

2020 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：最前線の動向と新たな展開

科学研究費補助金 基盤研究 (A) 20H00576 「大規模複雑データの理論と方法論の革新的展開」(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) 19K22837 「超高次元データによる個別化モデリングへの挑戦」(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)
石井 晶 (東京理科大学)

記

日時: 2020 年 12 月 4 日 (金) ~ 5 日 (土)

場所: つくば国際会議場 大会議室 102 (<https://www.epochal.or.jp/ja/>)

〒305-0032 茨城県つくば市竹園 2-20-3

プログラム

12 月 4 日 (金)

10:00 ~ 10:10 開会

10:10 ~ 10:50 中山 優吾 (京都大・情報学研究科)

矢田 和善 (筑波大・数理物質系)

青嶋 誠 (筑波大・数理物質系)

高次元カーネル主成分分析に基づく異常値の検出

11:00 ~ 11:40 藤澤 健吾 (東京理科大・理工学研究科)

田畑 耕治 (東京理科大・理工学部)

順序カテゴリをもつ二元分割表における独立性の必要十分条件

11:50 ~ 12:30 庄 建倉 (統計数理研究所)

Model Improvement Through Residual Analysis

12:30 ~ 14:00 昼食

14:00 ~ 14:40 永井 勇 (中京大・教養教育研究院)

rank 落ちを回避する新たな推定法とその応用

14:50 ~ 15:30 柳本 武美 (統計数理研究所)

ベイズ型対数尤度に基づくモデルの信用集合

15:45 ~ 16:25 竹内 努 (名古屋大・理学研究科)

Analysis of Spatially Resolved Galaxy Spectra as a High-Dimensional Low-Sample Size Data Problem

16:35 ~ 17:15 竹内 努 (名古屋大・理学研究科)

Topological Data Analysis of Galaxy Distribution: Detection of the Birth Cry of Cosmic Structures

12月5日(土)

9:00 ~ 9:40 矢田 和善 (筑波大・数理物質系)

青嶋 誠 (筑波大・数理物質系)

Sparse PCA by the noise-reduction methodology

9:50 ~ 10:30 沈 迅 (東京農工大・工学部)

庄 建倉 (統計数理研究所)

Decimating Nonlinear Response in State Space Models for Battery Capacity Estimation

10:40 ~ 11:20 地寄 頌子 (大阪工業大・情報科学部)

宮本 暢子 (東京理科大・理工学部)

藤原 良叔 (筑波大学)

二部グラフ構造を持つ処理集合の実験計画とその深層学習への応用

11:35 ~ 12:15 大久保 祐作 (統計数理研究所)

マクロ進化の大規模種間比較研究における枝特異的方向性淘汰モデルの提案

12:25 ~ 13:05 島谷 健一郎 (統計数理研究所)

2次林再生における林冠個体の15年間変遷を観る統計モデル

13:05 ~ 13:10 閉会

高次元カーネル主成分分析に基づく異常値の検出

京都大学・情報学研究科 中山 優吾 (Yugo Nakayama)
筑波大学・数理物質系 矢田 和善 (Kazuyoshi Yata)
筑波大学・数理物質系 青嶋 誠 (Makoto Aoshima)

1 はじめに

本講演では、高次元小標本データに対する異常値の検出を考えた。高次元データの異常値の検出について、マハラノビス距離に基づく異常値の検出が盛んに研究されている。しかしながら、高次元小標本において標本共分散行列の逆行列が存在しないため、マハラノビス距離を単純に用いることができない。本講演では、高次元カーネル PCA を用いたクラスタリングを応用して、高次元小標本データに対して異常値検出を考えた。

2 主成分スコアにおける高次元一貫性

2 クラスの高次元クラスタリングを考えることで異常値の検出を提案した。2つの d 次元分布を Π_1, Π_2 と名付け、それぞれ平均 $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ と、共分散行列 $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ をもつと仮定する。ここで、 $\text{tr}(\boldsymbol{\Sigma}_1) \leq \text{tr}(\boldsymbol{\Sigma}_2)$ とする。それらの混合分布から $n (\geq 2)$ 個のデータを無作為に抽出し、データ行列を $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ とする。ただし、 $n_i = \#\{j | \mathbf{x}_j \in \Pi_i, j = 1, \dots, n\}$ とおく。ここで、 $\#A$ は集合 A の要素の個数とし、 $n_i \geq 1$ とする。

本節では、線形カーネル

$$k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$$

を用いたカーネル PCA を考える。 \mathbf{I}_n を n 次の単位行列、 $\mathbf{1}_n = (1, \dots, 1)^T$ として、 $\mathbf{P}_n = \mathbf{I}_n - n^{-1} \mathbf{1}_n \mathbf{1}_n^T$ とする。カーネル関数 $k(\mathbf{x}_j, \mathbf{x}_{j'})$ を (j, j') 成分にもつグラム行列 \mathbf{K} を考え、その中心化グラム行列を $\mathbf{K}_0 = \mathbf{P}_n \mathbf{K} \mathbf{P}_n$ とし、 \mathbf{K}_0 の固有値分解を

$$\mathbf{K}_0 = \sum_{i=1}^{n-1} \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i^T \quad (\hat{\mathbf{u}}_i = (\hat{u}_{i1}, \dots, \hat{u}_{in})^T, \quad \|\hat{\mathbf{u}}_i\|^2 = 1)$$

とする。 \mathbf{x}_j の (基準化した) 第 i 主成分スコアを $s_{ij} = \sqrt{n} \hat{u}_{ij}$ とおく。次を仮定する。

(A-i) $\text{Var}(\|\mathbf{x} - \boldsymbol{\mu}_i\|^2 | \mathbf{x} \in \Pi_i) = O\{\text{tr}(\boldsymbol{\Sigma}_i^2)\}$ as $d \rightarrow \infty$ for $i = 1, 2$.

Π_i が正規分布ならば、 $\text{Var}(\|\mathbf{x} - \boldsymbol{\mu}_i\|^2 | \mathbf{x} \in \Pi_i) = 2\text{tr}(\boldsymbol{\Sigma}_i^2)$ となるので、(A-i) は正規性の条件を緩めた仮定であることに注意する。さらに、 $\Delta_\mu = \|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2$ 、 $\Delta_\Sigma = |\text{tr}(\boldsymbol{\Sigma}_1) - \text{tr}(\boldsymbol{\Sigma}_2)|$ とおく。次を仮定する。

(A-ii) $\text{tr}(\boldsymbol{\Sigma}_i^2) / \Delta_\mu^2 = o(1)$ as $d \rightarrow \infty$ for $i = 1, 2$.

そのとき、次の定理が成り立つ。

定理 1 (Nakayama et al. (2020)). (A-i) と (A-ii) を仮定する. さらに, 次を仮定する.

$$(A\text{-iii}) \limsup_{d \rightarrow \infty} \frac{\Delta_{\Sigma}}{n_1 \Delta_{\mu}} < 1 \text{ when } n_2 \geq 2.$$

$d \rightarrow \infty$ のとき次が成り立つ.

$$s_{1j} = \begin{cases} \sqrt{n_2/n_1} + o_P(1) & \text{when } j = 1, \dots, n_1, \\ -\sqrt{n_1/n_2} + o_P(1) & \text{when } j = n_1 + 1, \dots, n. \end{cases} \quad (1)$$

注意 1. Yata and Aoshima (2019) は 定理 1 と同様の結果を $\Delta_{\Sigma}/\Delta_{\mu} = o(1)$ のもとで与えた. (A-iii) は $\Delta_{\Sigma}/\Delta_{\mu} = o(1)$ よりも緩い条件であることに注意する. $n_2 = 1$ のとき, (A-i) と (A-ii) のもとで (1) が成り立つ.

注意 2. もし $\text{tr}(\Sigma_1) > \text{tr}(\Sigma_2)$ なら, (A-i), (A-ii) と次の条件のもとで (1) が成り立つ

$$\limsup_{d \rightarrow \infty} \Delta_{\Sigma}/(n_2 \Delta_{\mu}) < 1 \text{ when } n_1 \geq 2.$$

定理 1 から (A-iii) のもとで, 第 1 主成分スコアの符号で高次元データを分類できることがわかる. しかし $\Delta_{\mu}/\Delta_{\Sigma}$ が小さい場合, 線形カーネルでは分類することはできず, 非線形カーネルを使う必要がある. 例えば, ガウシアンカーネル

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/\gamma) \quad (\gamma > 0)$$

を用いたカーネル PCA では, $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ ($\Delta_{\mu} = 0$) のときでさえも共分散行列間の距離によって一致性 (1) を与えることができる. Nakayama et al. (2020) は高次元におけるカーネル主成分スコアの漸近的性質を導出し, 平均ベクトル間の距離 Δ_{μ} だけでなく, 共分散行列間の距離 Δ_{Σ} にもよって (1) のような主成分スコアの高次元一致性をもつことを示した.

本講演では, 主成分スコアの一致性 (1) を異常値検出に応用した. 例えば, 異常値が 1 つと想定できる場合は, $n_1 = 1$ として考えることで (1) より異常値の検出が可能である. 当日は, 一般のカーネル関数の場合と異常値が複数ある場合についても言及し, 主成分スコアの符号に基づき異常値検出が可能であることを理論的に示し, その性能を数値実験と実データ解析を用いて確認した.

References

- [1] Nakayama, Y., Yata, K. and Aoshima, M. (2020). Clustering by principal component analysis with Gaussian kernel in high-dimension, low-sample-size settings, submitted.
- [2] Yata, K. and Aoshima, M. (2019). Geometric consistency of principal component scores for high-dimensional mixture models and its application. *Scand. J. Stat.*, 47: 899–921.

順序カテゴリをもつ二元分割表における 独立性の必要十分条件

東京理科大学大学院 藤澤 健吾
東京理科大学 田畑 耕治

1. 二元分割表におけるモデル

順序カテゴリをもつ $r \times c$ 分割表において, (i, j) セル確率を π_{ij} ($i = 1, \dots, r; j = 1, \dots, c$) とする. 独立 (I) モデルは次のように表される:

$$\pi_{ij} = \pi_{i+}\pi_{+j} \quad (i = 1, \dots, r; j = 1, \dots, c),$$

ただし, $\pi_{i+} = \sum_{t=1}^c \pi_{it}$, $\pi_{+j} = \sum_{s=1}^r \pi_{sj}$ である.

I モデルの当てはまりが悪い場合, I モデルよりも制約の弱いモデルを当てはめることに関心がある. Goodman (1979) は一様連関 (U) モデルを次のように提案した:

$$\pi_{ij} = \mu\alpha_i\beta_j\theta^{ij} \quad (i = 1, \dots, r; j = 1, \dots, c).$$

U モデルは $\theta = 1$ のとき, I モデルである. また Agresti (1983) は線形-線形連関 (LL) モデルを次のように提案した:

$$\pi_{ij} = \mu\alpha_i\beta_j\theta^{u_iv_j} \quad (i = 1, \dots, r; j = 1, \dots, c),$$

ただし $\{u_i\}$ は $u_1 < \dots < u_r$ を満たす行カテゴリに割り振られた既知のスコアであり, $\{v_j\}$ は $v_1 < \dots < v_c$ を満たす列カテゴリに割り振られた既知のスコアである.

$p = (p_{ij})$, $q = (q_{ij})$ を離散確率分布とする. 確率分布 p と q に対する f -ダイバージェンスは次のように表される (Csiszár and Shields, 2004):

$$I(p||q) = \sum_i \sum_j q_{ij} f\left(\frac{p_{ij}}{q_{ij}}\right),$$

ただし f は $f(1) = 0$, $f(0) = \lim_{t \rightarrow 0} f(t)$, $0 \cdot f(0/0) = 0$, $0 \cdot f(a/0) = a \lim_{t \rightarrow \infty} (f(t)/t)$ を満たす $(0, \infty)$ 上の凸関数である.

f を 2 回微分可能な狭義凸関数とする. Kateri and Papaioannou (1994) は f -ダイバージェンスに基づく線形-線形連関 (LL[f]) モデルを次のように提案した:

$$\pi_{ij} = \pi_{i+}\pi_{+j}F^{-1}(\alpha_i + \beta_j + \theta u_i v_j) \quad (i = 1, \dots, r; j = 1, \dots, c),$$

ただし, $F(t) = f'(t)$ である. $f(t) = t \log t$ のとき, LL[f] モデルは LL モデルである.

3. 独立性の必要十分条件

行スコアに対する確率変数を U , 列スコアに対する確率変数を V とする. このとき, $E(U) = \sum_i u_i \pi_{i+}$, $E(V) = \sum_j v_j \pi_{+j}$ である. Tomizawa *et al.* (2008) は独立性の必要十分条件に関する次の定理を与えた:

定理 1 独立性が成立するための必要十分条件は LL モデルと $Cov(U, V) = 0$ が同時に成立することである.

本報告では LL モデルの一般化である LL[f] モデルを用いて次の定理を与えた:

定理 2 独立性が成立するための必要十分条件は LL[f] モデルと $Cov(U, V) = 0$ が同時に成立することである.

3. モデルの直交性

n_{ij} を (i, j) セル観測度数, m_{ij} を (i, j) セル期待度数とする. また, $\{n_{ij}\}$ は総観測度数 $n = \sum_i \sum_j n_{ij}$ の多項分布に従うことを仮定する. モデル M に対する尤度比カイ二乗検定統計量は次のように表される:

$$G^2(M) = 2 \sum_{i=1}^r \sum_{j=1}^c n_{ij} \log \left(\frac{n_{ij}}{\hat{m}_{ij}} \right),$$

ただし, \hat{m}_{ij} は m_{ij} の最尤推定量である.

行スコア $\{u_i\}$ と列スコア $\{v_j\}$ に

$$E(U) = E(V) = 0$$

を仮定する. このとき次の定理を与えた:

定理 3 独立モデルのもとで, $G^2(I)$ は $G^2(LL[f])$ と $G^2(Cov(U, V) = 0)$ の和に漸近的に同等である.

参考文献

- [1] Agresti, A. (1983). A survey of strategies for modeling cross-classifications having ordinal variables. *Journal of the American Statistical Association*, **78**, 184–198.
- [2] Csiszár, I. and Shields, P. C. (2004). *Information Theory and Statistics: A tutorial*. Now Publishers, Hanover.
- [3] Goodman, L. A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, **74**, 537–552.
- [4] Kateri, M. and Papaioannou, T. (1994). f -divergence association models. *International Journal of Mathematical and Statistical Sciences*, **3**, 179–203.
- [5] Tomizawa, S., Miyamoto, N., and Sakurai, M. (2008). Decomposition of independence model and separability of its test statistic for two-way contingency tables with ordered categories. *Advances and Applications in Statistics*, **8**, 209–218.

Model improvement through residual analysis

Jiancang Zhuang

The Institute of Statistical Mathematics, Tokyo, Japan

Statistical modelling has been increasingly and widely used in many research fields. Associated statistical inference techniques, such as model selection, specification, and evaluation, parameter estimation, and testing goodness-of-fit have been developed. Among them residual analysis can be used to evaluate a model and assess the goodness-of-fit without proposing unnecessary new models whose implementation may include heavy programming and computation tasks (Zhuang, 2006; Baddeley et al., 2005). Residual analysis is also helpful for understanding the advantages of a model even if its overall performance is not ideal.

More recently, general principles of first-order residuals for spatial point processes (Papangelou intensity-based models), temporal/spatiotemporal point processes and moment intensity-based models were proposed by Baddeley et al. (2005), Zhuang (2006), and Guan et al. (2008), respectively. Zhuang (2015) summarized the general principles for constructing innovation based residuals. For hidden Markov models, Buckby et al. (2020) developed a similar innovation based residuals by extending the framework of residual analysis in Zhuang (2015).

1 Residual analysis for point processes

A *point process* N is a random counting measure on \mathbb{R}^d satisfying that $N(A)$ for every bounded set $A \subset \mathbb{R}^d$. In this article, all point processes are assumed to be simple; that is, each point process can be represented by a sum of unit Dirac measures at distinct locations.

For a point process N on \mathbb{E} , its *first-order moment intensity*, provided that it exists, is defined on \mathbb{E} by

$$\mu(x) = \lim_{\delta \downarrow 0} \frac{N(B(x, \delta))}{\ell(B(x, \delta))}. \quad (1)$$

where ℓ is the Lebesgue measure, and $B(x, \delta)$ represents the ball centered at x and with radius δ . The following fact holds.

Conditional intensities When a temporal axis is embedded in \mathbb{R}^d , i.e., the other $d - 1$ dimension are regarded as spatial directions or marks. The conditional intensity function $\lambda(t, x)$, $t \in \mathbb{R}$ and $x \in \mathbb{R}^{d-1}$, is defined as

$$\lambda(t, x) = \lim_{\delta_t \downarrow 0, \delta_x \downarrow 0} \frac{\mathbf{E}[N([t, t + \delta_t] \times B(x, \delta_x)) \mid \mathcal{H}_t]}{\delta_t \ell(B(x, \delta_x))} \quad (2)$$

where \mathcal{H}_t represent the observational history of N up to t but not including t .

Papangelou intensity The Papangelou intensity is defined by, if it exists,

$$\lambda_p(x) = \lim_{\delta_x \downarrow 0} \frac{\mathbf{E}[N(B(x, \delta_x)) \mid \mathcal{E}(B(x, \delta_x))]}{\delta_x \ell(B(x, \delta_x))} \quad (3)$$

where $\mathcal{E}(B(x, \delta_x))$ represent the observation on N outside of the close ball $B(x, \delta_x)$.

The *first-order residuals* for moment intensity models, conditional intensity models, and Papangelou intensity models are defined by

$$R_m(B, f) = \int_B [\hat{f}(x) N(dx) - \hat{f}(x) \hat{\mu}(x) \ell(dx)], \quad (4)$$

$$R_h(B, f) = \int_B [\hat{f}(x) N(dx) - \hat{f}(x) \hat{\lambda}(x) \ell(dx)], \quad (5)$$

$$R_p(B, f) = \int_B [\hat{f}(x) N(dx) - \hat{f}(x) \hat{\lambda}_p(x) \ell(dx)], \quad (6)$$

respectively, where B is a Borel set, and f is a deterministic function in (4), a predictable function in (5), and an exvisible function in (6). If f also depends on the estimated parameters, then \hat{f} is obtained by substituting the estimated parameters in f . In practice, it is reasonable to assume that

$$R_{\{m, h, p\}}(B, f) \approx 0, \quad (7)$$

when the estimated model is a good approximation of the true model.

2 Residual analysis for Hidden Markov models

Give an HMM whose the observations are realised from the m members of a family of distributions, $f(x_t | S_t = i)$, with $i = 1, \dots, m$, where x_t is the observed value at time $t = 1, \dots, T$, S_t is the state of the Markov chain at time t and m is the finite number of states with a stationary transitive matrix

$$\Gamma_{ij} = P(S_t = j | S_{t-1} = i). \quad (8)$$

In an HMM, the forecasted distribution of X_t , given the previous observations is,

$$f(x_t | \mathcal{H}_t) = f(x_t | X_1, \dots, X_{t-1}) = \sum_{i=1}^m \frac{\sum_{k=1}^m \alpha_{t-1}(k) \Gamma_{ki} f(x_t | S_t = i)}{\sum_{k=1}^m \alpha_{t-1}(k)}, \quad (9)$$

where $\alpha_t(k)$ is the forward probability defined by

$$\alpha_t(k) = P(X_1, \dots, X_t, S_t = k) \quad (10)$$

Using the expectation calculated above, the predictive residual with a predictable function $h(t, x)$, where h is a predictable function of t for every fixed x , becomes,

$$R_n = \sum_{t=1}^n (h(t, X_t) - \int h(t, x_t) f(x_t | \mathcal{H}_t) dx_t). \quad (11)$$

For the exvisive residuals, we consider the forecasted distribution of X_t given all of the other observations rather than observations prior to time t ,

$$f(x_t | \mathcal{E}_t) = \sum_{i=1}^m \frac{\sum_{k=1}^m \alpha_{t-1}(k) \Gamma_{ki} f(x_t | S_t = i) \beta_t(i)}{\sum_{k=1}^m \alpha_{t-1}(k) \Gamma_{ki} \beta_t(i)} \quad (12)$$

where $\beta_t(i) = P(X_{t+1}, \dots, X_T | S_t = i)$ is the backward probability. Given a function $h(t, x)$ which is exvisible as a function of t for every x , the exvisible residual becomes

$$R_n^e = \sum_{t=1}^n \left[h(t, X_t) - \int h(t, x_t) f(x_t | \mathcal{E}_t) dx_t \right]. \quad (13)$$

It is easy to see $\mathbf{E}[R_n] = 0$ and $\mathbf{E}[R_n^e] = 0$.

3 Summary

Based on the theory of residual analysis for point processes and hidden Markov models, the stochastic reconstruction technique is illustrated with two examples: stochastic reconstruction of a Hawkes point process model for the robbery related violence in Castellon, Spain, and residual analysis to the results from fitting a zero-inflation HMM to the tremor data in the Kii region, Japan. Residual analysis helps us to find a proper direction in which to improve on the model formulation, in a data-driven manner.

References

- Baddeley, A., Turner, R., Møller, J., and Hazelton, M. (2005). Residual analysis for spatial point processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(5):617–666.
- Buckby, J., Wang, T., Zhuang, J., and Obara, K. (2020). Model checking for hidden Markov models. *Journal of Computational and Graphical Statistics*, 0(0):1–16.
- Guan, Y., Waagepetersen, R., and Beale, C. M. (2008). Second-order analysis of inhomogeneous spatial point processes with proportional intensity functions. *Journal of the American Statistical Association*, 103(482):769–777.
- Zhuang, J. (2006). Second-order residual analysis of spatiotemporal point processes and applications in model evaluation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(4):635–653.
- Zhuang, J. (2015). Weighted likelihood estimators for point processes. *Spatial Statistics*, 14(B):166–178.

rank 落ちを回避する 新たな推定法とその応用

※まだアイデア段階の話

永井 勇†

† 中京大学 教養教育研究院

2020年12月4日
大規模複雑データの理論と方法論：
最前線の動向と新たな展開

1 / 34

ここからの話

- モデルと推定量
 - モデルと最小二乗推定量など
 - 問題点と解決法
- 提案するモデルと最小二乗推定量など
 - リスク関数の変形
 - 提案するモデルでの推定量と前提条件
 - 利点と問題点
 - 問題点に関する考察と現状
 - 問題点について (前半)
 - 問題点について (後半)
- まとめなど
 - 提案するモデルと推定量と利点

4 / 34

rank(A) < k の時の推定

- Ξ の推定量をどう求めるか？
- Moore-Penrose 型一般逆行列を使う
 $\hat{\Xi} = (A'A)^+ A'Y$
 ((M)⁺: M の Moore-Penrose 型一般逆行列 (see e.g., Schott, 2005, Sec. 5.2))
- 罰則付推定法を用いる
 $\hat{\Xi} = \arg \min \{ \text{Risk}(\hat{\mu}, \Xi) + \text{Pen}(\Xi|\lambda) \}$
 例 Lasso 型の罰則付推定 (Tibshirani, 1996)
 $p = 1$ の場合、 $\text{Pen}(\Xi|\lambda) = \lambda \sum_{i=1}^k |\Xi_{(i)}|$
 (Ξ_(i) は Ξ の第 i 成分)

⇒ もっとシンプルに

7 / 34

Risk(μ, Ξ) の変形 (1/3)

- Y = 1_nμ' + ∑_{i=1}^r A_iΞ_i + ε のリスク
- Risk(μ, Ξ) = tr{ (Y - 1_nμ' - AΞ)Σ⁻¹(Y - 1_nμ' - AΞ)' }
 = tr{ (Y - 1_nμ' - ∑_{i=1}^r A_iΞ<sub>i})Σ⁻¹(Y - 1_nμ' - ∑_{i=1}^r A_iΞ<sub>i})' }
 $\stackrel{\text{def.}}{=} \text{Risk}(\mu, \Xi_1, \dots, \Xi_r)$
 ⇒ (μ̂, Ξ̂₁, ..., Ξ̂_r) = arg min_{μ, Ξ₁, ..., Ξ_r} Risk(μ, Ξ₁, ..., Ξ_r) を求める</sub></sub>

注意 A₁'1_n = 0_k より、書き換えたモデルでも μ̂ は前と同じ
 ⇒ Ξ₁, ..., Ξ_r に関する項に着目

10 / 34

rank(A_i) = k_i でどうやって推定するか

- Ξ̂_i = (A_i'A_i)⁻¹ A_i' (Y - ∑_{j>i}^r A_jΞ̂_j) (i = 1, ..., r) をどうやって推定していくか
- Ξ̂₁, ..., Ξ̂_r を求める方法
 - Ξ̂_r = (A_{r}'A_r)⁻¹ A_{r}'Y を求める}}
 - Ξ̂_{r-1} = (A_{r-1}'A_{r-1})⁻¹ A_{r-1}'(Y - A_rΞ̂_r) を求める}}
 - 同様に Ξ̂₁ まで

rank(A_i) = k_i (i = 1, ..., r) はどれくらい緩いのか？

13 / 34

提案する推定量の利点

- rank(A_i) に関する前提条件はほぼ皆無
- n < k = ∑_{i=1}^r k_i でも推定可能
 - 逆行列を用いて推定量が得られるための前提条件は n ≥ max(k₁, ..., k_r)
 - r = k とすると n ≥ 2
 - Moore-Penrose 型一般逆行列・罰則付推定を用いず rank(A) < k でも推定できる
- 相関の高い A の列の組が別々の A_i になるように分割すれば安定した推定量に
 - 相関の低い変数の組ごとに A を分割
- 分割+罰則付推定でグループごとに罰則付加

16 / 34

概要

- 多変量線形回帰モデル
- ⇒ 未知の行列に対してよく使われる推定量；最小二乗推定量
- よく置かれる仮定；説明変数行列が列フルランク
- ← サンプル数 ≥ 説明変数行列の列数
- サンプル数 < 説明変数行列の列数の場合は？
 - Moore-Penrose 型一般逆行列を用いる (see e.g., Schott, 2005, Sec. 5.2)
 - 罰則付推定を用いる (Lasso 回帰など, Tibshirani, 1996)
 - 今回；上の二つを使わない推定法の提案
 → シンプルな形への変形など

2 / 34

モデルと仮定

モデル 多変量線形回帰モデル

$$Y = 1_n \mu' + A \Xi + \mathcal{E}$$

$$\begin{cases} Y; & n \times p \text{ 目的変数行列} \\ 1_n; & n \text{ 次元の 1 ベクトル} \\ \mu; & p \text{ 次元未知ベクトル} \\ A; & n \times k \text{ 説明変数行列} \\ \Xi; & k \times p \text{ 未知回帰係数行列} \\ \mathcal{E}; & n \times p \text{ 誤差行列} \end{cases}$$

仮定 A₁'1_n = 0_k, n ≥ 2, A の各列 ≠ 0_n, 0_r; r 次元ゼロベクトル, E[ε] = 0_n, 0_r, Cov[vec(ε)] = Σ ⊗ I_n, Σ は p × p 未知正定値行列
 ⇒ μ, Ξ の推定

5 / 34

ここからの話

- モデルと推定量
 - モデルと最小二乗推定量など
 - 問題点と解決法
- 提案するモデルと最小二乗推定量など
 - リスク関数の変形
 - 提案するモデルでの推定量と前提条件
 - 利点と問題点
 - 問題点に関する考察と現状
 - 問題点について (前半)
 - 問題点について (後半)
- まとめなど
 - 提案するモデルと推定量と利点

8 / 34

Risk(μ, Ξ) の変形 (2/3)

- Risk(μ, Ξ₁, ..., Ξ_r) の Ξ₁, ..., Ξ_r に関する項
 $-2 \sum_{i=1}^r \text{tr}(A_i' Y \Sigma^{-1} \Xi_i) + \sum_{i=1}^r \sum_{j>i}^r \text{tr}(A_i \Xi_i \Sigma^{-1} \Xi_j' A_j')$
 ⇒ Ξ_i のために Ξ_j (j ≠ i) が全部必要
- tr(A_iΞ_iΣ⁻¹Ξ_j'A_j') = tr(A_jΞ_jΣ⁻¹Ξ_i'A_i') より
 上の式 (Risk(μ, Ξ₁, ..., Ξ_r) の Ξ₁, ..., Ξ_r に関する項)
 $= -2 \sum_{i=1}^r \text{tr}(A_i' Y \Sigma^{-1} \Xi_i) + \sum_{i=1}^r \text{tr}(A_i \Xi_i \Sigma^{-1} \Xi_i' A_i')$
 $+ 2 \sum_{i=1}^r \sum_{j>i}^r \text{tr}(A_i \Xi_i \Sigma^{-1} \Xi_j' A_j')$ (ここで ∑_{j>i} (*) = 0)

11 / 34

rank(A_i) = k_i はどれくらい緩い？

- rank(A_i) = k_i (i = 1, ..., r) はどれくらい緩い仮定か？
- rank(A) = k のための前提条件

$$n \geq k = \sum_{i=1}^r k_i$$
- rank(A_i) = k_i (i = 1, ..., r) のための前提条件

$$n \geq k_i \quad (i = 1, \dots, r)$$

$$\Leftrightarrow n \geq \max(k_1, \dots, k_r)$$

14 / 34

提案する推定量の問題点

- 推定量の代入を繰り返す形
 - Ξ_r → Ξ_{r-1} → ... → Ξ₂ → Ξ₁
- rank(A_i) = k_i でも E[Ξ̂_i] ≠ Ξ_i
- E[Ξ̂] = Ξ (rank(A) = k の場合)
- A₁, ..., A_r の分割に依存
 - 変数選択での組み合わせ問題と同様
- 罰則付推定の場合、パラメータの最適化をどうする？
 - 多変量一般化リッジ回帰による推定 (Yanagihara, 2009) を使った場合、罰則パラメータが陽に求まるのか？

⇒ これらの問題点についての考察

17 / 34

目次

- モデルと推定量
 - モデルと最小二乗推定量など
 - 問題点と解決法
- 提案するモデルと最小二乗推定量など
 - リスク関数の変形
 - 提案するモデルでの推定量と前提条件
 - 利点と問題点
 - 問題点に関する考察と現状
 - 問題点について (前半)
 - 問題点について (後半)
- まとめなど
 - 提案するモデルと推定量と利点

3 / 34

最小二乗推定量など

- μ, Ξ の最小二乗推定量; μ̂, Ξ̂, (μ̂, Ξ̂) = arg min_{μ, Ξ} Risk(μ, Ξ), Risk(μ, Ξ) $\stackrel{\text{def.}}{=} \text{tr}\{(Y - 1_n \mu' - A \Xi) \Sigma^{-1} (Y - 1_n \mu' - A \Xi)'\}$.
- nμ̂ = Y'1_n, A'AΞ̂ = A'Y の解
- n ≥ 2 より μ̂ = Y'1_n/n, rank(A) = k ならば Ξ̂ = (A'A)⁻¹ A'Y.
- rank(A) = k のための前提条件

$$n \geq k$$
- rank(A) < n < k の場合の Ξ の推定量は？

6 / 34

もっとシンプルに考える

- rank(A) < k のときに Ξ の推定量をどう求めるか？ (Moore-Penrose 型一般逆行列や罰則付推定を使わない方法)
- 再掲
 モデル; Y = 1_nμ' + AΞ + ε
 ● A = (A₁, ..., A_r), Ξ = (Ξ₁, ..., Ξ_r)', k ≥ r, A<sub>i}; n × k_i 説明変数行列, Ξ<sub>i}; k_i × p 未知行列,
 $k = \sum_{i=1}^r k_i, k_i \geq 1, A_1'1_n = 0_{k_1}, (\dots, A_r'1_n = 0_{k_r}).$
 ⇒ Y = 1_nμ' + ∑_{i=1}^r A_iΞ_i + ε
 ● Ξ₁, ..., Ξ_r の推定 ⇔ Ξ の推定</sub></sub>

9 / 34

Risk(μ, Ξ) の変形 (3/3)

- Risk(μ, Ξ₁, ..., Ξ_r) の Ξ₁, ..., Ξ_r に関する項
 $-2 \sum_{i=1}^r \text{tr}(A_i' Y \Sigma^{-1} \Xi_i) + \sum_{i=1}^r \text{tr}(A_i \Xi_i \Sigma^{-1} \Xi_i' A_i')$
 $+ 2 \sum_{i=1}^r \sum_{j>i}^r \text{tr}(A_i \Xi_i \Sigma^{-1} \Xi_j' A_j')$ の最小化
 ⇒ A_i'A_iΞ̂_i = A_i' (Y - ∑_{j>i}^r A_jΞ̂_j) の解 (i = 1, ..., r)
- rank(A_i) = k_i なら Ξ̂_i = (A_i'A_i)⁻¹ A_{i}' (Y - ∑_{j>i}^r A_jΞ̂_j) (i = 1, ..., r)}

12 / 34

条件の緩さの極端な例

- n = 2 でも推定可能
- 従来 (A'A)⁺ を使用, 罰則付推定を使用
- 提案 max(k₁, ..., k_r) = 1 になるように分割
 ⇔ A = (a₁, ..., a_k) の形へ (つまり r = k) (仮定より a_i ≠ 0_n)
 → rank(a_i) = 1 (⇔ a_i'a_i ≠ 0) より
 $\hat{\Xi}_i = (a_i' a_i)^{-1} a_i' \left(Y - \sum_{j>i}^r a_j \hat{\Xi}_j \right) \quad (i = 1, \dots, r)$
 (Ξ̂_i (i = 1, ..., r) は p 次元ベクトル)
 → Ξ̂_r から順にシンプルに全部求める

15 / 34

問題点について (1/12)

- 推定量の代入を繰り返す形
 → さらにモデルを書き換えてシンプルな形へ
 ● 不偏性が示せない
 → 上のモデルでの不偏性を示す
- A₁, ..., A_r の分割に依存
 → 情報量規準により選択？
- 多変量一般化リッジ回帰による罰則付推定を考える
 → 最適な罰則パラメータが陽に求まる形にできるか？
 → 複数の Ξ_i に罰則付推定を行い、最適な罰則パラメータを陽に求めるには？

18 / 34

問題点について (2/12) -モデルの書き換えなど-

モデル $Y = 1_n\mu' + A\Xi + \mathcal{E}$
 $= 1_n\mu' + \sum_{i=1}^r A_i\Xi_i + \mathcal{E}$
 $\Rightarrow B_i \stackrel{\text{def}}{=} \left(I_n - \sum_{j=1}^{i-1} P_{B_j} \right) A_i \quad (i = 1, \dots, r)$
 ただし $B_1 = A_1$, $P_M = M(M'M)^+M'$ を表す
 ● モデルをこの B_i を使って次の形で考え直す;
 $Y = 1_n\mu' + \sum_{i=1}^r B_i\Gamma_i + \mathcal{E}$
 $\rightarrow \Gamma_1, \dots, \Gamma_r$ の推定へ
 (μ の最小二乗推定量は前と同じ ($\cdot B_1\mu = 0_{k_1}$)) 19 / 34

問題点について (5/12) -不偏性-

モデル $Y = 1_n\mu' + \sum_{i=1}^r B_i\Gamma_i + \mathcal{E}$ での $\hat{\Gamma}_1, \dots, \hat{\Gamma}_r$ の不偏性
 ● $E[B_i\hat{\Gamma}_i] = B_i\Gamma_i \quad (i = 1, \dots, r)$
 ($\because B_i'1_n = 0_{k_i} \quad (i = 1, \dots, r), B_i'B_j = 0_{k_i, k_j} \quad (i \neq j)$)
 ● $\text{rank}(B_i) = k_i$ の場合; $E[\hat{\Gamma}_i] = \Gamma_i$
 問題 $Y = 1_n\mu' + \sum_{i=1}^r A_i\Xi_i + \mathcal{E}$ での $\hat{\Xi}_i$ の不偏性?
 ● $\text{rank}(A_i) = k_i$ でも $E[\hat{\Xi}_i] \neq \Xi_i$
 ● $\text{rank}(A_i) = k_i$ で $A_i'A_j = 0_{k_i, k_j} \quad (i \neq j)$ とすると
 $E[\hat{\Xi}_i] = \Xi_i$
 ● 結局, 上の B_i を使うモデルへしたのと同じ 22 / 34

問題点について (8/12) -罰則付推定と最適化 (1/4)-

● $Y = 1_n\mu' + \sum_{i=1}^r A_i\Xi_i + \mathcal{E}$ での
 多変量一般化リッジ回帰による推定 (Yanagihara ら, 2009)
 ● i 番目だけに入れる場合, $\hat{\Xi}_i$ だけ次に変化;
 $\hat{\Xi}_i(\theta_i) = (A_i'A_i + Q_i \text{diag}(\theta_i) Q_i')^{-1} A_i' \left(Y - \sum_{j>i} A_j \hat{\Xi}_j \right)$
 ここで $\theta_i = (\theta_{i1}, \dots, \theta_{ik_i})'$, $\theta_{ij} \geq 0 \quad (j = 1, \dots, k_i)$ が
 リッジパラメータ, $\text{diag}(\theta_i)$ は θ_i の要素を対角に持つ
 対角行列, Q_i は $A_i'A_i$ を対角化する直交行列
 $\Rightarrow Cp(\theta_i) \stackrel{\text{予想}}{=} \text{定数} + \sum_{j=1}^{k_i} \varphi_{ij}(\theta_{ij}) \hat{\Xi}_i, \dots, \hat{\Xi}_i, \Sigma^{-1}$ の推定量
 陽に求まると予想 25 / 34

問題点について (11/12) -罰則付推定と最適化 (4/4)-

● $Y = 1_n\mu' + \sum_{i=1}^r B_i\Gamma_i + \mathcal{E}$ での
 多変量一般化リッジ回帰による推定 (Yanagihara ら, 2009) を
 複数の Γ_i に入れる場合
 ● $\hat{\Gamma}_i(\lambda_i) = (B_i'B_i + P_i \text{diag}(\lambda_i) P_i')^{-1} B_i' Y \quad (i = 1, \dots, r)$
 ここで $\lambda_i = (\lambda_{i1}, \dots, \lambda_{ik_i})'$, $\lambda_{ij} \geq 0 \quad (j = 1, \dots, k_i, i = 1, \dots, r)$,
 $\text{diag}(\lambda_i)$ は λ_i の要素を対角に持つ対角行列, P_i は $B_i'B_i$ を対角化
 する直交行列であり, $\lambda = (\lambda_1', \dots, \lambda_r')$ とする
 $\Rightarrow Cp(\lambda) \stackrel{\text{予想}}{=} \text{定数} + \sum_{i=1}^r \sum_{j=1}^{k_i} \psi_{ij}(\lambda_{ij}) \Sigma^{-1}$ の推定量
 陽に求まると予想 28 / 34

提案する推定の利点などのまとめ (1/2)

● $Y = 1_n\mu' + A\Xi + \mathcal{E} = 1_n\mu' + \sum_{i=1}^r A_i\Xi_i + \mathcal{E}$
 利点 Ξ や $\Xi_i \quad (i = 1, \dots, r)$ の推定に関する前提条件
 Ξ の推定量が $(A'A)^{-1}A'Y$ となる前提条件
 $n \geq k = \sum_{i=1}^r k_i$
 Ξ_i の推定量が $(A_i'A_i)^{-1}A_i'Y$ となる前提条件
 $n \geq \max(k_1, \dots, k_r)$ 31 / 34

終

問題点について (3/12) -モデルの書き換えなど-

モデル $Y = 1_n\mu' + \sum_{i=1}^r B_i\Gamma_i + \mathcal{E}$ での $\Gamma_1, \dots, \Gamma_r$ の推定
 リスク 元のモデルと同様
 $\text{Risk}(\mu, \Gamma_1, \dots, \Gamma_r) =$
 $\text{tr} \left\{ \left(Y - 1_n\mu' - \sum_{i=1}^r B_i\Gamma_i \right) \Sigma^{-1} \left(Y - 1_n\mu' - \sum_{i=1}^r B_i\Gamma_i \right)' \right\}$
 $\rightarrow \Gamma_1, \dots, \Gamma_r$ に関する項に着目;
 $-2 \sum_{i=1}^r \text{tr} (B_i'Y \Sigma^{-1} \Gamma_i) + \sum_{i=1}^r \text{tr} (B_i\Gamma_i \Sigma^{-1} \Gamma_i' B_i')$
 ($\because B_i'B_j = 0_{k_i, k_j} \quad (i \neq j), B_i'1_n = 0_{k_i} \quad (i = 1, \dots, r)$)
 $\rightarrow \Gamma_i$ の最小二乗推定量; $B_i'B_i\hat{\Gamma}_i = B_i'Y$ の解 20 / 34

問題点について (6/12) -ここまでの話と残り-

● 推定量の代入を繰り返す形
 \rightarrow さらにモデルを書き換えてシンプルな形へ
 ● 不偏性が示せない
 \rightarrow 上のモデルでの不偏性を示す
 ● A_1, \dots, A_r の分割に依存
 \rightarrow 情報量規準により選択?
 ● 罰則付推定を考える
 \rightarrow 最適な罰則パラメータが陽に求まる形にできるか?
 \rightarrow 複数の Ξ_i に罰則付推定を行い,
 最適な罰則パラメータを陽に求めるには? 23 / 34

問題点について (9/12) -罰則付推定と最適化 (2/4)-

● $Y = 1_n\mu' + \sum_{i=1}^r A_i\Xi_i + \mathcal{E}$ での
 多変量一般化リッジ回帰による推定 (Yanagihara ら, 2009) を
 複数の Ξ_i に入れる場合
 ● $\hat{\Xi}_i(\theta_i) = (A_i'A_i + Q_i \text{diag}(\theta_i) Q_i')^{-1} A_i' \left(Y - \sum_{j>i} A_j \hat{\Xi}_j(\theta_j) \right)$
 ここで $\theta_i = (\theta_{i1}, \dots, \theta_{ik_i})'$, $\theta_{ij} \geq 0 \quad (j = 1, \dots, k_i, i = 1, \dots, r)$,
 $\text{diag}(\theta_i)$ は θ_i の要素を対角に持つ対角行列, Q_i は $A_i'A_i$ を対角化
 する直交行列であり, $\theta = (\theta_1', \dots, \theta_r')$ とする
 $\Rightarrow Cp(\theta) \stackrel{\text{予想}}{=} \text{定数} + \sum_{i=1}^r \sum_{j=1}^{k_i} \varphi_{ij}(\theta_{ij}) \hat{\Xi}_i(\theta_i) (\ell \neq i), \Sigma^{-1}$ の推定量
 陽に求まらない 26 / 34

問題点について (12/12) -罰則付推定のまとめ-

● $Y = 1_n\mu' + \sum_{i=1}^r A_i\Xi_i + \mathcal{E}$ の場合の予想
 ● i 番目だけに罰則 \rightarrow 最適なパラメータ; 陽に求まる
 ● 複数の罰則 \rightarrow 最適なパラメータ; 陽に求まらない
 ● i 番目のパラメータ $\leftrightarrow j$ 番目のパラメータ
 ● $Y = 1_n\mu' + \sum_{i=1}^r B_i\Gamma_i + \mathcal{E}$ の場合の予想
 ● i 番目だけに罰則 \rightarrow 最適なパラメータ; 陽に求まる
 ● 複数の罰則 \rightarrow 最適なパラメータ; 陽に求まる
 ● i 番目のパラメータ $\leftrightarrow j$ 番目のパラメータ
 問題1 Σ^{-1} の推定量 29 / 34
 問題2 複数の Ξ_i への罰則付推定と最適化

提案する推定の利点などのまとめ (2/2)

● $\text{rank}(A_i)$ に関する前提条件はほぼ皆無
 ● $n < k = \sum_{i=1}^r k_i$ でも推定可能
 ● 逆行列を用いて推定量が得られるための前提条件は
 $n \geq \max(k_1, \dots, k_r)$
 ● $r = k$ とすると $n \geq 2$
 ● Moore-Penrose 型一般逆行列・罰則付推定を用いず
 $\text{rank}(A) < k$ でも推定できる
 ● 相関の高い A の列の組が別々の A_i になるように
 分割すれば安定した推定量に
 ● 相関の低い変数の組ごとに A を分割
 ● 分割+罰則付推定でグループごとに罰則付加 32 / 34

問題点について (4/12) -シンプルな形での推定-

モデル $Y = 1_n\mu' + \sum_{i=1}^r B_i\Gamma_i + \mathcal{E}$ での $\Gamma_1, \dots, \Gamma_r$ の推定
 元のモデル $Y = 1_n\mu' + \sum_{i=1}^r A_i\Xi_i + \mathcal{E}$
 変数変換 $B_i = \left(I_n - \sum_{j=1}^{i-1} P_{B_j} \right) A_i$ (ただし $B_1 = A_1$),
 $P_M = M(M'M)^+M'$ を表す
 $\rightarrow \Gamma_i$ の最小二乗推定量; $\text{rank}(B_i) = k_i$ なら
 $\hat{\Gamma}_i = (B_i'B_i)^{-1} B_i' Y \quad (i = 1, \dots, r)$
 ● $\text{rank}(B_i) < k_i$ なら $B_i'B_i\hat{\Gamma}_i = B_i'Y$ の解
 疑問 $\text{rank}(A_i) = k_i$ ならば $\text{rank}(B_i) = k_i$? 21 / 34

問題点について (7/12) -分割に関する問題-

● $Y = 1_n\mu' + \sum_{i=1}^r A_i\Xi_i + \mathcal{E}$
 ● 情報量規準による分割の選択を考える
 例 PMSE (予測平均二乗誤差) の推定量; C_p 規準
 $\rightarrow r, k_1, \dots, k_r, A_1, \dots, A_r$ に含まれる列の選択
 問題1 最適化するパラメータが多い
 問題2 A_1, \dots, A_r に含まれる列の選択
 \leftarrow 組み合わせ数が非常に多い
 (従来の変数選択に加え, 組み合わせの内容まで必要)
 ● $Y = 1_n\mu' + \sum_{i=1}^r B_i\Gamma_i + \mathcal{E}$ としても同様 24 / 34

問題点について (10/12) -罰則付推定と最適化 (3/4)-

● $Y = 1_n\mu' + \sum_{i=1}^r B_i\Gamma_i + \mathcal{E}$ での
 多変量一般化リッジ回帰による推定 (Yanagihara ら, 2009)
 ● i 番目だけに入れる場合, $\hat{\Gamma}_i$ だけ次の形に;
 $\hat{\Gamma}_i(\lambda_i) = (B_i'B_i + P_i \text{diag}(\lambda_i) P_i')^{-1} B_i' Y$
 ここで $\lambda_i = (\lambda_{i1}, \dots, \lambda_{ik_i})'$, $\lambda_{ij} \geq 0 \quad (j = 1, \dots, k_i)$ が
 リッジパラメータ, $\text{diag}(\lambda_i)$ は λ_i の要素を対角に持つ
 対角行列, P_i は $B_i'B_i$ を対角化する直交行列
 $\Rightarrow Cp(\lambda_i) \stackrel{\text{予想}}{=} \text{定数} + \sum_{j=1}^{k_i} \psi_{ij}(\lambda_{ij}) \Sigma^{-1}$ の推定量
 陽に求まると予想 27 / 34

ここからの話

- モデルと推定量
 - モデルと最小二乗推定量など
 - 問題点と解決法
- 提案するモデルと最小二乗推定量など
 - リスク関数の変形
 - 提案するモデルでの推定量と前提条件
 - 利点と問題点
 - 問題点に関する考察と現状
 - 問題点について (前半)
 - 問題点について (後半)
- まとめなど
 - 提案するモデルと推定量と利点

参考文献

1. Schott, J. R. (2005). *Matrix Analysis for Statistics* (Second Edition).
2. Tibshirani, R. (1996). Regression Shrinkage and Selection via the lasso. *J. Roy. Stat. Soc. B*, 58, 267-288.
3. Yanagihara, H., Nagai, I. and Satoh, K. (2009). A bias-corrected C_p criterion for optimizing ridge parameters in multivariate generalized ridge regression. *Japanese J. Appl. Statist.*, 38, 151-172 (in Japanese).

ベイズ型対数尤度に基づくモデルの信用集合

柳本 武美: 統計数理研究所

1. 序

K 次元多項分布 $Mn(\mathbf{p})$ は自然母数を $\log p_k, k = 1, \dots, K$ と置いた曲指数型分布と見なせる。多項分布モデルに還元させるためには、自然母数と結びつくようにする。そのためにモデルの尤度ではなくモデルの対数尤度を定義する必要がある。異なる K 個の事前分布をもつベイズモデル対して、そのベイズ型対数尤度を対数尤度の事後平均により定義する。 K 個のベイズモデルは、分類問題の K 個のクラスと定式化してそのクラスに事前確率を仮定する。この定式化により、事後確率が誘導されモデルの信用集合が得られる。信用集合の導入により得られたモデルの評価が可能になる。例えば、仮説検定とモデル選択の手続きが統一的に得られる。単にモデルの最適な分類を超えた幅広い適用が期待される。

2. ベイズ型対数尤度

標本密度として指数型分布を仮定する。

$$p(\mathbf{x}|\theta) = \exp\{n(\bar{\mathbf{i}}\theta - M(\theta))\} a(\mathbf{x}) \quad (1)$$

$\mathbf{x} \in \mathcal{X} \subset R^n, \theta \in \Theta \subset R^p$. 以降では、標本密度は固定して議論する。異なる K 個の事前密度を仮定して、それらを $\pi_k(\theta) k = 1, \dots, K$ ($K \geq 2$) とする。これから K 個のベイズモデル。

$$M_k = (p(\mathbf{x}|\theta), \pi_k(\theta)), k = 1, \dots, K \quad (2)$$

が得られる。ベイズ型対数尤度は、対数尤度の事後平均 $E_{\text{post}}\{\log p(\mathbf{x}|\theta)\}$ が基本となる。しかし、このままでは標本 \mathbf{x} の二重使用になっているのでその偏りを補正する必要がある。

その補正のためには $E_{\text{post}}\{\log p(\mathbf{y}|\theta)\}$ を評価する。簡単のために漸近近似が使われるが、有限標本での妥当性に問題が生じる。。そこでこの事後平均が最適な予測子と関係することに注意する。予測子を $p(\mathbf{y}|\mathbf{x}) \propto \exp(E_{\text{post}}[\log p(\mathbf{y}|\theta)])$ で定義すると、この予測子は $D(p(\mathbf{y}|\mathbf{x}), p(\mathbf{y}|\theta))$ の事後平均を最小にする意味で最適な予測子である。更にその予測子は θ の事後平均 $\hat{\theta}$ を plug-in して最適な予測子が得られる。つまり $p_e(\mathbf{y}|\mathbf{x}) = p(\mathbf{y}|\hat{\theta})$ となる (Yanagimoto-Ohnishi, 2009, JSPI)。これからモデル M_k のベイズ型対数尤度 $l(M_k) = l(k)$ を

$$l(k) = E_{\text{post}}[\log p(\mathbf{x}|\theta) - D(p_e(\mathbf{y}|\mathbf{x}), p(\mathbf{y}|\theta))] \quad (3)$$

で定義する。この尤度は次の別表現が可能である。

$$\log p_e(\mathbf{x}|\mathbf{x}) - 2 E_{\text{post}}[D(p_e(\mathbf{y}|\mathbf{x}), p(\mathbf{y}|\theta))]$$

3. モデルの事前分布と信用集合

尤度ではなくて対数尤度を扱った理由のは K 次元多項分布 $Mn(\mathbf{p})$ の構造にある。多項分布は自然母数を $\log \mathbf{p} = (\log p_1, \dots, \log p_K)$ と置いた正值測度空間の部分空間と見なせる。解析的には曲指数分布族と同じように扱える。そこで、ベイズ型対数尤度を用いて

$$\log p_k = l(k), k = 1, \dots, K$$

と置いて、多項分布の推定問題に還元させる。母数 p は K 次元標準単錐 S^{K-1} の要素である。その単錐の頂点全体を V^{K-1} と書くと、 K 個の各々の頂点は分類問題におけるクラスに対応する。この問題ではモデルに対応する。そこでモデルの事前確率を $\pi(k) = q_k, q \in S^{K-1}$ とする。事前確率は $q_k = 1/K$ が default 選択になるが、問題に応じて自由に選ぶことができる。これから事後確率

$$\pi(k|\mathbf{x}) = \frac{\exp(\ell(k))q_k}{\sum_h \exp(\ell(h))q_h}$$

が得られる。

事後確率から信用水準、信用区間が導かれる。本節のような手続きは無用を感じるかも知れない。ごく常識的は

$$\hat{k} = \underset{k}{\text{Argmax}} \pi(k|\mathbf{x})$$

として \hat{k} を求めれば十分と考えるからである。しかし実際には $\pi(\hat{k}|\mathbf{x})$ の値の大きさが分類結果の解釈に大きな影響を及ぼす。例えば、この値が $1/K$ に近いと分類の意味は弱い。また、信用水準 β を与えておいて、適当な r に対して HPD 信用集合は $C = \{k | \pi(k|\mathbf{x}) \geq r\}$ が

$$\sum_{k \in C} \pi(k|\mathbf{x}) \geq \beta$$

を満たすように選ぶことで定義される。実用上の分類問題で必要になる情報を提供する。。

4. 適用

4.1. モデル選択

モデル選択は、文字通りのモデルの選択と選択したモデルの下での母数の推定と推定した母数を plug-in した予測子の構成の二面がある。しかし、その理論では両者は一体的である。伝統的には *IC と表され多くの手法、AIC, TIC, DIC, WAIC などが提案されている。大雑把近似では $(-1/2)*IC$ がベイズ型対数尤度に相当する。尤度法におけるモデル選択では (2) ではなく母数の集合 Θ_k により定義される。(1) は

$$p(\mathbf{x}|\theta) = \exp\{n(\bar{t}\theta - M(\theta))\} a(\mathbf{x}), \theta \in \Theta_k \subset \Theta$$

と表される。モデルは $\tilde{M}_k = (p(\mathbf{x}|\theta), \Theta_k)$ である。事後確率に相当する量の研究が行われ Burnham-Anderson (2002) はそのレビューを与えている。モデルの事前分布は、モデルの重みとされる。定義の理論体系が曖昧である。

提案法の特徴は

- 1) 関連した予測子が標本サイズ有限の下で最適性がある、(AIC, TIC, DIC と比較して)
- 2) 漸近的近似は援用しない
- 3) 観察予測子 $p(\mathbf{x}|\mathbf{x})$ ではなく対数尤度を基本とする
- 4) モデルを与えると一意に定まる、(DIC と比較して)

4.2. 仮説検定

3 節でも議論したようにベイズ理論における信頼区間に対応する信用区間を定義した。従って、モデルの事前確率を選ぶことにより、仮説検定を説明できる、例えば Ogura-Yanagimoto (2016, Stat. Med)。その結果、モデル選択の理論で強調されている仮説検定との乖離がなくなる。既存の研究では、ベイズ因子とか *IC によるモデルの選択が仮説検定よりも優れていることを主張されていた。本研究では、優劣を論じないで両者を統一的に俯瞰する。

Analysis of Spatially Resolved Galaxy Spectra as a High-Dimensional Low-Sample Size Data Problem

(高次元小標本データ問題としての銀河の分光マップの解析)

竹内 努^{1,2}, 河野 海¹, 中西 康一郎³, 矢田 和善, 青嶋 誠⁴, 石井 晶⁵, 江頭 健斗⁶

1. 名古屋大学理学研究科素粒子宇宙物理学専攻,
2. 統計数理研究所統計的機械学習研究センター
3. 国立天文台 ALMA プロジェクト
4. 筑波大学数理物質系
5. 東京理科大学理工学部情報科学科
6. 筑波大学 理工情報生命学術院

銀河とは、星と星間物質(ガスと星間塵)、暗黒物質からなる巨大な天体である。銀河は現在観測可能な宇宙に数千億個存在しており、我々の目に見える波長(可視光線)での宇宙の姿を形作っている。しかし宇宙誕生当時、物質分布はほぼ一様であった。銀河は平均からわずかに密度が高い領域が重力によって成長し、合体成長を経て現在の姿へと進化してきたのである。銀河進化は周囲の銀河の密度やガス密度など、銀河の置かれた環境にも強く依存する極めて複雑な過程である。そして銀河の形成と進化は 138 億年の宇宙進化の歴史の中でも非常に重要な現象のひとつであると考えられている。

銀河の進化を駆動する重要な内的物理過程が星形成である。星形成は星間物質が重力によって凝集し、電離ガスから原子ガス、そして分子ガスへと冷却することで分子雲、さらに高密度のガス塊を形成し、核融合が始まるという過程である。星になる母体が分子雲であることから、分子雲の物理状態を解明することが銀河進化の重要な課題となっている。観測的にも星形成率がガス密度の冪乗に比例するという経験則(Kennicutt-Schmidt 則: Schmidt 1959; Kennicutt 1989, 1998)が知られており、星間分子の素過程の解明はこの経験則を物理的に説明するための手がかりとしても重視されている。

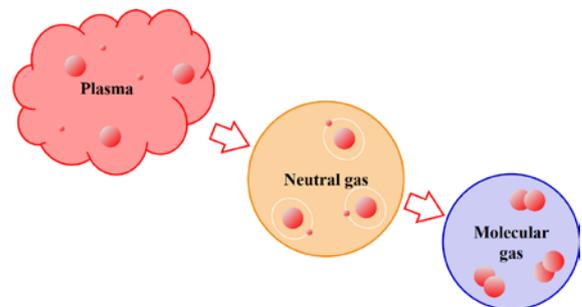


図 1: 電離ガス(プラズマ)から中性ガス雲、分子雲への凝集。

実験室実験の不可能な遠方宇宙を対象とする天文学において、天体の分子の物理情報を得る唯一の手段が分光観測である。分子の特徴的量子状態に対応する電磁波放射(輝線放射)は主として電波に現れるため、分子雲の物理は電波天文学で発展してきた。分子の電波輝線を観測できる、現在最も大規模かつ高性能の電波観測装置が Atacama Large Millimeter/Submillimeter Array (ALMA)である。ALMA を始めとする現在の天体分光観測装置は、天体の星間物質が含む原子、分子、イオン、そしてダスト(星間塵)について圧倒的な質・量の情

報を提供する。しかし詳細な分光観測は一般的に非常に時間がかかり、天体をマッピングして独立な観測点を多数得ることは容易ではない。その結果、データの持つ波長方向の次元 d に対してサンプル数 n が非常に小さい ($n \ll d$) という状況が生じる。伝統的な天文学では、このような状態は**不良設定問題**とされ、その解決は諦めざるを得ないとされてきた。物理的考察のためにはデータ d の情報を大幅に捨て、 n より小さくして解析する他ないと考えられていたのである。

当然このような無駄は避け、情報量をフルに活かせる方法が望ましい。 $n \ll d$ となるデータは高次元小標本(HDLSS)データと呼ばれる。天文学以外の分野、たとえばゲノム解析では遺伝子の標本数 $n \sim 100$ に対し塩基配列の次元数が $d \sim 10^5$ といった問題は特殊ではない。HDLSS の問題を解決するためにここ 10 年で発展した統計学の方法が**高次元統計学**であり、現在も続々と新しい知見が得られている(e.g., Aoshima 2018)。本研究では、天体の各点を ALMA によって分光観測した探査(マッピング)型のデータに対して高次元統計学の方法を適用した新しい解析を行っている。上記のように、分光マッピングには大変な観測的コストがかかり、ALMA の場合現状では最大級でも $n \sim 200$ 程度の観測点が得られる程度であるのに対して波長(振動数)分散方向には $d \sim 2000$ の情報がある。よって、ALMA の分光マッピングは典型的な HDLSS データである。従来の天文学では、分光データから物理の第一原理から有用であるとわかっているいくつかの輝線のみを抽出し、それを用いて分類を行っていた。しかし、ALMA の観測によって、たとえ隣り合った星間分子雲の領域であっても分子輝線は大幅に異なった様相を示すことが判明し(図 2)、古典的方法の限界が明らかになっている。巨大すぎる情報量のため、分子雲の進化と星形成の関係が全く見えてきていないのである。高次元統計学の方法によって、かつて絶望的と考えられていた、星間物質の輝線の全ての情報を活かした分類が可能になると期待される。もちろん、天文学での分光データに対する高次元統計学の適用は前例がなく、天体分光固有のデータの問題あり、その可能性は未知数の部分も大きい。本講演では問題意識と目標、そして研究の現状について述べる。

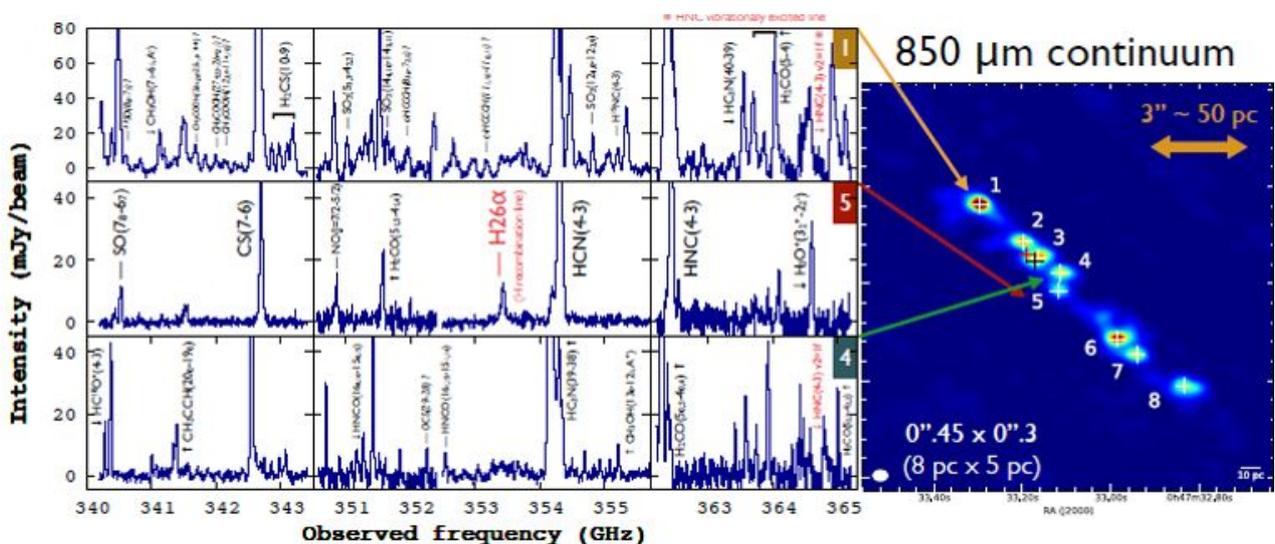


図 2: ALMA による爆発的星形成銀河 NGC253 中心部の星形成領域の分光マッピングの例(Ando et al. 2017 より改変)。右パネルが観測領域のマップ、左パネルが観測点に対応する分子輝線スペクトルデータを示す。

Topological Data Analysis of Galaxy Distribution: Detection of the Birth Cry of Cosmic Structures

(銀河分布の位相的データ解析: 宇宙構造の産声の検出)

竹内 努^{1,2}, 河野 海¹, Suchetha COORAY¹, 西澤 淳^{3,1}, 村上 広椰¹

1. 名古屋大学理学研究科素粒子宇宙物理学専攻,
2. 統計数理研究所統計的機械学習研究センター
3. 名古屋大学高等研究院

銀河とは、星と星間物質(ガスと星間塵)、暗黒物質からなる巨大な天体である。銀河は現在観測可能な宇宙に数千億個存在しており、我々の目に見える波長(可視光線)での宇宙の姿を形作っている。しかし宇宙誕生当時、物質分布はほぼ一様であった。銀河は平均からわずかに密度が高い領域が重力によって成長し、合体成長を経て現在の姿へと進化してきたのである(図 1)。銀河進化は周囲の銀河の密度やガス密度など、銀河の置かれた環境にも強く依存する極めて複雑な過程である。そして銀河の形成と進化は 138 億年の宇宙進化の歴史の中でも非常に重要な現象のひとつであると考えられている。

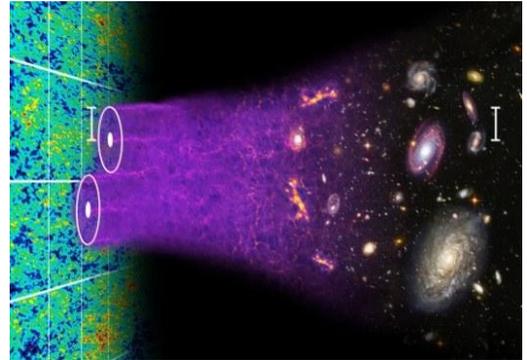


図 1: 一様な物質分布から銀河が形成され、進化する様子。

我々は 130 数億年にわたる銀河の形成・進化という複雑な物理現象を、これまでの宇宙物理学の方法とはまったく異なる、データ科学の最新手法である位相的データ解析 (topological data analysis: TDA) に基づくアプローチによって新たな角度から解明することを試みている。本研究は特に、銀河形成以前の通常物質(宇宙物理学ではバリオンと呼ぶ)が光との相互作用によって生じる粗密波である「バリオン音響振動」 (baryon acoustic oscillation: BAO) が銀河分布に残す痕跡について扱う。バリオンと光の相互作用が切れる宇宙の晴れ上がり(脱結合)時に物質が放射した熱放射は宇宙膨張によって波長が引き伸ばされ、現在は電波(ミリ波)の波長で宇宙マイクロ波背景放射(cosmic microwave background: CMB)として観測される。BAO は CMB のゆらぎに特徴的なスケールとして現れ、ゆらぎをフーリエ変換したパワースペクトルにおいて最も顕著なピークとして観測される(図 2)。このスケールは共動距離で観測した際に、どの赤方偏移においてもおよそ 148 Mpc (1 Mpc = 3.08×10^{24} cm) の一定値をとる。この性質から、BAO シグナルの観測的スケールは宇宙論パラメータの推定の最も信頼

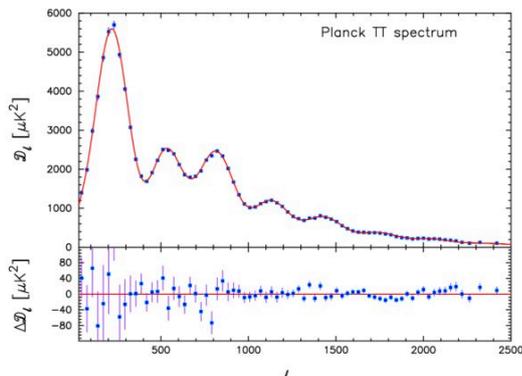


図 2: Planck 衛星が観測した CMB のパワースペクトル。最も顕著なピークが BAO である。

図 1: 一様な物質分布から銀河が形成され、進化する様子。

できる指標のひとつとなっている。

従来, BAO の検出には銀河分布の2点相関関数やパワースペクトルが用いられてきたが, BAO シグナルは銀河の重力的クラスタリングによって生じるシグナルに対して 1/100 程度と微弱であるため, 統計的に有意な結果を得るためには大規模かつサンプリングレートの高い銀河探査が必要であり, 現存する最大の銀河探査であるスローンデジタルスカイサーベイ (Sloan Digital Sky Survey: SDSS) によって初めてそのシグナルが検出された. この規模の探査は決して容易ではなく, 独立な, そしてより有効な検出方法が望まれている。

我々は, 銀河の空間分布がもつ位相幾何学的情報を抽出することで BAO シグナルの検出や宇宙大規模構造を解析することを目的として, パーシステントホモロジー (PH) の計算を行った. PH はトポロジカルデータ解析 (TDA) の中心的方法論であり, ホモロジー群に対して関心のあるパラメータに関するフィルトレーションによって定義される. さらに, 穴の位置や形状などの宇宙論的に重要な情報を得ることができる点が特徴である. 本研究では, 共動距離 (宇宙膨張の影響を取り除いた本質的な距離) をパラメータにとることにより, 空間に対する解像度を变化させた際に生じる構造を次元ごとに解析することができる。

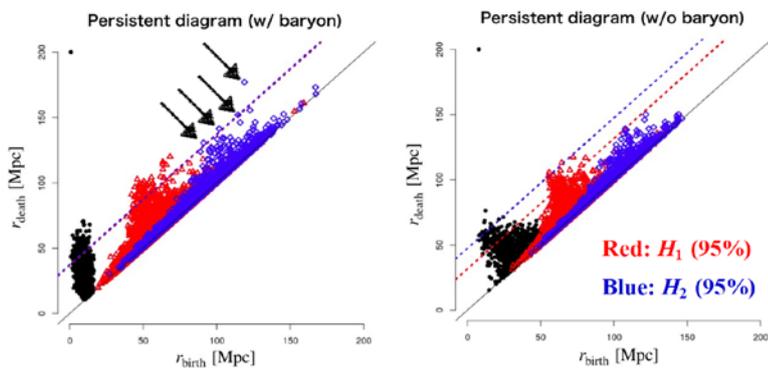


図 3: 宇宙論的シミュレーションデータから求めたパーシステント図. 左がバリオン有, 右がなしの場合を示す。

り, BAO の理論と整合的であった. さらに, SDSS eBOSS サーベイによって得られた実際のクエーサー観測データに対しても同様の解析を行い, H_2 に対して $\bar{r} = 146.6 \pm 2.0$ [Mpc] という推定値を得た. また本研究では PH の逆解析も行い, 得られた H_2 に対応する実際の宇宙の構造を可視化することに成功した (図 4). 銀河分布に対しては世界初の例である. これらの計算結果に H_1, H_0 についての考察を加味し, 銀河の空間分布に関して得られた結果を報告する。

我々はまず TDA の BAO 検出に対する有用性の検証のため, バリオン有り・無しの二通りの N 体シミュレーションデータに対して PH の計算を行った. その結果, バリオン有の場合にのみ有意な 2 次ホモロジー (H_2), すなわち中空のシェル構造を検出した (図 3). また, 検出された H_2 構造の半径の平均値は $\bar{r} = 150.16 \pm 8.46$ [Mpc] であり,

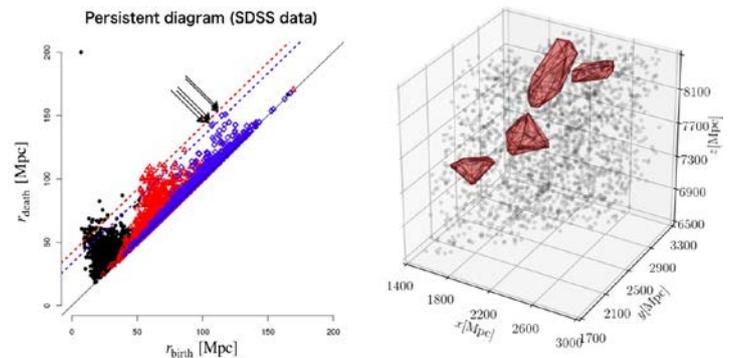


図 4: SDSS eBOSS データから求めた PH (左), およびその逆解析によって得られた銀河分布の対応構造。

Sparse PCA by the noise-reduction methodology

Kazuyoshi Yata

Institute of Mathematics, University of Tsukuba
and

Makoto Aoshima

Institute of Mathematics, University of Tsukuba

High-dimension, low-sample-size (HDLSS) data situations occur in many areas of modern science such as genomics, medical imaging, text recognition, finance, and so on. In recent years, substantial work has been done on HDLSS asymptotic theory, where the sample size n is fixed or $n/d \rightarrow 0$ as the data dimension $d \rightarrow \infty$. Ahn et al. (2007), Aoshima and Yata (2011), Hall, Marron and Neeman (2005), Jung and Marron (2009), Yata and Aoshima (2012) explored several types of geometric representations on HDLSS data. Jung and Marron (2009) investigated into consistency properties both of the eigenvalues and principal component (PC) directions for the sample covariance matrix. Yata and Aoshima (2012) gave consistent estimators both for the eigenvalues and PC directions together with the PC scores by developing a new PCA method called the *noise-reduction methodology*. On the hand, sparse PCA (SPCA) methods were studied by several references. Zou and Hastie (2006) and Lee, Huang and Hu (2010) considered a regularized SPCA (RSPCA) based on L1 penalties in high-dimensional settings. Johnstone and Lu (2009) proposed a thresholded SPCA (TSPCA) and showed a consistency property of the TSPCA when $n/d \rightarrow 0$. Shen, Shen and Marron (2013) showed that the PC directions by the RSPCA and TSPCA have the consistency property when $d \rightarrow \infty$ while n is fixed. In addition, Paul and Johnstone (2007) developed an augmented SPCA method and Ma (2013) proposed an iterative thresholding procedure for PC directions.

In this talk, we investigated the TSPCA in high-dimensional settings. We first proposed a new sparse estimator of the PC directions and showed that it

holds the consistency property without any threshold values. We also proposed a shrinkage PC directions and apply it to the clustering.

References

- Ahn, J., Marron, J.S., Muller, K.M., Chi, Y.-Y. (2007). The high-dimension, low-sample-size geometric representation holds under mild conditions. *Biometrika* **94**, 760-766.
- Aoshima, M., Yata, K. (2011). Two-stage procedures for high-dimensional data. *Sequential Analysis (Editor's special invited paper)* **30**, 356-399.
- Hall, P., Marron, J.S., Neeman, A. (2005). Geometric representation of high dimension, low sample size data. *J. R. Statist. Soc. Ser. B* **67**, 427-444.
- Johnstone, I., Lu, Y. (2009). On consistency and sparsity for principal components analysis in high dimensions, *J. Amer. Statist. Assoc.* **104**, 682-693.
- Jung, S. and Marron, J. S. (2009). PCA consistency in high dimension, low sample size context. *Ann. Statist.* **37**, 4104-4130.
- Lee, S., Huang, J. Z., Hu, J. (2010). Sparse logistic principal components analysis for binary data. *Ann. Appl. Statist.* **4**, 1579-1601.
- Ma, Z. (2013). Sparse principal component analysis and iterative thresholding. *Ann. Statist.* **41**, 772-801.
- Paul, D., Johnstone, I. (2007). Augmented sparse principal component analysis for high dimensional data. *Technical Report*, UC Davis.
- Shen, D., Shen, H., Marron, J.S. (2013). Consistency of sparse PCA in high dimension, low sample size contexts. *J. Multivariate Anal.* **115**, 317-333.
- Yata, K., Aoshima, M. (2012). Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. *J. Multivariate Anal.* **105**, 193-215.
- Zou, H., Hastie, R. (2006). Sparse principal component analysis. *J. Comput. Graph. Statist.* **15**, 265-286.

Decimating Nonlinear Response in State-Space Models for Battery Capacity Estimation

Xun Shen (Tokyo University of Agriculture and Technology)

Jiancang Zhuang (Institute of Statistical Mathematics)

1 Introduction

A discrete state space model is used to perform the State of Charge (SoC) estimation, which is written as:

$$(1) \quad \begin{bmatrix} U_{e,t+1} \\ SoC_{t+1} \end{bmatrix} = \begin{bmatrix} e^{-\frac{\Delta\tau}{R_e C_e}} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} U_{e,t} \\ SoC_t \end{bmatrix} + \begin{bmatrix} 1 - e^{-\frac{\Delta\tau}{R_e C_e}} \\ \frac{\eta_c \Delta\tau}{C_n} \end{bmatrix} i_t + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix},$$

$$(2) \quad U_t = U_{OCV,0} + K_0 SoC_t + s(SoC_t) + i_t R_o + U_{e,t} + w_t.$$

Here, SoC_t denotes the SoC at time step t . U_t denote the terminal voltage. U_e is the voltage of the equivalent RC circuit. R_e and C_e are the equivalent resistance and capacity for the characteristics include polarization, diffusion, and so on. R_o is the internal ohmic resistance. i is the charging current. $U_{OCV,0} + K_0 SoC_t + s(SoC_t)$ is the Open Circuit Voltage of the battery. $s(SoC_t)$ is the **unknown nonlinear part**. Besides, $v_t = [v_{1,t}, v_{2,t}]^T$ and w_t are the process and the measurement noises which are Gaussian white noise with covariance Q and R , respectively.

In this work, we approximate $s(SoC_t)$ by a non-parametric model and decimate it from the model. Then, we can use the linear Kalman filter to achieve SoC estimation.

2 Proposed Algorithm

The state-space model for SoC estimation can be generally written as

$$(3) \quad x_{t+1} = Fx_t + Hu_t + v_t,$$

$$(4) \quad y_t = C + Gx_t + Ju_t + s(x_t).$$

We introduce non-parametric model to approximate $s(x_t)$. Since $s(x_t)$ is from $\mathbb{R}^k \rightarrow \mathbb{R}^d$, we can split $s(x_t)$ as

$$(5) \quad s(x_t) = [s_1(x_t), \dots, s_d(x_t)]^T.$$

The non-parametric model used to approximate $s(x_t)$ is defined as $\hat{s}_{np}(x_t)$. We can also split

$$(6) \quad \hat{s}_{np}(x_t) = [\hat{s}_{np,1}(x_t), \dots, \hat{s}_{np,d}(x_t)]^T.$$

Each $\hat{s}_{\text{np},j}(x_t), j \in 1, \dots, d$ can be written as

$$(7) \quad \hat{s}_{\text{np},j}(x_t) = \sum_{i=1}^{N_{\text{np}}} \beta_{ij} h(x_t, a_{ij}, b_{ij})$$

where N_{np} denotes the number of hidden nodes of the non-parameter model, $h(\cdot)$ denotes the activation function, and β_{ij} denotes the weight connecting the i -th hidden node and the output, $a_{ij} = [a_{ij,1}, \dots, a_{ij,k}]$ represents the weight vector towards x_t , and b_{ij} is the scalar threshold of the i -th hidden node. $\beta_j = [\beta_{1j}, \dots, \beta_{N_{\text{np}}j}]^T$ denotes the weight vector connecting the hidden nodes and the output. According to the universal approximation theorem (??), $\forall \epsilon_S > 0, \exists N_{\text{pn}} \in \mathbb{N}^+, \beta, b \in \mathbb{R}^{N_{\text{np}}}, a \in \mathbb{R}^{N_{\text{np}} \times k}$, and $\exists g : \mathbb{R}^k \times \mathbb{R}^{N_{\text{np}}} \times \mathbb{R}^{N_{\text{np}} \times k} \rightarrow \mathbb{R}$, such that

$$(8) \quad \|\hat{s}_{\text{np},j} - s_j(x_t)\| \leq \epsilon_S, \forall x_t \in \mathbb{R}^k.$$

Then, we apply iterative algorithms to calibrate the parameter and estimate the SoC which is summarized as

1. Initialize iteration number $n_{\text{ite}} = 0$, parameter decision $\hat{\theta}_m^{(n_{\text{ite}})}, \hat{\theta}_s^{(n_{\text{ite}})} = \mathbf{0}$, and relaxation factor $\alpha^{(n_{\text{ite}})} = 1$;

2. Calculate $\hat{s}_{\text{np}}^{(n_{\text{ite}})}(\hat{x}_{t|t}^{(n_{\text{ite}})})$ based on $\hat{\theta}_s^{(n_{\text{ite}})}$ and obtain linear part of output $y_{\text{linear},t}^{(n_{\text{ite}})}$ as

$$(9) \quad y_{\text{linear},t}^{(n_{\text{ite}})} = y_t - \alpha^{(n_{\text{ite}})} \hat{s}_{\text{np}}^{(n_{\text{ite}})}(\hat{x}_{t|t}^{(n_{\text{ite}})}) - \hat{C}^{(n_{\text{ite}})} - \hat{J}^{(n_{\text{ite}})} u_t.$$

(For $n_{\text{ite}} = 0$, use $\hat{s}_{\text{np}}^{(n_{\text{ite}})}(\hat{x}_{t|t}^{(n_{\text{ite}})}) = \mathbf{0}$.)

3. Implement Kalman filter (?) based on $\hat{\theta}_m^{(n_{\text{ite}})}, U_T = \{u_0, u_1, \dots, u_T\}, Y_{\text{linear},T}^{(n_{\text{ite}})} = \{y_{\text{linear},0}^{(n_{\text{ite}})}, y_{\text{linear},1}^{(n_{\text{ite}})}, \dots, y_{\text{linear},T}^{(n_{\text{ite}})}\}$ and obtain the update of state estimation $\hat{x}_{t|t}^{(n_{\text{ite}})}$, nonlinear residual $\hat{e}_{\text{nl},t}^{(n_{\text{ite}})} = y_t - \hat{C}^{(n_{\text{ite}})} - \hat{J}^{(n_{\text{ite}})} u_t - \hat{G}^{(n_{\text{ite}})} \hat{x}_{t|t}^{(n_{\text{ite}})}$ for all t ;

4. Update the parameter decision $\hat{\theta}_s^{(n_{\text{ite}}+1)}$ by

$$(10) \quad \hat{\theta}_s^{(n_{\text{ite}}+1)} = \min_{\hat{\theta}_s} \sum_t \|\hat{e}_{\text{nl},t}^{(n_{\text{ite}})} - \alpha^{(n_{\text{ite}})} \hat{s}_{\text{np}}(\hat{x}_{t|t}, \hat{\theta}_s)\|^2.$$

5. Update $\hat{\theta}_m^{(n_{\text{ite}}+1)}, \alpha^{(n_{\text{ite}})}$ by

$$(11) \quad \{\hat{\theta}_m^{(n_{\text{ite}}+1)}, \alpha^{(n_{\text{ite}}+1)}\} = \max_{\hat{\theta}_m, \alpha} \log p_{y_0}(y_0) + \sum_{t=1}^T \log p_{y_t}(y_t | \hat{\theta}_m, \alpha, \hat{\theta}_s^{(n_{\text{ite}}+1)}, Y_{t-1}, U_{t-1})$$

or equivalently

$$(12) \quad \{\hat{\theta}_m^{(n_{\text{ite}}+1)}, \alpha^{(n_{\text{ite}}+1)}\} = \min_{\hat{\theta}_m, \alpha} \sum_{t=1}^T \|y_t - \hat{y}_t(\hat{\theta}_m, \alpha, \hat{\theta}_s^*, Y_{t-1}, U_{t-1})\|^2.$$

6. Set $n_{\text{ite}} = n_{\text{ite}} + 1$. If $n_{\text{ite}} > N_{\text{ite}}$, stop and output $\hat{\theta}_m^* = \hat{\theta}_m^{(n_{\text{ite}})}, \hat{\theta}_s^* = \hat{\theta}_s^{(n_{\text{ite}})}, \alpha^* = \alpha^{(n_{\text{ite}})}$. Otherwise, go back to step 2.

Gradient descent algorithm can be used to solve Eq. (10). Besides, the EM algorithm is efficient to solve (12).

二部グラフ構造を持つ処理集合の実験計画とその深層学習への応用

地寄 頌子 (大阪工業大)

藤原 良 (筑波大学)

宮本 暢子 (東京理科大学)

本報告では、完全二部グラフの辺を処理集合に持つようなブロック計画を提案する。実験計画では普通は処理の集合とブロック (処理の部分集合) の集まりでブロック計画を構成する。ここでは処理集合は単なる集合ではなく、その中に構造を持っている場合を考える。いま完全二部グラフの辺の集合が処理集合であるとする。そのとき処理パラメータの推定精度を上げるために処理のペアをバランスさせようとする、複雑な問題が生じる。完全二部グラフ $K_{n,m}$ の二つの頂点集合を V_1, V_2 とし、以下のような条件を満たす部分グラフの集合 $\mathcal{B} = \{B_1, B_2, \dots, B_N\}$ を考える。

条件

- (1) \mathcal{B} のどの部分グラフ B_i も V_1, V_2 の点を全て含む
- (2) \mathcal{B} の中に $K_{n,m}$ のすべての辺はちょうど μ 回ずつ現れる
- (3) 任意の $x_1, x_2 \in V_1, (x_1 \neq x_2), y_1 \in V_2$ に対して、二つの辺 $(x_1, y_1), (x_2, y_1)$ が同時に含まれる部分グラフは \mathcal{B} 内に λ_{21} 個存在する
- (4) 任意の $x_1 \in V_1, y_1, y_2 \in V_2, (y_1 \neq y_2)$ に対して、二つの辺 $(x_1, y_1), (x_1, y_2)$ が同時に含まれる部分グラフは \mathcal{B} 内に λ_{12} 個存在する
- (5) 任意の $x_1, x_2 \in V_1, (x_1 \neq x_2)$ かつ $y_1, y_2 \in V_2, (y_1 \neq y_2)$ に対して、二つの辺 $(x_1, y_1), (x_2, y_2)$ が同時に含まれる部分グラフは \mathcal{B} 内に λ_{22} 個存在する

これらの条件を確認するために部分グラフ B_1, B_2, \dots, B_N に対応する $(n \times m)$ -結合行列をそれぞれ H_1, H_2, \dots, H_N とする。結合行列 H_1, H_2, \dots, H_N に対して、 $(N \times m)$ -行列 X_j ($j = 1, 2, \dots, n$) をその第 i 行が H_i の第 j 行となっている行列であるとする。それらを結合した $(N \times nm)$ -行列 $X = [X_1 | X_2 | \dots | X_n]$ を計画行列という。このとき、上の条件 (2) から (5) が満たされれば、情報行列 $X^t X$ の小行列に次のように現れる: (a) 小行列 $X_i^t X_i$ の対角成分は全て μ (条件 2), 非対角成分は全て λ_{21} (条件 3). (b) 小行列 $X_i^t X_j$, ($i \neq j$) の対角成分は全て λ_{12} (条件 4), 非対角成分は全て λ_{22} (条件 5). 理想的には $\lambda_{21} = \lambda_{12} = \lambda_{22} (= \rho)$, すなわち $X^t X = \mu I + \rho(J - I)$ の形になる時、推定量の精度が最適になる。そうでない場合には、非対角成分に現れる値の種類ができるだけ少ない方が良くとされている。(長尾, 栗木 [3])

定義 1 ((r, λ) -design [2]) V を m 点集合とし、 \mathcal{B} を V の部分集合からなる集合族 (ブロック集合と呼ぶ) とする。 (V, \mathcal{B}) が次の条件を満たすとき、 (r, λ) -design という。

- V の各要素は必ず \mathcal{B} の r 個のブロックに含まれる
- V の任意の異なる 2 つの要素を同時に含むブロックは、 \mathcal{B} の中に必ず λ 個存在する

(r, λ) -design は点の数を m , ブロック数を n とすると、 $(n \times m)$ -結合行列 H で表現できる。そして、 V 上の巡回写像を P_0, P_2, \dots, P_{m-1} であるとし、 m 個の design としては同型な結合行列を生成することができるとする。これらを基本結合行列とする。

$$H_i = H P_i, \quad i = 0, 1, \dots, m-1$$

これらの基本結合行列からできる $(m \times nm)$ -計画行列 $X = [X_0 | X_1 | \dots | X_{n-1}]$ は条件の (1) から (4) を満たす。ただし (5) の条件は満たさない。(5) の条件を満たすために、次の概念を導入する。

定義 2 (Difference matrix [1]) \mathbf{Z}_m 上の $(k \times m\gamma)$ -行列を $D = [d_{ij}]$ とする. 任意の $1 \leq i < j \leq k$ について, \mathbf{Z}_m 上の多重集合

$$\{d_{ih} - d_{jh} \pmod{m} \mid h = 1, 2, \dots, m\gamma\}$$

が \mathbf{Z}_m の要素をちょうど γ 回ずつ含むとき, D を (m, k, γ) -difference matrix (DM) という.

先ほど巡回的に作られた $H_i = HP_i$ ($i = 0, 1, \dots, m-1$) を基に計画行列 $X = [X_0|X_1|\dots|X_{m-1}]$ を DM を用いて次のように拡大する. ここで, X_i は各 H_j ($j = 0, 1, \dots, m-1$) の i 行目からなる行列である. (H_i, X_i の行列の行を, 上から 0 行目, 1 行目, ... と数える.)

$$\mathcal{X}_1 = \begin{bmatrix} X_{d_{11}} \\ X_{d_{12}} \\ \vdots \\ X_{d_{1(m\gamma)}} \end{bmatrix}, \quad \mathcal{X}_2 = \begin{bmatrix} X_{d_{21}} \\ X_{d_{22}} \\ \vdots \\ X_{d_{2(m\gamma)}} \end{bmatrix}, \dots, \quad \mathcal{X}_k = \begin{bmatrix} X_{d_{k1}} \\ X_{d_{k2}} \\ \vdots \\ X_{d_{k(m\gamma)}} \end{bmatrix}.$$

ただし, $D = [d_{ij}]$ は (n, k, γ) -DM とする. このとき, 拡大された計画行列 $\mathcal{X} = [\mathcal{X}_1|\mathcal{X}_2|\dots|\mathcal{X}_k]$ は条件 (2) から (5) を満たす. 拡大された計画行列 \mathcal{X} でも, 各行は一つの実験パターンを表している. 拡大前の計画行列 X の第 i 行は H_i の結合行列に対応していたが, 拡大された場合は変形された結合行列の集合になっている. その時の新しい結合行列のサイズは $k \times m$ となっている, (これを実験結合行列と呼ぶこととする).

ここでパラメータを整理すると,

- 基本結合行列 H_i のサイズは $n \times m$
- H_i を生成する巡回群の位数は m
- Difference matrix (m, k, γ) -DM のサイズは \mathbf{Z}_m 上の $k \times m\gamma$
- 拡大計画行列のサイズは $m^2\gamma \times mk$, この時の実験結合行列のサイズは $k \times m$
- (r, λ) -design と (m, k, γ) -DM を使って情報行列を作った場合, $\mu = rm\gamma$, $\lambda_{21} = m\lambda\gamma$, $\lambda_{12} = \lambda_{22} = m\gamma + \lambda\gamma(m-1)$ なる計画行列を作ることが可能である.

本報告では, 応用事例や具体例を示すとともに, difference matrix の要素や基本結合行列を生成する際に用いる巡回群について議論した.

謝辞 本研究は JSPS 科研費 19K11866 の助成を受けたものです.

参考文献

- [1] T. Beth, D. Jungnickel, and H. Lenz. *Design Theory*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2 edition, 1999.
- [2] D. Stinson. *Combinatorial designs: constructions and analysis*. Springer Science & Business Media, 2007.
- [3] 長尾壽夫 and 栗木進二. **数理統計学**. データサイエンス・シリーズ. 共立出版, 2006.

マクロ進化の大規模種間比較研究における枝特異的方向性淘汰モデルの提案

大久保祐作¹²

1. 統計数理研究所モデリング研究系 特任研究員
2. データサイエンス共同利用基盤施設データ同化研究支援センター 特任研究員

・背景

種間系統比較法(phylogenetic comparative method; PCM)とは生物形質のマクロ進化を分析する統計的手法の総称である。PCMは進化生物学で提案され(Felsenstein 1985)、祖先種の状態や進化速度などのパラメータ推定や、回帰分析において進化史の共有に起因する相関を補正したりする目的で広く応用されている。特に後者については、環境変動の与える生物への影響を評価する際に複数の種を跨ぐ大規模なデータセット(例えば > 1000 種)でメタ解析を行う手法が広く普及したことに伴い近年急速に利用が広がっている(中川 & 久保 2016)。

既存のPCMには大きく分けガウス過程によるアプローチとシミュレーションによるアプローチの2つが知られてきた。ガウス過程によるアプローチでは、過去から現在に至る生物の進化をWiener過程やOrnstein-Uhlenbeck過程などのよく知られた確率過程としてモデル化する。このアプローチは推定量の信頼性が比較的評価しやすく、計算も容易なことから長くPCMで主流であった。しかし、モデルの柔軟性が不十分で実データへの適合や予測精度に問題があることが指摘されている(Losos 2011)。また、モデルの構造にマイクロ進化など他の生物学領域で知られた知見との乖離があり推定結果の解釈が困難であることも知られてきた(Pennell et al. 2013)。特にデータセットが大規模になると多様な進化パターンを遂げた種を内包するため、生物学的な背景理論を取り入れることのできる既存のモデルより柔軟性が高いモデルが求められてきた。

一方シミュレーションによるアプローチでは、集団遺伝学など関連領域の知見を取り入れた柔軟なモデルが可能となるが、閉じた形式の尤度が得られないことが多い。例えばKutsukake & Innan(2013; 2014)は系統樹上の一部に特異的な方向性選択が生じる過程をシミュレーションし、近似ベイズ計算(Approximate Bayesian Computation; Beaumont et al. 2002)を用いて計算する手法を提案した。この手法はガウス過程アプローチに比べ極めて柔軟な手法として評価されてきた(Pennell et al. 2013; Garamszegi 2014 など)一方で、尤度ベースの手法と比べ 1) 統計的信頼性が評価しにくい 2) 高い計算コストを伴う 3) 現状では各データセットに合わせた専用ソフトウェアが必要で、実務家にとって利用しづらい等の問題がある。

・提案手法

そこで本研究では、Kutsukake & Innan(2013; 2014)のシミュレーションモデルをガウス過程モデルとして再定式化した“枝特異的方向性淘汰モデル(branch-specific directional selection;

BSDS)モデル”を提案する。このモデルは2次オーダーで元のシミュレーションと等価で、高次は進化史が十分長ければ0に収束することを明らかにした。

・数値実験

ランダムに生成された系統樹のもとで Kutsukake & Innan(2013; 2014)のモデルに従う生物進化のシミュレーションを行い、生成されたデータに提案手法を適用した。その結果、祖先種における形質値、進化速度、淘汰の強さなどのパラメータを不偏に推定できることを明らかにした。一方、既存のガウス過程モデルと近似ベイズ計算による方法では、それぞれ共通祖先形質と淘汰圧を過大に推定することを明らかにした。

・実データ分析

霊長類のデータを用いて、ヒト(*Homo sapience*)の脳サイズがチンパンジー(*Pan troglodytes*)と分岐して以降にどの程度の淘汰圧を受けたか推定した。その結果、脳サイズが大きくなる方向に生じた変異が他の種に比べ1.73倍(95%確信区間 [1.62-1.87])多く生じており強い方向性をもった淘汰が生じたと推定された。

・今後の課題

本研究では系統樹上の方向性淘汰を生じた枝が既知であるという仮定で淘汰圧を推定する手法を検討し、提案したBSDSモデルが不偏な推定を可能にすることを明らかにした。しかし、より大規模なモデルについては事前に変数選択が必要になるなど、この仮定が不適当な場合がありうると考えられる。

一般に大規模なモデルを扱う場合には、近年スパース推定により必要な変数の選択とパラメータ推定を同時に行うアプローチが有効であると知られている。ところがPCMの場合、系統樹という複雑に構造化されたモデルを扱うため、よく知られたLASSOやGroup LASSOなどの手法では良い推定を与えない。今後、系統樹の構造を適切に反映した罰則項の検討や、その罰則項に対応する事前分布のモデル化などを検討する必要がある。

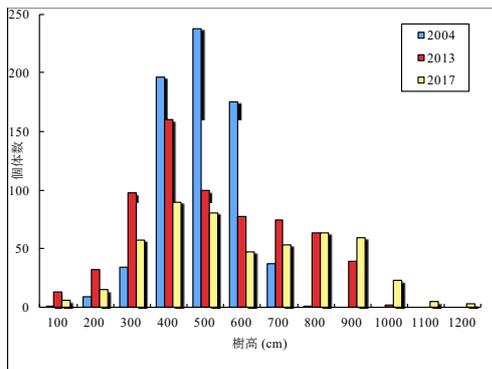
2次林再生における林冠個体の15年間変遷を観る統計モデル

島谷健一郎（統計数理研究所）

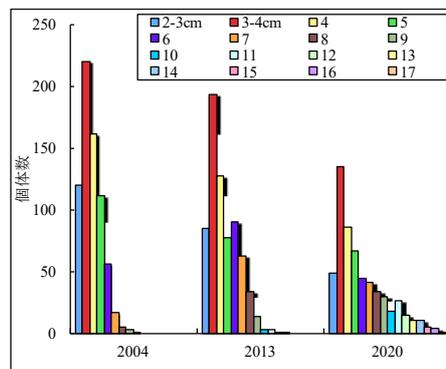
本研究では、1970–80年代に青森に設定されたブナ天然更新指標林で得ている、2005年からの15年間の樹木の動態データを分析した。

幅5m、長さ170mの調査地を設け、その中の胸高直径2.5cm以上の高木種について、樹種、胸高直径、樹高、階層（目視により1＝林冠層、2＝亜林冠層、3＝中—低木層、4＝倒れなど、5＝枯死）を4回、計測した。樹木の位置は、170mを5mごとに区切ったコード番号で与えた。計測したのは約900本である。その樹高と直径の変化を以下に示す。

樹高分布の変化



直径分布の変化



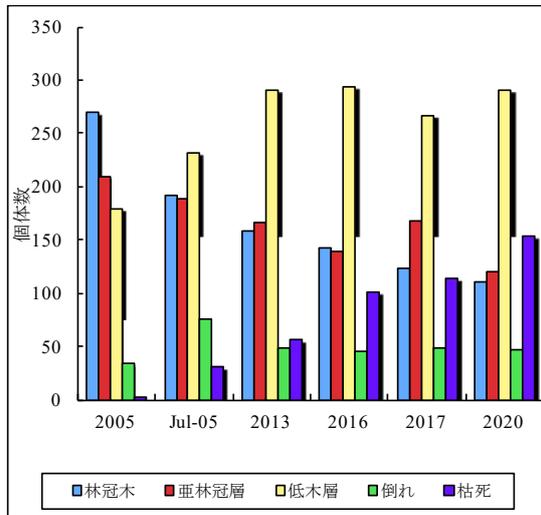
概して針葉樹、特に人工林では、最初に高かった個体が良好な光環境により成長が促進され、最初に出遅れた個体はその後塵を拝する。しかし、天然性広葉樹林では、その限りでなく、以下のような逆転現象も観測される。

個体2と個体3は同じくらいの樹高だったが、直径で劣っていた個体2が樹高でも直径でも階層でも先を進んでいる。個体1は個体2より高く直径はほぼ同じだったが、樹高でも階層でも後塵を拝している。

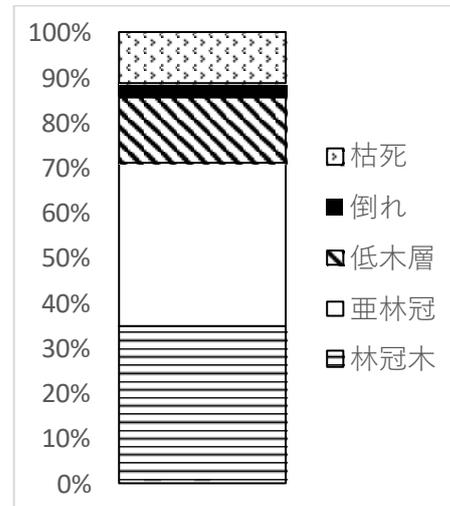
個体番号	樹高 (cm)			階層 (1 = 林冠、2 = 亜林冠)						直径 (cm)			
	2004	2013	2017	2005	2010	2013	2016	2017	2020	2004	2010	2016	2020
1	735	864	965	1	1	1	2	1	2	5.8	7.1	8.1	9.0
2	682	950	1042	1	1	1	1	1	1	5.9	7.2	9.2	11.0
3	694	795	920	1	2	2	2	2	2	6.2	7.3	8.3	8.8

当初多数観察された林冠層の個体は、年と共に競合により亜林冠層あるいは枯死といった運命をたどった。

階層分布の変化



2004年の林冠木の2020年の運命



本研究では、直径-樹高の関係式 (アロメトリー) に着目した。ずっと林冠層だったブナ111本を抽出した。直径と樹高については古くから様々な回帰式が提案されている。その主な目的は材積、バイオマス、2酸化炭素吸収量の推定などに置かれている。そのため、樹高測量はスナップショットで広い森林の代表的な樹高の個体に絞って行われることも多い。一方、本調査では、ブナ2次林の更新過程を詳細に追跡するべく、個体単位で樹高を正確に測量してきた。

直径 (x)-樹高 (y)アロメトリーでしばしば使われる数式に、Ogawa (1965) がある。

$$\frac{1}{y} = \frac{1}{ax^b} + \frac{1}{H_{Max}}$$

H_{max} は潜在的にとりうる樹木の樹高の最大値である。

この式は、単に当てはまりの良さから使われているのではなく、この式が、直径と樹高それぞれの相対成長が比例する、ただし樹高の成長は最大値が近づくとつれて漸減する、という以下の微分方程式から導かれるという特徴がある。

$$\frac{1}{y} \frac{dy}{dt} = b \frac{1}{x} \frac{dx}{dt} \left(1 - \frac{y}{H_{Max}}\right)$$

そこで、本研究では、5回の計測データを用いて、結果としての回帰式に加え、相対成長の関係についても吟味した。

結果として、相対成長比例定数 (b) は個体及び年による変動が見られ、上式が成り立っているとは言い難いが、決して見当違いな関係式にも見えなかった。個体や環境(金林個体との競合、密度など)による変異を適切に分散パラメータとして表現する階層モデル (Ogawa (1965) 式を状態方程式とし樹高などを観察モデルとする状態空間モデル) で定式化する方向性が伺えた。