

高次元時系列の線型判別分析

早稲田大学・応用数理学科 劉 言

早稲田大学・数学応用数理専攻, 統計数理研究所 長幡 英明

早稲田大学・理工学研究所 谷口 正信

1. はじめに

多変量時系列の母平均に対する線型判別分析はフィッシャーの判別基準から構成される. 変量の次元が観測数よりも十分に小さい時, 誤判別率が漸的に0に近づく. 一方で, 機械学習で直面する変量の次元が観測数よりも大きい状況では, フィッシャーの判別基準がたちまち上手くいかなくなる. 一方, 変量間の相関を0にした判別基準がより良い性質を示す. 本発表では, 高次元時系列の判別分析問題を取り扱い, 高次元判別基準の誤判別率が漸的に0となる条件を詳説する.

2. 高次元時系列データの線型判別分析

平均 μ で $p \times p$ 共分散行列関数 $\mathbf{R}(t) = \{R_{ij}(t); i, j = 1, \dots, p\}$ をもつ高次元定常時系列 $\{\mathbf{X}(t) = (X_1(t), \dots, X_p(t))'; t \in \mathbb{Z}\}$ は, 以下の二カテゴリーのいずれかに属するとする:

$$\pi_1: \mu = \mu^{(1)}, \quad \mathbf{R}(t) = \mathbf{R}^{(1)}(t), \quad \pi_2: \mu = \mu^{(2)}, \quad \mathbf{R}(t) = \mathbf{R}^{(2)}(t).$$

また, この二カテゴリーからそれぞれ観測数 n_1 と n_2 のトレーニングサンプル $\mathbf{X}^{(1)}$ と $\mathbf{X}^{(2)}$ が観測されている. 観測数 n の高次元時系列 $\mathbf{X} \equiv \{\mathbf{X}(1), \dots, \mathbf{X}(n)\}$ の判別基準を

$$\Gamma \equiv \Gamma(\mathbf{X}, \mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \left(\bar{\mathbf{X}} - \frac{\bar{\mathbf{X}}^{(1)} + \bar{\mathbf{X}}^{(2)}}{2} \right)' (\bar{\mathbf{X}}^{(2)} - \bar{\mathbf{X}}^{(1)}) \quad (2.1)$$

とし, $\Gamma < 0$ のとき, \mathbf{X} を π_1 , $\Gamma \geq 0$ のとき, \mathbf{X} を π_2 と分類する. Liu et al. [3] では, (2.1) に基づく高次元時系列の線型判別分析について, 十分大きな p に対し, $n_2 = c_0 n_1$ と $c_1 p^\eta < \|\mu^{(1)} - \mu^{(2)}\|^2 < c_2 p^\eta$ の二条件の下で, 判別基準の誤判別率が漸的に0となる (p, n, n_1, η) についての十分条件を与えた. しかし, 判別基準 Γ を利用する際, バイアスの問題を抱える. このバイアスを修正するため, Künsch [2] に従い, $l_1 = c_3 n_1^\alpha$ という部分ブロックによるジャックナイフ型の修正判別基準 Γ_{Jack} を導入した. この枠組で, $(p, n, n_1, \eta, \alpha)$ について誤判別率が漸的に0となるための十分条件を与えた. さらに, 部分ブロックの長さ l_1 に関する α の選び方について, ジャックナイフの対応するカーネルで分類し, 正規過程と非正規過程のそれぞれの条件の下での最適な α を求めた.

本発表はこれらの結果に加えて, 基準 (2.1) が漸的に0になるための十分条件より緩い, スパース性のある共分散行列関数の場合, 最適な高次元時系列判別基準に関する諸結果を紹介する.

References

- [1] AOSHIMA, M. and YATA, K. (2014). A distance-based, misclassification rate adjusted classifier for multiclass, high-dimensional data. *Annals of the Institute of Statistical Mathematics* **66** 983–1010.
- [2] KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The annals of Statistics* 1217–1241.
- [3] LIU, Y., NAGAHATA, H., UCHIYAMA, H. and TANIGUCHI, M. (2017). Discriminant and cluster analysis of possibly high-dimensional time series data by a class of disparities. *Communications in Statistics-Simulation and Computation*. DOI:10.1080/03610918.2016.1263732.