ISSN　0385-2148

研究所報

No.56

様々な多様体上における統計的推測

２０２２年 3月

法政大学
日 本 統 計 研 究 所

はじめに

今回の研究集会では、様々な多様体上の統計的推測について各分野で活躍している研究者らに最近の研究発表をしていただいた。

最初に、鶴田靖人先生(長野県立大学)に方向データのためのカーネル密度推定について近年の研究について報告していただいた。次に、明石郁哉先生(東京大学)に裾の重い時変自己回帰モデルに対する頑健な推測手法についての紹介をしていただきながら、最新の研究成果について講演していただいた。塩濱敬之先生(南山大学)にはディスク上の分布の混合分布について講演していただき、その適用例として、画像データへの応用を紹介していただいた。白石博先生(慶應義塾大学)にはランダムフォレストを用いた時系列分位点回帰について講演していただいた。最後に蛭川潤一先生(新潟大学)には自己重み付きLAD推定量に関する紹介を含めた講演をしていただいた。

本研究集会でも講演中に活発なディスカッションがされ、発表後も近年の各研究分野での発展について情報交換ができ、大変有意義な研究集会となった。

2022年 3月　日本統計研究所

目次

方向データのためのカーネル密度推定量のバイアス修正

長野県立大学グローバルマネジメント学部助教　鶴田靖人

1　はじめに

方向データは，d 次元球面 $\mathbb{S}^d := \{\boldsymbol{x} \in \mathbb{R}^{d+1} : ||\boldsymbol{x}|| = 1\}$ 上に観測点が位置するデータである．方向データの例として，次元 $d = 2$ のときは過去の地磁気のデータなどがあり，$d \geq 3$ のときはテキストデータ（の各文書が含む単語総数を 1 に基準化したもの）がある．また，$d = 1$ のときは円周上のデータと呼ばれており，例えば風向などの角度データがある．カーネル密度推定法などのノンパラメトリック密度推定法は柔軟な推定が可能であるという利点がある．方向データのためのカーネル密度推定量の初期の研究は，Beran (1979), Hall et al. (1987) や Bai et al. (1989) がある．方向データ $X_1, \ldots, X_n$ は独立同一分布 $f(x) \quad (x \in \mathbb{S}^d)$ に従うとする．f についてのカーネル密度推定量は,

$$\hat{f}_\kappa(x) := \frac{1}{n}\sum_{i=1}^{n} K_\kappa(x^{\mathrm{T}} X_i),$$

ただし，$K_\kappa(x^{\mathrm{T}} X_i) : \mathbb{S}^d \to \mathbb{R}$ は球面上のカーネル関数，$\kappa := \kappa(n)$ は $\lim_{n\to\infty} \kappa = \infty$ かつ $\lim_{n\to\infty} n^{-1}\kappa^{d/2} = 0$ を満たす平滑化パラメータとする．ここで，$\eta \in \mathbb{S}^d$ とおく．カーネル関数を $K_\kappa(x^{\mathrm{T}}\eta) := C_\kappa(L)^{-1} L(\kappa(1 - x^{\mathrm{T}}\eta))$ と定義する．ただし，$L(r) : [0, \infty) \to \mathbb{R}$ は微分可能とし，$C_\kappa(L) := \int_{\mathbb{S}^d} L(\kappa(1 - x^{\mathrm{T}}\eta))w_d(x)$ は基準化定数を表す．よく用いられる誤差基準は，平均積分二乗誤差 (MISE) $\mathrm{MISE}[\hat{f}_\kappa(x)] := \int_{\mathbb{S}^d} \mathrm{E}[(\hat{f}_\kappa(x) - f(x))^2] w_d(dx)$ である（ただし，w_d は d 次元球面上のルベーグ測度を表す）．f が十分に滑らかであるという仮定の下で，MISE の収束レートは $O(n^{-4/(4+d)})$ であることが知られている (Hall et al. 1987)．つまり，次元 d が増加することで MISE の収束レートは悪化する（これを次元の呪いという）.

本研究の目的は，次元の呪いの影響を減らすためにカーネル密度推定量 $\hat{f}$ のバイアスを修正し，MISE の収束レートを改良することである．そのために, 実数空間上のバイアス修正法である Jones and Foster 型修正法 (Jones and Foster 1993) と Terrell and Scott 型修正法 (Terrell and Scott 1980) を $d \geq 2$ 次元の球面上のカーネル密度推定量に適用し，そのバイアスを修正可能なことを示す．本研究の提案手法は，Tsuruta and Sagae (2017) が円周上 ($d = 1$) のカーネル密度推定量のために提案したバイアス修正法を $d \geq 2$ 次元に拡張したものである.

2　提案手法

期待値 $\mathrm{E}[[\hat{f}_\kappa(x)]$ は，f をテイラー展開すると κ のべき乗に L のモーメント $\mu_l(L) := \int_0^\infty L(r) r^{(l+d-2)/2} dr$ をかけた項の線形和になる．したがって，次式のように低次のモーメントが 0 になるようなカーネル関数はバイアスを修正する.

$$\mu_0(L) \neq 0, \quad \mu_l(L) = 0, \quad l = 2, 4, 6, \ldots, p-2, \quad \mu_p(L) \neq 0, \tag{1}$$

ただし，$p \geq 2$ は偶数である．(1) を満たすカーネル関数を p 次オーダーカーネルと呼ぶ (Tsuruta and Sagae 2017)．p 次オーダーカーネルは以下のような性質を持つ．

定理 1. 緩やかな仮定の下で，p 次オーダーカーネルを採用したカーネル密度推定量のバイアスと分散はそれぞれ，$\mathrm{bias}[\hat{f}_\kappa(x)] = O(\kappa^{-p/2})$ と $\mathrm{Var}[\hat{f}_\kappa(x)] = O(n^{-1}\kappa^{d/2})$ となる．また，平滑化パラメータとして $\kappa_* = O(n^{2/(2p+d)})$ を選択すると $\mathrm{MISE}[\hat{f}_\kappa(x)]$ の収束レートは $O(n^{-2p/(2p+d)})$ となる．

p 次オーダーカーネルの構成法として，以下のような Jones and Foster 型修正法を提案する．

定義 1. p 次オーダーカーネルの関数 L を $L_{[p]}$ と表す．このとき，$p+2$ 次オーダーカーネルの関数 L を次式のように定義する．

$$L_{[p+2]}(r) := \frac{p+d}{p} L_{[p]}(r) + \frac{2}{p} r L'_{[p]}(r).$$

Jones and Foster 型修正法は低次のモーメントが 0 となるように 2 つの関数 L を足し合わせたものである．低次のモーメントが 0 となるように 2 つの異なるカーネル密度推定量をかけ合わせることでもバイアスを修正することができる．Terell and Scot 型のバイアス修正法を次式のように定義する．

定義 2.

$$\tilde{f}_\kappa(x) = \left[\hat{f}_\kappa(x)\right]^{1/(1-a)} \left[\hat{f}_{a\kappa}(x)\right]^{-a/(1-a)},$$

ただし，　$a \in (0,1)$ とする．

Terell and Scot 型のバイアス修正法は以下の性質を持つ．

定理 2. 緩やかな仮定の下で，定義 *2* で与えたカーネル密度推定量のバイアスと分散はそれぞれ，$\mathrm{bias}[\tilde{f}_\kappa(x)] = O(\kappa^{-2})$ と $\mathrm{Var}[\tilde{f}_\kappa(x)] = O(n^{-1}\kappa^{d/2})$ となる．また，平滑化パラメータとして $\kappa_* = O(n^{2/(8+d)})$ を選択すると $\mathrm{MISE}[\tilde{f}_\kappa(x)]$ の収束レートは $O(n^{-8/(8+d)})$ となる．

したがって，Terell and Scot 型のバイアス修正法は MISE の収束レートが 4 次オーダーカーネルに対応するカーネル密度推定量であると言える．

参考文献

ZD Bai, C Radhakrishna Rao, and LC Zhao. Kernel estimators of density function of directional data. *Multivariate Statistics and Probability*, 27:24–39, 1989.

Rudolf Beran. Exponential models for directional data. *The Annals of Statistics*, 7(6):1162–1178, 1979.

Peter Hall, GS Watson, and Javier Cabrera. Kernel density estimation with spherical data. *Biometrika*, 74(4):751–762, 1987.

MC Jones and PJ Foster. Generalized jackknifing and higher order kernels. *Journal of Nonparametric Statistics*, 3(1):81–94, 1993.

George R Terrell and David W Scott. On improving convergence rates for nonnegative kernel density estimators. *The Annals of Statistics*, 8(5):1160–1163, 1980.

Yasuhito Tsuruta and Masahiko Sagae. Higher order kernel density estimation on the circle. *Statistics & Probability Letters*, 131:46–50, 2017.

裾の重い時変自己回帰モデルに対する頑健な推測手法について

明石郁哉 (東京大学)

Abstract

本講演では, 裾の厚い誤差項を持つ自己回帰モデルの係数について頑健な推測手法を提案する. 特に自己加重法と呼ばれる手法を最小絶対誤差回帰と併せて用いることにより, モデルの無限分散性を制御し漸近正規分布を持つ推定量を構成する. また, 漸近分散の推定を回避しつつ係数の信頼領域を構成するために, 一般化経験尤度統計量を用いた推定手法を構成する.

1 モデル

観測系列 $Y_{1,n},...,Y_{n,n}$ は, p 次の時間変動自己回帰モデル

$$Y_{t,n} = \beta(t/n)^\top X_{t-1,n} + \varepsilon_t \quad (t = 1,...,n) \tag{1}$$

から生成されたものとする. ここで, $\beta(u) := (\beta_1(u),...,\beta_p(u))^\top$ は $[0,1]$ における滑らかな関数, $X_{t-1,n} := (Y_{t-1,n},...,Y_{t-p,n})^\top$ であり, $\{\varepsilon_t : t \in \mathbb{Z}\}$ は独立同分布な誤差過程で, 原点で正の値を持つ確率密度関数をもつとする. 本研究の特徴として, 誤差変数 ε_t は中央値がゼロであることを仮定するが, 有限分散は仮定しない. そのため, モデル (1) は無限分散・有限分散モデル両方を許容する広範なものとなる.

2 自己加重法による推定量の構成

局所定常過程における時間変動係数の推定量としては, Dahlhaus and Rao (2006) が提案した時間方向でのカーネル型推定量が考えられる. 本研究ではさらに, Ling (2005) の提案した自己加重法を用いて統計量頑健化を行う. 具体的には, ある $u_0 \in (0,1)$ における係数 $\beta(u_0)$ の推定量として, 目的関数

$$\sum_{t=p+1}^{n} w_{t-1,n} K\left(\frac{t-nu_0}{nh}\right) |Y_{t,n} - b^\top X_{t-1,n}| \tag{2}$$

を最小化する b を推定量として用いることで自己加重型の推定量を構成する. ここで K は有界な台を持つカーネル関数, h は $h \to 0$ $(n \to \infty)$ であるバンド幅, $w_{t-1,n}$ は $X_{t-1,n}$ の可測関数であり, 自己加重 (self-weight, Ling (2005)) と呼ばれる. $w_{t-1,n}$ を適切に選択することで, $X_{t-1,n}$ の無限分散を制御し, 推定量の漸近展開に現れる項を 2 乗可積分なマルチンゲール差分アレイに変換することができる. 本講演ではさらに, ペナルティ部分 $|Y_{t,n} - b^\top X_{t-1}|$ の部分を 1 次線形近似したものに置き換えて, モデル (1) の係数に対する局所線形近似自己加重型推定量を構成し, 推定量の一致性と漸近正規性を示した.

3　一般化経験尤度法による信頼領域の構成

モデル係数の点推定量が構成できた場合, 次に興味があるのは $\beta(u_0)$ の信頼領域の構成である. 推定量の漸近正規性が保証されているので, 漸近分散の推定に基づく手法が考えられるが, 推定量の漸近分散には誤差過程の確率密度関数が含まれ, 推定のためにはさらなるカーネル密度推定量を用いる必要があり, 煩雑である. また本講演の設定では, 誤差過程の分布は有限次元のパラメータで規定されているとは限らないため尤度の計算も直接は不可能である. そこで本講演では, Owen (1988) により提案された経験尤度法に基づく信頼区間の構成を行う. いま, ある分布関数 F に依存するパラメータ θ (母平均, 高次モーメントなど) の真値 θ_0 の推測に興味があり, 制約条件 $E_F(m(Z,\theta_0)) = 0$ が成り立っているとする. ここで, Z は分布関数 F を持つ確率変数, E_F は F に基づく期待値, $m : \mathcal{Z} \times \Theta \to \mathbb{R}$ は解析者が定める可測関数, $\mathcal{Z}$ は分布関数 F の台, Θ は候補のパラメータ集合である. F から独立同分布に観測系列 $Z_1, ..., Z_n$ が得られた場合, 候補の分布 F の下での非母数的な尤度は

$$L_n(F) := \prod_{i=1}^{n} \left\{F(Z_i) - F^-(Z_i)\right\} \quad \left(F^-(z) := \lim_{u\uparrow z} F(u)\right)$$

で与えられる. そこで制約条件 $E_F(m(Z,\theta)) = 0$ の下での経験尤度比を,

$$R_n(\theta) := \frac{\sup\left\{L_n(F) : F \in \mathbb{F}_n(\theta)\right\}}{L_n(\hat{F}_n)} \tag{3}$$

で定める. ここで $\hat{F}_n$ は観測系列 $Z_1, ..., Z_n$ に基づく経験分布関数であり, $\mathbb{F}_n(\theta)$ は分布関数の族で

$$\mathbb{F}_n(\theta) := \left\{\sum_{i=1}^{n} v_i \mathbb{I}(Z_t \le x) : v_i \in [0,1] \ (\forall i \in \{1, ..., n\}), \sum_{i=1}^{n} v_i m(Z_t, \theta) = 0, \sum_{i=1}^{n} v_i = 1\right\}$$

と定義される. Lagrange の未定乗数法より, $-2\log R_n(\theta)$ は

$$-2\log R_n(\theta) = 2 \sup_{\lambda \in \hat{\Lambda}_n^{\mathrm{EL}}(\theta)} \sum_{i=1}^{n} \log(1 - \lambda m(Z_i, \theta))$$

と表される. ここで $\hat{\Lambda}_n^{\mathrm{EL}}(\theta) = \{\lambda : \lambda \in \mathbb{R}, \lambda m(Z_i,\theta) < 1 \ (\forall i \in \{1, ..., n\})\}$ である. さらに Smith (1997) は経験尤度統計量を拡張し, 一般化経験尤度 (GEL) 統計量を提案した. GEL 法の特徴として, 分布に対する緩い仮定の下で, 統計量は漸近的に通常のカイ 2 乗分布に収束することが挙げられる. そこで本講演では, 自己加重型のモーメント関数

$$m_{t,n}(b) = \frac{1}{h^{1/2}} K\left(\frac{t - nu_0}{nh}\right) w_{t-1,n} \mathrm{sign}(Y_{t,n} - b^\top X_{t-1,n}) \quad (b \in \mathbb{R}^p)$$

を定義し, GEL 統計量

$$r_n^*(b) := 2 \sup_{\lambda \in \hat{\Lambda}_n^{\rho}(b)} \sum_{t=p+1}^{n} \rho(\lambda m_{t,n}(b))$$

を定める. ここで ρ は解析者が定める関数で, $\rho(0) = 0$, $\rho'(0) = \rho''(0) = -1$ を満たす凸関数であり, 定義域 $\mathcal{V}_\rho$ を持つ関数である. また $\hat{\Lambda}_n^{\rho}(b) = \{\lambda : \lambda \in \mathbb{R}, \lambda m_{t,n}(b) \in \mathcal{V}_\rho \ (\forall t \in \{p+1, ..., n\})\}$ である. $m_{t,n}(b)$ は自己加重型最小絶対値回帰の目的関数 (2) の勾配とと

らえることができるので, GEL 法の目的関数として自然なものである. 結果として, 真値 $\beta(u_0)$ の下で $r_n^*(\beta(u_0)) \xrightarrow{d} \chi_p^2 \ (n \to \infty)$ を示した. この結果に基づき, $\mathbb{R}^p$ における $\beta(u_0)$ の信頼係数 $(1-\delta)$ の信頼領域を

$$\hat{C}_n := \{b : b \in \mathbb{R}, r_n^*(b) \leq \chi_{1-\delta}^2\} \tag{4}$$

と提案した. ここで χ_α^2 はカイ 2 乗分布の α 分位数である. 統計量の計算や漸近分布の分位数の計算には, 局外母数（誤差過程の裾指数など）の推定が不要であり, $\hat{C}_n$ は標本から構成可能な頑健な信頼領域となっている.

講演では, 有限標本における提案推定量・統計量の振る舞いを数値実験により確認した. 特に最小二乗回帰に基づく推定量との比較を行い, 良好な結果を確認した.

References

Dahlhaus, R. and Rao, S. S. (2006). Statistical inference for time-varying ARCH processes. *Annals of Statistics*, Vol. 34, pp. 1075–1114.

Ling, S. (2005). Self-weighted least absolute deviation estimation for infinite variance autoregressive models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, Vol. 67, pp. 381–393.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, pp. 237–249.

Smith, R. J. (1997). Alternative semi-parametric likelihood approaches to generalised method of moments estimation. *The Economic Journal*, Vol. 107, pp. 503–519.

Estimating Finite Mixtures of Disc Distribution and Its Applications to Image Recognition

Takayuki Shiohama
Department of Data Science, Faculty of Science and engineering
Nanan University

Abstract

Analyzing data on HSV color space can provide useful information for image segmentation and classification. These data take values on a special manifold such that the pairs of the hue and saturation variables are located on the unit disc, and the trivariate variables including the value have data on the inside of the cylinder. Implementing these special geometric manifolds into statistical data analysis needs special attention. In this study, we consider the mixtures of the disc distribution and provide some methodologies for estimating model parameters. These estimation techniques are applied for image recognitions using cifer10 datasets.

1 Introduction

Several machine learning techniques on image processing have widely evolved in the last two decades. Various algorithms are applied to image recognition, including object recognition and face recognition, image search, and image manipulation. The most successful algorithm is deep learning or the Convolutional Neural Networks (CNNs) as it provides a method that can extract image features and applies these features on automated unsupervised classification. AlexNet (Krizhevsky et al., 2012) and ResNet (He et al., 2016) are some of the popular CNNs, that have improved performances of image recognition by increasing network depths.

In general, image processing is mainly investigated on the RGB color space which represents images as a three-dimensional Cartesian coordinate system. Each of the pixels has value on the three-dimensional cube. It is often reported that the disadvantages of the RGB color space are the redundancy in color representation. Some approaches on reducing redundancy in color image date are found in, for example, Bhurchandi et al. (2000) and Marguier (2010). In this study, we investigate image classification based on HSV color space instead of using RGB color space. Data on HSV color space are consists of special manifolds known as a cylinder or a disc. Statistical analysis on these manifolds needs to pay special attention. Applications on the image segmentation based on the HSV color space can be found in Cantrell et al. (2010), Ganesan et al. (2014), and the references therein.

There are few distributions defined on the unit disc and its extension for higher-dimensional space. The first distribution on the unit disc is proposed byJones (2004) and its extension is investigated by Wang and Shimizu (2012). On the other hand, the copula type distribution proposed by Wehrly and Johnson (1980) can be applied for fitting data on the unit disc. Applying these existing distributions for fitting HSV color data in image processing does not provide satisfactory fitting performance, since these distributions do not have peaks on the origin. It is often the case that the image data contains white pixels located at the origin of the unit dice. To overcome this shortcoming, we consider practical distribution and its mixtures for analyzing image data on HSV color space.

The rest of this paper is organized as follows. Section 2 introduces the distributions on the unit disc. In Section 3, we introduce a finite mixture of disc distributions and discuss the parameter estimation procedures for it. Section 4 illustrates the image data analysis using Cifer10 datasets. Finally, in Section 5, we conclude the paper with discussions and further research topics.

References

Bhurchandi, K. M., Nawghare, P., and Ray, A. (2000). An analytical approach for sampling the rgb color space considering physiological limitations of human vision and its application for color image analysis. In *Proceedings of Indian Conference on Computer Vision, Graphics and Image Processing*, pages 44–49. Citeseer.

Cantrell, K., Erenas, M., de Orbe-Payá, I., and Capitán-Vallvey, L. (2010). Use of the hue parameter of the hue, saturation, value color space as a quantitative analytical parameter for bitonal optical sensors. *Analytical chemistry*, 82(2):531–542.

Ganesan, P., Rajini, V., Sathish, B., and Shaik, K. B. (2014). Hsv color space based segmentation of region of interest in satellite images. In *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, pages 101–105. IEEE.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778.

Jones, M. (2004). The möbius distribution on the disc. *Annals of the Institute of Statistical Mathematics*, 56(4):733–742.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25.

Marguier, J. (2010). Exploiting redundancy in color images. Technical report, EPFL.

Wang, M. and Shimizu, K. (2012). On applying möobius transformation to cardioid random variables. *Statistical Methodology*, 9:604–614.

Wehrly, T. E. and Johnson, R. A. (1980). Bivariate models for dependence of angular observations and a related markov process. *Biometrika*, 67(1):255–256.

Time Series Quantile Regressions by using Random Forest
(ランダムフォレストを用いた時系列分位点回帰)

Ryotaro Shibuki* , Tomoshige Nakamura† , Hiroshi Shiraishi‡

In this paper, we discuss an estimation procedure of conditional quantile by using random forests in time series setting. Our study is an extension of the quantile random forest (QRF) by Meinshausen (2006), the generarized random forest (GRF) by Athey et. al (2019) and random forest in time series setting by Davis and Nielsen (2020).

Model Let $(\varepsilon_t)_{t\geq 1}$be a sequence of i.i.d. random variables with $\mathbb{E}[\varepsilon_t] = 0$ and $\mathbb{E}[\varepsilon_t^4] < \infty$, and fix an integer $p \geq 1$. Given a measurable function $g : \mathbb{R}^p \to \mathbb{R}$, define the process $(Y_t)_{t\geq 1}$ recursively by

$$Y_t = g(X_t) + \varepsilon_t, \quad \boldsymbol{X}_t = (Y_{t-1}, \ldots, Y_{t-p}). \tag{1}$$

In addition to the initial data $\eta = (Y_0, Y_{-1}, \cdots, Y_{1-p})$, suppose that we have T observations $Y_1, \ldots, Y_T$ from the model (1) available and that we group them in input-output pairs, $\mathcal{D}_T = \{(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_T, Y_T)\}$. For each fixed value $\tau \in (0,1)$, we seek forest-based (function) estimator of $q_*^\tau : \mathbb{R}^p \to \mathbb{R}$ defined by a solution of local estimating equation of the form

$$\Psi^\tau(q_*^\tau, \boldsymbol{x}) := \mathbb{E}[\psi^\tau_{q_*^\tau}(Y_t)|\boldsymbol{X}_t = \boldsymbol{x}] = 0, \quad \text{for all } \boldsymbol{x} \in \mathbb{R}_t^p \tag{2}$$

where $\psi_q^\tau(y) = \tau - \mathbf{1}_{\{y\leq q\}}$. Let $q_*^\tau \equiv (q_*^\tau(\boldsymbol{x}))_{\boldsymbol{x}\in\mathbb{R}^p}$ be the solution of (2) under the model (1). We assume that there exists $q_*^\tau(\boldsymbol{x})$ for all $\boldsymbol{x} \in \mathbb{R}^p$ and $\tau \in (0,1)$.

Double sample We next define our random forests following Athey et. al (2019). Our random forests consists of the double sample trees, which are regression trees based on two subsamples $\mathcal{I}_s$ and $\mathcal{J}_s$ from sample $\mathcal{D}_T$.

Definition 1. *(Double Sample) Suppose that sample $\mathcal{D}_T$ is available and the sub-sample size $s = s(T)$ with $s \leq T$ is provided. Let*

$$\mathcal{A}_s := \left\{ A = \{A^{\mathcal{I}}, A^{\mathcal{J}}\},\ A^{\mathcal{I}}, A^{\mathcal{J}} \subset \{1, 2, \ldots, T\} \middle| A^{\mathcal{I}} \cap A^{\mathcal{J}} = \emptyset,\ |A^{\mathcal{I}}| = \left\lfloor \frac{s}{2} \right\rfloor,\ |A^{\mathcal{J}}| = \left\lceil \frac{s}{2} \right\rceil \right\}$$

For any $A = \{A^{\mathcal{I}}, A^{\mathcal{J}}\} \in \mathcal{A}_s$, we define two sub-samples $\mathcal{I}_s$ and $\mathcal{J}_s$ by $\mathcal{I}_s = \mathcal{D}_{A^{\mathcal{I}}}, \mathcal{J}_s = \mathcal{D}_{A^{\mathcal{J}}}$ where $\mathcal{D}_{\boldsymbol{\theta}^\cdot} = \{(\boldsymbol{X}_t, Y_t)\}_{t\in A^\cdot}$.

Splitting rule We next define splitting rule in order to construct the double-sample regression trees following Wager and Athey (2018).

Definition 2. *(Splitting rule) Given subsample $\mathcal{J}_s$ in Definition 1, we define a sequence of partitions $\mathcal{P}_0, \mathcal{P}_1, \ldots$ by starting form $\mathcal{P}_0 = \{\mathbb{R}^p\}$ and then, for each $\ell \geq 1$, construct $\mathcal{P}_\ell$ from $\mathcal{P}_{\ell-1}$ by replacing one set (parent node) $P \in \mathcal{P}_{\ell-1}$ by (child node) $C_1 := \{\boldsymbol{x} = (x_1, \ldots, x_p) \in P \subset \mathbb{R}^p : x_\xi \leq \zeta\}$ and $C_2 := \{\boldsymbol{x} = (x_1, \ldots, x_p) \in P \subset \mathbb{R}^p : x_\xi > \zeta\}$, where the split direction $\xi \in \{1, \ldots, p\}$ is randomly chosen (i.e., random split) .*

*Graduate School of Science and Technology, Keio University
†Faculty of Science and Technology, Keio University
‡Faculty of Science and Technology, Keio University

Double-sample regression trees A given partition Λ of $\mathbb{R}^p$ is called "recursive" if $\Lambda = \mathcal{P}_\ell$ for some $\ell \geq 0$, where $\mathcal{P}_0, \ldots, \mathcal{P}_\ell$ are obtained as above. Note that the splitting rules determining how to choose node, direction and position of a split may depend on the data $\mathcal{D}_T$, the double sampling procedure $A \in \mathcal{A}_s$ and the sequence of independent random splittings $\boldsymbol{\xi} = \{\xi_i\}_{i=1,\ldots,\ell}$ with $\xi_i \overset{i.i.d.}{\sim} \Xi$. By using the recursive partition Λ, we define our double-sample regression trees.

Definition 3. *Given a recursive partition $\Lambda(A, \boldsymbol{\xi}, \mathcal{D}_T) = \{L_1, \ldots, L_{|\Lambda|}\}$ and a fixed $\boldsymbol{x} \in \mathbb{R}^d, q \in \mathbb{R}, \tau \in (0,1)$, our double-sample regression tree $\mathcal{T}(q, \boldsymbol{x}; A, \boldsymbol{\xi}, \mathcal{D}_T)$ is defined by*

$$\mathcal{T}(q, \boldsymbol{x}; A, \boldsymbol{\xi}, \mathcal{D}_T) = \sum_{t \in A^{\mathcal{I}}} \frac{\mathbf{1}_{\{\boldsymbol{X}_t \in L_A(\boldsymbol{x})\}}}{|L_A(\boldsymbol{x})|} \psi_q^\tau(Y_t)$$

where $L_A(\boldsymbol{x}) = \{\boldsymbol{X}_t : \boldsymbol{X}_t \in \tilde{L}(\boldsymbol{x})\} \cap \mathcal{I}_s^X$, $\mathcal{I}_s^X = \{\boldsymbol{X}_t\}_{t \in \mathcal{I}_s}$ and $\tilde{L}(\boldsymbol{x}) \in \Lambda(A, \boldsymbol{\xi}, \mathcal{D}_T)$ is a leaf containing $\boldsymbol{x}$ (i.e., $\boldsymbol{x} \in \tilde{L}_A(\boldsymbol{x})$).

Random Forests According to Wager and Athey (2018), the predictor $\mathcal{T}$ defined by Definition 3 is called "k- PNN predictor" if the assumption (A-3) is satisfied. Then, we define our random forests following Athey et. al (2019).

Definition 4. *For a fixed $\boldsymbol{x} \in \mathbb{R}^d, q \in \mathbb{R}, \tau \in (0,1)$, our forest score is defined by*

$$\Psi_T^\tau(q, \boldsymbol{x}) := \frac{1}{|\mathcal{A}_s|} \sum_{A \in \mathcal{A}_s} \mathcal{T}(q, \boldsymbol{x}; A, \boldsymbol{\xi}, \mathcal{D}_T) =: \sum_{t=1}^{T} \alpha_t(\boldsymbol{x}) \psi_q^\tau(Y_t)$$

where $\alpha_t(\boldsymbol{x}) = \frac{1}{|\mathcal{A}_s|} \sum_{A \in \mathcal{A}_s} \alpha_{A,t}(\boldsymbol{x})$ and $\alpha_{A,t}(\boldsymbol{x}) = \frac{\mathbf{1}_{\{\boldsymbol{X}_t \in L_A(\boldsymbol{x})\}}}{|L_A(\boldsymbol{x})|}$.

Quantile estimator By using the above random forest, we can define an estimator of the conditional quantile as follows.

Definition 5. *For each $\tau \in (0,1)$ and given $X_t = x$, we define an estimator of $q_*^\tau(x)$ by*

$$\hat{q}_T^\tau(x) \in \arg\min_{q \in \mathbb{R}} \left\{ \left\| \sum_{t=1}^{T} \alpha_t(x) \psi_q^\tau(Y_t) \right\|_2 \right\}.$$

Theorem 1. *Under some regularity conditions and subsample size $s(T)$ satisfies $s(T)/T \to 0$ and $s(T) \to \infty$ as $T \to \infty$. For each $\tau \in (0,1)$ and $\boldsymbol{x} \in \mathbb{R}^p$, any sequence of estimators $\hat{q}_T^\tau(\boldsymbol{x})$ converges in probability to $q_*^\tau(\boldsymbol{x})$, that is,*

$$\hat{q}_T^\tau(\boldsymbol{x}) \xrightarrow{p} q_*^\tau(\boldsymbol{x}) \quad \text{as } T \to \infty.$$

参考文献

An, H. Z., and F. C. Huang. (1996). The geometrical ergodicity of nonlinear autoregressive models. *Statist. Sinical*, 6(4), 943-956.

Athey, P., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *Annals of Statistics*, 47(2), 1148-1178.

Davis, Richard A. and Nielsen, Mikkel S. (2020). Modeling of time series using random forests: theoretical developments. *Electronic Journal of Statistics*, 14(2), 3644-3671.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research (JMLR)*, 7, 983-999.

Wager, S. and Athey, P. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523), 1228-1242.

The self-weighted LAD estimator for unit root process with locally stationary innovations

Junichi Hirukawa[1]

[1]*Niigata University*

1 Introduction

In most cases, the limiting distributions of the integrated time series are given by certain functionals of Brownian motion. The main stream of this research area is based on the least squares estimators and those resutls on unit root estimator can be found in Phillips (1987a, 1987b), Perron (1988), Phillips and Durlauf (1986), and references therein. Phillips (1990) and Chan and Tran (1989) extended it for infinite variance innovation cases, in which the limiting distributions are given by the functionals of Lévy processes. The least absolute deviations (LAD) estimation of the autoregressive parameter in unit root processes with infinite variance innovations are considered in Knight (1989, 1991) and Davis, Knight and Liu (1992). The limiting distribution of the LAD estimators are functionals of a bivariate Brownian motion or combination of Brownian motion and Lévy process. Herce (1996) derived the limiting distribution of LAD estimator of the unit root process when the innovation process is given by the general linear (MA (∞)) processes. To undertake statistical inference for infinite variance autoregressive models, Ling (2005) proposed a self-weighted least absolute deviation estimator and showed that this estimator is asymptotically normal. Although the unit root inferences under stationary assumption on the innovation processes are well established, empirical studies show that the innovation processes are not constant in time. One of the most important classes of nonstationary processes has been formulated in a rigorous asymptotic framework by Dahlhaus (1996a,b, 1997 and 2000), called locally stationary processes. Locally stationary processes have time-varying spectral densities whose spectral structures smoothly change in time. In this talk, we derive the asymptotic distribution of the self weighted LAD estimator of the first-order autoregressive parameter under the unit root hypothesis with locally stationary and α-stable innovation processes.

The original random walk process of i.i.d. innovations is given by

$$r_t = r_{t-1} + \varepsilon_t, \quad r_0 = 0, \tag{1}$$

where $\{\varepsilon_t\}$ is a sequence of independent, identically distributed random variables. We consider two different cases. That is, (i) $\{\varepsilon_t\}$ is mean zero and has the finite variance $\mathrm{Var}(\varepsilon_t) = \sigma^2 < \infty$. In this case, we set $a_T = \sqrt{T}$. (ii) $\{\varepsilon_t\}$ belongs to the α-stable domain of attraction $\mathcal{D}(\alpha)$, $0 < \alpha < 2$. We also assume that $E(\varepsilon_t) = 0$ for $1 < \alpha < 2$, ε_t is symmetric for $\alpha = 1$ and the normalizing constant is given by a_T. For the random walk process (1), we construct the partial sum process

$$D_T(s) := \frac{1}{a_T} \sum_{t=1}^{[sT]} \varepsilon_t. \tag{2}$$

Then, this partial sum process satisfies the FCLT

$$D_T(s) \Rightarrow D(s) = \sigma W_1(s) \tag{3}$$

for the finite variance case and the stable law

$$D_T(s) \Rightarrow D(s) = S_\alpha(s) \tag{4}$$

for α-stable case with $0 < \alpha < 2$, where $\Rightarrow$ denotes the convergence in distribution in $D(0,1)$ with J_1-topology, $W_1(\cdot)$ is the standard Brownian motion and $S_\alpha(\cdot)$ is the stable process with index α.

2 The self-weighted LAD estimator for unit root process with locally stationary innovations

Now, we consider the univariate time series $y_{t,T}$, generated according to

$$y_{t,T} = \beta_0 y_{t-1,T} + u_{t,T}, \quad \beta_0 = 1,\ t = 1, 2, \ldots,$$

where we assume that $y_{0,T} := y_0 \equiv 0$ to simplify the notation, and $\{u_{t,T}\}$ is a time varying MA (∞) process.

We adopt the self-weighted least absolute deviations estimator (SWLADE) defined as

$$\widehat{\beta}_{SWLAD} := \arg\min_{\beta\in\mathbb{R}} \{L_T(\beta)\}$$

with

$$L_T(\beta) := \frac{1}{a_T\sqrt{T}} \sum_{t=1}^{T} w_{t-1,T} \left| y_{t,T} - \beta y_{t-1,T} \right|,$$

where $w_{t-1,T}$ is a given real and positive function of $y_{t-1,T}$.

Then, we can show that the following result.

Theorem 1. *It holds that, as $T \to \infty$,*

$$a_T\sqrt{T}\left(\widehat{\beta}_{SWLAD} - 1\right) \Rightarrow \frac{A}{2B}.$$

References

ASTRAUSKAS, A. (1983). Limit theorems for sums of linearly generated random variables. *Lithuanian Mathematical Journal* **23**, 127–134.

AVRAM, F. AND TAQQU, M. S. (1992). Weak Convergence of Sums of Moving Averages in the α-Stable Domain of Attraction. *The Annals of Probability* **20**, 483–503.

BEVERIDGE, S AND NELSON, S. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary Economics* **7**, 151–174.

CHAN, N. H. AND TRAN, L. T. (1989). On the First-Order Autoregressive Process with Infinite Variance. *Econometric Theory* **5**, 354–362.

DAHLHAUS, R (1996a). On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Processes and their Applications* **62**, 139–168.

DAHLHAUS, R (1996b). Maximum likelihood estimation and model selection for locally stationary processes. *Journal of Nonparametric Statistics* **6**, 171–191.

DAHLHAUS, R (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics* **25**, 1–37.

DAHLHAUS, R (2000). A likelihood approximation for locally stationary processes. *The Annals of Statistics* **6**, 1762–1794.

DAVIS, R.A., KNIGHT, K. AND LIU, J. (1992). M-estimation for autoregressions with infinite variance. *tochastic Processes and their Applications* **40**, 145–180.

DAVIS, R. AND RESNICK, S. (1985). Limit Theory for Moving Averages of Random Variables with Regularly Varying Tail Probabilities. *The Annals of Probability* **13**, 179–195.

HERCE, M. A. (1996). Asymptotic Theory of "LAD" Estimation in a Unit Root Process with Finite Variance Errors. *Econometric Theory* **12**, 129–153.

KNIGHT, K. (1989). Limit Theory for Autoregressive-Parameter Estimates in an Infinite-Variance Random Walk. *The Canadian Journal of Statistics* **17**, 261–278.

KNIGHT, K. (1991). Limit Theory for M-Estimates in an Integrated Infinite Variance. *Econometric Theory* **7**, 200–212.

KNIGHT, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *The Annals of Statistics* **26**, 755–770.

LING, S. (2005). Self-Weighted Least Absolute Deviation Estimation for Infinite Variance Autoregressive Models. *Journal of the Royal Statistical Society. Series B* **67**, 381–393.

PERRON, P. (1988). Trends and random walks in macroeconomic time series: Further evidence from a new approach. *Journal of Economic Dynamics and Control* **12**, 297–332.

PHILLIPS, P. C. B. (1987a). Time Series Regression with a Unit Root. *Econometrica* **55**, 277–301.

PHILLIPS, P. C. B. (1987b). Towards a Unified Asymptotic Theory for Autoregression. *Biometrika* **74**, 535–547.

PHILLIPS, P. C. B. (1990). Time Series Regression with a Unit Root and Infinite-Variance Errors. *Econometric Theory* **6**, 44–62.

PHILLIPS, P. C. B. AND DURLAUF,S. N. (1986). Multiple Time Series Regression with Integrated Processes. *The Review of Economic Studies* **53**, 473–495.

RESNICK, S. AND GREENWOOD, P (1979). A bivariate stable characterization and domains of attraction. *Journal of Multivariate Analysis* **9**, 206-221.

SAMARAKOON, D. M. M. AND KNIGHT, K (2009). A Note on Unit Root Tests with Infinite Variance Noise. *Econometric Reviews* **28**, 314–334.

研究所報（最近刊行分）

号数	タイトル	刊行年月日
32	ミクロデータとその利用	2004.04.20
33	International Symposia on Population Census and Micro Data Archives	2005.01.10
34	政府統計の二次的利用	2005.04.20
35	ジェンダー（男女共同参画）統計	2007.02.20
36	人口センサスの現状と新展開	2007.04.01
37	統計における官学連携	2007.04.20
38	ジェンダー（男女共同参画）統計　Ⅱ	2009.02.10
39	社会生活基本調査とその利用	2010.01.15
40	地方統計の現状と課題	2010.09.15
41	Exploring Potential of Individual Statistical Records	2011.11.05
42	観光統計	2013.02.05
43	国民経済計算関連統計の新たなる展開	2014.01.30
44	タウンページデータによる事業所立地分析	2014.02.15
45	フィンランドのビジネス・レジスター	2015.03.20
46	19 世紀ドイツ営業統計史研究	2015.07.20
47	地方統計と統計 GIS	2016.01.25
48	首都圏の人口移動	2017.03.10
49	宿泊業及び飲食業の実証分析	2018.08.01
50	サービス分野の生産物分類	2019.01.31
51	全市区町村産業連関表（平成 23 年表）の推計	2019.10.15
52	商業統計調査	2021.01.31
53	産業連関表から供給・使用表へ	2021.03.31
54	統計的モデリング	2021.11.30
55	数理生態学とデータサイエンス	2022.01.31

研　究　所　報　No.56
2022 年 3 月 31 日

発行所　法政大学　日本統計研究所
〒194-0298 東京都町田市相原町 4342
Tel 042-783-2325,6
Fax 042-783-2332
jsri@adm.hosei.ac.jp
発行人　菅　幹雄

BULLETIN
OF
JAPAN STATISTICS RESEARCH INSTITUTE

No.56 **March 2022**

Statistical Inference on various manifolds

CONTENTS

Edited by
JAPAN STATISTICS RESEARCH INSTITUTE
HOSEI UNIVERSITY
TOKYO, JAPAN