, Z_i^2, \ldots, Z_i^p)^T$ は有界な p-次元共変量ベクトル, $\beta_0 \in \mathbb{R}^p$ は未知のパラメータとする. $T_0 := \{j : \beta_0^j \neq 0\}$, T_0 の要素の個数を S と置くとき, $p = p_n \gg n,\ S \ll n$ なる高次元・スパースな設定の下で β_0 および, 累積ベースラインハザード関数

$$\Lambda_0(t) := \int_0^t \lambda_0(t), \quad t \in [0,1]$$

の推定問題を本講演では扱った. 規格化された Cox の対数部分尤度は,

$$l_n(\beta) := \frac{1}{n}\sum_{i=1}^n \int_0^1 \{Z_i^T\beta - \log(S_n^{(0)}(\beta,u))\}dN_i(u),$$

で与えられる. $l_n(\beta)$ の gradient を $U_n(\beta)$, Hessian を $-J_n(\beta)$ とする. すなわち,

$$U_n(\beta) := \dot{l}_n(\beta) = \frac{1}{n}\sum_{i=1}^n \int_0^1 \left(Z_i - \frac{S_n^{(1)}}{S_n^{(0)}}(\beta,u)\right) dN_i(u)$$

$$J_n(\beta) := -\ddot{l}_n(\beta) = \frac{1}{n}\int_0^1 \left[\frac{S_n^{(2)}}{S_n^{(0)}}(\beta,u) - \left(\frac{S_n^{(1)}}{S_n^{(0)}}\right)^{\otimes 2}(\beta,u)\right] d\bar{N}(u)$$

とおく. ただし, $S_n^{(k)}(\beta,u) = \sum_{i=1}^n Y_i(u)\exp(Z_i^T\beta)Z_i^{\otimes k}$, $\bar{N}(u) = \sum_{i=1}^n N_i(u),$ とした. さらに,

$$\sup_{\beta\in\mathbb{R}^{p_n}}\sup_{t\in[0,\tau]}\left\|\frac{1}{n}S_n^{(l)}(\beta,t) - s_n^{(l)}(\beta,t)\right\|_\infty \to^p 0,\ l=0,1,2,$$

を満たす $\mathbb{R}$-値関数 $s_n^{(0)}(\beta,t)$, $\mathbb{R}^{p_n}$-値関数 $s_n^{(1)}(\beta,t)$, $p_n \times p_n$ 行列値関数 $s_n^{(2)}(\beta,t)$ の存在を仮定し, $p_n \times p_n$ 行列 $I_n(\beta)$ を次のように定義する.

$$I_n(\beta) := \int_0^1 \left[\frac{s_n^{(2)}}{s_n^{(0)}}(\beta,u) - \left(\frac{s_n^{(1)}}{s_n^{(0)}}\right)^{\otimes 2}(\beta,u)\right] s_n^{(0)}(\beta,\ u)\lambda_0(u)du.$$

以上の準備に基づいて, β_0 の推定量 $\hat{\beta}_n$ を次で定義する.

$$\hat{\beta}_n := \arg\min_{\beta\in\mathcal{B}_n}\|\beta\|_1, \quad \mathcal{B}_n := \{\beta\in\mathbb{R}^{p_n} : \|U_n(\beta)\|_\infty \le \gamma_n\}.$$

ただし, γ_n は tuning parameter である. この推定量は, 適当な正則条件下で, 一致性を満たすことが示される (Fujimori and Nishiyama, 2017). さらに, 本講演では次のように T_0 の推定量 $\hat{T}_n$ を導入した.

$$\hat{T}_n = \{j : |\hat{\beta}_n^j| > \gamma_n\}.$$

S が n に依存しない定数であるとき, 次の意味で $\hat{\beta}_n$ の変数選択の意味の一致性が示される.

Theorem 1. 適当な正則条件の下で, 次が成立する.

$$\lim_{n\to\infty} P(\hat{T}_n = T_0) = 1.$$

$\hat{T}_n$ を用いることにより, β_0 の漸近正規推定量を次のように構成することができる.

Theorem 2. 次の方程式の解として, 推定量 $\hat{\beta}_n^{(2)}$ を定義する.

$$U_n(\beta_{\hat{T}_n}) = 0, \quad \beta_{\hat{T}_n^c} = 0. \tag{1}$$

ただし, ベクトル $v \in \mathbb{R}^p$, 集合 $T \subset \{1, 2, \ldots, p\}$ に対し, v_T は成分を T に制限して構成される $|T|$-次元 *sub-vector* としている. 適当な正則条件下で, $\hat{\beta}_n^{(2)}$ は次を満たす.

$$\sqrt{n}(\hat{\beta}^{(2)}_{n\hat{T}_n} - \beta_{0T_0})1_{\{\hat{T}_n = T_0\}} \to^d N(0, \mathcal{I}^{-1}).$$

ただし, $\mathcal{I}$ は行列 $I_n(\beta_0)$ の行と列を T_0 で制限して構成された *sub-matrix* である.

さらに, $\hat{\beta}_n^{(2)}$ を用いて次のように累積ベースラインハザード関数の Breslow 型の推定量を定義する.

$$\hat{\Lambda}(t) := \int_0^t \frac{d\bar{N}(s)}{\sum_{i=1}^n Y_i(s)\exp(\hat{\beta}_n^{(2)T} Z_i(s))}, \quad t \in [0, 1].$$

このとき, $\hat{\Lambda}(t)$ は次の意味で漸近正規推定量であることが示される.

Theorem 3. 適当な正則条件下で, $\sqrt{n}(\hat{\beta}^{(2)}_{n\hat{T}_n} - \beta_{0T_0})1_{\{\hat{T}_n = T_0\}}$ と, 次で定まる確率過程は $t \in [0, 1]$ ごとに漸近的に独立である.

$$\left[\sqrt{n}\{\hat{\Lambda}(t) - \Lambda_0(t)\} + \sqrt{n}\int_0^t (\hat{\beta}^{(2)}_{n\hat{T}_n} - \beta_{0T_0})^T \frac{s^{(1)}}{s^{(0)}}(\beta_{0T_0}, s)\lambda_0(s)ds\right] 1_{\{\hat{T}_n = T_0\}}$$

さらに, 後者の確率過程は漸近的に, 分散関数が次で与えられる *Gaussian martingale* と同様の分布に従う.

$$\int_0^t \frac{\lambda_0(s)}{s^{(0)}(\beta_{0T_0}, s)} ds.$$

ただし, $s^{(0)}(\beta_{0T_0}, s) = s_n^{(0)}(\beta_{0T_0}, s)$, $s^{(1)}(\beta_{0T_0}, t) := s^{(1)}_{nT_0}(\beta_{0T_0}, t)$ としている.

なお, 本講演はプレプリント Fujimori (2017) に基づくものである.

参考文献

[1] Fujimori, K. and Nishiyama, Y. (2017). The l_q consistency of the Dantzig selector for Cox's proportional hazards model. *J. Statist. Plann. Inference.* 181, 62-70.

[2] Fujimori, K. (2017). Cox's proportional hazards model with a high-dimensional and sparse regression parameter. arXiv:1710.10416[math.ST]. *Submitted.*

再生核ヒルベルト空間における正規性の検定

島根大・総合理工学研究科　牧草 夏実
島根大・総合理工学研究科　数理科学領域　内藤 貫太

1　はじめに

P を確率分布とするとき，データ $X_1, \ldots, X_n \overset{i.i.d.}{\sim} P$ に基づく

$$\text{帰無仮説 } H_0 : P = N(m_0, \Sigma_0) \ \text{ v.s } \ \text{対立仮説 } H_1 : P \neq N(m_0, \Sigma_0)$$

の検定を “正規性の検定” という．ここで，$m_0 = \mathbb{E}[X_1]$，$\Sigma_0 = V[X_1]$ である．1 変量正規性の検定としては，シャピロ・ウィルク検定，コルモゴロフ・スミノフ検定，歪度と尖度によるオムニバス検定などがよく知られており，多変量でも同様に様々な検定方法が議論されている．一方，高次元データに対する正規性の検定方法については，あまり議論がなされておらず，その検定方法について数理的な一般化を考え，ヒルベルト空間に値を取る確率変数に対する正規性の検定を考える．

2　ヒルベルト空間における正規分布

ヒルベルト空間 $\mathcal{H}$ に値を取る確率変数 X が正規分布に従うとは，任意の $h \in \mathcal{H}$ に対して，$\mathcal{H}$ における内積 $\langle X, h \rangle_{\mathcal{H}}$ が 1 変量の正規分布の確率変数となることである．また，$\mathbb{E}\left[\|X\|_{\mathcal{H}}^2\right] < \infty$ とすると，任意の $h, h' \in \mathcal{H}$ に対し，$\langle m, h \rangle_{\mathcal{H}} = \mathbb{E}_X\left[\langle X, h \rangle_{\mathcal{H}}\right]$，$\langle \Sigma h, h' \rangle_{\mathcal{H}} = \mathrm{cov}\left(\langle X, h \rangle_{\mathcal{H}}, \langle X, h' \rangle_{\mathcal{H}}\right)$ を満たす $m \in \mathcal{H}$ と作用素 $\Sigma \in HS(\mathcal{H})$ が一意に存在し，m，Σ をそれぞれ，X の期待値，共分散作用素といい，X の従う正規分布を $N(m, \Sigma)$ と表す．ここで，$HS(\mathcal{H})$ は $\mathcal{H}$ から $\mathcal{H}$ へのヒルベルトシュミット作用素の空間である．

3　検定統計量の構築

$\mathcal{H}$ 上の 2 つの確率分布 P と $N(m_0, \Sigma_0)$ との違いは

$$\Delta(P, N(m_0, \Sigma_0)) = \sup_{f \in \mathcal{F}} \left| \mathbb{E}_{X \sim P}\left[f(X)\right] - \mathbb{E}_{Y \sim N(m_0, \Sigma_0)}\left[f(Y)\right] \right| \tag{1}$$

によって測られることが [1] により知られている．ここで，$\mathcal{F}$ は $\mathcal{H}$ 上の実数値関数のクラスである．characteristic kernel k の再生核ヒルベルト空間 $H(k)$ に対し，$\mathcal{B}_1(k) = \{f \in H(k) \mid ||f||_{H(k)} \leq 1\}$ を $H(k)$ の単位球とする．$\mathbb{E}_{X \sim P}[\sqrt{k(X, X)}] < \infty$ かつ $\mathbb{E}_{X \sim N(m_0, \Sigma_0)}[\sqrt{k(X, X)}] < \infty$ を仮定し，(1) の $\mathcal{F}$ として $\mathcal{B}_1(k)$ を用いることで，P と $N(m_0, \Sigma_0)$ の MMD（Maximum Mean Discrepancy）

$$\Delta(P, N(m_0, \Sigma_0)) = ||\mu(P) - \mu(N(m_0, \Sigma_0))||_{H(k)}$$

によってそれらの違いを見ることができる．ここで，$\mu(P) = \mathbb{E}_{X \sim P}\left[k(X, \cdot)\right]$，$\mu(N(m_0, \Sigma_0)) = \mathbb{E}_{X \sim N(m_0, \Sigma_0)}\left[k(X, \cdot)\right]$ は確率分布 P，$N(m_0, \Sigma_0)$ のカーネル k による埋め込みである．$\mathcal{H}$ におけ

るデータ $X_1, \ldots, X_n \overset{i.i.d.}{\sim} P$ を用いて，$\Delta^2 = ||\mu(P) - \mu(N(m_0, \Sigma_0))||^2_{H(k)}$ は

$$\widehat{\Delta}^2 = \left|\left| \frac{1}{n} \sum_{i=1}^{n} k(\cdot, Y_i) - \mu(N(\widehat{m}, \widehat{\Sigma})) \right|\right|^2_{H(k)}$$

によって推定される．ここで，

$$\widehat{m} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

$$\widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \widehat{m})^{\otimes 2} = \frac{1}{n} \sum_{i=1}^{n} \langle X_i - \widehat{m}, \cdot \rangle_{\mathcal{H}} (X_i - \widehat{m})$$

である．さらに，$\mu(N(m, \Sigma))$ が (m, Σ) に関して連続であるとき，$\widehat{\Delta}^2$ は Δ^2 の一致推定量となることが [2] によって知られている．

本発表では，この検定統計量 $\widehat{\Delta}^2$ の漸近挙動について報告を行った．特に，帰無仮説 H_0 : $P = N(m_0, \Sigma_0)$ のもとで，退化 V 統計量の結果に帰着させることで $\widehat{\Delta}^2$ の漸近分布が，独立な自由度 1 の χ^2 分布の重みつき無限和の形で得られること ([3] 参照)，また，対立仮説 $H_1 : P \neq N(m_0, \Sigma_0)$ のもとでの $\widehat{\Delta}^2$ の漸近分布が，平均が 0 の正規分布になっていること，局所対立仮説 $P = P_n = (1 - 1/\sqrt{n})N(m_0, \Sigma_0) + (1/\sqrt{n})Q$ のもとでの $\widehat{\Delta}^2$ の漸近分布が，独立な非心 χ^2 分布の重みつき無限和の形になっていることについても報告を行った．

参考文献

[1] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schlkopf, and A. Smola (2007). A kernel method for the two-sample-problem. *Advances in Neural Information Procesing Systems*, volume 19 of *MIT Press*, *Cambridge*.

[2] J. Kellner and A. Celisse (2015). A One-Sample Test for Normality with Kernel Methods, *arXiv preprint arXiv:1507.02904v1*.

[3] R. J. Serfling (1980). *Approximation theorems of mathematical statistics*, JOHN WILEY & SONS INC.

Support recovery of adaptive generalized lasso under high-dimensionality

東京工業大学工学院経営工学系　片山 翔太

1　はじめに

近年，遺伝情報やファイナンスなどの分野において，変数の次元が大きないわゆる高次元データが頻繁に得られている．線形回帰分析においては，変数選択と推定値の計算を同時に行う Lasso (Tibshirani, 1996) が，高次元データ解析に対して大きな注目を集めており，Friedman et al. (2007) や Bickel et al. (2009) をはじめとした一連の研究でその有用性が計算・理論の両方面から確認されてきている．Lasso の背後にある本質的な想定はスパース性である．これは，説明変数の多くは結果変数に効果がない，つまり，回帰係数の多くは 0 である (スパース) といったものである．高次元データ解析においてこのスパース性は自然な想定であるものの，説明変数間のグループ関係など，これだけでは捕まえられない情報も多い．そこで，Tibshirani and Taylor (2011) では，回帰係数の線形変換に対してスパース性を想定した，Generalized Lasso と呼ばれる方法を提案している．これは，Lasso のスパース制約を一般化しており，より複雑な情報を抽出することが可能になっている．しかしながら，その一般化に伴って制約も複雑になっているため，実際の推定値があまりスパースにならないという問題点がある．そこで本報告では，スパース制約にデータ依存の重みを付与した，Adaptive Generalized Lasso (AGL) を提案し，高次元データにおけるその理論的性質を明らかにする．

2　Adaptive Generalized Lasso

線形回帰モデルを考える．

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}.$$

ここで，$\boldsymbol{y} = (y_1, \ldots, y_n)^T$ は結果変数，$\boldsymbol{X} = (x_{ij}) \in \mathbb{R}^{n\times p}$ は説明変数行列，$\boldsymbol{\beta}^* = (\beta_1^*, \ldots, \beta_p^*)^T$ は回帰係数，$\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$ は誤差を意味する．このとき，Adaptive Generalized Lasso (AGL) は次のように定式化される．

$$\underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} L(\boldsymbol{\beta}) := \frac{1}{2n}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \|\boldsymbol{\Lambda}_1\boldsymbol{\beta}\|_1 + \|\boldsymbol{\Lambda}_2\boldsymbol{C}\boldsymbol{\beta}\|_1 \tag{2.1}$$

ここで，$\boldsymbol{C}$ は既知の $q\times p$ 行列であり，$\boldsymbol{\Lambda}_1 = \lambda_1\mathrm{diag}(w_{11}, \ldots, w_{1p})$, $\boldsymbol{\Lambda}_2 = \lambda_2\mathrm{diag}(w_{21}, \ldots, w_{2q})$ はそれぞれ $\boldsymbol{\beta}$, $\boldsymbol{C}\boldsymbol{\beta}$ に対する重み行列である．目的関数 $L(\boldsymbol{\beta})$ の第 3 項において回帰係数の線形変換 $\boldsymbol{C}\boldsymbol{\beta}$ をスパースにしている．行列 $\boldsymbol{C}$ の取り方によって，Generalized Fused Lasso (Hoefling, 2010) など様々な制約が表現できる．

式 (2.1) における目的関数は，

$$L(\boldsymbol{\beta}) = \frac{1}{2n}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \|\boldsymbol{\Lambda}\boldsymbol{D}\boldsymbol{\beta}\|_1,$$

と表すこともできる．ここで，

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_1 & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{\Lambda}_2 \end{pmatrix}, \quad \boldsymbol{D} = \begin{pmatrix} \boldsymbol{I}_p \\ \boldsymbol{C} \end{pmatrix}$$

である．明らかに $\mathrm{rank}(\boldsymbol{D}) = p$ であり，行列 $\boldsymbol{D}$ は最大列階数である．重み行列 $\boldsymbol{\Lambda}$ は Generalized Lasso を用いて定める．まず，

$$\tilde{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\mathrm{argmin}} \frac{1}{2n}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|_2^2 + \lambda_0\|\boldsymbol{D}\boldsymbol{\beta}\|_1, \tag{2.2}$$

を計算し，これを用いて重みを，

$$w_{1j} = \begin{cases} \frac{1}{|\tilde{\beta}_j|}, & \tilde{\beta}_j \neq 0 \\ n^\alpha & \tilde{\beta}_j = 0 \end{cases}, \quad w_{2k} = \begin{cases} \frac{1}{|(\boldsymbol{C}\tilde{\boldsymbol{\beta}})_k|}, & (\boldsymbol{C}\tilde{\boldsymbol{\beta}})_k \neq 0 \\ n^\alpha & (\boldsymbol{C}\tilde{\boldsymbol{\beta}})_k = 0, \end{cases}$$

のように定義する．ここで $\alpha > 0$ は，理論の導出の際に必要な調節パラメータである．実際上は $n^\alpha \approx \max_{1\leq j\leq p+q} 1/|(\boldsymbol{D}\tilde{\boldsymbol{\beta}})_j|$ となるように定めておけば良い．

3　理論的性質

真のサポートを $F = \mathrm{supp}(\boldsymbol{D}\boldsymbol{\beta}^*)$ とし，その補集合を $\bar{F}$ とする．いくつかの条件の下で，AGL の変数選択に関する重要な性質

$$(\boldsymbol{D}\hat{\boldsymbol{\beta}})_j = 0, \quad j \in \bar{F}$$

が示される．これは，AGL が真に 0 であるシグナルを完全に特定できることを示している．例えば行列 $\boldsymbol{C}$ の第 1 行が，$(1, -1, 0, \ldots, 0)$ で与えられるとき，すなわちスパース制約として $|\beta_1 - \beta_2|$ を考えているとき，真の回帰係数が $\beta_1^* = \beta_2^*$ を満たしているならば，AGL はこれを復元できる．

本報告では，上記の変数選択に関する性質を示すために必要な条件について，具体例な行列 $\boldsymbol{C}$ を挙げながら詳細に議論する．また AGL の応用として，グラフィカルモデルにおけるカテゴリ間の差異検出についても説明する．

High Dimensional Limit Theorems on the Stiefel Manifold

Yasuko Chikuse

Professor Emeritus, Kagawa University

1 Introduction

The Stiefel manifold is the space whose points are k-frames in the m-dimensional real space, and is represented by the matrix space of $m \times k$ matrices X such that $X'X = I$. The case $k = 1$ and the case $k = m$ are respectively the unit hypersphere and the orthogonal group. On the manifold, some fundamental distributions, matrix Langevin and matrix Bingham distributions, are introduced and utilized for the statistical inferences. See Chikuse (2003) and those references cited in the book.

However, the density functions of these distributions on the manifold are expressed in terms of the hypergeometric functions with matrix arguments. This fact makes the analyses on the manifold difficult, and we must rely on the asymptotic solutions for the analyses, e.g., for large sample size, for high concentrations and for high dimension.

This paper deals with the high dimensional limit behaviors of the statistics on the manifold. Stam (1982) considered the limit behaviors of some statistics for the uniform distribution only on the hypersphere ($k = 1$). This paper extends his results to some of the non-uniform distributions, 4 kinds of distributions constructed from the matrix Langevin and matrix Bingham distributions, on the general Stiefel manifold. We derive the high dimensional limit normality of the suitably normalized sample mean matrix (First Theorem) and the high dimensional limit orthogonality (Second Theorem) of the iid samples, for these 4 kinds of distributions.

References

[1] Chikuse, Yasuko (2003). Statistics on Special Manifolds, Lecture Notes in Statistics, Vol. 174, Springer.

[2] Stam, A.J. (1982). Limit theorems for uniform distributions on spheres in high dimensional Euclidean spaces, J. Appl. Prob., 19, 221-228.

複素モーメント型並列固有値解法とその応用

今倉 暁 ⋆

⋆ 筑波大学・システム情報系，人工知能科学センター

E-mail: imakura@cs.tsukuba.ac.jp

1 はじめに

本稿では, 複素平面上の指定された領域 $\Omega \subset \mathbb{C}$ 内部の固有値および対応する固有ベクトルを求める一般化固有値問題

$$A\boldsymbol{x}_i = \lambda_i B\boldsymbol{x}_i, \quad A, B \in \mathbb{C}^{n\times n}, \quad \boldsymbol{x}_i \in \mathbb{C}^n \setminus \{\boldsymbol{0}\}, \quad \lambda_i \in \Omega \subset \mathbb{C} \tag{1}$$

を解くことを考える. ここで, 領域 Ω の境界上の点 z において $zB - A$ は正則であるとする. このような内部固有値問題 (1) は電子状態計算や振動解析など様々な応用分野で現れる.

一般化固有値問題 (1) に対する有力な解法として, 2003 年に櫻井・杉浦によって複素モーメント型並列固有値解法が提案された [10]. その基本となるアイディアは, 一般化固有値問題 $A\boldsymbol{x} = \lambda B\boldsymbol{x}$ の固有値 λ を極に持つ有理関数 $g(z) := \boldsymbol{u}^{\mathrm{H}}(zB-A)^{-1}B\boldsymbol{v}, \boldsymbol{u}, \boldsymbol{v} \in \mathbb{C}^n$ を考え, 領域 Ω 内部の極を Cauchy の積分公式に基づく Kravanja らの手法 [8] により求めるというものである. Kravanja らの手法 [8] を用いることで, 対象の一般化固有値問題 (1) は, Hankel 行列を係数に持つ小規模な一般化固有値問題に帰着される.

複素モーメント型固有値解法はその並列性の高さから注目を集めており, 櫻井・杉浦のアイディアの直接的な改良法 [2–4,6,11] や加速部分空間法に基づく FEAST 法 [9] およびその改良法 [1,7,12] など活発に研究が進められている.

本講演では, 代表的な複素モーメント型固有値解法のアルゴリズムを紹介し, これらの解法の部分空間射影法の観点で Fig. 1 のような関係性を持つことを示す [5]. また, 電子状態計算から現れる一般化固有値問題に対する数値計算例を紹介する.

参考文献

[1] S. Güttel, E. Polizzi, T. Tang, G. Viaud, Zolotarev quadrature rules and load balancing for the FEAST eigensolver, SIAM J. Sci. Comput., **37**(2015), A2100–A2122.

[2] T. Ikegami, T. Sakurai, U. Nagashima, A filter diagonalization for generalized eigenvalue problems based on the Sakurai-Sugiura projection method, J. Comput. Appl. Math., **233**(2010), 1927–1936.

[3] T. Ikegami, T. Sakurai, Contour integral eigensolver for non-Hermitian systems: a Rayleigh-Ritz-type approach, Taiwan. J. Math., **14** (2010), 825–837.

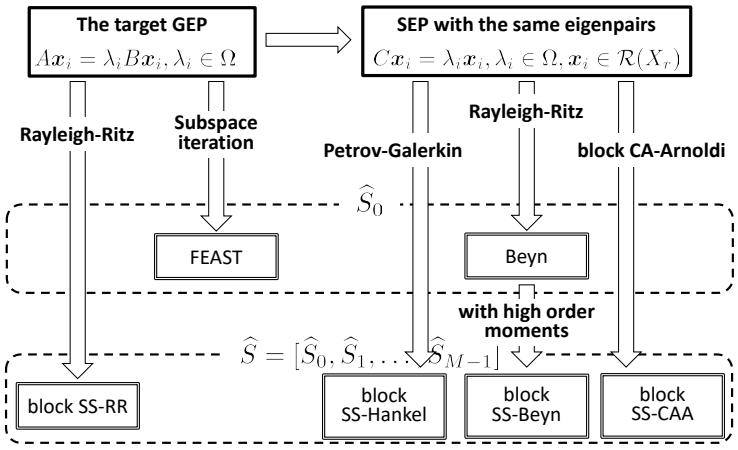


Fig. 1: A map of contour integral-based eigensolvers.

[4] A. Imakura, L. Du, T. Sakurai, A block Arnoldi-type contour integral spectral projection method for solving generalized eigenvalue problems, Applied Mathematics Letters, **32** (2014), 22–27.

[5] A. Imakura, L. Du, T. Sakurai, Relationships among contour integral-based methods for solving generalized eigenvalue problems, J. Ind. Appl. Math., **33** (2016), 721–750.

[6] A. Imakura, T. Sakurai, Block Krylov-type complex moment-based eigensolvers for solving generalized eigenvalue problems, Numer. Alg., **75**(2017), 413–433.

[7] J. Kestyn, V. Kalantzis, E. Polizzi, Y. Saad, PFEAST: a high performance sparse eigenvalue solver using distributed-memory linear solvers, SC'16 proceeding of the International Conference for High Performance Computing, Networking, Storage and Analysis, **16**, 2016.

[8] P. Kravanja, T. Sakurai, M. van Barel, On locating clusters of zeros of analytic functions, BIT, **39** (1999), 646–682.

[9] E. Polizzi, A density matrix-based algorithm for solving eigenvalue problems, Phys. Rev. B, **79** (2009), 115112.

[10] T. Sakurai, H. Sugiura, A projection method for generalized eigenvalue problems using numerical integration, J. Comput. Appl. Math., **159** (2003), 119–128.

[11] T. Sakurai, H. Tadano, CIRR: a Rayleigh-Ritz type method with counter integral for generalized eigenvalue problems. Hokkaido Math. J., **36** (2007), 745–757.

[12] P. T. P. Tang, E. Polizzi, FEAST as a subspace iteration eigensolver accelerated by approximate spectral projection, SIAM J. Matrix Anal. Appl., **35**(2014), 354–390.

深層学習の Wasserstein 幾何学的解析にむけた取組み *

園田　翔 †　村田　昇
早稲田大学　先進理工学部

深層ニューラルネットの輸送解析では，ニューラルネットの中間層を輸送写像とみなし，輸送の性質によってニューラルネットを分類する。ニューラルネットのパラメータは値が異なっていても同じ写像を表すことがあり，取扱が難しい。輸送軌道に着目することで，ニューラルネットの写像としての性質を調べるのに有効である。本講演では，Gaussian denoising autoencoder (DAE) と呼ばれる深層ニューラルネットの一種が，データ分布のエントロピーを減らす方向に輸送する写像であることを説明した。本研究の最終目的は，一般の雑音分布による DAE や，教師あり学習による深層ニューラルネットを輸送写像として記述することである。

1　Denoising Autoencoder

Denoising Autoencoder （DAE） [Vincent et al., 2008] とは，訓練データにわざと雑音を加え，雑音を除去するようにニューラルネットを訓練するオートエンコーダーの亜種である。DAE の学習問題は $\mathbb{E}_\pi \mathbb{E}_{\nu_t} |\boldsymbol{g}(\boldsymbol{x}+\boldsymbol{\varepsilon}) - \boldsymbol{x}|^2$ を目的関数とする最適化問題（変分問題）と等価である。変分計算を用いてこの変分問題を解くと，停留点 $\boldsymbol{g}_t^*(\boldsymbol{x}) = \boldsymbol{x} - \frac{1}{\nu_t * \pi(\boldsymbol{x})} \int_{\mathbb{R}^m} \boldsymbol{\varepsilon} \nu_t(\boldsymbol{\varepsilon}) \pi(\boldsymbol{x}-\boldsymbol{\varepsilon}) \mathrm{d}\boldsymbol{\varepsilon}$ が得られる。正規雑音の場合には，Stein の恒等式を用いてさらに $\boldsymbol{g}_t^*(\boldsymbol{x}) = \boldsymbol{x} + t \nabla \log[\pi * \nu_t](\boldsymbol{x})$ と簡約化できる [Sonoda and Murata, 2016]。DAE の学習手続きは**平均の推定**と等価であることに注意すると，DAE 輸送写像 $\boldsymbol{g}_t^*(\boldsymbol{x})$ は事後平均（posterior mean）$\mathbb{E}[\boldsymbol{x} \mid \widetilde{\boldsymbol{x}}]$ であることも，直接計算によって確かめられる。

2　輸送に伴うデータ分布の変化

一般の輸送写像 $\boldsymbol{g}_t : \mathbb{R}^m \to \mathbb{R}^m$ とデータ分布 π_0 が与えられた場合に，データ点 $\boldsymbol{x} \sim \pi_0$ を輸送に伴って変形して得られるデータ $\boldsymbol{x}' = \boldsymbol{g}_t(\boldsymbol{x})$ が従うデータ分布を π_t と書く。確率変数の変数変換を計算することで，π_t の時間発展法則は連続方程式 $\partial_t \pi_t(\boldsymbol{x}) = -\nabla \cdot [\pi_t(\boldsymbol{x}) \boldsymbol{v}_t(\boldsymbol{x})]$ に従うことが分かる。また，ベクトル場 $\boldsymbol{v}_t$ がポテンシャル関数 $V_t : \mathbb{R}^m \to \mathbb{R}$ の勾配とし

* 科研費シンポジウム「大規模複雑データの理論と方法論，及び，関連分野への応用」
† sho.sonoda@aoni.waseda.jp

て与えられるとき，Wasserstein 幾何学の基本的な結果 [Villani, 2009, Ex.15.10] により，∇V_t による連続方程式は，適当な汎関数 $\mathcal{F}$ による Wasserstein 勾配流 π_t に対応付けられる。連続方程式を特徴付ける速度場 $\boldsymbol{v}_t$ は時間に依存するが，Wasserstein 勾配流を特徴付けるポテンシャル汎関数 $\mathcal{F}$ は時間に依存しないので，複雑な輸送過程を扱う場合にも威力を発揮することが期待される。

特に正規雑音 DAE の場合，輸送の初速度ベクトルはスコア $\nabla \log \pi_0(\boldsymbol{x})$ で与えられる。これを連続方程式に代入すると，データ分布の変形法則は逆拡散方程式 $\partial_t \pi_{t=0}(\boldsymbol{x}) = -\triangle \pi_0(\boldsymbol{x})$ になることが分かる。さらに，逆熱方程式は Shannon-Boltzmann エントロピーを汎関数とする Wasserstein 勾配流 $\frac{\mathrm{d}}{\mathrm{d}t}\pi_t = \mathsf{grad}\,\mathcal{H}[\pi_t]$ であることから，正規雑音 DAE はエントロピーを減らすようにデータ点を輸送する写像であることが分かる。

一回の正規雑音 DAE を短時間 Δt に抑えて，得られたデータ分布に対して再び正規雑音 DAE を適用することで，L 層の深層 DAE $\boldsymbol{g}^{L:1} := \boldsymbol{g}^L \circ \cdots \circ \boldsymbol{g}^1$ が得られる。特に，各層において初速はエントロピー勾配流に従う。こうして得られる $\mathcal{W}_2(\mathbb{R}^m)$ 上の軌跡は，エントロピー勾配流の折れ線近似（接線近似，Euler 近似）である。さらに，$\Delta t \to 0$ の無限小極限では，連続無限層の DAE $\boldsymbol{g}_t$ が得られる。連続無限層正規雑音 DAE は，任意の時刻 $t \geq 0$ においてエントロピー勾配流となる。

3　一般の深層ニューラルネットの輸送解析に向けて

一般の深層ニューラルネットに対する輸送解析の展望を紹介する。

まず，Tsallis エントロピー $\mathcal{H}^q[\pi] := -\int_{\mathbb{R}^m} \frac{\pi^q(x)-\pi(x)}{q-1}\mathrm{d}x$ に対応する連続方程式は “逆” 多孔媒質方程式（backward porous medium equation）$\partial_t \pi_t = -\triangle \pi_t^q$ である。このとき，雑音分布は q-正規分布になることが予想されるが，証明には至っていない。

次に，教師有り学習では，各データ点 $\boldsymbol{x}$ にラベルが付与されている。輸送に伴い，同じラベル同士の点は近づき，異なるラベル同士の点は遠ざかることが期待される。このような輸送現象は多成分系の拡散方程式 $\partial_t \boldsymbol{\mu}_t = \triangle \boldsymbol{\mu}_t + \mathbf{R}(\boldsymbol{\mu}_t)$ として記述できる。

また，ResNet や Highway Network に見られるスキップコネクションは，ConvNet を著しく深くするためのヒューリスティクスとして不可欠である。スキップコネクションは輸送写像を陽に表した形式 $\boldsymbol{x} + \mathbf{f}(\boldsymbol{x})$ であるから，ConvNet においても輸送解析が展開できることを示唆している。

参考文献

Sho Sonoda and Noboru Murata. Decoding Stacked Denoising Autoencoders. 2016.

Cédric Villani. *Optimal Transport: Old and New*, volume 338. Springer-Verlag Berlin Heidelberg, 2009. doi: 10.1007/978-3-540-71050-9.

Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and Composing Robust Features with Denoising Autoencoders. In *ICML-08*, pages 1096–1103, Helsinki, Finland, 2008. ACM.

Locally Robust Methods and Near-Parametric Asymptotics*

内藤貫太（島根大学大学院総合理工学研究科数理科学領域）

はじめに. 密度推定における局所ロバスト推定量について議論する．$X_1, \dots, X_n \sim_{i.i.d.} f$ とし，d 変量密度関数 f の推定を問題とする．$t \in \mathbb{R}^d$ について $f(t)$ の値を推定することになる．核型推定量，スプライン，ウェーブレットなどのノンパラメトリックアプローチは，この問題への一定の解を与えている．この問題へのパラメトリックアプローチは，いわゆるプラグ・インで実装される．すなわち，p 次元パラメータ $\theta \in \Theta \subset \mathbb{R}^p$ で規定される適当なパラメトリックモデル $g(t, \theta)$ を準備し，θ の推定量 $\hat{\theta}$ をプラグ・インし，$f(t)$ の推定量 $\hat{f}(t)$ を $\hat{f}(t) = g(t, \hat{\theta})$ で構成する．このアプローチは，点 t における推定量だけでなく同時に大域的な推定量を与えているのがわかる．

一方，パラメトリックモデルにおけるパラメーターの推定に幾分拘り，自然な形でノンパラメトリックに似たアプローチが同期する手法が [2] で議論されており，これは局所尤度のバージョンである．このアプローチでは，θ の推定量を点 t に依存させる形で構成し，それを局所推定量 $\hat{\theta}(t)$ としてプラグ・インした $\hat{f}(t) = g(t, \hat{\theta}(t))$ を局所密度推定量とする．局所的なパラメータ推定の際に核関数が用いられることにより，ノンパラメトリックアプローチの要素が組み込まれることになっている．

ダイバージェンス. 密度関数間の「隔たり」を測る道具として，ダイバージェンスを用意する．ロバスト化はダイバージェンスの段階で組み込む．Box-Cox 変換

$$G_\lambda(x) = \begin{cases} \dfrac{1}{\lambda}(x^\lambda - 1) & , \quad \lambda > 0 \\ \log x & , \quad \lambda = 0 \end{cases}$$

を用いて，$U_\lambda(x) = x(G_\lambda(x) - 1)$ とする．ここで，$0 < \lambda < 1$ としておく．この U_λ を用いた

$$D_\lambda(f, g_\theta) = \int_{\mathbb{R}^d} \{U_\lambda(f(x)) - U_\lambda(g(x, \theta)) - (f(x) - g(x, \theta))U'_\lambda(g(x, \theta))\}\, dx \tag{1}$$

を，パワーダイバージェンスと呼ぶ．ここで，U'_λ は導関数，$g_\theta(\cdot) = g(\cdot, \theta)$ である．U_λ は凸であり，これは Bregman ダイバージェンスの 1 つである．非負性 $D_\lambda(f, g_\theta) \geq 0$ から，$D_\lambda(f, g_\theta)$ を最小にする θ が望ましい．$\lambda \searrow 0$ のとき，$U_\lambda(x) \to x \log x - x$ となり，パワーダイバージェンスはカルバック・ライブラーダイバージェンスを含むことに注意する．対数関数がべき関数に代わることで，推定のロバスト化が図られる（[1]）.

局所化. 密度の値を推定したい点 t に依存させる形で θ の推定を行う．このような t での局所化は，核関数を組み込むことで実装される．非負単峰可積分の核関数 $K(z), z \in \mathbb{R}^d$ は原点対称で $K(0) = 1$ を満たすとする．これを用いると，パワーダイバージェンス (1) の t-局所化は，

$$D_{\lambda,t}(f, g_\theta) = \int_{\mathbb{R}^d} K\left(\frac{x - t}{h}\right) \{U_\lambda(f(x)) - U_\lambda(g(x, \theta)) - (f(x) - g(x, \theta))U'_\lambda(g(x, \theta))\}\, dx \tag{2}$$

---

* Spiridon Penev(University of New South Walwes, Sydney, Australia) との共同研究である.

となる．ここで，$h > 0$ はスカラーバンド幅で，局所化の程度を制御する．核関数の介在で，点 t に近い部分でパワーダイバージェンス (1) が機能し，t から離れた部分では機能しないことになる．つまり，推定量を構成したい点 t の近傍での情報が重要となっている．

パラメータの推定量． 経験分布関数を F_n とする．パワーダイバージェンス (1)，(2) において，θ に依らない部分は落とし，正負を変えて得られる，

$$\rho_\lambda(t,x,\theta) = \left(\frac{\lambda+1}{\lambda}\right) K\left(\frac{x-t}{h}\right)\{g(x,\theta)^\lambda - 1\} - \int_{\mathbb{R}^d} K\left(\frac{s-t}{h}\right) g(s,\theta)^{\lambda+1} ds,$$

$$\rho_\lambda(x,\theta) = \left(\frac{\lambda+1}{\lambda}\right)\{g(x,\theta)^\lambda - 1\} - \int_{\mathbb{R}^d} g(s,\theta)^{\lambda+1} ds$$

が推定方程式で重要となる．べきパラメータ λ を明示して，従来の（大域的な）推定量 $\hat{\theta} = \hat{\theta}_\lambda$ は

$$\hat{\theta}_\lambda = arg\max_{\theta\in\Theta} \int_{\mathbb{R}^d} \rho_\lambda(x,\theta) dF_n(x) \tag{3}$$

で得られる．一方，点 t の周りの情報のみで作られる局所推定量 $\hat{\theta}(t)$ は

$$\hat{\theta}_\lambda(t) = arg\max_{\theta\in\Theta} \int_{\mathbb{R}^d} \rho_\lambda(t,x,\theta) dF_n(x) \tag{4}$$

で与えられる．

密度推定量． 大域推定量 $\hat{\theta}_\lambda$ のプラグ・インで得られる密度推定量を $g_{\hat{\theta}_\lambda}(t) = g(t,\hat{\theta}_\lambda)$ とする．局所推定量 $\hat{\theta}_\lambda(t)$ をプラグ・インした $g(t,\hat{\theta}_\lambda(t))$ は，密度関数から逸脱してしまうので，正規化した

$$\hat{f}(t) = \frac{g(t,\hat{\theta}_\lambda(t))}{\displaystyle\int_{\mathbb{R}^d} g(x,\hat{\theta}_\lambda(x))dx} \tag{5}$$

が提案される密度推定量となる．$h \to \infty$ のとき，$\rho_\lambda(t,x,\theta) \to \rho_\lambda(x,\theta)$ であるから，h が大きい時には，(3) の大域推定量 $\hat{\theta}_\lambda$ と (4) の局所推定量 $\hat{\theta}_\lambda(t)$ の違いはなくなっていく．このことに注意すると，大域推定量のプラグ・インによる $g(t,\hat{\theta}_\lambda)$ と局所推定量による (5) の $\hat{f}(t)$ の違いを見る必要がある．局所化のメリットはあるのであろうか？

本講演においては，$\hat{f}$ の密度推定量としての良さについて報告するとともに，局所化のメリットについての知見を与えた．シミュレーションおよび実データへの適用による数値的検証についても報告を行った．

参考文献

[1] Basu, A., Harris, I., Hjort, N. and Jones, M.C. (1998) Robust and Efficient Estimation by Minimizing a Density Power Divergence. *Biometrika*, **85**, 549–559.

[2] Eguchi, S. and Copas, J. (1998) A Class of Local Likelihood Methods and Near-Parametric Asymptotics. *J. R. Statist. Soc. B*, **60**, 709–724.

多項出現確率の推定法としての深層学習分類器

柳 本 武 美: 統計数理研究所

1. 序

確率モデルを仮定してデータを確率モデルからのランダムな実現値であると見なした上で、個々のデータを分類することが目的であるとすると、その問題は統計的問題である。この問題は分類問題とも見なされるし、より一般的には推定問題とも見なすことができる。機械学習の手法も原則的には同じである場合が多いが、従来その違いが強調されることが多かった。しかし、深層学習の基本的な骨格は統計的方法と馴染みやすい上に、違いが強調されることも少ない。一方で、汎化誤差がリスクと同じように使われていたり、交叉エントロピーと尤度との関係が明瞭でないなど、重要な点での違いも見受けられる。そこで、統計的方法と見なして深層学習をより深く理解することを目指す。

この発表では、データ集合の育成を視野に入れて、活性化関数としての softmax, ReLU 関数をより良い推定量を導出するための技法であるとの視点から考察する。

2. Softmax 関数

K 分類問題は、学習データ集合 $\boldsymbol{D} = (X_1, t_1, \cdots, X_n, t_n)$、$X_i$ は個別データの特性値で $t_i \in \{1, \cdots, K\}$ はその属性値、から試験データ $\tilde{X}$ の属性値を定める分類方式を設定することである。ここで $\boldsymbol{X} = (X_1, \cdots .X_n)$ と書く。深層学習が L 層から構成されて、第 L 層が出力層であるとき、重み W と偏り $\boldsymbol{b}$ を推定して、$K \times n$ 行列 $\boldsymbol{Y}$ を $\boldsymbol{Y} = W\boldsymbol{X}^{L-1} + \boldsymbol{b}$ により求める。この操作は本質的に回帰分析である。得られた K 次元ベクトル $\boldsymbol{Y}$ を活性化関数 softmax 関数 $a(\boldsymbol{Y})$ で変換する。第 k-成分は

$$a(\boldsymbol{Y})_k \;=\; \frac{\exp(\boldsymbol{Y}_k)}{\sum \exp(\boldsymbol{Y}_h)} \tag{1}$$

で表される。

2.1. 多項出現確率の推定

試験データ $\tilde{X}$ は中間層を経て K 次元ベクトル $\tilde{Y}$ を求め、(2) 式の softmax 関数で変換する。これから、K 次元多項分布の出現確率の推定値 $\hat{p}_k = \exp(\tilde{Y}_k)/\sum \exp(\tilde{Y}_h)$ を得る。この推定値から特定の属性値を決定することが出来る。分類問題を多項分布の推定に還元したことになるが、この還元は必ずしも必然ではない。本来の目的は分類だから、SVM などの適用も考えられる (例えば Karpathy, 2017)。

以下では、softmax 関数が多項分布の推測で指数分布族の微分幾何学的な双対構造に対応している点を指摘する。従って活性化関数の元来の意味づけから離れて、専ら統計的推定の方法として理解できる。また、多項出現確率の推定への還元が、データ集合の育成への寄与を促すことができることを注意する。

2.2. e 因子と m 因子

この変換は、多項分布での一般化線形モデルにおけるにおいて現れることは Yang (2016) でも指摘されている。多項分布 $p(\boldsymbol{x}|\,\boldsymbol{p}) \propto \prod p_k^{x_k}$ での回帰モデルでは自然リンク回帰モデル

$$\log \boldsymbol{p} \;=\; Z\beta \tag{2}$$

が広く用いられる。条件付き最尤推定量の適用が可能になる。$K = 2$ の場合の logit モデルがその

典型である。このとき、標本和 $x_1 + x_2$ を与えたときの条件付き分布は傾向母数にのみ依存する事実が利用される。() 式の左辺 $\boldsymbol{\theta} = \log \boldsymbol{p}$ は自然母数であり、平均母数 $\boldsymbol{p}$ と双対構造をもつ。

演者らは平均母数の Bayes 推定において、自然母数に事前分布を仮定した下での事後平均の有用性を調べている。指数分布族では、Kullback-Leibler divergence $\mathrm{D}(p(\boldsymbol{y}|\boldsymbol{x}), p(\boldsymbol{y}|\boldsymbol{p}))$ を損失としたときの事後平均を最小にする予測子 $p_e(\boldsymbol{y}|\boldsymbol{x})$ が plug-in 予測子 $p(\boldsymbol{y}|\hat{\boldsymbol{\theta}})$ と表されるからである (Yanagimoto and Ohnishi 2009)。実際多項出現確率の推定でも良好な Bayes 推定量が得られつつある。この場合にも、平均母数 m-スコアで評価するより、自然母数 e スコアで評価する方が性能が良いと見込まれる。Takano *et al.* (2016) は有限混合モデルを例にして双対モデルにおける e 混合モデルの有用性を論じている。従って、ニューロンの機能に似せた非線形関数として登場した活性化関数とは役割が大きく異なる。

2.3. データ集合の育成

分類問題を多項出現確率の推定に還元する長所として、データ集合の育成のための情報が得られることがある。推定した確率から分類する場合には $\mathrm{Argmax}\, \hat{p}_k$ を用いるしかない。しかしその最大化された値と他の値とを比べることが出来る。データ集合を改善するために、学習データの一部を除いたり後に属性値が確定したときに試験データを追加する際に、推定値は基本的な情報源になる。

この視点をより具体的に議論するために、学習到達度確認試験における項目プールの育成の問題を例に取り上げる。受験者の能力に応じた問題を構成するためには、問題の困難度だけが分かればよい。しかし、項目反応モデルで識別度をも推定しておくと、当該問題が項目プールを構成するために適しているかを判断する基礎データとなる。受験者評価のを切れ味を向上させるよう項目プールを育成させる。

アカデミアでの深層学習の研究は、データ集合の個別データを収集する費用が小さい場合が多い。そのために、データ集合を効率よく育成する視点が弱い。実際、Goodfellow ら (2016) の monograph でもデータについては anomality とか augmentation が紹介されているに過ぎない。

3. ReLU 関数

活性化関数

$$a(y) \;=\; \mathrm{Max}\{y, 0\} \tag{3}$$

は ReLU 関数と呼ばれる。この関数はシグモイド関数 $a_S(y) = e^y/(1 + e^y)$ より性能が良い活性化関数として提案されている。統計的推測ではシグモイド関数が現れるのは、目的変数が実数の場合には先ず現れない。現れるのは二項分布の回帰モデルの場合である。一方、ReLU 関数は、制約付き最尤推定量、経験ベイス推定量における超母数の推定でしばしば現れる。計算上の便宜が重要視されてきたが、生体による高度な認識の側面からの議論も求められる。

一方でこの活性化関数は、生体の反応率の近似など様々な実際に観測されるデータを記述するモデルとして用いられる。回帰モデルでは hockey stick regression 関数が使われる (Yanagimoto and Yamamoto, 1979)。健康科学・環境系の研究は hockey stick 回帰モデルでは閾値に特別な意味を求める。しかし、深層学習では中間層で用いられるのでその危惧はない。

文献

1) Goodfellow, I., Bengio. Y. it et al. (2016). MIT press, Cambridge. 2) Karpathy, A. (2017). http://cs231n.github.io/ 3) Takano, K., Hino, H. *et al.* (2016). *Neur. Comput.*, **28**, 2687-2725. 4) Yanagimoto, T. and Ohnishi (2009). *J. Stat. Plann. Inf.*, **139**, 3064-3075. 5) Yanagimoto, T. and Yamamoto, E. (1979). *Env. Health Persp.*, 32, 193-199. 6) Yang, M. (2016). http://mikkyang.com/tag/deep

大規模グラフクラスタリングの高速化

塩川浩昭
筑波大学 計算科学研究センター

1 背景

構造的類似度に基づいたグラフクラスタリング手法 SCAN[4] は近年注目を集めているクラスタリング手法である．しかしながら，SCAN はその計算量の大きさから大規模なグラフ構造を対象としたクラスタリングは難しい．SCAN では，全てのエッジに対して構造的類似度を計算する．そのため，グラフ構造中のエッジ数を $|E|$ とした時に，$O(|E|)$ の時間計算量が生じる．またグラフ構造中のノード数を $|V|$ としたとき $|E| \approx |V|^2$ となるような場合，SCAN の時間計算量は最悪計算量 $O(|V|^2)$ となる．したがって，近年増加する大規模なグラフ構造のクラスタリングに膨大な処理時間を要することになる．

本研究では従来よりも大規模なグラフ構造に対して適用可能にするため，SCAN と同一のクラスタ集合 C，ハブ集合 H，外れ値集合 O を高速に抽出するクラスタリング手法を提案する．現実のグラフ構造は高いクラスタ性を持つことから，クラスタを形成しやすく密にエッジで接続した部分グラフ構造を内包していると考えられる．そこで本研究では，全てのエッジに対して構造的類似度の計算を必要とした SCAN を高速化するために，高いクラスタ性により密なエッジの接続を有する部分グラフ構造の計算を可能な限り回避するような手法を考える．提案手法では，最短ホップ数が 2 となるような部分ノード集合を抽出し，抽出した部分ノードに含まれるノードに接続したエッジについてのみ構造的類似度計算を行う．本稿では，本手法で抽出する最短ホップ数が 2 となる部分ノード集合を 2-hop away ノードと呼ぶ．2-hop away ノードに接続するエッジについてのみ構造的類似度の計算を行うことで，高いクラスタ性により密なエッジの接続を有するサブグラフ構造を少ない計算回数でクラスタリングする．従来手法である SCAN では，全てのエッジについて構造的類似度計算を行う必要があったのに対し，提案手法は 2-hop away ノードに接続したエッジのみ構造的類似度計算を行う．ゆえに，クラスタリング全体で計算されるエッジの本数を削減することができる．

2 提案手法 SCAN++

本節では提案手法 SCAN++ について概説する．まず本稿ではグラフのクラスタ性に着目する．グラフのクラスタ性とは，エッジで接続する 2 つのノード u，v と，同じくエッジで接続する 2 つのノード v，w が存在するときに，ノード u，w がエッジで接続している割合を表したものである．このことから，グラフ構造中の多くのノードは高い確率でその隣接ノード同士がエッジで接続されていると考えられる．また同様に，エッジで接続した隣接ノード同士は他のノードから共に接続されている確率も高い．すなわち，あるノードの隣接ノード集合は他のノードの隣接ノード集合である可能性が高く，パラメータ μ が適切に設定されている状態を仮定すると，隣接ノード集合に隣接するノードについてのみ構造的類似度を計算することで，隣接ノード集合に対する構造的類似度計算を補完することができると考えられる．

そこで提案手法では，時間計算量 $O(|V|^2)$ を削減するために，最短ホップ数が 2 となるような部分ノード集合を抽出し，抽出した部分ノード集合に含まれるノードに接続したエッジについてのみ構造的類似度計算を行う．本稿では最短ホップ数が 2 となるような部分ノード集合を 2-hop away ノード集合と呼ぶ．このような方式を採ることで，2-hop away ノード集合から最短ホップ数が 1 となるノード集合に接続したエッジに対する構造的類似度計算を削減することができる．

提案手法は 2 つの利点を有する．1 つ目は，現実世界に数多く存在する複雑ネットワークに対して，従来手法 SCAN よりも高速にクラスタリングできるという点である．複雑ネットワークは，先に述べた様に高いクラスタ性を有することが知られている．このような特性をもつ現実のグラフデータに提案手法を用いることで，数多くのエッジに対する構造的類似度計算を少数のエッジに対する構造的類似度計算により補完し，計算量を削減することができる．具体的には，ノード数が $|V|$ となるグラフ構造に対して $O(|V|)$ の時間計算量でクラスタリングを実行することができる．

2 つ目は，クラスタリング結果の正確性である．提案手法は，従来手法の SCAN と異なり，一部のエッジに対してのみ構造的類似度の計算を行うが，出力されるクラスタ，ハブ，および外れ値は同一のパラメータに対して同一の結果を

表 1　Real-world datasets

	|V|	|E|	グラフの種類	クラスタ係数
condmat	23,133	186,936	Social network	0.6334
slashdot	77,360	905,468	Social network	0.0555
amazon	334,863	925,872	Web graph	0.3967
dblp	317,080	1,049,866	Social network	0.6324
road	1,379,917	3,843,320	Road network	0.0470
google	875,713	5,105,039	Web graph	0.5143
cnr	325,557	5,477,938	Web graph	0.5586
skitter	1,696,415	11,095,298	Computer network	0.2581
uk-2002	18,520,486	298,113,762	Web graph	0.6891
webbase	118,142,155	1,019,903,190	Web graph	0.5533

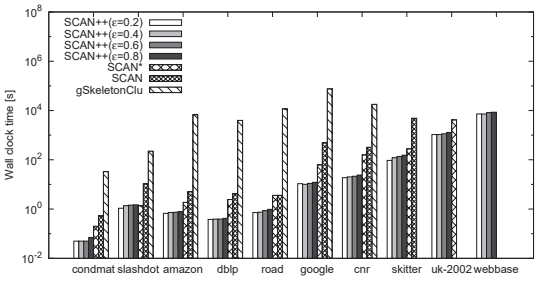


図 1　実行時間の比較

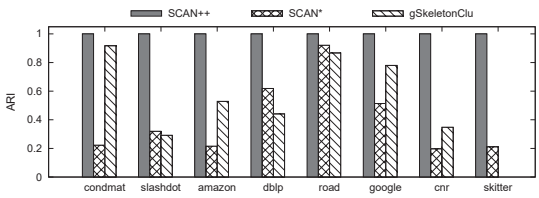


図 2　ARI

出力する．その理由として 2-hop away ノード集合のクラスタ包含性という特性が挙げられる．提案手法は，計算量削減のため，2-hop away ノード集合を逐次的に選択していくが，選択したノード集合とその隣接ノードから構成されるサブグラフ構造内に structure-connected クラスタが完全に包含される特性を有する．言い換えると，提案手法で選択した 2-hop away ノード集合で到達不可能なノードは structure-connected クラスタに含まれないということが保証されている．ゆえに，クラスタの正確性が保証されている．

3　評価実験

提案手法の有効性を評価するために，我々の提案したその高速化手法および Xu らによる SCAN[4] および SCAN の改良手法 LinkSCAN* [3]，gSkeletonClu [1] に対し，処理の高速性およびクラスタリング結果の正確性の観点から比較評価を行う．本実験には CPU が Intel Xeon Quad–Core L5640，メモリが 144GB の Linux サーバを利用する．また，提案手法および SCAN は gcc–g++_4.1.2 を用いて実装した．本実験で用いたデータセットを表 1 に示す．

実験結果を図 1 に示す．縦軸が対数表示となっていることに注意されたい．図 1 に示したように，いずれのデータセットに対しても提案手法は従来手法 SCAN に対して 20 倍程度の高速化に成功している．特にクラスタ係数の大きなデータセットである Google では計算時間を 69.9% 短縮しており，最も大きく計算時間を短縮できている．

各データセットに対する調整ランド指数 ARI の比較結果を表 2 に示す．表 2 に示すように，同一のデータセットに対して同一のパラメータが与えられる時，提案手法の ARI は従来手法 SCAN の示す ARI と一致する．すなわち，提案手法は正解ラベルに対して従来手法 SCAN と同等の ARI を示すクラスタリング結果を出力していることがわかる．

参考文献

[1] J. Huang, H. Sun, Q. Song, H. Deng, and J. Han. Revealing density-based clustering structure from the core-connected tree of a network. IEEE Trans. Knowl. Data Eng., 25(8):1876–1889, 2013.

[2] L. Hubert and P. Arabie. Comparing partitions. Journal of classification, 2(1):193–218, 1985.

[3] S. Lim, S. Ryu, S. Kwon, K. Jung, and J.-G. Lee. LinkSCAN*: Overlapping Community Detection Using the Link-space Transformation. In Proc. ICDE, pages 292–303, 2014.

[4] X. Xu, N. Yuruk, Z. Feng, and T. A. J. Schweiger. Scan: a structural clustering algorithm for networks. In KDD, pages 824–833, 2007.

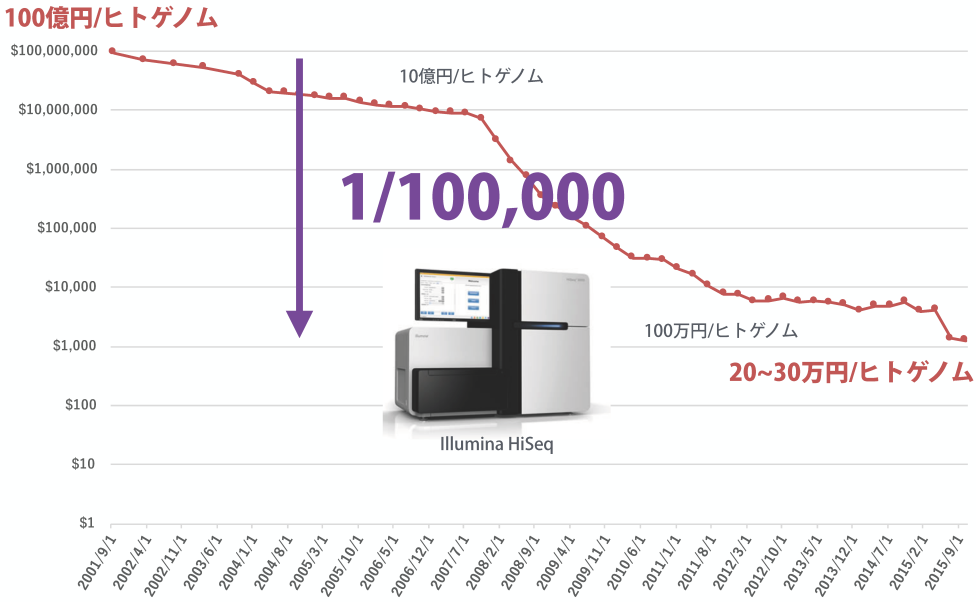
全ゲノムシーケンシング時代を勝ち抜く計算科学技術

仲木　竜
株式会社 Rhelixa 代表取締役 CEO

あらゆる生物は生体の設計図である固有の分子「ゲノム」を持っている。我々の生体活動を維持する分子機構は複雑かつ精巧である一方で、ゲノムはシンプルな構造をしていることが知られている。ゲノムはアデニン（A)、シトシン（C)、グアニン（G)、チミン（T）のたった 4 種類の分子で構成され、ヒトにおいてはそれが 30 億個連なってゲノムができている。すなわちゲノムとは文字列情報であり、その特殊な配列パターンが我々の生体のあり方を規定している。ゲノムを読み解くには、それら配列情報の中に潜むルールを解析する必要がある。

ヒトゲノムを構成する ACGT の配列は２０００年に初めて解読され、特定の領域に絞り機能を演繹していた従来の解析から、全ゲノムを同時に解析しより重要な因子を絞っていく網羅的な解析が主流となってきた。さらに２０００年以降、ゲノムを読み取る技術は革新的な飛躍を遂げ、これより数年のうちに、ヒトゲノム解読コストは 1 万円以下になると言われている。

一方で、未だヒトゲノムには未知の領域が数多く残っている。その原因として、多次元少数サンプル、生物・実験的なバイアスやノイズ、異なる性質を持つデータの統合、といった大規模ゲノムデータ特有の性質がある。株式会社 Rhelixa は、これらゲノムデータより有用な情報を引き出すための独自の計算アルゴリズムを開発・応用している。さらにゲノムデータが増加していく中、どのような技術と視点を持つことで「全ゲノムシーケンシング時代」を勝ち抜いていけるかを議論する。



ゲノムシーケンシングのコストダウン

素粒子実験における多変量解析・機械学習・深層学習などのビッグデータ解析 – LHC-ATLAS 実験を例に

筑波大学 数理物質系物理学域・宇宙史研究センター (TCHoU)　大川英希

多変量解析や機械学習などのビッグデータ解析は、素粒子実験において、非常に様々な形で用いられてきた。深層学習の導入は、まだ黎明期にあるが、今後重要な役割を果たすと予想される。本講演では、Large Hadron Collider (LHC) における ATLAS 実験、特にクォークとグルーオン由来のジェットの識別を例に、素粒子実験におけるビッグデータ解析の現状について述べる。

1　LHC-ATLAS 実験の概要

Large Hadron Collider (LHC) は、スイス・ジュネーブの欧州原子核研究機構 (CERN) にある、周長約 27km の重心系エネルギー $\sqrt{s} = 13$ TeV を誇る世界最大・最高エネルギーの陽子・陽子衝突型加速器である。LHC には、4 つの衝突点があり、そのうちの一つに ATLAS 実験がある。素粒子標準理論の検証や、それを超える物理に由来する新現象や新粒子の探索を行っている。2012 年に、LHC の ATLAS・CMS 両実験は、ヒッグス粒子を発見し、理論提唱者であるヒッグスとアングレールの、2013 年ノーベル物理学賞受賞を後押しした。

ATLAS 実験を含む、素粒子高エネルギー実験は、本シンポジウムに関係した、以下の特徴を持っている。(i) 高統計、高次元事象のビッグデータサイエンスである。(ii) 検出器の応答から粒子の運動量やエネルギーを再構成する、パターン認識であり、ある種の画像認識である。(iii) 巨大なノイズ・背景事象から、少数のシグナル事象を分離する局面に頻繁に遭遇する。

ATLAS 実験では、様々な粒子の再構成・同定において、機械学習は長らく使用されており、ニューラルネットや Boosted-Decision Tree (BDT) は、素粒子実験において市民権を得ている。深層学習の活用については、いまだ黎明期にあるものの、活気を帯び始めている。本講演では、ATLAS 実験の中でも、特に、クォークとグルーオン由来のジェットの識別における、BDT や深層学習の適用について述べる。

2　クォークとグルーオンの識別

クォークとグルーオンは、LHC において非常に高い断面積で生成される。これらの粒子は、どちらも単独では存在できず (後述の “カラー” の閉じ込め) に、それらの複合体である多数のハドロン粒子 (π・K 中間子など) が生成されることが知られている。一つのクォークやグルーオンから生成した多数の粒子が、同一方向に集中的に放出されるため、それらをまとめて “ジェット” と呼ばれる塊として再構成することがしばしば行われる。その際に、ジェットの起源となった粒子がクォークであるのかグルーオンであるのかを知ることが、物理解析の中で重要になる局面がしばしば存在する。

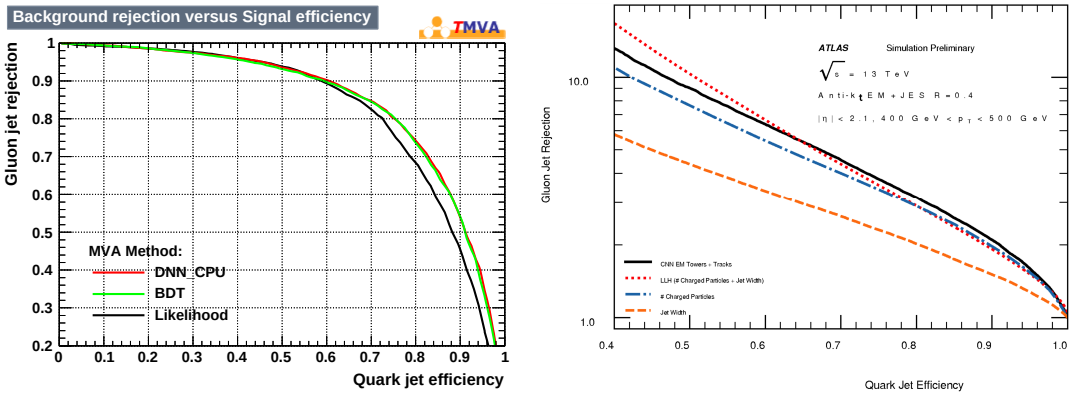
クォークとグルーオンは、異なる “カラー荷” と呼ばれる量子数 (クォーク: 4/3, グルーオン: 3) を持っており、そのために、グルーオンは、クォークに比べてより多くの生成粒子を持ち、それらの粒子の空間的広がりも大きくなる傾向がある。その性質を用いて、クォークとグルーオン由来のジェットを識別するために有効である変数が、様々に考案されてきた。特に、識別に有用な変数としては、内部飛跡検出器で再構成される荷電粒子の数 (n_{track})、ジェットの横方向に対する広がり (内部飛跡検出器を用いた場合 w_{track}、カロリメータのクラスタでは w_{calo})、又、ジェット内のトラックやクラスタの運動量の二点相関関数である C_β ($\beta = 0.2$ が通常用いられる) などが挙げられる。

しかし、どの変数においても、クォークとグルーオンジェット間の違いはそれほど有意ではなく、一変数で効果的な分離を行うことが困難であった [1–3]。そのため、次節に述べるように、多変量解析による機械学習や深層学習を用いて、識別能力の向上を目指す研究 [4–6] が様々に行われている。

2.1　多変量解析・機械学習

モンテカルロジェネレータ Pythia8 の情報を用いて、多変量解析の性能を評価したものが、図 1(a) である (ジェットの横運動量は、150< p_{T} < 200 GeV を考慮)。ATLAS 検出器シミュレーションを通していないため、[1–3] や次節とは、直接比較できないが、一変数よりも優れた識別能力を持っている。

入力変数としては、前節で示した荷電粒子数 (n_{track})、広がり、C_β に加えて、Les Houches Angularity (LHA) や破砕関数 (fragmentation function; $p_{\mathrm{T}}D$)、そしてジェットの質量の、計 11 個の変数を考慮した。



(a) (b)

図 1: Pythia8 のジェネレータ情報を用いて評価した、多変量解析を用いたクォークとグルーオンの識別能力を示す ROC カーブ (a) と深層学習による Jet imaging とジェットの内部変数を用いた場合の識別能力の比較 (b)。

図 1(a) では、尤度関数 (Likelihood)、BDT、フィードフォワードニューラルネットワークを用いた深層学習 (図中の DNN_CPU) による識別についての、各々の受信者動作特性 (ROC) 曲線を示している。Likelihood は、変数間の相関を考慮しないために、若干性能が落ちる。興味深いことに、考慮した変数に対しては、BDT と深層学習では、ほとんど結果が変わらなかった。これは、深層学習の本領発揮できる情報 (ジェットの内部構造における n 点相関など) を、十分に与えていないためであると考えられる。

入力変数の最適化の詳細については、現在調査中であるが、深層学習については、多変数の処理よりも画像処理に適しているとの示唆も得られている。この観点から行った手法について、次節で述べる。

2.2 画像認識を用いた深層学習

ジェットの内部変数を用いた識別については、上述のとおりであるが、その性能には限界がある。近年、内部変数による多変量解析ではなく、検出器 (内部飛跡検出器とカロリメータ) の情報を「画像」として捕らえること (通称 jet imaging) によって、性能が向上することが示された [6]。画像認識を行う際には、主成分分析よりは Fisher の線形判別分析が、又、更には深層学習がより優れた性能を示すことがわかった。

現在、ATLAS 実験では、jet imaging の際に、畳み込みニューラルネットワーク (CNN) を用いている。プーリング層での逆伝搬については、最大値プーリング (Max-pooling) を採用している。学習は、Keras に導入されている Adam optimizer を用いて、交差エントロピー最小化により行っている。

識別能力を示すクォークジェットの同定効率と、それに対応するグルーオンジェットの除去率を示したのが図 1(b) で、深層学習を用いた jet imaging と、ジェットの内部変数を用いて一変数又は多変数で識別を行った場合の比較を示している。深層学習による jet imaging が、最も優れた性能を持っていることがわかる。

3 まとめ

LHC-ATLAS 実験、特にクォークとグルーオン由来のジェットの識別を例に、素粒子実験における多変量解析・機械学習・深層学習などのビッグデータ解析手法の現状について述べた。深層学習については、まだ黎明期にあり、今後、適用法の詳細について様々な変更・改善が進んでいくと予想される。又、高次元統計解析手法 [7] の素粒子実験への適用可能性については、現在調査中である。

参考文献

[1] G. Aad *et al.* [ATLAS Collaboration], Eur. Phys. J. C 74:3023 (2014).

[2] M. Aaboud *et al.* [ATLAS Collaboration], ATLAS Conference Note, ATLAS-CONF-2016-034 (2016).

[3] M. Aaboud *et al.* [ATLAS Collaboration], ATLAS Note, ATL-PHYS-PUB-2017-009 (2017).

[4] 佐藤真一, 陣内修, 大川英希, 日本物理学会 2016 年秋季大会の講演, 2016 年 9 月.

[5] H. Okawa, talk at Theoretical and Experimental Issues on Jet Structure at Hadron Colloiders, Kavli-IPMU, the University of Tokyo, January 2017.

[6] M. Aaboud *et al.* [ATLAS Collaboration], ATLAS Note, ATL-PHYS-PUB-2017-017 (2017).

[7] 青嶋誠, 矢田 和善, 日本統計学会誌, 43 (2013) 123.

ディープラーニングのビジネス適用に向けた取り組み

小副川 健[1]

[1] 株式会社 富士通研究所

1　はじめに

本講演では, 富士通グループにおいて, ディープラーニングをビジネスに適用している取り組みについて紹介する. ディープラーニングは (人工) ニューラルネットワークの層を深くしたものであり, 単純な教師あり学習の解法として用いた場合でも, 他の手法を大きく上回る精度を出す場合があり, 人工知能の認識機能を支える重要な技術として期待されている. ディープラーニングの強力な特長として, これまでは知見を持った人間が設計する必要があった特徴量を自動で抽出できる, というものがある. この特長にうまく合致する応用に適用できた場合は, 膨大な時間と人手が必要だった技術開発のコストが大幅に下がり, また精度も大幅に向上する可能性があるため, ビジネス側からの期待が非常に大きい. このような状況の中, 富士通, 富士通研究所ではビジネス上の課題に対してディープラーニングを適用し, ビジネス課題を解決する例を生み出し始めている.

本講演の前半ではいくつかの事例を紹介し, ビジネス可能性の大きさや, その上での技術開発のポイントについて述べる. 後半は, 島津製作所との共同研究で, 筆者が開発を担当する, クロマトグラム分析にディープラーニングを適用した事例の紹介を通して, ビジネス領域においてこのような取り組みがどのように進むか述べる.

2　先行事例について

講演の前半は, 現在研究中のものも含め, 以下に挙げる例について紹介する:

- スマート都市監視
- 路面下空洞探査

- 3 次元形状検索
- 体操競技の採点支援
- 橋梁モニタリングシステム
- グラフデータのディープラーニングによるマルウェア検知

3　クロマトグラム分析自動化に向けたピークピッキング自動化

分子をイオン化し, その質量数や数を測定することで, 物質に何がどれだけ含まれているかなどを分析する質量分析では, 昨今の技術発展により分析に費やす労力や時間が改善され, 大量のデータを高速に計測することが可能となっている. 一方で, データ解析工程において自動化の難しいところがあり, 人手で行う部分がボトルネックになっている. 特に, クロマトグラム中のピークを検出・判定する作業では, パラメータを設定したアルゴリズムによりピークの始点と終点を求めることができるが, ノイズや複雑なピーク形状変化などの影響で, 自動で求めたピークのうち, およそ 30%程度は修正の必要があると言われており, 実際最終的には分析の技術を持った人間による目視の確認及び調整が行われている.

本研究ではこの, クロマトグラム中のピークを特定する処理にディープラーニングを適用し [1], [2], 実験データに対して F 値 0.915 を達成した. 本講演では, この取り組みの紹介を通して, ディープラーニングによるビジネス課題の解決プロセスがどのようなものであるか述べる.

参考文献

[1] 小副川健, 他. 大量クロマトグラム処理に向けた *Deep Learning* による自動ピークピッキング法 *(1)*〜アルゴリズム開発〜. 第 11 回メタボロームシンポジウム, 2017.

[2] 金澤真司, 他. 大量クロマトグラム処理に向けた *Deep Learning* による自動ピークピッキング法 *(2)*〜性能評価〜. 第 11 回メタボロームシンポジウム, 2017.

Divergence に基づく局所密度推定

島根大学総合理工学研究科　川村健太
島根大学総合理工学研究科　内藤貫太

はじめに. 密度推定におけるナイーブなパラメトリックアプローチは，有限次元パラメータ θ で規定される密度関数を考え，それに θ の推定量 $\hat{\theta}_n$ をプラグインするものである．この $\hat{\theta}_n$ は密度を推定したい点に依存しない大域的なものであり，$\hat{\theta}_n$ をプラグインして得られる密度推定量は，局所的には柔軟さに欠くことになる．したがって，密度を推定したい点に依存させる形で θ の推定量を構成し，それをプラグインする方法が魅力的となる．我々は特に，Bregman divergence と核関数を用いた局所最適化により構成されるパラメータの局所推定量とその漸近的挙動について考えていく．

Bregman divergence とパラメトリックモデル. U を凸関数とし，u をその導関数とする．また，f を $\mathbb{R}^d$ 上の未知の密度関数とし，f を推定するために用いるパラメトリックモデルを指数型分布族 $\{g_\theta(x) : \theta \in \Theta \subset \mathbb{R}^p\}$ とする．このとき，f と g_θ の違いは，Bregman divergence

$$D_U(f, g_\theta) = \int_{\mathbb{R}^d} \{U(f(x)) - U(g_\theta(x)) - u(g_\theta(x))(f(x) - g_\theta(x))\}dx$$

で測られる．ここで，U^* を U の凸共役とすると，$D_U(f, g_\theta) = D_{U^*}(u(g_\theta), u(f))$ が成り立つ ([1] の 2.1 節参照)．我々は，簡単のため，f と g_θ の違いを測るために $D_U(f, g_\theta)$ の代わりに $D_{U^*}(u(g_\theta), u(f))$ を用いる．

大域推定量. θ_0 を Θ の中で $D_{U^*}(u(g_\theta), u(f))$ を最小にするパラメータとする．このとき，θ_0 はパラメトリックモデル $\{g_\theta : \theta \in \Theta \subset \mathbb{R}^p\}$ の中で D_{U^*} の意味での f の最良近似を与えるので，f を推定するパラメトリックなアプローチでは，θ_0 を推定することが問題となる．実際，θ_0 は，f からのデータ $X_1, \ldots, X_n$ が与えられたとき，$D_{U^*}(u(g_\theta), u(f))$ の経験版である

$$\ell_n(\theta) = -\frac{1}{n}\sum_{i=1}^{n} u(g_\theta(X_i)) + \int_{\mathbb{R}^d} U^*(u(g_\theta(x)))dx \tag{1}$$

を最小にするパラメータ $\hat{\theta}_n$ により推定される．

局所推定量. 点 $t \in \mathbb{R}^d$ における $f(t)$ を推定するために，t に依存させる形で θ_0 の推定量を構成する．このような局所化は，$D_{U^*}(u(g_\theta), u(f))$ に核関数を組み込むことにより実装される．考える局所化した Bregman divergence は

$$D_{U^*}(u(g_\theta), u(f) \,|\, t) = \int_{\mathbb{R}^d} K\left(\frac{x-t}{h}\right) \{U^*(u(g_\theta(x))) - U^*(u(f(x))) - f(x)(u(g_\theta(x)) - u(f(x)))\}dx$$

となる．ここで，$K(z)$ は原点対称かつ単峰な可積分関数であり，h は Bandwidth として局所化の程度を制御する．このとき，$\theta_0(t)$ を Θ の中で $D_{U^*}(u(g_\theta), u(f)|t)$ を最小にするパラメータとすると，$\theta_0(t)$ はパラメトリックモデルの中で，t における D_{U^*} の意味での f の最良近似を与える．

そして，その推定量は $D_{U^*}(u(g_\theta), u(f)\,|\,t)$ の経験版である

$$\ell_{t,n}(\theta) = -\frac{1}{n}\sum_{i=1}^{n} K\left(\frac{X_i - t}{h}\right) u(g_\theta(X_i)) + \int_{\mathbb{R}^d} K\left(\frac{x-t}{h}\right) U^*(u(g_\theta(x)))dx \tag{2}$$

を最小にするパラメータ $\hat{\theta}_n(t)$ により構成される．

問題. 本研究の主目的は，$\hat{\theta}_n$ と $\hat{\theta}_n(t)$ の挙動を明らかにすることである．ここで，我々は特に $K(0) = 1$, かつ $h \to \infty$ という設定を考える．この設定の下では，各 $\theta \in \Theta$ に対して (2) の $\ell_{t,n}(\theta)$ は (1) の $\ell_n(\theta)$ へ漸近する．したがって，$h \to \infty$ のとき $\hat{\theta}_n(t)$ も $\hat{\theta}_n$ へ漸近することが予想される．そこで我々は，$n, h \to \infty$ のとき $\hat{\theta}_n(t)$ が θ_0 へ確率収束することの厳密な証明を与える．

主結果. 大域推定量 $\hat{\theta}_n$ に関する Theorem 1 は M-推定量の一致性に関するものである ([2] 参照).

Theorem 1　Θ をコンパクト集合とし，U を二回連続微分可能であるとする．また，$\int_{\mathbb{R}^d} \sup_{\theta\in\Theta} |u(g_\theta(x))|f(x)dx < \infty$ であり，$D_{U^*}(u(g_\theta), u(f))$ の最大点は唯一であるとする．このとき，$\hat{\theta}_n \xrightarrow{P} \theta_0$ が成り立つ.

局所推定量 $\hat{\theta}_n(t)$ の漸近挙動について述べたのが次の Theorem 2 となる.

Theorem 2　Theorem 1 の仮定に加えて，任意の $\varepsilon > 0$ に対して，

$$\sup_{\theta\in\Theta} \int_{\{\|x\|>L\}} |U^*(u(g_\theta(x)))|dx < \varepsilon$$

を満たす $L > 0$ が存在するとする．このとき，任意の $t \in \mathbb{R}^d$ に対して，$n, h \to \infty$ において $\hat{\theta}_n(t) \xrightarrow{P} \theta_0$ が成り立つ.

この二つの結果から，$\hat{\theta}_n$ と $\hat{\theta}_n(t)$ が $n, h \to \infty$ において，挙動が一致することが示される．本講演においては，Theorem 2 を得るために必要な Lemma について概説するとともに，$\hat{\theta}_n$ と $\hat{\theta}_n(t)$ の漸近挙動を確認するシミュレーション結果について報告を行った.

参考文献

[1] G. Raskutti and S. Mukherjee(2015), The Information Geometry of Mirror Descent. *IEEE Transactions on Information Theory*, 61(3), 1451-1457.

[2] A. W. van der Vaart and J. A. Wellner(1996), *Weak convergence and empirical processes.* Springer.

大規模複雑データの理論と方法論, 及び, 関連分野への応用

分散共分散行列の逆行列における縮小推定法の提案

中京大学 国際教養学部 国際教養学科 永井 勇

(講演日; 2017/12/3)

1 導入

n 個の個体から得られた $n \times p$ の多変量データ $\boldsymbol{Y}$ の分析を考える. このとき, 各行の真の分散共分散行列を $\boldsymbol{\Sigma}$ (未知, 正定値行列) とする. $\boldsymbol{\Sigma}$ の不偏推定量は $\boldsymbol{S} = \boldsymbol{Y}'\{\boldsymbol{I}_n - \mathbf{1}_n(\mathbf{1}_n'\mathbf{1}_n)^{-1}\mathbf{1}_n'\}\boldsymbol{Y}/(n-1)$ で得られる (Rencher & Christensen (2012) など参照). ここで, $\mathbf{1}_q$ は全ての要素が 1 の q 次元ベクトルである. 多変量データの分析の多くの場面では, 推定量やモデルの良さを測るために $\boldsymbol{S}^{-1}$ が必要なため, $\boldsymbol{S}$ が正定値行列であることを仮定している. しかしながら, $\boldsymbol{Y}$ の各列の相関が強い場合や標本数 n が次元数 p より小さい場合は, $\boldsymbol{S}^{-1}$ が不安定になることや $\boldsymbol{S}^{-1}$ がそもそも存在しないという問題が起きる. 本講演では, これらの問題について考える.

これらの問題に対して, $\boldsymbol{S}$ の対角成分だけを用いる手法が提案されている (Srivastava *et al.*, 2013). 一方で, Chanohara *et al.* (2017) などでは, $\boldsymbol{\Sigma}$ の Cholesky 分解に着目し, その分解に用いられる要素の推定の際に罰則を付ける推定法が用いられている. また, Kubokawa and Srivastava (2008), Chen *et al.* (2011) や Wang *et al.* (2015) ではリッジタイプの罰則を用いた推定法が提案されている. しかしながら, これらの罰則を用いた推定においては, 罰則パラメータの最適化のための反復計算が必要となることが多い.

そこで反復計算の回避のために, 多変量線形回帰モデルにおいては一般化リッジ回帰を用いた罰則付推定法が提案された (Nagai *et al.*, 2012). この手法を用いた場合, 最適な罰則パラメータが陽に求まり, さらに説明変数の固有値の中で不要なものに関する回帰係数の推定量をゼロに縮小するという特徴がある. 本講演では, この手法をリッジタイプの罰則を用いた $\boldsymbol{S}$ の推定法へ拡張し, 後述する二乗ロスなどの様々なロス関数において, 最適なパラメータが陽に求まることを示し, 特徴を述べる.

2 罰則付推定量

2.1 リッジタイプの罰則付推定量

上述したように, $\boldsymbol{S}^{-1}$ が不安定になることや存在しないという問題を解決するために, リッジタイプの罰則付推定量が提案されている. Wang *et al.* (2015) では, 二つのパラメータ $\lambda > 0$ と $\beta > 0$ を用いて $\lambda(\boldsymbol{S} + \beta\boldsymbol{I}_p)^{-1}$ という罰則付推定量が用いられ, λ と β の最適化が考えられている. また, $\lambda = 1$ とすると Chen *et al.* (2011) で用いられている推定量である. これらの推定量は, $\boldsymbol{S} + \beta\boldsymbol{I}_p$ の固有値は $\boldsymbol{S}$ の固有値よりも大きな値となるため, $\boldsymbol{S}$ の逆行列よりも $\boldsymbol{S} + \beta\boldsymbol{I}_p$ の逆行列のほうが安定する形となっている. また, Wang *et al.* (2015) の推定量は $\beta\boldsymbol{I}_p$ により $\boldsymbol{S}$ の逆行列を縮小し, λ でさらなる調整を行っている推定量と考えることができる.

これらのパラメータ λ や β の最適化を考えると, リッジ回帰による推定と同様に, 反復計算などが必要となる. そこで本講演では, 一般化リッジ回帰による推定と同様に, パラメータを増やすことでさらに柔軟な推定量を構築すると同時に反復計算を回避した推定法を提案する.

2.2　提案する一般化リッジタイプの罰則付推定量

リッジタイプの罰則付推定量において, パラメータ最適化のための反復計算が必要であるという問題を回避するために, 一般化リッジ回帰による推定と同様の手法を用いた推定量を本講演で提案する. つまり, 正のパラメータ λ と非負のパラメータ θ_i $(i=1,\ldots,p)$ を用いて次の推定量を提案する;

$$\hat{\boldsymbol{S}}^{-1}(\lambda,\boldsymbol{\theta}) \stackrel{\text{def.}}{=} \lambda(\boldsymbol{S}+\boldsymbol{Q}\boldsymbol{\Theta}\boldsymbol{Q}')^{-1},$$

ここで, $\boldsymbol{\Theta}=\mathrm{diag}(\boldsymbol{\theta})$, $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_p)'$, さらに $\boldsymbol{Q}$ は $\boldsymbol{S}$ の固有値 $d_1,\ldots,d_p$ $(d_i\geq 0,\ i=1,\ldots,p)$ を対角に並べた $\boldsymbol{D}$ を用いて, $\boldsymbol{Q}'\boldsymbol{S}\boldsymbol{Q}=\boldsymbol{D}$ となる直交行列である. この $\hat{\boldsymbol{S}}^{-1}(\lambda,\boldsymbol{\theta})$ を書き換えると, $\hat{\boldsymbol{S}}^{-1}(\lambda,\boldsymbol{\theta})=\lambda\boldsymbol{Q}(\boldsymbol{D}+\boldsymbol{\Theta})^{-1}\boldsymbol{Q}'$ となる. よって, この推定量 $\hat{\boldsymbol{S}}^{-1}(\lambda,\boldsymbol{\theta})$ は $\boldsymbol{D}$ の逆行列をそのまま考えるのではなく $\boldsymbol{D}+\boldsymbol{\Theta}$ の逆行列を考えることで, $\boldsymbol{S}^{-1}$ の不安定さなどを回避した推定量である. また, $\boldsymbol{\theta}=\beta\mathbf{1}_p'$ とすると Wang *et al.* (2015) などのリッジタイプの罰則付推定量と一致する. 本講演では, λ を固定した下で後述する二乗ロスなどのロス関数を最小にする $\boldsymbol{\theta}$ が陽に求まることを示す.

2.3　二乗ロス関数

λ を固定した下での $\boldsymbol{\theta}$ の最適化のために, Wang *et al.* (2015) などで推定量の良さを測るために使われる二乗ロス関数を考える. $\hat{\boldsymbol{S}}^{-1}(\lambda,\boldsymbol{\theta})$ の二乗ロスは, 次の形で定義される;

$$\mathrm{Loss}^{[2]}(\hat{\boldsymbol{S}}^{-1}(\lambda,\boldsymbol{\theta}))=\frac{1}{p}\mathrm{tr}\left\{\left(\hat{\boldsymbol{S}}^{-1}(\lambda,\boldsymbol{\theta})\boldsymbol{\Sigma}-\boldsymbol{I}_p\right)^2\right\}.$$

この関数を最小にする $\boldsymbol{\theta}$ を λ を固定した下で求めると, 陽に求まることを本講演で示す. また本講演では, $\boldsymbol{\Sigma}$ に $\boldsymbol{S}$ を代入したプラグイン型の最適化法を提案する. さらに, James and Stein (1961) などで使われているロス関数やそれぞれのロス関数を最小にする $\boldsymbol{\theta}$ の導出などについても触れる.

数値実験などによる比較など

他のロス関数を最小にする $\boldsymbol{\theta}$ との数値実験を通じた比較や特徴については当日の講演で報告する.

引用文献:

[1] Chanohara, N., Nakagawa, T. and Wakaki, H. (2017). Estimation of covariance matrix via shrinkage Cholesky factor. *Hiroshima Statistical Research Group: Technical Report*, TR 17-03.

[2] Chen, L. S., Paul, C., Prentice, R. L., and Wang, P. (2011). A regularized Hotelling's T^2 test for pathway analysis in proteomic studies. *J. Am. Stat. Assoc.*, **106**, 1345–1360.

[3] James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.*, **1**, 361–379.

[4] Kubokawa, T. and Srivastava, M. S. (2008). Estimation of the precision matrix of a singular Wishart distribution and its application in high-dimensional data. *J. Multivariate Anal*, **99**, 1906–1928.

[5] Nagai, I., Yanagihara, H. and Satoh, K. (2012). Optimization of ridge parameters in multivariate generalized ridge regression by plug-in methods. *Hiroshima Math. J.*, **42**, 301–324.

[6] Rencher, A. C. and Christensen, W. F. (2012). *Methods of Multivariate Analysis* (Third Edition).

[7] Srivastava, M. S., Katayama, S. and Kano, Y. (2013). A two sample test in high dimensional data. *J. Multivariate Anal.*, **114**, 349–358.

[8] Wang, C., Pan, G., Tong, T., and Zhu, L. (2015). Shrinkage estimation of large dimensional precision matrix using random matrix theory. *Statistica Sinica*, **25**, 993–1008.

多次元分割表の完全独立性検定における変換検定統計量の構築について

北海道教育大学・札幌　種市　信裕
北海道教育大学・釧路　関谷　祐里
数学利用研究所　外山　淳

1 はじめに: 多項分布モデルを想定した多次元分割表（M 次元分割表）において, 完全独立性帰無仮説の検定を考える. 検定統計量として, 対数尤度比統計量や Pearson X^2 統計量を含む ϕ-ダイバージェンス統計量の族を考える. 完全独立性帰無仮説のもとでの検定統計量の分布の漸近展開に基づく近似式を導出した. さらに, その近似式を用いて標本数が余り大きくない場合であっても, 極限カイ二乗分布への近似が良い変換統計量の構築をおこなった.

2 多次元分割表とモデル: ある整数 M $(M \geq 3)$ に対して, 標本数 n を固定した M 次元の $J_1 \times \cdots \times J_M$ 分割表多項分布モデルを考える. つまり, $(j_1, \ldots, j_M)$ $(j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$ セルにおける度数を $X_{j_1 \cdots j_M}$ ただし, $\sum_{j_1=1}^{J_1} \cdots \sum_{j_M=1}^{J_M} X_{j_1 \cdots j_M} = n$ とし, $X^* = (X_{1 \cdots 11}, \ldots, X_{1 \cdots 1J_M}, \ldots, X_{J_1 \cdots J_M})^T$ とすると X^* は多項分布 $\mathrm{Mult}_K(n, p^*)$ ただし, $K = \prod_{m=1}^{M} J_m$, $p^* = (p_{1 \cdots 11}, \ldots, p_{1 \cdots 1J_M}, \ldots, p_{J_1 \cdots J_M})^T$, $0 < p_{j_1 \cdots j_M} < 1 \quad (j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$ および $\sum_{j_1=1}^{J_1} \cdots \sum_{j_M=1}^{J_M} p_{j_1 \cdots j_M} = 1$ に従うとする.

3 完全独立性の帰無仮説と検定統計量: 完全独立性の帰無仮説と検定統計量を表すために次のように記号を定義する. $a_{\cdot(m,j_m)} = \sum_{j_1=1}^{J_1} \cdots \sum_{j_{m-1}=1}^{J_{m-1}} \sum_{j_{m+1}=1}^{J_{m+1}} \cdots \sum_{j_M=1}^{J_M} a_{j_1 \cdots j_m \cdots j_M}$ $(j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$, ただし $a_{j_1 \cdots j_m \cdots j_M}$ $(j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$ は M 個の添え数を持つ系列とする. 以下で用いられる記号 $p_{\cdot(m,j_m)}$ と $X_{\cdot(m,j_m)}$ は $a_{\cdot(m,j_m)}$ と同じ方法で定義されているものとする. 本報告においては, M 次元分割表の完全独立性帰無仮説

$$H_0^M : p_{j_1 \cdots j_M} = Q(j_1, \ldots, j_M)\ (j_m = 1, \ldots, J_m;\ m = 1, \ldots, M),$$

ただし, $Q(j_1, \ldots, j_M) = \prod_{m=1}^{M} p_{\cdot(m,j_m)}$ $(j_m = 1, \ldots, J_m;\ m = 1, \ldots, M)$ を考察した. 帰無仮説 H_0^M を検定するための ϕ-ダイバージェンスに基づく検定統計量 C_ϕ^M は

$$C_\phi^M = 2n \sum_{j_1=1}^{J_1} \cdots \sum_{j_M=1}^{J_M} \hat{Q}(j_1, \ldots, j_M) \phi\left(\frac{\hat{p}_{j_1 \cdots j_M}}{\hat{Q}(j_1, \ldots, j_M)} \right),$$

と与えられる, ただし, $\hat{Q}(j_1, \ldots, j_M) = \prod_{m=1}^{M} \hat{p}_{\cdot(m,j_m)}$, $\hat{p}_{j_1 \cdots j_M} = n^{-1} X_{j_1 \cdots j_M}$ $(j_m = 1, \ldots, J_m ;m = 1, \ldots, M)$, $\hat{p}_{\cdot(m,j_m)} = n^{-1} X_{\cdot(m,j_m)}$ $(j_m = 1, \ldots, J_m; m = 1, \ldots, M)$, $\phi(t)$ は $\phi(1) = \phi'(1) = 0$ および $\phi''(1) = 1$ を満たす $(0, \infty)$ で定義される実凸関数とする.

4 H_0^M のもとでの検定統計量の分布の近似: 帰無仮説 H_0^M のもとで, 検定統計量 C_ϕ^M は自由度 $L = K - \sum_{m=1}^{M} J_m + M - 1$ のカイ二乗分布を極限分布として持つ. Yarnold [6] は多項分布の適合度検定統計量としての Pearson X^2 統計量の単純帰無仮説のもとでの漸近展開に基づく近似を提案した. この近似の考え方をもとに Taneichi and Sekiya [4] は, $M = 2$ の場合, つまり 2 次元分割表における独立性検定統計量 C_ϕ^2 の漸近展開に基づく分布の近似 $P\{C_\phi^2 \leq z | H_0^2\} \approx K_1^\phi(z) + K_2^\phi(z)$ を与えた. ここで, $K_1^\phi(z)$ は多変量エッジワース展開による項, $K_2^\phi(z)$ は不連続性を考慮した離散項である. 本報告では, この近似を一般の M 次元分割表の完全独立性検定に拡張した. その結果, $\phi(t)$ が 6 回連続微分可能であるとき, K_1^ϕ 項は,

$$K_1^\phi(z) = P\{\chi_L^2 \leq z\} + n^{-1} \sum_{j=0}^{3} d_j^\phi P\{\chi_{L+2j}^2 \leq z\} + O(n^{-2}) \tag{1}$$

という形式で評価されることがわかった. ただし, χ^2_L は自由度 L のカイ二乗分布に従う確率変数を表す. また, 係数 $d^\phi_j (j=0,1,2,3)$ に関して $\sum_{j=0}^{3} d^\phi_j = 0$ が成り立つ.

５ ϕ-ダイバージェンスに基づく検定統計量の改良: 本報告では, 少ない標本数でも C^M_ϕ より速く極限分布に近づく検定統計量の構築を目指す. 導出された離散項 $K^\phi_2(z)$ の評価式は非常に複雑である. また, 報告者らが研究をおこなって来た, 多項モデル [3], ２次元分割表モデル [4], 一般化線型モデル [5] などにおいての検定統計量の分布の漸近展開に基づく近似の研究において, 離散項の省略はそれほど多くの誤差を生まないとの知見がある. これらのことより C^M_ϕ の分布に対して，(1) 式で与えられる $K^\phi_1(z)$ の近似式のみを用いて，小標本におけるカイ二乗近似の改良を考える. (1) 式にバートレット修正および改良変換の構築と漸近展開式との関係の理論 (e.g., Fujikoshi [2], Cordeiro and Ferrari [1]) を適用した. (1) 式において, $\phi'''(1) = -1$ かつ $\phi^{(4)}(1) = 2$ が成り立つ場合には $d^\phi_1 = -d^\phi_0$ かつ $d^\phi_2 = d^\phi_3 = 0$ が成り立つ. このことは, バートレット修正が可能なことを意味する. よってこの場合には, C^M_ϕ にバートレット修正を施した変換統計量 $(C^M_\phi)_B = \{1 + 2d^\phi_0 (nf)^{-1}\} C^M_\phi$ を構築できる. (1) 式においてその他の場合には, 改良変換統計量 ([2])

$$
\begin{aligned}
(C^M_\phi)_I \quad = \quad & (n\alpha+\beta)^2 \log\left[1 + \frac{1}{(n\alpha)^2}\left\{ C^M_\phi + \frac{1}{n\alpha}((C^M_\phi)^2 + \gamma (C^M_\phi)^3) \right.\right. \\
& \left.\left. + \frac{1}{(n\alpha)^2}\left(\frac{1}{3}(C^M_\phi)^3 + \frac{3\gamma}{4}(C^M_\phi)^4 + \frac{9\gamma^2}{20}(C^M_\phi)^5 \right) \right\}\right],
\end{aligned}
$$

ただし, $\alpha = -L(L+2)\{2(d^\phi_2 + d^\phi_3)\}^{-1}$, $\beta = -(L+2)d^\phi_0\{2(d^\phi_2 + d^\phi_3)\}^{-1}$, $\gamma = d^\phi_3\{(L+4)(d^\phi_2 + d^\phi_3)\}^{-1}$, および, バートレット型変換統計量 ([1])

$$
(C^M_\phi)_{CF} = \left\{ 1 + \frac{2d^\phi_0}{nL} + \frac{2(d^\phi_0 + d^\phi_1)}{nL(L+2)} C^M_\phi + \frac{2(d^\phi_0 + d^\phi_1 + d^\phi_2)}{nL(L+2)(L+4)} (C^M_\phi)^2 \right\} C^M_\phi
$$

を構築できる. これらの変換統計量は, 仮に (1) ではなく C^M_ϕ の漸近展開式の係数を用いたとすると, $P\{(C^M_\phi)_\xi \le z\} = P\{\chi^2_L \le z\} + O(n^{-2})$ $(\xi = B, I, CF)$ と評価される.

参考文献

[1] Cordeiro, G. M. and Ferrari, S. L. P.: *Biometrika*, **78**, (1991), 573–582.

[2] Fujikoshi, Y.: *Journal of Multivariate Anal.*, **72**, (2000), 249–263.

[3] Taneichi, N., Sekiya, Y. and Suzukawa, A.: *Journal of Multivariate Anal.,* **81**(2), (2002), 335–359.

[4] Taneichi, N. and Sekiya, Y.: *Journal of Multivariate Anal.*, **98**, (2007), 1630–1657.

[5] Taneichi, N., Sekiya, Y. and Toyama, J.: *Journal of Multivariate Anal.*, **123**, (2014), 311–329.

[6] Yarnold, J. K.: *Ann. Math. Statist.*, **43**, (1972), 1566–1580.

要約　なぜ癌の遺伝子解析は30年以上成功しなかったのか？

成蹊大学　名誉教授　新村秀一（しんむら　しゅういち）

〒277-0042 千葉県柏市逆井 1-8-7-301

sshinmura@gmail.com; Tel/Fax:04-7172-7493

Summary:　多くの世界中の研究者が、30年以上癌の遺伝子解析を行い明白な結果を得られなかった。しかし、筆者は2015年10月25日に富山市で開催された統計シンポジュームで米国の主要な6研究グループが、Microarrayデータを分析した論文を発表し、これらのデータが公開されていることを知った。28日にダウンロードしShippらのデータを改定IP-OLDF（RIP）で判別すると、最小誤分類数(Minimum Number of Misclassifications, MNM)が0である上に、m僅か32個の遺伝子の係数だけが0でないことが分かった。すなわちLASSOを含め、多くの研究で種々のFeature Selectionが研究されているが、何もしないで自然に癌遺伝子の選択が行えたことになる。さらにこれらの遺伝子を全体から省いて判別すると、また少数の組の遺伝子がMNM=0であることが分かった。最終的にこれを繰り返すことで、遺伝子空間は少数個の癌遺伝子を含む少数個のSmall Matryoshka(SM)と呼ぶ排他的な和集合と、残りの部分空間はMNMが1以上の和集合に分離されることが分かった。そこでMatryoshuka Feature Selection Method（Method 2)を開発した。そして、数理計画法ソフトのLINGOで6種類全てのSMを12月20日に見つけた。すなわち、僅か54日でこの問題（問題5）を解決したことになる。癌の遺伝子空間は高次元であり、信号と雑音が混じっているため統計分析は困難であるといわれている。そして、これを分離するための工学的な種々のフィルタリング手法が提案されている。LINGOでMethod2を行えば、このフィルタリングも自然に簡単に行える。

　何故、癌の遺伝子解析は非常に容易であるにもかかわらず、30年以上かけても良い結果がでず、多くの研究者が永遠に無理と考えるようになっていたのか？筆者は、それは単に統計的判別関数が全く役に立たないためであったと考える。これらを、“Small N Large P Problem”” NP-Hard”” 多重解の問

題”、“雑音を含む高次元データの困難さ”等のキーワードを取り上げて、説明したい。