多次元時系列の変化点検出のための Selective Inference

梅津 佑太 [†]　　竹内 一郎 [†]

[†] 名古屋工業大学 大学院工学研究科

脳波やパネルデータなどを始め, 同時に観測された複数の時系列から構造の変化を捉えることは重要な課題として知られている. これまで, アルゴリズミックなものや漸近理論に基づくものなど, 様々な変化点検出手法が提案されてきた [1, 2, 4, 6]. 一方で, 漸近理論に基づく統計的変化点検出では, 行列の各成分が独立に正規分布に従うという仮定がしばしば用いられている. しかしながら, 時系列データに対する独立性の仮定は比較的厳しいものに思われる. そこで, 本研究では独立性の仮定を緩和するため, 近年 Lee et al. (2016) によって提案された selective inference [3] のアイデアを用いて多次元時系列の平均構造の変化を検出するための手法を提案する.

時点 $t = 1, \ldots, T$ での観測値を $\boldsymbol{y}_t \in \mathbb{R}^N$ とし, multivariate normal mean model $Y = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T) \sim \mathrm{N}_{N\times T}(M, \Xi \otimes \Sigma)$ を考える. ただし, $\Xi \in \mathbb{R}^{T\times T}$, $\Sigma \in \mathbb{R}^{N\times N}$ はそれぞれ既知の系列内共分散行列及び系列間共分散行列とする. また, $M = (\boldsymbol{m}_1, \ldots, \boldsymbol{m}_T)$ は平均構造であり, K 個の変化点 $\tau_1, \ldots, \tau_K$ を持つとする. つまり, 各変化点 τ_j で $\boldsymbol{m}_{\tau_j} \neq \boldsymbol{m}_{\tau_j+1}$ である. 本研究では, 観測された系列に対して, 時点 t を中心とする長さ $2h$ の窓 $W_h(t) = \{t-h+1, \ldots, t+h\}$ を考え, 以下の local hypothesis testing を定式化する:

$$\mathrm{H}_{0,t} : \boldsymbol{m}_u = \boldsymbol{m}_{u+1}, \quad {}^{\forall}u \in W_h(t) \qquad \text{vs.} \qquad \mathrm{H}_{1,t} : \boldsymbol{m}_t \neq \boldsymbol{m}_{t+1}. \tag{$*$}$$

直感的には, 検定問題 $(*)$ において帰無仮説 $\mathrm{H}_{0,t}$ が棄却された場合, 窓の中心 t が変化点であることが示唆される. 検定問題 $(*)$ を定式化するために, (多変量) cumulative sum statistics [5]

$$\boldsymbol{s}_u(t) = \left\{ \frac{2h}{(h-t+u)(h+t-u)} \right\}^{1/2} \sum_{v=t-h+1}^{u} \left(\boldsymbol{y}_v - \frac{1}{2h} \sum_{w \in W_h(t)} \boldsymbol{y}_w \right), \quad u \in W_h(t) \setminus \{t+h\}$$

を適当な写像 $\mathcal{F}$ で aggregate し, local change point estimates [7] によって変化点の推定量を構成する:

$$\hat{\mathcal{T}} = \left\{ t \in \{h, \ldots, T-h\} \mid \mathcal{F}(\boldsymbol{s}_t(t)) = \max_{u \in W_h(t) \setminus \{t+h\}} \mathcal{F}(\boldsymbol{s}_u(t)) \right\}.$$

local change point estimates によって選択された各変化点 $t \in \hat{\mathcal{T}}$ に対して, 検定統計量 $\mathcal{F}(\boldsymbol{s}_t(t))$ が閾値を超えた場合に帰無仮説 $\mathrm{H}_{0,t}$ を棄却する検定を構築することを考える. しかしながら, データに基づき検定すべき $|\hat{\mathcal{T}}|$ 個の仮説を選択しているため, selection bias を考慮した上で妥当な推論を行わなければならない. selective inference は, 仮説の選択手順を selection event として条件付けることで selection bias を調整するための手法である. 本研究では, local change point estimates の選択手順 $\{\mathcal{T} = \hat{\mathcal{T}}\}$ を selective event として条件付けることを考える. このとき, 適当なクラスの aggregation function $\mathcal{F}$ に対して, selective inference の意味での type I error をコントロールできることを示すことができる. シミュレーション結果については当日報告する.

References

[1] Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection, *The Annals of Statistics*, **42**, 2243–2281.

[2] Jirak, M. (2015). Uniform change point tests in high dimension, *The Annals of Statistics*, **43**, 2451–2483.

[3] Lee, J. D., Sun, D. L., Sun, Y., and Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso, *The Annals of Statistics*, **44**, 907–927.

[4] Li, S., Xie, Y., Dai, H., and Song, L. (2015). M-statistic for kernel change-point detection, In *Advances in Neural Information Processing Systems*, 3366–3374.

[5] Page, E. S. (1954). Continuous inspection schemes, *Biometrika*, **41**, 100–115.

[6] Wang, T. and Samworth, R. J. (2016). High-dimensional changepoint estimation via sparse projection, *arXiv preprint arXiv:1606.06246*.

[7] Yau, C. Y. and Zhao, Z. (2016). Inference for multiple change points in time series via likelihood ratio scan statistics, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **78**, 895-916.