報告書

Recent Developments of Statistical Theory in Statistical Science

Date: October 27, 2016 - October 29, 2016
Location: Hokkaido University
Graduate School of Economics and Business Administration, 3rd floor

Leader: Masanobu Taniguchi (Waseda University)

Supported by
(1) Kiban (A-15H02061) M. Taniguchi, Research Institute for Science & Engineering, Waseda University
(2) Houga (26540015) M. Taniguchi, Research Institute for Science & Engineering, Waseda University

Organizers: Yoshihide Kakizawa (Hokkaido University), Akio Suzukawa (Hokkaido University), Fumiya Akashi (Waseda University)

本シンポジウムにて１９件の講演があった.
いずれの講演も最先端をいくハイレベルなもので, 活発な質疑が交わされた.

2016年１０月２７日
講演 1

Statistical inference for quantiles in frequency domain

Yan Liu

Waseda University

Department of Applied Mathematics

Abstract

We consider the estimation and testing problems of quantiles in frequency domain. For second order stationary process, the spectral distribution function is uniquely determined by the autocovariance function of the process. We first define the quantiles of the spectral distribution function. The asymptotic distribution of the naive quantile estimator is shown to be non-Gaussian. This result is different from that considered in time domain. We recover the asymptotic normality of quantile estimation by smoothing the periodogram. Besides, we consider the quantile tests in frequency domain from our estimation procedure. Strong statistical power is shown in our numerical studies. The power of our proposed statistic under local alternatives is also discussed.

keywords: quantile estimator, frequency domain, asymptotic properties of estimators

1. Quantiles in frequency domain

Suppose $\{X_t\,;\, t \in \mathbb{Z}\}$ is a second order stationary process. From Herglotz's theorem, there uniquely exists a right continuous, nondecreasing and bounded function $F(\lambda)$ on $\Lambda \equiv [-\pi, \pi]$ with $F(-\pi) = 0$ such that

$$R(h) = \int_{-\pi}^{\pi} e^{ih\lambda} dF(\lambda), \quad (h \in \mathbb{Z}).$$

The function $F(\lambda)$ is called as the spectral distribution. If $F(\lambda)$ is absolutely continuous with respect to the Lebesgue measure, the spectral density function $f(\lambda)$ is uniquely defined almost everywhere. In the following, we define the quantile of the spectral distribution function $F(\lambda)$ through an objective function $S(\theta)$, i.e.,

$$S(\theta) = \int_{-\pi}^{\pi} \rho_p(\lambda - \theta) F(d\lambda), \tag{1.1}$$

where $\rho_\tau(x) = x(\tau - \mathbb{1}(x < 0))$. Under this formulation, we obtain the following theorem.

Theorem 1.1. *Suppose $\{X_t;\, t \in \mathbb{Z}\}$ is a zero mean second order stationary process with spectral distribution function $F(\lambda)$. Define $S(\theta)$ by (1.1). Then the pth quantile λ_p of the spectral distribution function $F(\lambda)$ is uniquely defined by the minimizer of $S(\theta)$, i.e.,*

$$\lambda_p = \min_{\theta \in \Lambda} S(\theta). \tag{1.2}$$

2. Estimation for the quantiles in the frequency domain

Suppose $\{X_t\,;\,1 \leq t \leq n\}$ is the observation stretch of the process. Let us define the periodogram $I_{n,X}(\omega)$ based on the observations by $I_{n,X}(\omega) = \left|\sum_{t=1}^n X_t e^{it\omega}\right|^2/(2\pi n)$. Then, the estimator for the quantiles in the frequency domain is defined by

$$\hat{\lambda}_n = \arg\min_{\theta\in\Lambda} \int_{-\pi}^{\pi} \rho_p(\omega-\theta) I_{n,X}(\omega) d\omega. \tag{2.1}$$

Under regularity conditions, the consistency of estimator (2.1) is shown.

Theorem 2.1. *Suppose $\{X_t;\, t \in \mathbb{Z}\}$ satisfies regularity conditions. Then we have $\hat{\lambda}_p \xrightarrow{\mathcal{P}} \lambda_p$.*

However, the estimator (2.1) is shown to be asymptotically mixed normal.

Theorem 2.2. *Suppose $\{X_t;\, t \in \mathbb{Z}\}$ satisfies regularity conditions. For $-\pi < \lambda_p < 0$,*

$$\sqrt{n}(\hat{\lambda}_n - \lambda_p) \to_d \mathcal{MN}(0, \mathscr{E}^{-2}\sigma^2),$$

where $\mathscr{E}$ is an exponential distributed random variable with mean $f(\lambda)$ and

$$\begin{aligned}\sigma^2 = \pi p^2 \int_{-\pi}^{\pi} f(\omega)^2 d\omega + 2\pi(1-4p)\int_{-\pi}^{\lambda_p} f(\omega)^2 d\omega + 2\pi\Big\{\int_{-\pi}^{\lambda_p}\int_{-\pi}^{\lambda_p} Q(\omega_1,\omega_2,-\omega_2) d\omega_1 d\omega_2 \\ + \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} p^2 Q(\omega_1,\omega_2,-\omega_2) d\omega_1 d\omega_2 - 2p\int_{-\pi}^{\lambda_p}\int_{-\pi}^{\pi} Q(\omega_1,\omega_2,-\omega_2) d\omega_1 d\omega_2\Big\},\end{aligned}$$

where $Q(\omega_1,\omega_2,\omega_3)$ is the fourth order spectral density.

To overcome this difficulty, we consider the following modified estimator $\hat{\lambda}_p^*$, i.e.,

$$\hat{\lambda}_p^* = \arg\min_{\theta\in[-\pi,\pi]} \int_{-\pi}^{\pi} \rho_p(\omega-\theta)\hat{f}(\omega) d\omega, \tag{2.2}$$

where $\hat{f}(\omega)$ is a quantity that the periodogram $I_{n,X}(\lambda)$ smoothed by $\phi(\omega)$.

Theorem 2.3. *Suppose $\{Y_t;\, t \in \mathbb{Z}\}$ follows a sinusoidal models with a second order stationary process $\{X_t\}$, whose spectral density is defined by $f_X(\omega)$. Then for continuity point λ_p ($-\pi < \lambda_p < 0$),*

$$\sqrt{n}(\hat{\lambda}_p^* - \lambda_p) \to_d \mathcal{N}(0, \sigma^2), \tag{2.3}$$

where

$$\begin{aligned}\sigma^2 = f_X(\lambda_p)^{-2}\Big[\pi p^2 \int_{-\pi}^{\pi} \phi(\omega)^2 f_Y(\omega) f_X(\omega) d\omega + 2\pi(1-4p)\int_{-\pi}^{\lambda_p} \phi(\omega)^2 f_Y(\omega) f_X(\omega) d\omega \\ + 2\pi\Big\{\int_{-\pi}^{\lambda_p}\int_{-\pi}^{\lambda_p} \phi(\omega)^2 Q_X(\omega_1,\omega_2,-\omega_2) d\omega_1 d\omega_2 + \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} p^2 \phi(\omega)^2 Q_X(\omega_1,\omega_2,-\omega_2) d\omega_1 d\omega_2 \\ - 2p\int_{-\pi}^{\lambda_p}\int_{-\pi}^{\pi} \phi(\omega)^2 Q_X(\omega_1,\omega_2,-\omega_2) d\omega_1 d\omega_2\Big\}\Big],\end{aligned}$$

where $f_Y(\omega)$ is a formal spectral of $\{Y_t\}$ and $Q_X(\omega_1,\omega_2,\omega_3)$ is the fourth order spectral of $\{X_t\}$.

We applied this result to quantile test for goodness of fit. Strong statistical power is shown in our numerical studies. The power of our proposed statistic under local alternatives is also discussed.

Quantile Hodrick–Prescott Filtering

Hiroshi Yamada

Department of Economics, Hiroshima University, Japan

1 Introduction

Quantile regression was introduced in the seminal work by Koenker and Bassett (1978) and widely applied in econometrics. Hodrick–Prescott (HP) (1997) filtering is used frequently to estimate trend components of macroeconomic time series. In this paper, we consider a filtering method that combines these two statistical tools. We refer to it as quantile HP (qHP) filtering. qHP filtering enables us to obtain not only the median trend, which is more robust to outliers than HP filtering, but also other quantile trends, which may provide a deeper understanding of time series properties. As in the case of HP filtering, it requires selection of tuning parameter. We propose a method for selecting it, which enables us to compare trends from (q)HP filtering. As an empirical example, we present estimated quantile trends of Japan's industrial index of production (IIP).

2 Quantile HP filtering

HP filtering is defined as follows:

$$\widetilde{\boldsymbol{x}} = \underset{x_1,\ldots,x_T\in\mathbb{R}}{\arg\min} \sum_{t=1}^{T}(y_t-x_t)^2 + \lambda\sum_{t=3}^{T}(\Delta^2 x_t)^2 = \underset{\boldsymbol{x}\in\mathbb{R}^T}{\arg\min}\ \|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{D}\boldsymbol{x}\|_2^2 = (\boldsymbol{I}_T+\lambda\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{y}, \quad (1)$$

where $\lambda\in\mathbb{R}_{>0}$ is a tuning parameter, $\Delta^2 x_t = \Delta x_t - \Delta x_{t-1} = x_t - 2x_{t-1} + x_{t-2}$, $\boldsymbol{y} = [y_1,\ldots,y_T]'$, $\boldsymbol{x} = [x_1,\ldots,x_T]'$, $\boldsymbol{I}_T$ is an identity matrix of size T, and $\boldsymbol{D}\in\mathbb{R}^{(T-2)\times T}$ is a second-order difference matrix such that $\boldsymbol{D}\boldsymbol{x} = [\Delta^2 x_3,\ldots,\Delta^2 x_T]'$. Here, for a vector $\boldsymbol{a} = [a_1,\ldots,a_n]$, $\|\boldsymbol{a}\|_p = (|a_1|^p+\cdots+|a_n|^p)^{1/p}$.

Letting $\rho_\tau(\,\cdot\,)$ be the tilted absolute value function such that:

$$\rho_\tau(u) = \begin{cases}\tau|u|, & \text{if } u\geq 0,\\ (1-\tau)|u|, & \text{otherwise,}\end{cases}$$

where $\tau\in(0,1)$, qHP filtering is defined by replacing $\sum_{t=1}^T(y_t-x_t)^2$ in (1) with $\sum_{t=1}^T\rho_\tau(y_t-x_t)$:

$$\underset{x_1,\ldots,x_T\in\mathbb{R}}{\text{minimize}} \sum_{t=1}^{T}\rho_\tau(y_t-x_t) + \psi\sum_{t=3}^{T}(\Delta^2 x_t)^2, \quad (2)$$

where $\psi\in\mathbb{R}_{>0}$ is a tuning parameter. (2) is a kind of squared ℓ_2-norm penalized quantile regression. Since $\rho_\tau(y_t-x_t) = 0.5|y_t-x_t| + (\tau-0.5)(y_t-x_t)$, (2) can be represented as

$$\underset{x_1,\ldots,x_T\in\mathbb{R}}{\text{minimize}}\ 0.5\sum_{t=1}^{T}|y_t-x_t| + (\tau-0.5)\sum_{t=1}^{T}(y_t-x_t) + \psi\sum_{t=3}^{T}(\Delta^2 x_t)^2,$$

or in matrix notation,

$$\underset{\boldsymbol{x}\in\mathbb{R}^T}{\text{minimize}}\ 0.5\|\boldsymbol{y}-\boldsymbol{x}\|_1 + (\tau-0.5)\boldsymbol{\iota}'(\boldsymbol{y}-\boldsymbol{x}) + \psi\|\boldsymbol{D}\boldsymbol{x}\|_2^2, \quad (3)$$

where $\boldsymbol{\iota}\in\mathbb{R}^T$ is a vector of ones. We denote the solution of (3) by $\widehat{\boldsymbol{x}}$ and refer to it as qHP trend.

Letting $\boldsymbol{\theta}\in\mathbb{R}^T$ be such that $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{\theta}$, where

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0\\ 1 & 1 & 0 & & \vdots & \vdots\\ 1 & 2 & 1 & \ddots & \vdots & \vdots\\ 1 & 3 & 2 & \ddots & 0 & \vdots\\ \vdots & \vdots & \vdots & \ddots & 1 & 0\\ 1 & T-1 & T-2 & \ldots & 2 & 1\end{bmatrix}\in\mathbb{R}^{T\times T},$$

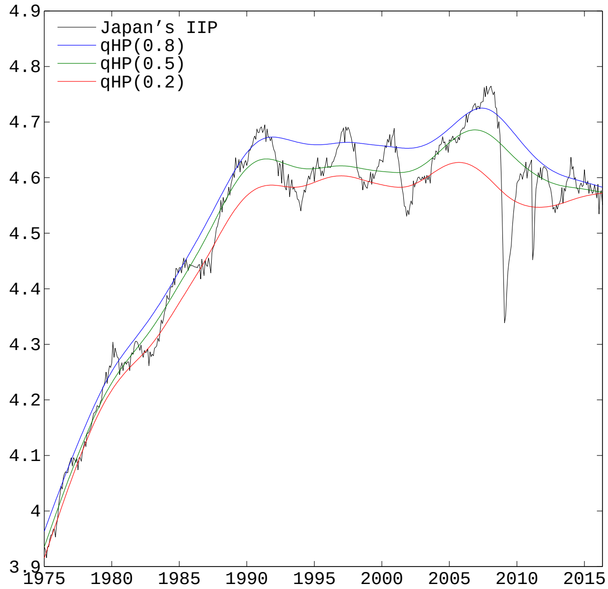

Figure 1: Japan's IIP and its qHP trends ($\lambda = 133107.94$).

qHP filtering, (3), can be represented as

$$\underset{\boldsymbol{\theta}\in\mathbb{R}^T}{\text{minimize}}\ 0.5\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{\theta}\|_1 + (\tau-0.5)\boldsymbol{\iota}'(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{\theta}) + \psi\|\boldsymbol{J}\boldsymbol{\theta}\|_2^2, \tag{4}$$

where $\boldsymbol{J} = [\boldsymbol{0}, \boldsymbol{I}_{T-2}] \in \mathbb{R}^{(T-2)\times T}$. It is notable that the first column of $\boldsymbol{A}$ is $\boldsymbol{\iota}$ and the first entry of $\boldsymbol{\theta}$ is not penalized, which indicates that τ-quantile of qHP residuals, $\boldsymbol{y}-\widehat{\boldsymbol{x}}$, approximately equals zero.

(3) is also represented as the following constrained minimization problem:

$$\underset{\boldsymbol{x}\in\mathbb{R}^T}{\text{minimize}}\ 0.5\|\boldsymbol{y}-\boldsymbol{x}\|_1 + (\tau-0.5)\boldsymbol{\iota}'(\boldsymbol{y}-\boldsymbol{x}), \quad \text{subject to}\ \|\boldsymbol{D}\boldsymbol{x}\|_2^2 \le c, \tag{5}$$

where $c \in \mathbb{R}_{>0}$ is a tuning parameter that corresponds to ψ in (3). Then, instead of selecting the value of ψ, we select the value of c as $c = \|\boldsymbol{D}\widetilde{\boldsymbol{x}}\|_2^2$ so that we obtain $\widehat{\boldsymbol{x}}$ such that $\|\boldsymbol{D}\widehat{\boldsymbol{x}}\|_2^2 = \|\boldsymbol{D}\widetilde{\boldsymbol{x}}\|_2^2$. The corresponding value of ψ to $c = \|\boldsymbol{D}\widetilde{\boldsymbol{x}}\|_2^2$ can be expressed as

$$\psi = \{0.5\|\boldsymbol{y}-\widehat{\boldsymbol{x}}\|_1 + (\tau-0.5)\boldsymbol{\iota}'(\boldsymbol{y}-\widehat{\boldsymbol{x}})\} \,/\, \{2(\boldsymbol{y}-\widehat{\boldsymbol{x}})'\boldsymbol{D}'\boldsymbol{D}\widehat{\boldsymbol{x}}\}\,, \tag{6}$$

which becomes $\psi = \|\boldsymbol{y}-\widehat{\boldsymbol{x}}\|_1 / \{4(\boldsymbol{y}-\widehat{\boldsymbol{x}})'\boldsymbol{D}'\boldsymbol{D}\widehat{\boldsymbol{x}}\}$ when $\tau = 0.5$.

References

1. Hodrick, R. J. and E. C. Prescott, 1997, Postwar U.S. business cycles: An empirical investigation, *Journal of Money, Credit and Banking*, 29, 1, 1–16.

2. Koenker, R. and G. Bassett, 1978, Regression quantiles, *Econometrica*, 46, 1, 33–50.

Modifying Gamma Kernel Density Estimator by Reducing Variance

Rizky Reza Fauzi, Graduate School of Mathematics Kyushu University
Yoshihiko, Maesono, Faculty of Mathematics Kyushu University

We discuss a new kernel type estimator for nonnegatively supported density function $f_X(x)$, using pdf of gamma distribution. Chen(2000, Ann.Inst.Stat.Math.) introduced two gamma kernels which are $Gamma\left(\frac{x}{h}+1,h\right)$ and $Gamma(\rho_h(x),h)$ densities. The order of convergence of variances are $O\left(\frac{1}{n\sqrt{h}}\right)$ in the interior and $O(\frac{1}{nh})$ near boundary. Under some conditions for x and h, Chen showed his estimators having $O(n^{-\frac{4}{5}})$ for the optimal mean squared error.

Rosenblatt (1956, A.M.S.) and Parzen (1962, A.M.S.) introduced kernel density estimator as

$$\hat{f}_X(x)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right),$$

where $K(u)$ is a kernel function which satisfies $\int K(u)\mathrm{du}=1$; and h is called as *bandwidth* satisfying $h\to 0$ and $nh\to\infty$ when $n\to\infty$. Under some regularity conditions for $f_X(x)$, $K(u)$ and h, we have

$$\begin{aligned}bias[\hat{f}_X(x)] &= \frac{h^2}{2}f_X''(x)\int u^2K(u)\mathrm{du}+\mathrm{o(h^2)},\\ var[\hat{f}_X(x)] &= \frac{f_X(x)}{nh}\int K^2(u)\mathrm{du}+\mathrm{o}\left(\frac{1}{\mathrm{nh}}\right),\\ MSE[\hat{f}_X(x)] &= O(n^{-\frac{4}{5}}),\ \text{when h is optimum.}\end{aligned}$$

However, if we deal with nonnegative support distribution, the standard KDE will suffer the boundary bias problem. The interval $[0,h]$ is called as boundary region, and the point greater than h is called as interior point. In the boundary region, the standard KDE $\hat{f}_X(x)$ usually underestimates $f_X(x)$, because standard KDE puts some weights on the negative axis as well. Or, mathematically, if we use a symmetric kernel supported on $[-1,1]$, we have

$$\begin{aligned}bias[\hat{f}_X(x)] &= \left[\int_{-1}^{c}K(u)\mathrm{du}-1\right]f_X(x)-hf_X'(x)\int_{-1}^{c}uK(x)\mathrm{du}\\ &\quad+\frac{h^2}{2}f_X''(x)\int_{-1}^{c}u^2K(u)\mathrm{du}+\mathrm{o(h^2)},\end{aligned}$$

when $x=ch$, $0\le c\le 1$. Which means $\lim_{n\to\infty}bias[\hat{f}_X(x)]=\left[\int_{-1}^{c}K(u)\mathrm{du}-1\right]f_X(x)$ (not consistent).

To overcome this problem, Chen (2000) introduced gamma kernel for the first time by using the pdf of $Gamma\left(\frac{x}{h}+1,h\right)$, that is

$$\hat{f}_1(x)=\frac{1}{n}\sum_{i=1}^{n}\frac{X_i^{\frac{x}{h}}e^{-\frac{X_i}{h}}}{\Gamma\left(\frac{x}{h}+1\right)h^{\frac{x}{h}+1}}$$

and for some $\kappa > 0$, we have

$$\begin{aligned}
bias[\hat{f}_1(x)] &= \left[f'_X(x) + \frac{1}{2}xf''_X(x)\right]h + o(h), \\
var[\hat{f}_1(x)] &= \frac{f_X(x)}{2\sqrt{\pi x}n\sqrt{h}}, \quad \frac{x}{h} \to \infty \\
&= \frac{\Gamma(2\kappa+1)f_X(x)}{2^{2\kappa+1}\Gamma^2(\kappa+1)nh}, \quad \frac{x}{h} \to \kappa.
\end{aligned}$$

Chen's gamma KDEs obviously solved the boundary bias problem. However, Chen's gamma KDEs also got some problems. We need a condition $\frac{x}{h} \to \kappa$ as we can see at the variance formulas, the variance also depends on a factor $\frac{1}{\sqrt{x}}$ in the interior point, which means the variance becomes much larger when x is small even though still not in the boundary. Zhang (2010, Stat.Prob.Let.) showed the MSE is $O(n^{-\frac{2}{3}})$ when x is close to the boundary (worse than the standard KDE).

Here, we tried to define another gamma KDE, as the density of $Gamma\left(\frac{1}{\sqrt{h}}, x\sqrt{h}+h\right)$, that is

$$\begin{aligned}
\hat{f}_h(x) &= \frac{1}{n}\sum_{i=1}^{n} \frac{X_i^{\frac{1}{\sqrt{h}}-1} e^{-\frac{X_i}{x\sqrt{h}+h}}}{\Gamma\left(\frac{1}{\sqrt{h}}\right)(x\sqrt{h}+h)^{\frac{1}{\sqrt{h}}}}, \\
bias[\hat{f}_h(x)] &= \left[f'_X(x) + \frac{1}{2}x^2 f''_X(x)\right]\sqrt{h} + o(\sqrt{h}), \\
var[\hat{f}_h(x)] &= \frac{R^2\left(\frac{1}{\sqrt{h}}-1\right)f_X(x)}{2(x+\sqrt{h})\sqrt{\pi(1-\sqrt{h})}R\left(\frac{2}{\sqrt{h}}-2\right)n\sqrt[4]{h}}, \quad x > h \\
&= \frac{R^2\left(\frac{1}{\sqrt{h}}-1\right)f_X(x)}{2(c\sqrt{h}+1)\sqrt{\pi(1-\sqrt{h})}R\left(\frac{2}{\sqrt{h}}-2\right)nh^{\frac{3}{4}}}, \quad x \le h,
\end{aligned}$$

where $c = \frac{x}{h}$ and $R(z) = \frac{\sqrt{2\pi}z^{z+\frac{1}{2}}}{e^z\Gamma(z+1)}$, $z \ge 0$.

By modifying it with similar technique as geometric extrapolation, we define the modified gamma kernel density estimator as $\tilde{f}_X(x) = [\hat{f}_h(x)]^2[\hat{f}_{4h}(x)]^{-1}$. And we can get the bias and the variance are

$$\begin{aligned}
bias[\tilde{f}_X(x)] &= -2\left[b(x) - \frac{a(x)}{2f_X(x)}\right]h + o(h) \\
var[\tilde{f}_X(x)] &= var[2\hat{f}_h(x) - \hat{f}_{4h}(x)],
\end{aligned}$$

where

$$\begin{aligned}
a(x) &= f'_X(x) + \frac{1}{2}x^2 f''_X(x) \\
b(x) &= \left(x+\frac{1}{2}\right)f''_X(x) + x^2\left(\frac{x}{3}+\frac{1}{2}\right)f'''_X(x)
\end{aligned}$$

Improvement of kernel estimators using boudary bias reduction methods

Taku MORIYAMA (Graduate School of Mathematics, Kyushu University)*

1. Introduction

In this talk, I discuss improvement of kernel estimators using boundary bias reduction methods. As we know, kernel estimators are biased (as we call boundary bias) if the true density has finite or semi-finite support. When the support is known, we have various methods to improve the estimators. On the other hand, there are few ways to adjust them to the unknown support. Hall and Park (2002) introduced a way to estimate the density whose support is estimated by order statistics. I propose a new way to estimate the unknown support using a boundary bias reduction method which is used to correct the distribution estimator. The asymptotic properties and some simulation results of kernel estimators using the proposed boundary estimator are shown. Furthermore, I discuss the necessity of giving appropriate (bounded) support of the estimated density even if the information is not available.

Let $X_1, \cdots, X_n$ be i.i.d. from F and the density be f. The kernel density estimator (K.E.) and cumulative estimator are

$$\widehat{f}(x) = \frac{1}{h}\int K\left(\frac{x-y}{h}\right) dF_n(y), \quad \text{and} \quad \widehat{F}(x) = \int W\left(\frac{x-y}{h}\right) dF_n(y)$$

where $F_n(y) = n^{-1}\sum_{i=1}^n I(X_i \leq y)$ and W is a integral of kernel function K : $W(z) = \int_{-\infty}^{z} K(u)du$. K is usually assumed its nonnegativity and symmetricity. h is a bandwidth which satisfies $h \to 0$ and $nh \to \infty$. They have consistency when $supp(f) = (-\infty, \infty)$ holds. Otherwise (e.g. $supp(f) = (-\infty, U]$), the density estimator $\widehat{f}(x)$ loses it because of the following boundary bias near the boundary

$$\begin{aligned} Bias[\widehat{f}(U)] &= \int_{-\infty}^{\infty} K\left(\frac{U-y}{h}\right) f(y)dy - f(U) \\ &= \int_{0}^{\infty} K(z)f(U-hz)dz - f(U) = -\frac{f(U)}{2} + O(h) = O(1) \end{aligned}$$

The bias of $\widehat{F}(U)$ is of order $O(h)$ which is also lower order.

In case the boundary is known, various ways to improve it has been discussed well. However, there are few ways to adjust them to the unknown support. I think even in case we have no information in advance, we should estimate the (compact) support of f appropriately and use the kernel type estimator reducing the boundary bias. A reasonable approach is to use an estimator reducing the boundary bias derived from the estimated support. Hall & Park(2002) discusses boundary estimator based on sample maximum $X_{(n)}$ (minimum $X_{(1)}$). I propose a new boundary estimator fitting the boundary bias reduction method which is what we want to use.

2. Proposed method

We assume that $supp(f) = [L, U]$, L is known, U is unknown and $f(U) > 0$. The boundary estimator $u = \widehat{u}$ using the estimator $\widehat{F}_u^{\dagger}$, $\widehat{f}_u^{\dagger}$ reducing boundary bias derived

Keywords: boundary bias, kernel density estimator, kernel cumulative estimator.

* e-mail: moritaku3542168@gmail.com

from the support $[L, u]$ is defined as follows.

$$\widehat{F}_u^\dagger(X_{(n)}) = \frac{n}{n+1} \quad , \quad \widehat{F}_u^\dagger(z) = \int_{-\infty}^{z} \widehat{f}_u^\dagger(x)dx$$

Define the estimator of U as the solution $u = \widehat{u}$, the distribution estimator as $\widehat{F}_{\widehat{u}}^\dagger$ and the density estimator as $\widehat{f}_{\widehat{u}}^\dagger$. The equation $\widehat{F}_u^\dagger(X_{(n)}) = n/(n+1)$ derives from the property of maximum estimation in $U(0,1)$ because $F(X) \overset{d}{=} U(0,1)$ and $F(X_{(n)}) \approx n/(n+1)$ holds. The general solution $u = \widehat{u}$ of (1) is not given as an explicit formula, but in many cases $\widehat{u}$ can be seen as a M estimator. Using boundary kernel method (represented by $\widehat{F}_u^{[BK]}$) (Tenreiro(2013)) and reflection method ($\widehat{F}_u^{[R]}$) (Silverman(1986)), asymptotic properties and simulation results are shown.

3. Asymptotic properties

We can prove the following asymptotic properties of them using the properties of M estimator.

$$\widehat{u} \xrightarrow{p} U, \quad \widehat{F}_{\widehat{u}}^{[BK]}(\cdot) = F_U(\cdot) + O_P(\underset{\sim}{h}^2 + n^{-1/2})$$

We can find that the bias term becomes higher order and $\widehat{f}_{\widehat{u}}^{[BK]}(\cdot)$ recovers its consistency.

$$\widehat{u} \xrightarrow{p} U, \quad \widehat{F}_{\widehat{u}}^{[R]}(\cdot) = F_U(\cdot) + O_P(\underset{\sim}{h}^2 + n^{-1/2})$$

$\widehat{f}_{\widehat{u}}^{[R]}(\cdot)$ also recovers the goodness of kernel density estimator. Their proofs are omitted here.

I claim that the proposed method is superior in sense of precise estimation of tail behavior at least in case the support seems compact. Some numerical experiments are given in the day.

References

[1] Hall, P., & Park, B. U. (2002). New methods for bias correction at endpoints and boundaries. *Annals of Statistics*, 1460-1479.

[2] Tenreiro, C. (2013). Boundary kernels for distribution function estimation. *REVSTAT Statistical Journal*, 11(2), 169-190.

[3] Ćwik, J., & Mielniczuk, J. (1993). Data-dependent bandwidth choice for a grade density kernel estimate. *Statistics & Probability Letters*, 16(5), 397-405.

On the extension of Lo's modified R/S statistic against to locally stationary short-range dependence

Junichi Hirukawa, Minehiro Takahashi, Ryota Tanaka (Niigata University)

1. Introduction

To detect long-range or " strong " dependence, Mandelbrot has suggested using the range over standard deviation or R/S statistic, also called the " rescaled range, " which was developed by Hurst in his studies of river discharges. The R/S statistic is the range of partial sums of deviations of a time series from its mean, rescaled by its standard deviation. Lo's modified R/S statistics can distinguish between long-range dependence and short-range dependence. Although Lo's method of test has been used various fields of the studies, the time series is assumed to be stationary processes. The stationary process has been widely used for many statistical model in time series analysis. Although these models have an important role, stationarity is not sufficient to describe the real world. In order to develop the asymptotic theory for non-stationary models, the important model of a non-stationary process is proposed in Dahlhaus, called locally stationary processes. In this paper, we apply Lo's modified R/S statistics to locally stationary processes.

2. Lo's modified R/S statistic for locally stationary processes.

2.1. **Under null hypothesis.** Define locally stationary general linear processes $\{u_{j,T}\}$ as

$$u_{j,T} = \sum_{k=0}^{\infty} b_k \left(\frac{j}{T} \right) \varepsilon_{j-k} = b \left(\frac{j}{T}, L \right) \varepsilon_j,$$

where

$$b \left(\frac{j}{T}, L \right) = \sum_{k=0}^{\infty} b_l \left(\frac{j}{T} \right) L^l, \quad L \text{ is the lag operator,}$$

$\varepsilon_t \overset{i.i.d.}{\sim} (0,1)$ and $b_l(u)$ is twice continuously differentiable with

$$\sup_{u\in[0,1]} \sum_{l=0}^{\infty} |l| \left| b_l^{(i)}(u) \right| < \infty, \quad i = 0, 1, 2.$$

Let

$$Q_T^{loc} := \max_{1\le k\le T} \sum_{j=1}^{k} \left(y_{j,T} - \overline{y}_T \right) - \min_{1\le k\le T} \sum_{j=1}^{k} \left(y_{j,T} - \overline{y}_T \right),$$

where

$$y_{j,T} := \frac{u_{j,T}}{b\left(\frac{j}{T}, 1\right)}, \quad \overline{y}_T = \frac{1}{T} \sum_{j=1}^{T} y_{j,T}.$$

Then we have

$$\frac{1}{\sqrt{T}} Q_T^{loc} \overset{d}{\Longrightarrow} V,$$

where V is the range of a Brownian bridge on the unit interval.

We assume the terminal values $\{u_{s,T} \mid T+1 \le s \le T+n-1\}$ can be available as $u_{s,T} = \sum_{k=0}^{\infty} b_k(1)\, \varepsilon_{s-k}$. Define

$$\widehat{\sigma}_{j,T,n}(q)^2 := \frac{1}{n} \sum_{l=-q}^{q} \omega_l(q) \sum_{i=\max\{0,l\}}^{\min\{n-1,n-1+l\}} \left(u_{j+i,T} - \overline{u}_{j,T,n} \right) \left(u_{j+i-l,T} - \overline{u}_{j,T,n} \right)$$

where

$$\omega_l(q) = 1 - \frac{|l|}{q+1},\ \overline{u}_{j,T,n} = \frac{1}{n}\sum_{i=0}^{n-1} u_{j+i,T},\ u_{s,T} = \sum_{k=0}^{\infty} b_k\left(\frac{s}{T}\right)\varepsilon_{s-k}$$

and $q << n << T$. Let

$$\widehat{y}_{j,T} := \frac{u_{j,T}}{\widehat{\sigma}_{j,T,n}(q)},\quad \widetilde{y}_T = \frac{1}{T}\sum_{j=1}^{T}\widehat{y}_{j,T}$$

$$\widehat{Q}_{n,T}^{loc} := \max_{1\le k\le T}\sum_{j=1}^{k}\left(\widehat{y}_{j,T} - \overline{y}_T\right) - \min_{1\le k\le T}\sum_{j=1}^{k}\left(\widehat{y}_{j,T} - \overline{y}_T\right).$$

Then, we can see that

$$\widehat{\sigma}_{j,T,n}(q) \xrightarrow{p} b\left(\frac{j}{T}, 1\right).$$

Theorem 2.1 (Locally stationary modified R/S statistic)**.**

$$\frac{1}{\sqrt{T}}\widehat{Q}_T^{loc} \overset{d}{\Longrightarrow} V.$$

2.2. **Under alternative hypothesis.** Let ε_t satisfy the following equation

$$(1-L)^d\varepsilon_t = \eta_t, \qquad \eta_t \sim i.i.d.(0,1),$$

where L is the lag operator and

$$u_{s,T} = \sum_{l=0}^{\infty} b_l\left(\frac{s}{T}\right)\varepsilon_{s-l}.$$

Now, we consider weak convergency of partial sum process to fractional Brownian bridge. Let

$$X_T^{(d)}(t) := \frac{1}{T^{d+\frac{1}{2}}}\sum_{j=1}^{[tT]}\left(y_{j,T} - \overline{y}_T\right)$$

Then, using continuous mapping theorem, we can obtain

$$X_T^{(d)}(t) \overset{d}{\Rightarrow} B_{d+\frac{1}{2}}(t) - tB_{d+\frac{1}{2}}(1) := B_{d+\frac{1}{2}}^{\circ}(t),$$

where $B_{d+\frac{1}{2}}(t) = W_H(\tau)$ is fractional Brownian motion and $B_{d+\frac{1}{2}}^{\circ}(t) = W_H^{\circ}(\tau)$ is so-called fractional Brownian bridge with $H = d + \frac{1}{2}$.

Theorem 2.2. *(a)* $\displaystyle\max_{1\le k\le T}\frac{1}{T^{d+\frac{1}{2}}}\sum_{j=1}^{k}(y_{j,T} - \overline{y}_T) \Rightarrow \max_{0\le t\le 1} B_{d+\frac{1}{2}}^{\circ}(t) \equiv M_{d+\frac{1}{2}}^{\circ},$

(b) $\displaystyle\min_{1\le k\le T}\frac{1}{T^{d+\frac{1}{2}}}\sum_{j=1}^{k}(y_{j,T} - \overline{y}_T) \Rightarrow \min_{0\le t\le 1} B_{d+\frac{1}{2}}^{\circ}(t) \equiv m_{d+\frac{1}{2}}^{\circ},$

(c) $\displaystyle\frac{1}{T^{d+\frac{1}{2}}}Q_{n,T}^{loc} \Rightarrow M_{d+\frac{1}{2}}^{\circ} - m_{d+\frac{1}{2}}^{\circ}.$

(d) $\displaystyle\widehat{y}_{j,T} = \frac{b\left(\frac{j}{T},1\right)}{\widehat{\sigma}_{j,T,n}(q)}y_{j,T} \overset{p}{\to} \begin{cases}\infty & \text{for } d>0,\\ 0 & \text{for } d<0,\end{cases}$

(e) $\displaystyle\frac{1}{\sqrt{T}}\widehat{Q}_{n,T}^{loc} \overset{p}{\to} \begin{cases}\infty & \text{for } d>0,\\ 0 & \text{for } d<0.\end{cases}$

Self-normalized and random weighting approach to likelihood ratio test for the model diagnostics of stable processes

Fumiya Akashi (Waseda University)
Jianqing Fan (Princeton University)

1. Fundamental Settings

Let $\{X_t : t \in \mathbb{Z}\}$ be a symmetric α-stable (sαs) linear process generated as

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad (t \in \mathbb{Z}), \tag{1}$$

where $\{Z_t : t \in \mathbb{Z}\}$ is a sequence of independent and identically distributed sαs-random variables, whose characteristic function is given as $\mathrm{E}[\exp(\mathbf{i}uZ_t)] = \exp(-\sigma|u|^{\alpha})$. Here $\sigma \in (0, \infty)$ and $\alpha \in (0, 2]$ are, respectively, called the scale and the tail-index of the sαs distribution. Especially, the tail-index α controls the behavior of the tail probability of X_t, and the model does not have finite variance when $\alpha < 2$. To deal with such infinite variance model, let us consider the frequency-domain representation of (1), which is the power transfer function

$$f(\omega) = \left| \sum_{j=0}^{\infty} \psi_j \exp(-\mathbf{i} j\omega) \right|^2 .$$

This talk focuses on the estimation problem of the *pivotal quantity* of the model (1) defined as

$$\theta_0 := \arg\min_{\theta\in\Theta} \int_0^{2\pi} \frac{f(\omega)}{g(\omega;\theta)} d\omega, \tag{2}$$

where $g : [-\pi, \pi] \times \mathbb{R}^p \to \mathbb{R}^1$ is a user-specified score function and does not necessarily coincide with the true power transfer function $f(\omega)$. By choosing g appropriately, our framework can grasp various important problems such as variable selection or test for independence of the model. To make inference for the pivotal quantity (2), we consider two typical statistics

$$D_n(\theta) := \frac{1}{n} \sum_{j=1}^{n} \frac{\tilde{I}_n(\lambda_j)}{g(\lambda_j;\theta)} \quad \text{and} \quad T_n(\theta) := \min_{\eta\in\Theta} D_n(\eta) - D_n(\theta) \quad (\lambda_j = 2\pi j/n),$$

where $\tilde{I}_n(\omega) := |\sum_{t=1}^n X_t \exp(-\mathbf{i}t\omega)|^2 / \sum_{t=1}^n X_t^2$ is a self-normalized periodogram (c.f. [3]). Under some regularity conditions, it is show that

$$x_{n,\alpha} D_n(\theta_0) \xrightarrow{\mathcal{L}} v \quad \text{and} \quad x_{n,\alpha}^2 T_n(\theta_0) \xrightarrow{\mathcal{L}} V^\top W^{-1} V \quad (x_{n,\alpha} = (n/\log n)^{1/\alpha}),$$

where v is a sum of sαs-random variables, V is a sum of sαs-random vectors and W is a constant matrix. Since the rate of convergence $x_{n,\alpha}$ involves the unknown tail-index α, we consider the self-normalized statistics as

$$\tilde{D}_n(\theta) := D_n(\theta)/d_n^{1/2} \quad \text{and} \quad \tilde{T}_n(\theta) := T_n(\theta)/d_n,$$

This work was supported by Grant-in-Aid for Young Scientists (B) (16K16022, Fumiya Akashi).

where

$$d_n = n^{-1} \sum_{j=1}^{n} \left\{K_{n,j} - (j/n)K_{n,n}\right\}^2 \quad \text{and} \quad K_{n,j} = \frac{1}{j} \sum_{k=1}^{j} \tilde{I}_n(\lambda_k) d\omega.$$

As a remarkable feature, the rate of convergence of the self-normalized factor d_n is $x_{n,\alpha}^2$; that is, there exists some non-degenerate random variable d such that $x_{n,\alpha}^2 d_n \xrightarrow{\mathcal{L}} d$. Therefore, $\tilde{D}_n(\theta_0)$ and $\tilde{T}_n(\theta_0)$ also have the non-degenerate limit distributions v/d and $V^\top W^{-1} V/d$, respectively, without any additional normalization.

2. Main Result

In order to make the inference for θ_0, we next approximate the limit distribution v/d and $V^\top W^{-1} V/d$ directly by frequency domain bootstrap method proposed by [1]. Since the model does not have the finite variance, we make use of the results [1] and [3], and the proposed self-normalized frequency domain bootstrap procedure is constructed as follows:

Step 1. Calculate the self-normalized periodogram ordinate $\tilde{I}_n(\lambda_j)$ and its consistent estimator $\hat{f}_n(\lambda_j)$ for each $j = 1, \ldots, n$, where $\hat{f}_n(\omega)$ is constructed by Theorem 4.3 of [3].

Step 2. Generate the bootstrap samples $\{\varepsilon_j^* : j = 1, \ldots, n\}$ from the empirical distribution of $\{\tilde{\varepsilon}_j : j = 1, \ldots, n\}$, where $\tilde{\varepsilon}_j = \hat{\varepsilon}_j / \hat{\varepsilon}_\bullet$, $\hat{\varepsilon}_j = \tilde{I}_n(\lambda_j)/\hat{f}_n(\lambda_j)$ and $\hat{\varepsilon}_\bullet = \sum_{j=1}^n \hat{\varepsilon}_j$.

Step 3. Define bootstrap periodogram ordinate $\tilde{I}_{n,j}^* := \hat{f}_n(\lambda_j)\varepsilon_j^*$.

Iterating the procedure above, we obtain B-times replication, and therefore we can approximate the distributions of $\tilde{D}_n(\theta_0)$ and $\tilde{T}_n(\theta_0)$ by the empirical distribution functions

$$\hat{F}_{n,\tilde{D}}(x) := \frac{\#\left\{\tilde{D}_n^*(\tilde{\theta}_n) \leq x\right\}}{B} \quad \text{and} \quad \hat{F}_{n,\tilde{T}}(x) := \frac{\#\left\{\tilde{T}_n^*(\tilde{\theta}_n) \leq x\right\}}{B},$$

where $\tilde{D}_n^*(\theta)$ and $\tilde{T}_n^*(\theta)$ are generated by replacing $\tilde{I}_n(\lambda_j)$ in $\tilde{D}_n(\theta)$ and $\tilde{T}_n(\theta)$, respectively, by $\tilde{I}_{n,j}^*$, and

$$\tilde{\theta}_n := \arg\min_{\theta \in \Theta} \sum_{j=1}^{n} \frac{\hat{f}(\lambda_j)}{g(\lambda_j; \theta)}.$$

Theorem 1. *Under some regularity conditions,*

$$\left|\hat{F}_{n,\tilde{D}}(x) - P\left(\tilde{D}_n(\theta_0) \leq x\right)\right|, \quad \left|\hat{F}_{n,\tilde{T}}(x) - P\left(\tilde{T}_n(\theta_0) \leq x\right)\right| \xrightarrow{\mathcal{P}} 0$$

for each x as n and $B \to \infty$.

By Theorem 1, we can construct robust inference for the pivotal quantity θ_0 of the infinite variance model. This talk also provides some simulation results, and we observe that the proposed method shows appropriate finite sample performance for heavy-tailed observations.

References

[1] Dahlhaus, R. and Janas, D. (1996). A frequency domain bootstrap for ratio statistics in time series analysis. *The Annals of Statistics*. 24(5), 1934-1963.

[2] Fan, J. and Zhang, W. (2004). Generalized likelihood ratio tests for spectral density. *Biometrika*. 91(1), 195-209.

[3] Klüppelberg, C. and Mikosch, T. (1996) Self-normalized and randomly centered spectral estimates. *Athens Conference on Applied Probability and Time Series Analysis*. Springer New York.

Multi-step circular Markov processes with canonical vine representations*

Hiroaki Ogata

Tokyo Metropolitan University

Let $f_1(\theta)$ and $f_2(\eta)$ be arbitrary circular density functions on $\Pi := [0, 2\pi)$, and $F_1(\theta)$ and $F_2(\eta)$ be their distribution functions. Also, let $g(\cdot)$ be an arbitrary density function on Π. Then, Wehrly and Johnson (1980) proposed circular bivariate density function

$$f(\theta, \eta) = 2\pi g[2\pi\{F_1(\theta) - qF_2(\eta)\}]f_1(\theta)f_2(\eta) \tag{1}$$

having marginal densities $f_1(\theta)$ and $f_2(\eta)$. Here, $q \in \{-1, 1\}$ is a given non-random constant. Apparently, this is associated with the copula representation. Consider a pair of linear random variables (X, Y) and its joint density $f_{X,Y}(\cdot, \cdot)$ is expressed as

$$f_{X,Y}(x, y) = c(F_X(x), F_Y(y))f_X(x)f_Y(y)$$

where $c(\cdot, \cdot)$ is a copula density, $f_X(\cdot)$ and $f_Y(\cdot)$ are densities of X and Y, respectively. The function $2\pi g[2\pi\{F_1(\theta) - qF_2(\eta)\}]$ in (1) corresponds to the copula density.

Now, let $\{\Theta_t\}_{t\geq 1}$ be a sequence of random variables on Π. Using (1), Wehrly and Johnson (1980) naturally defined the stationary Markov process on the unit circle as

$$p(\theta_1) = f(\theta_1), \quad p(\theta_t|\theta_{t-1}, \ldots, \theta_1) = p(\theta_t|\theta_{t-1}) = 2\pi g[2\pi\{F(\theta_t) - qF(\theta_{t-1})\}]f(\theta_t)$$

where $f(\cdot)$ and $g(\cdot)$ are arbitrary densities on Π, and $F(\theta) = \int_0^\theta f(\xi)\, d\xi$. Then, $p(\theta_1)$ is the initial distribution and $p(\theta_t|\theta_{t-1})$ is the stationary trantision density. This stationary circular Markov process can be extended to k-step circular Markov processes[1] by use of the canonical vine copula (Aas et al. (2009)). If the process $\{\Theta_t\}_{t\geq 0}$ has 2-step Markov property[2], the joint density has the expression

$$\begin{aligned} f(\theta_1, \ldots, \theta_n) &= f(\theta_1)f(\theta_2|\theta_1)f(\theta_3|\theta_2, \theta_1)f(\theta_4|\theta_3, \theta_2, \theta_1)\cdots f(\theta_n|\theta_{n-1}, \ldots, \theta_1) \\ &= f(\theta_1)f(\theta_2|\theta_1)f(\theta_3|\theta_2, \theta_1)f(\theta_4|\theta_3, \theta_2)\cdots f(\theta_n|\theta_{n-1}, \theta_{n-2}) \end{aligned} \tag{2}$$

*The research reported herein was supported by JSPS KAKENHI Grant Numbers 26870655.

[1]The term "k-step Markov" is used here in the sense that $p(\theta_t|\theta_{t-1}, \ldots) = p(\theta_t|\theta_{t-1}, \ldots, \theta_{t-k})$.

[2]For the simplicity, we consider 2-step Markov process here. The extension to k-step Markov process is straightforward.

In general, a conditional density can be decomposed as

$$f(x|\boldsymbol{v}) = \frac{f(x, v_j|\boldsymbol{v}_{-j})}{f(v_j|\boldsymbol{v}_{-j})} = c_{x,v_j|\boldsymbol{v}_{-j}}\{F(x|\boldsymbol{v}_{-j}), F(v_j|\boldsymbol{v}_{-j})\} \cdot f(x|\boldsymbol{v}_{-j}) \tag{3}$$

for a vetor $\boldsymbol{v}$.[3] Here, v_j is one arbitrarily chosen component of $\boldsymbol{v}$ and $\boldsymbol{v}_{-j}$ denotes the v-vector, excluding this component. The function $c_{x,v_j|\boldsymbol{v}_{-j}}$ is an appropriate pair-copula density, applied to the transformed variables $F(x|\boldsymbol{v}_{-j})$ and $F(v_j|\boldsymbol{v}_{-j})$. Then, (2) is

$$f(\theta_1)c_{2,1}\{F(\theta_2), F(\theta_1)\}f(\theta_2)\left\{\prod_{i=1}^{n-2} c_{3,2|1}\{F(\theta_{i+2}|\theta_i), F(\theta_{i+1}|\theta_i)\}c_{3,1}\{F(\theta_{i+2}), F(\theta_i)\}f(\theta_{i+2})\right\}$$

In the circular process, the pair-copula functions are written in the form $2\pi g[2\pi\{F_1(\theta) - qF_2(\eta)\}]$, and this leads to the final expression

$$\begin{aligned} & f(\theta_1, \ldots, \theta_n) \\ = \;& (2\pi)^{2n-3}\left\{\prod_{i=1}^{n} f(\theta_i)\right\} g_{2,1}\Big[2\pi\{F(\theta_2) - q_{2,1}F(\theta_1)\}\Big] \\ & \left\{\prod_{i=1}^{n-2} g_{3,1}\Big[2\pi\{F(\theta_{i+2}) - q_{3,1}F(\theta_i)\}\Big]\right\}\left\{\prod_{i=1}^{n-2} g_{3,2|1}\Big[2\pi\{F(\theta_{i+2}|\theta_i) - q_{3,2|1}F(\theta_{i+1}|\theta_i)\}\Big]\right\}. \end{aligned}$$

Here, $f(\cdot)$ is the common marginal circular density and $g.[\cdot]$'s are arbitrary circular densities. Moreover,

$$f(\theta_{i+2}|\theta_i) = \frac{f(\theta_{i+2}, \theta_i)}{f(\theta_i)} = 2\pi\, g_{3,1}\Big[2\pi\{F(\theta_{i+2}) - q_{3,1}F(\theta_i)\}\Big]f(\theta_{i+2})$$

leads to the expression

$$F(\theta_{i+2}|\theta_i) = 2\pi\int_0^{\theta_{i+2}} g_{3,1}\Big[2\pi\{F(t) - q_{3,1}F(\theta_i)\}\Big]f(t)\,dt.$$

Similary,

$$F(\theta_{i+1}|\theta_i) = 2\pi\int_0^{\theta_{i+1}} g_{2,1}\Big[2\pi\{F(t) - q_{2,1}F(\theta_i)\}\Big]f(t)\,dt.$$

References

Aas K, Czado C, Frigessi A, Bakken H (2009) Pair-copula constructions of multiple dependence. Insur Math Econ 44:182-198

Wehrly TE, Johnson RA (1980) Bivariate models for dependence of angular observations and a related Markov process. Biometrika 67:255–256

[3]See also Aas et al. (2009).

On improvement of generalized $\phi-$divergence goodness-of-fit statistics for GLIM of binary data

Nobuhiro Taneichi, Kagoshima University
Yuri Sekiya, Hokkaido University of Education
Jun Toyama, The Institute for the Practical Application of Mathematics

We consider generalized linear models (Nelder and Wedderburn [3]) in which the response variables are measured on a binary scale. Let N independent random variables $Y_\alpha,\ \alpha = 1, \dots, N$ corresponding to the number of successes in N different subgroups be distributed according to binomial distributions $B(n_\alpha, \pi_\alpha)$, $\alpha = 1, \dots, N$. If we use a monotone and differentiable function g as a link function, we obtain a generalized linear model for binary data as follows.

(1) $$g(\pi_\alpha) = \boldsymbol{x}'_\alpha \boldsymbol{\beta} \ (\alpha = 1, \dots, N),$$

where $\boldsymbol{x}_\alpha = (x_{\alpha 1}, \dots, x_{\alpha p})'(\alpha = 1, \dots, N)$ are covariate vectors and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is an unknown parameter vector and $p < N$. We consider a minimum ϕ^*-divergence estimator of model (1) and also consider a ϕ-divergence goodness-of-fit test statistic based on the estimator. Let y_α $(\alpha = 1, \dots, N)$ be an observed value of Y_α $(\alpha = 1, \dots, N)$, then the minimum ϕ^*-divergence estimator of model (1) is given by

$$\hat{\boldsymbol{\beta}}^{g\phi^*} = \arg \min_{\beta \in \Theta} D_{\phi^*},$$

where

$$D_{\phi^*} = \frac{1}{N} \sum_{\alpha=1}^{N} n_\alpha \left\{ \pi_\alpha(\boldsymbol{\beta}) \phi^* \left(\frac{\dfrac{y_\alpha}{n_\alpha}}{\pi_\alpha(\boldsymbol{\beta})} \right) + (1 - \pi_\alpha(\boldsymbol{\beta})) \phi^* \left(\frac{1 - \dfrac{y_\alpha}{n_\alpha}}{1 - \pi_\alpha(\boldsymbol{\beta})} \right) \right\},$$

where ϕ^* is a real convex function in $(0, \infty)$ satisfying $\phi^*(1) = \phi^{*'}(1) = 0,\ \phi^{*''}(1) = 1$, $0\phi^*(0/0) = 0,\ 0\phi^*(x/0) = \lim_{u \to \infty} \phi^*(u)/u$, and Θ is an open subset of R^p (Pardo [4]). The maximum likelihood estimator is a special case of the minimum ϕ^*-divergence estimator. In order to test the null hypothesis,

(2) $$H_0^g : \pi_\alpha = \pi_\alpha(\boldsymbol{\beta}) = g^{-1}(\boldsymbol{x}'_\alpha \boldsymbol{\beta}) \ (\alpha = 1, \dots, N),$$

we consider the family of ϕ-divergence statistics based on the minimum ϕ^*-divergence estimator

(3) $$C_{\phi\phi^*} = 2 \sum_{\alpha=1}^{N} n_\alpha \left\{ \hat{\pi}_\alpha^{g\phi^*} \phi \left(\frac{\dfrac{Y_\alpha}{n_\alpha}}{\hat{\pi}_\alpha^{g\phi^*}} \right) + (1 - \hat{\pi}_\alpha^{g\phi^*}) \phi \left(\frac{1 - \dfrac{Y_\alpha}{n_\alpha}}{1 - \hat{\pi}_\alpha^{g\phi^*}} \right) \right\},$$

where $\hat{\pi}_\alpha^{g\phi^*} = \pi_\alpha(\hat{\boldsymbol{\beta}}^{g\phi^*})$ $(\alpha = 1, \dots, N)$, $\hat{\boldsymbol{\beta}}^{g\phi^*} = (\hat{\beta}_1^{g\phi^*}, \dots, \hat{\beta}_p^{g\phi^*})'$ is the minimum ϕ^*-divergence estimator of $\boldsymbol{\beta}$ under H_0^g given by (2) and ϕ satisfies the same conditions of

ϕ^* (Pardo [4]). The family of test statistics C_ϕ given by (7) in Taneichi *et al.* [6] is $C_{\phi\phi^*}$ when using the maximum likelihood estimator, and therefore the family of statistics given by (3) includes that of C_ϕ.

Under H_0^g, all members of the class of statistics $C_{\phi\phi^*}$ have a χ^2_{N-p} limiting distribution, assuming a suitable condition. Using the results, we can use $C_{\phi\phi^*}$ as a goodness-of-fit test statistic for model (1).

However, in the case in which all $n_\alpha, \alpha = 1, \ldots, N$ are not large enough, such an approximation by a χ^2_{N-p} limiting distribution to the distribution of D under H_0 becomes poor. So, there are risks that the hypothesis test based on large sample theory will give results opposite to those of an exact test. In this presentation, in order to reduce the risks, we propose a new transformed statistics $C^I_{\phi\phi^*}$ of $C_{\phi\phi^*}$ whose speed of convergence to a chi-square distribution is quicker than that of $C_{\phi\phi^*}$. To construct $C^I_{\phi\phi^*}$, we use the following procedure. First, we consider the asymptotic expansion of the original statistics $C^I_{\phi\phi^*}$, which is developped by Yarnold [7], Siotani and Fujikoshi [5], and Taneichi *et al.* [6]. Next, we obtain transformed statistics $\tilde{C}^I_{\phi\phi^*}$ by performing improved transformation ([1], [2]) to $C_{\phi\phi^*}$ on the basis of continuous term of the asymptotic expansion. As a special case of $\phi-$divergence statistics, we consider power divrgence statistics and executing a Monte Carlo simulation. By the Monte Carlo simulation, we find that the performance of transformed statistics $\tilde{C}^I_{\phi\phi^*}$ is much better than original statistics $C_{\phi\phi^*}$ in models given by the logit link, probit link and complementary log-log link. We also find that the power of statistics $\tilde{C}^I_{\phi\phi^*}$ is almost the same as the original statistics $C_{\phi\phi^*}$.

References

[1] Fujikoshi, Y. (2000). Transformations with improved chi-squared approximations, *J. Multivariate Anal.*, **72**, 249–263.

[2] Kakizawa, Y. (1996). Higher order monotone Bartlett-type adjustment for some multivariate test statistics, *Biometrika*, **83**, 923–927.

[3] Nelder, J. A. and Wedderburn, R. W. M. (1972) *Generalized linear models. J. R. Statist. Soc. A*, **135**, 370–384.

[4] Pardo, L. (2006). *Statistical inference based on divergence measures*, Chapman & Hall/CRC.

[5] Siotani, M. and Fujikoshi Y. (1984) *Asymptotic approximations for the distributions of multinomial goodness-of-fit statistics. Hiroshima Math. J.*, **14**, 115–124.

[6] Taneichi, N., Sekiya, Y. and Toyama, J. (2014) *Transformed goodness-of-fit statistics for a generalized linear model of binary data. J. Multivariate Anal.*, **123**, 311–329.

[7] Yarnold, J. K. (1972) *Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set. Ann. Math. Statist.*, **43**, 1566–1580.

The l_q consistency of the Dantzig selector for Cox's proportional hazards model

Kou Fujimori* and Yoichi Nishiyama

Waseda University

Let $t \rightsquigarrow N_i(t),\ i = 1, 2, \ldots, n, , t \in [0, \tau]$ for fixed $\tau > 0$ be counting processes which do not have simultaneous jumps. Suppose that the intensity of $N_i(t)$ has the form

$$\lambda_i(t, Z_i) = \alpha_0(t) Y_i(t) \exp(Z_i^T \beta_0),$$

where $Y_i(t)$ is the at risk process, $\alpha_0(t)$ is a nuisance baseline hazard function, $Z_i = (Z_i^1, Z_i^2, \ldots, Z_i^{p_n})$ is a covariate vector, and $\beta_0 \in \mathbb{R}^{p_n}$ is an unknown parameter. We are interested in the estimation problem of β_0 in a high dimensional and sparse setting, *i.e.*, $p_n \gg n$ and the number S of nonzero components of β_0 is relatively small. The log-partial likelihood is introduced by

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

Let the gradient of $l_n(\beta)$ be $U_n(\beta)$, and Hessian of $l_n(\beta)$ be $-J_n(\beta)$, *i.e.*,

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

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

where $S_n^k(\beta, u) = \sum_{i=1}^{n} Y_i(u) \exp(Z_i^T \beta) Z_i^{\otimes k}$ and $\bar{N}(u) = \sum_{i=1}^{n} N_i(u)$. Moreover, we assume that there exist $\mathbb{R}$-valued function $s_n^0(\beta, t)$, $\mathbb{R}^{p_n}$-valued function $s_n^1(\beta, t)$ and $p_n \times p_n$ matrix-valued function $s_n^2(\beta, t)$ such that

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

We define the $p_n \times p_n$ matrix $I_n(\beta)$ by

$$I_n(\beta) := \int_0^{\tau} \left[\frac{s_n^2}{s_n^0}(\beta, u) - \left(\frac{s_n^1}{s_n^0} \right)^{\otimes 2} (\beta, u) \right] s_n^0(\beta,\ u) \alpha_0(u) du.$$

Now, we define the estimator of β_0 by

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

To prove the l_q consistency for $q \in [1, \infty]$, we define some factors for $I_n(\beta_0)$.

*wisteria_forest@toki.waseda.jp

Definition 1. *For every index set $T \subset \{1,\ 2,\ \cdots,\ p_n\}$ and $h \in \mathbb{R}^{p_n}$, h_T is a $\mathbb{R}^{|T|}$ dimensional sub-vector of h constructed by extracting the components of h corresponding to the indices in T. Define the set C_T by $C_T := \{h \in \mathbb{R}^{p_n} : \|h_{T^c}\|_1 \le \|h_T\|_1\}$. We introduce the following factors.*

(A) **Compatibility factor**

$$\kappa(T_0; I_n(\beta_0)) := \inf_{0 \neq h \in C_{T_0}} \frac{S^{\frac{1}{2}}(h^T I_n(\beta_0) h)^{\frac{1}{2}}}{\|h_{T_0}\|_1}.$$

(B) **Weak cone invertibility factor**

$$F_q(T_0; I_n(\beta_0)) := \inf_{0 \neq h \in C_{T_0}} \frac{S^{\frac{1}{q}}(h^T I_n(\beta_0) h)^{\frac{1}{2}}}{\|h_{T_0}\|_1 \|h\|_q}, \quad q \in [1, \infty),$$

$$F_\infty(T_0; I_n(\beta_0)) := \inf_{0 \neq h \in C_{T_0}} \frac{(h^T I_n(\beta_0) h)^{\frac{1}{2}}}{\|h\|_\infty}.$$

(C) **Restricted eigenvalue**

$$RE(T_0; I_n(\beta_0)) := \inf_{0 \neq h \in C_{T_0}} \frac{(h^T I_n(\beta_0) h)^{\frac{1}{2}}}{\|h\|_2},$$

where $T_0 := \{j : \beta_{0j} \neq 0\}$.

Using these factors, we obtain the following results. Hereafter, put $\log p_n = O(n^\zeta)$, $0 < \zeta < \alpha$ and $\gamma = \gamma_{n,p_n} = K_1 \log(1 + p_n)/n^\alpha$, where $0 < \alpha \le 1/2$ and $K_1 > 0$ is a constant.

Theorem 2. *Assume that $\liminf_{n\to\infty} \kappa(T_0; I_n(\beta_0)) > 0$. Under some regularity conditions, following (i), (ii), (iii) and (iv) hold true for some positive constants K_2, K_3, K_4 and the random sequence $\epsilon_n = o_p(1)$.*

(i) *It holds that*

$$\lim_{n\to\infty} P\left(\|\hat{\beta}_n - \beta_0\|_2^2 \ge \frac{K_2 \gamma_{n,p_n} + K_3 \epsilon_n}{RE^2(T_0; I_n(\beta_0))}\right) = 0,$$

(ii) It holds that

$$\lim_{n\to\infty} P\left(\|\hat{\beta}_n - \beta_0\|_1 \ge \frac{4K_4 S \gamma_{n,p_n}}{\kappa^2(T_0; I_n(\beta_0)) - 4S\epsilon_n}\right) = 0,$$

(iii) For all $q \in [1, \infty)$, it holds that

$$\lim_{n\to\infty} P\left(\|\hat{\beta}_n - \beta_0\|_q \ge \frac{2S^{\frac{1}{q}}\epsilon_n}{F_q(T_0; I_n(\beta_0))} \cdot \frac{2K_4 S \gamma_{n,p_n}}{\kappa^2(T_0; I_n(\beta_0)) - 2S\epsilon_n} + \frac{2K_4 S^{\frac{1}{q}} \gamma_{n,p_n}}{F_q(T_0; I_n(\beta_0))}\right) = 0.$$

(iv) It holds that

$$\lim_{n\to\infty} P\left(\|\hat{\beta}_n - \beta_0\|_\infty^2 \ge \frac{K_2 \gamma_{n,p_n} + K_3 \epsilon_n}{F_\infty^2(T_0; I_n(\beta_0))}\right).$$

References

[1] Fujimori, K. and Nishiyama, Y. (2016). The l_q consistency of the Dantzig Selector for Cox's Proportional Hazards Model. *To appear in J. Statist. Plann. Inference.*

講演１０

Shrinkage estimators of Poisson means based on prior information and its applications to multiplicative Poisson models

目白大学　張　元宗,　　慶応義塾大学　篠崎　信雄

0.1　はじめに

$X_1, \cdots, X_p$ を互いに独立にポアソン分布 $P_o(\lambda_i), i = 1, \cdots, p(\geq 2)$, にしたがう確率変数とする。標準化２乗誤差損失関数（normalized squared error loss）

$$L(\hat{\lambda}, \lambda) = \sum_{i=1}^{p} \lambda_i^{-1}(\hat{\lambda}_i - \lambda_i)^2 \tag{1.1}$$

を基準としたとき、母平均 $\lambda = (\lambda_\mathbf{1}, \cdots, \lambda_\mathbf{p})'$ の同時推定問題に対して、Clevension & Zidek (1975) は原点に縮小する推定量

$$\hat{\lambda}_i^{CZ}(\mathbf{X}) = \left(1 - \frac{\varphi(Z)}{Z + p - 1}\right) X_i, \qquad i = 1, \ldots, p,$$

を提案した。ここで、$Z = \sum_{i=1}^{p} X_i$ である。$\varphi(Z)$ が非減少関数、$0 \leq \varphi(Z) \leq 2(p-1)$ ならばこの推定量が不偏推定量 $\mathbf{X} = (X_1, \cdots, X_p)'$ を改良することを示した。しかし、いくつかの λ_i が大きな値である場合、原点に縮小する推定量は大きな改良を与えるとは言えない。Ghosh, Hwang and Tsui(1983) および Tsui (1984,1986) は、指定した非負の整数点あるいは順序統計量に縮小する推定量を提案した。しかし、両論文で提案された推定量については、改善の余地がある。さらに、ポアソン母平均に simple tree order 制約がある場合に、 isotonic regression 推定量を改良する方法を提案する。また、multiplicative Poisson models での母平均の同時推定問題も取り上げ、順序統計量への縮小推定量を提案する。

0.2　事前情報への縮小

この節では、指定された非負な値及び順序統計量への縮小を論じる。応用例として、母平均に simple tree order 制約条件がある場合、isotonic regression 推定量を縮小する同時推定量を提案する。

1) 非負な値への縮小：$a_i \geq 0, i = 1, \cdots, p$ とし、部分集合 $\mathcal{C} = \{(x_1, \ldots, x_p) | x_i \geq a_i, i = 1, \ldots, p\}$ とそのインジケータ関数を $I_\mathcal{C}$ とする。a_i に縮小する推定量をつぎのように考える。

$$\hat{\lambda}_i(\mathbf{X}) = X_i - \varphi(Z_\mathcal{C}) \frac{(X_i - a_i)}{Z_\mathcal{C} + d} I_\mathcal{C}, \qquad i = 1, \ldots, p.$$

ここで、$Z_\mathcal{C} = \sum_{i=1}^{p}(X_i - a_i)$ であり、$d \geq 0$ である。

部分集合 $\mathcal{C}$ で (1.1) の損失関数の下で、$\mathbf{X}$ と $\hat{\lambda}(\mathbf{X}) = (\hat{\lambda}_1(\mathbf{X}), \ldots, \hat{\lambda}_p(\mathbf{X}))$ との平均損失の差を評価し、$\hat{\lambda}(\mathbf{X})$ が $\mathbf{X}$ を改良するための十分条件をつぎの定理で与える。

定理 2.1. $p \geq 2$ とし、$\varphi(\cdot)$ を非減少関数とする。損失関数 (1.1) の下で、$\hat{\lambda}(\mathbf{X})$ が $\mathbf{X}$ を改良するための十分条件は $0 \leq \varphi(\cdot) \leq 2(p-1)$, $d \geq \sup \varphi(\cdot)/2$ である。

次に、ある $k \geq 2$ に対して、部分集合 $\mathcal{C}_k = \{(x_1, \ldots, x_p) | x_i \geq a_i, i = 1, \ldots, k, x_j < a_j, j = k+1, \ldots, p\}$ とする。このような $2^p - p - 1$ 個の互いに排反な部分集合のそれぞれで a_i に縮小するような推定量を考える。つまり、$\mathbf{X} \in \mathcal{C}_k$ のとき、

$$\hat{\lambda}_i(\mathbf{X}) = \begin{cases} X_i - \varphi_k(Z_{\mathcal{C}_k}) \dfrac{(X_i - a_i)}{Z_{\mathcal{C}_k} + d_k}, & i = 1, \ldots, k, \\ X_i, & i = k+1, \ldots p, \end{cases}$$

を考える。ここで、$Z_{\mathcal{C}_k} = \sum_{i=1}^{k}(X_i - a_i)$ であり、 $d_k \geq 0$ である。各部分集合で $\mathbf{X}$ と $\hat{\lambda}(\mathbf{X})$ との平均損失の差を評価することで、$\hat{\lambda}(\mathbf{X})$ が $\mathbf{X}$ を改良するための十分条件をつぎの定理で与える。

定理 2.2 $\varphi_k(\cdot)$ を非減少関数とする。損失関数 (1.1) の下で、$\hat{\lambda}(\mathbf{X})$ が $\mathbf{X}$ を改良するための十分条件は $0 \leq \varphi_k(\cdot) \leq 2(k-1)$, $d_k \geq \sup \varphi_k(\cdot)/2$ である。

2) 順序統計量への縮小：$p \geq 3$ とし、最小値 $X_{(1)} = min\{X_1, \ldots, X_p\}$ に縮小するような推定量を

$$\hat{\lambda}_i(\mathbf{X}) = X_i - \varphi(W) \frac{X_i - X_{(1)}}{W + d}, \qquad i = 1, \ldots, p$$

を考える。ここで、$W = \sum_{k=1}^{p}(X_k - X_{(1)})$ である。

定理 2.3 $\varphi(\cdot)$ は非減少関数とする。損失関数 (1.1) の下で $\hat{\lambda}(\mathbf{X})$ が $\mathbf{X}$ を改良するための十分条件は $0 \leq \varphi(\cdot) \leq 2(p-2)$, $d \geq \sup \varphi(\cdot)/2$ である。

この十分条件は標本空間を p 個の互いに排反な部分集合に切り分け、各集合での平均損失の差を評価することで示される。

3) 応用例ー母平均に simple tree order 制約条件がある場合の同時推定: $X_i \sim P_o(\lambda_i), i=0,1,\ldots,p$ に従い、母数に simple tree order 制約条件、$\lambda_0 \le \lambda_i, i=1,\ldots,p$ がある場合、$\boldsymbol{\lambda}$ の isotonic regression 推定量は一般に次のように与えられる。

$$\hat{\lambda}_i^{st}(\mathbf{X}) = \begin{cases} X_i, & \text{for } i \in S^c \\ A_X(S), & \text{for } i \in S, \end{cases}$$

ここで、$S=\{0,1,\ldots,k\}$, $S^c=\{k+1,\ldots,p\}$ であり、$A_X(S)=\sum_{i\in S} X_i/(k+1)$, $X_i \ge A_X(S)$, $i \in S^c$ である。$p-k \ge 2$ ならば、$\hat{\lambda}_i^{st}(\mathbf{X})$ を次のように縮小する。

$$\hat{\lambda}_i^{m}(\mathbf{X}) = \begin{cases} X_i - \varphi_{p-k}(W_{S^c})\dfrac{X_i - A_X(S)}{W_{S^c}+d_{p-k}}, & \text{for } i \in S^c \\ A_X(S), & \text{for } i \in S, \end{cases}$$

ここで、$W_{S^c}=\sum_{i\in S^c}(X_i - A_X(S))$ である。

定理 2.4 損失関数 (1.1) の下で $\hat{\lambda}^m(\mathbf{X})$ が $\hat{\lambda}^{st}(\mathbf{X})$ を改良するための十分条件は $p-k \ge 2$, $0 \le \varphi_{p-k}(\cdot) \le 2(p-k-1)$、$\varphi_{p-k}(\cdot) \ge 0$ は非減少関数、$d_{p-k} \ge \sup \varphi_{p-k}(\cdot)/2$ である。

0.3 multiplicative Poisson models での母平均の同時推定問題への応用

本節では、multiplicative Poisson models での母平均の同時推定問題を考え、最尤推定量を順序統計量へ縮小する推定量を提案する。

multiplicative Poisson models $X_{i_1 i_2 \ldots i_J} \sim P_o(\lambda_{i_1 i_2 \ldots i_J})$ は下記のように表現される。$\lambda = \sum \lambda_{i_1 i_2 \ldots i_J}$ とすると、$\lambda_{i_1 i_2 \ldots i_J}$ は次のように表現される。

$$\lambda_{i_1 i_2 \ldots i_J} = \lambda \alpha_{1 i_1} \alpha_{2 i_2} \ldots \alpha_{J i_J}, i_j = 1,\ldots,I_j, j=1,\ldots,J,$$

ここで、

$$\alpha_{j i_j} > 0, \ \sum_{i_j=1}^{I_j} \alpha_{j i_j} = 1, \qquad j=1,\ldots J.$$

である。Hara, Takemura(2006) は $\boldsymbol{\lambda}=\{\lambda_{i_1 i_2 \ldots i_J}\}$ の最尤推定量

$$\hat{\lambda}^{MLE}_{i_1 i_2 \ldots i_J} = \begin{cases} \dfrac{\prod_j X^+_{j,i_j}}{(X^+)^{J-1}}, & \text{if} \quad X^+ \ne 0 \\ 0, & \text{if} \quad X^+ = 0 \end{cases}$$

を導出し、標準化 2 乗損失関数

$$L(\boldsymbol{\delta}, \boldsymbol{\lambda}) = \sum_{j=1}^{J} \sum_{i_j=1}^{I_j} \frac{1}{\lambda_{i_1 i_2 \ldots i_J}} (\delta_{i_1 i_2 \ldots i_J} - \lambda_{i_1 i_2 \ldots i_J})^2. \tag{3.1}$$

を基準としたとき、最尤推定量を原点に縮小する Clevension-Zidek タイプ推定量を提案した。

ここで、最尤推定量を順序統計量に縮小するような推定量を考える。k-th layout の周辺度数と総度数

$$X^+_{k,i_k} = \sum_{i_s : s \ne k} X_{i_1 i_2 \ldots i_K}, \qquad X^+ = \sum_{i_k=1}^{I_k} X^+_{k,i_k}$$

をとし、k-th layout の周辺度数のベクトル $\mathbf{X}^+_k = (X^+_{k,1},\ldots,X^+_{k,I_k})', k=1,\ldots,K$ とする。k-th layout の各要素 $\mathbf{X}^+_k = \{X^+_{k,i_k}, i_k = 1,\ldots,I_k\}$ を順序統計量への縮小推定量 $X^+_{k,(1)} \equiv \min\{X^+_{k,1},\ldots,X^+_{k,I_k}\}$, $\tilde{\lambda}^k = \{\tilde{\lambda}^k_{i_1 i_2 \ldots i_J}\}$,

$$\tilde{\lambda}^k_{i_1 i_2 \ldots i_J} = \frac{\prod_{\substack{j=1 \\ j \ne k}}^{J} X^+_{j,i_j}}{(X^+)^{J-1}} \left\{ X^+_{k,i_k} - \varphi_k(W_k) \frac{(X^+_{k,i_k} - X^+_{k,(1)})}{W_k + d_k} \right\}$$

を考える。ここで、

$$W_k = \sum_{i_k=1}^{I_k} (X^+_{k,i_k} - X^+_{k,(1)})$$

である。以下の定理が得られる。

定理 3.1 $I_k \ge 3$ をとし、$\varphi_k(\cdot)$ を非減少関数とする。損失関数 (3.1) の下で $\tilde{\lambda}^k$ が最尤推定量 $\hat{\lambda}^{MLE}$ を改良するための十分条件は

$$0 \le \varphi_k(\cdot) \le 2(I_k - 2), \ \ d_k \ge \frac{\varphi_k(\cdot)}{2}.$$

である。

招待
講演11

Bartlett correction of empirical likelihood with unknown variance

Kun Chen (Southwestern University of Finance and Economics)

Bartlett correction is one of the desirable features of likelihood inference, which allows constructions of confidence regions for parameters with improved coverage probabilities. In this paper we study Bartlett correction for frequency domain empirical likelihood based on the Whittle likelihood of linear time series models. Previous studies demonstrated the Bartlett correction of EL for independent observations, Gaussian short- and long-memory time series with known innovation variance. Nordman and Lahiri (2006) showed that frequency domain empirical log-likelihood ratio statistics does not have an ordinary χ^2-limit when the innovation is non-Gaussian with unknown variance, which restricts the use of empirical likelihood inference in time series. By profiling out the innovation variance from the Whittle likelihood function, we show that the empirical log-likelihood ratio statistic is χ^2-distributed and is Bartlett correctable. In particular, orders of the coverage error of confidence regions can be reduced from $1/n$ to $1/n^2$.

Locally stable regression without ergodicity and finite moments

Hiroki Masuda
Kyushu University

Model setup. Given a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbb{P})$ with $\mathcal{F}_t = \sigma(X_0)\vee \sigma(J_s; s\le t)$, we consider a univariate stochastic process process $X = (X_t)_{t\in[0,T]}$ defined through

$$X_t = X_0 + \int_0^t a(X_s, H_s; \alpha)ds + \int_0^t c(X_{s-}, H_{s-}; \gamma)dJ_s,$$

where: the initial random variable X_0 is $\mathcal{F}_0$-measurable; the driving process J is a pure-jump Lévy process independent of X_0 and having the Lévy-Khintchine representation

$$\mathbb{E}(e^{iuJ_t}) = \exp\left\{ t\bigg(\int_{|z|\le 1} (e^{iuz} - 1 - iuz)\nu(dz) + \int_{|z|>1} (e^{iuz} - 1)\nu(dz) \bigg) \right\}$$

for $t\in\mathbb{R}_+$ and $u\in\mathbb{R}$; the trend coefficient $a : \mathbb{R}\times\Theta_\alpha \to \mathbb{R}$ and scale coefficient $c : \mathbb{R}\times\Theta_\gamma \to \mathbb{R}$ are assumed to be known except for the p-dimensional parameter $\theta := (\alpha, \gamma) \in \Theta_\alpha\times\Theta_\gamma = \Theta \subset \mathbb{R}^p$, with $\Theta_\alpha \in \mathbb{R}^{p_\alpha}$ and $\Theta_\gamma \in \mathbb{R}^{p_\gamma}$ being bounded convex domains. Assume that the true value $\theta_0 = (\alpha_0, \gamma_0) \in \Theta$ exists, and that we only observe a discretized step process $X^{(n)}_t := X_{\lfloor t/h_n\rfloor h_n}$, $t\in[0,T]$, where the sampling step size $h_n = h := T/n \to 0$ for a fixed terminal sampling time T. Further, we assume the following regularity conditions.

(1) Regularity of the coefficients: (a, c) smooth enough, with $a(\cdot,\cdot,\alpha_0)$ and $c(\cdot,\cdot,\gamma_0)$ globally Lipschitz; $\forall(x, y)$, $\sup_\gamma |c(x, y, \gamma)|^{-1} \le K(1 + |x| + |y|)^K$ for some $K \ge 0$; the process $H = (H_t)_{t\in[0,T]}$ is a strong solution to a smooth stochastic differential equation driven by J and another Lévy process J' independent of J (the details omitted).

(2) Identifiability: $(a(\cdot,\cdot,\alpha),\, c(\cdot,\cdot,\gamma)) = (a(\cdot,\cdot,\alpha_0),\, c(\cdot,\cdot,\gamma_0)) \iff \theta = \theta_0$.

(3) $\sqrt{n}\int_{\mathbb{R}} |f_h(y) - \phi_\beta(y)|dy \to 0$, where f_h denotes the Lebesgue density of $\mathcal{L}(h^{-1/\beta}J_h)$ and ϕ_β the standard β-stable density corresponding to the characteristic function $u \mapsto e^{-|u|^\beta}$.

Some of them could be weakened in compensation for more complicated descriptions.

Results. We introduce the *stable quasi-maximum likelihood estimator (SQMLE)*:

$$\hat\theta_n = (\hat\alpha_n, \hat\gamma_n) \in \operatorname*{argmax}_{\theta\in\Theta} \sum_{j=1}^n \log\left\{ \frac{1}{h_n^{1/\beta} c_{j-1}(\gamma)} \phi_\beta\bigg(\frac{\Delta_j X - h a_{j-1}(\alpha)}{h^{1/\beta} c_{j-1}(\gamma)} \bigg) \right\},$$

where $c_{j-1}(\gamma) := c(X_{t_{j-1}}, H_{t_{j-1}}; \gamma)$ with a similar manner for $a_{j-1}(\alpha)$. For an a.s. positive definite $\mathcal{F}$-measurable random variable $A \in \mathbb{R}^p \otimes \mathbb{R}^p$, we denote by $MN_p(0, A(\omega))$ the mixed normal distribution $\mathcal{L}(A^{1/2}Z)$ where Z is a p-dimensional standard-normal random vector independent of $\mathcal{F}$ defined on an extension of the original probability space.

The next theorem clarifies the asymptotic distribution of the SQMLE.

Theorem. *Under the aforementioned setting, we have the asymptotic mixed normality*

$$\Big(\sqrt{n}h^{1-1/\beta}(\hat\alpha_n - \alpha_0),\ \sqrt{n}(\hat\gamma_n - \gamma_0) \Big) \xrightarrow{\mathcal{L}} MN_p\left(0,\, \Gamma_T(\theta_0; \beta)^{-1}\right),$$

where $\Gamma_T(\theta_0;\beta) := \operatorname{diag}\{C_\alpha(\beta)\Sigma_{T,\alpha}(\theta_0),\, C_\gamma(\beta)\Sigma_{T,\gamma}(\gamma_0)\}$ *with*

$$C_\alpha(\beta) = \int \left(\frac{\partial\phi_\beta}{\phi_\beta}(y) \right)^2 \phi_\beta(y)dy, \qquad C_\gamma(\beta) = \int \left(1 + y\frac{\partial\phi_\beta}{\phi_\beta}(y) \right)^2 \phi_\beta(y)dy,$$

$$\Sigma_{T,\alpha}(\theta_0) = \frac{1}{T}\int_0^T \frac{\{\partial_\alpha a(X_t, H_t; \alpha_0)\}^{\otimes 2}}{c^2(X_t, H_t; \gamma_0)}dt, \qquad \Sigma_{T,\gamma}(\gamma_0) = \frac{1}{T}\int_0^T \frac{\{\partial_\gamma c(X_t, H_t; \gamma_0)\}^{\otimes 2}}{c^2(X_t, H_t; \gamma_0)}dt.$$

Corollary. *Under the same situation, we have the asymptotic standard normality:*

$$\left(\hat{\Gamma}_{T,\alpha,n}^{1/2}\sqrt{n}h^{1-1/\beta}(\hat{\mu}_n-\mu_0),\,\hat{\Gamma}_{T,\gamma,n}^{1/2}\sqrt{n}(\hat{\sigma}_n-\sigma_0)\right)\xrightarrow{\mathcal{L}}N_p(0,I_p)$$

where $\hat{\Gamma}_{T,\alpha,n}=C_\alpha(\beta)\hat{\Sigma}_{T,\alpha,n}$ *and* $\hat{\Gamma}_{T,\gamma,n}:=C_\gamma(\beta)\hat{\Sigma}_{T,\gamma,n}$ *with*

$$\hat{\Sigma}_{T,\alpha,n}:=\frac{1}{n}\sum_{j=1}^{n}\frac{\{\partial_\alpha a_{j-1}(\hat{\alpha}_n)\}^{\otimes 2}}{c_{j-1}^2(\hat{\gamma}_n)},\qquad \hat{\Sigma}_{T,\gamma,n}:=\frac{1}{n}\sum_{j=1}^{n}\frac{\{\partial_\gamma c_{j-1}(\hat{\gamma}_n)\}^{\otimes 2}}{c_{j-1}^2(\hat{\gamma}_n)}.$$

Remarks.

(1) Under (3) the driving Lévy process J is locally (small-time) β-stable, that is, the limit distribution of $\mathcal{L}(h^{-1/\beta}J_h)$ for $h\to 0$ is standard β-stable. This property is satisfied by many specific Lévy processes such as the generalized hyperbolic, Student-t, Meixner, stable, and the (normal) tempered stable Lévy processes. Moreover, J can distributionally approximate a Wiener process by controlling dominating parameters in a suitable manner.

(2) The asymptotic distribution of $\hat{\theta}_n$ is normal if both $x\mapsto\frac{\partial_\gamma c(x,y,\gamma_0)}{c(x,y,\gamma_0)}$ and $x\mapsto\frac{\partial_\alpha a(x,y,\alpha_0)}{c(x,y,\gamma_0)}$ are non-random; this is the case especially if X is a Lévy process.

(3) The estimators $\hat{\alpha}_n$ and $\hat{\gamma}_n$ are asymptotically orthogonal, whereas not necessarily independent due to possible non-Gaussianity in the limit.

(4) For $\beta\in(1,2)$, we can rewrite Theorem as (recall that $h=T/n$)

$$\left(n^{1/\beta-1/2}(\hat{\alpha}_n-\alpha_0),\ \sqrt{n}(\hat{\gamma}_n-\gamma_0)\right)$$
$$\xrightarrow{\mathcal{L}}MN_p\left(0,\,\mathrm{diag}\big(T^{-2(1-1/\beta)}\{C_\alpha(\beta)\Sigma_{T,\alpha}(\theta_0)\}^{-1},\,\{C_\gamma(\beta)\Sigma_{T,\gamma}(\gamma_0)\}^{-1}\big)\right).$$

If fluctuation of X is virtually stable in the sense that both of the random time averages $\Sigma_{T,\alpha}(\theta_0)$ and $\Sigma_{T,\gamma}(\gamma_0)$ do not vary so much with the terminal sampling time T, then the asymptotic covariance matrix of $\hat{\alpha}_n$ would tend to get smaller (resp. larger) in magnitude for a larger (resp. smaller) T; this feature is non-asymptotic in T.

(5) Of special interest is the locally Cauchy case ($\beta=1$), where $\mathbb{H}_n$ is fully explicit. The corresponding asymptotic distribution of $(\sqrt{n}(\hat{\alpha}_n-\alpha_0),\ \sqrt{n}(\hat{\gamma}_n-\gamma_0))$ is the centered mixed-normal with the random covariance matrix given by

$$\mathrm{diag}\left\{\left(\frac{1}{2T}\int_0^T\frac{\{\partial_\alpha a(X_t,H_t,\alpha_0)\}^{\otimes 2}}{c(X_t,H_t,\gamma_0)^2}dt\right)^{-1},\left(\frac{1}{2T}\int_0^T\frac{\{\partial_\gamma c(X_t,H_t,\gamma_0)\}^{\otimes 2}}{c(X_t,H_t,\gamma_0)^2}dt\right)^{-1}\right\}.$$

This formally extends the i.i.d. model from the location-scale Cauchy population, where we have $\sqrt{n}$-asymptotic normality for the maximum-likelihood estimator.

(6) It is also possible to deduce large-time counterparts to Theorem and Corollary under the the ergodicity. In that case the asymptotic distribution is not mixed normal but normal, with the asymptotic covariance matrix taking a completely analogous form. The result formally extends the Gaussian quasi-likelihood estimation of diffusion.

(7) We conjecture that our SQMLE is asymptotically efficient in the present setting; it is the case for some special models.

See [1] for further and full details in absence of the covariate process H in (a,c), as well as for many relevant references and background materials.

References

[1] Masuda, H. (2016), Non-Gaussian quasi-likelihood estimation of SDE driven by locally stable Lévy process. *arXiv:1608.06758*

Recent developments in inferences for conditionally heteroscedastic location-scale time series models

Sangyeol Lee (Seoul National Univ.)

This study examines the asymptotic properties of a class of conditionally heteroscedastic location-scale time series models with innovations following a generalized asymmetric Student t distribution (ASTD) or an asymmetric exponential power distribution (AEPD). We first show the consistency and asymptotic normality of the conditional maximum likelihood estimator of the model parameters under certain regularity conditions. Then, based on the maximum likelihood estimator, we estimate conditional value-at-risk (VaR) and expected shortfall (ES) by using their closed forms induced from the model. Their performance is finally compared with that of conditional autoregressive VaR and expectile methods. To ensure the adequacy of the model in advance of the VaR and ES calculation, we develop an entropy-type goodness-of-fit test based on residuals and a residual-based cumulative sum test to conduct a parameter change test. To handle the former, we also investigate the asymptotic behavior of the residual empirical process.

Large sample behaviour of high dimensional autocovariance matrices when the dimension grows slower than the sample size

Arup Bose (Indian Statistical Institute)

Consider a sample of size n from a linear process of dimension p where $n, p \to \infty, p/n \to 0$. Let $\hat{\Gamma}_u$ be the sample autocovariance of order u. The existence of the limiting spectral distribution (LSD) of $\hat{\Gamma}_u + \hat{\Gamma}_u^*$, the symmetric sum of the sample autocovariance matrix $\hat{\Gamma}_u$ of order u, after appropriate centering and scaling, has been considered in the literature in exactly one article under appropriate (strong) assumptions on the coefficient matrices. Under significantly weaker conditions, we prove, in a unified way, that the LSD of any symmetric polynomial in these matrices such as $\hat{\Gamma}_u + \hat{\Gamma}_u^*$, $\hat{\Gamma}_u\hat{\Gamma}_u^*$, $\hat{\Gamma}_u\hat{\Gamma}_u^* + \hat{\Gamma}_k\hat{\Gamma}_k^*$, after suitable centering and scaling, exists and is non-degenerate. We use methods from free probability in conjunction with the method of moments to establish our results but unlike in the case $p/n \to y \in (0, \infty)$, the embedding technique does not work in this scenario. In addition, we are able to provide a general description for the limits in terms of some freely independent variables. The earlier result follows as a special case. We also establish asymptotic normality results for the traces of these matrices. We suggest statistical uses of these results in problems such as order determination of high-dimensional MA and AR processes and testing of hypotheses for coefficient matrices of such processes.

Monte Carlo filtering and data assimilation

Hans R. Kuñsch (Seminar for Statistics, ETH Zurich)*
Joint work with Marco Frei and Sylvain Robert

The problem of estimating the current state of a latent Markov process based on a sequence of partial and noisy observations up to the same time is called filtering in engineering and in statistics, and data assimilation in the geosciences. In the linear Gaussian case, the Kalman filter provides the exact solution in a recursive form, but for nonlinear and non-Gaussian cases one has to rely on approximations by recursive Monte Carlo algorithms. The two most widely used such algorithms are the particle filter and the ensemble Kalman filter. The former originated in statistics and engineering, the latter in the geosciences, and until recently there was little exchange of ideas between these two areas. In this talk, I will describe the basics of both algorithms, discuss their strengths and weaknesses and present some recent proposals that aim to combine their strengths.

Statistical inference for misspecified ergodic Lévy driven stochastic differential equation models

Yuma Uehara
Graduate school of Mathematics, Kyushu University

Currently, we can obtain high-frequency data stemming from time varying phenomena, such as log returns, spike noise of neurons and so on. To better describe their non-Gaussian random behavior, Lévy driven stochastic differential equation (SDE) model is regarded as one of good candidate models. In this study, we tackle the parametric estimation problem based on high-frequency samples in the case that the coefficients of Lévy driven SDE model are partly or fully misspecified. The outline of our result is presented below.

We suppose that the data generating model is the following one-dimensional SDE model:

$$dX_t = A(X_t)dt + C(X_{t-})dZ_t, \tag{0.1}$$

defined on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, \mathbb{P})$ where:

- X_0 is an $\mathcal{F}_0$-measurable initial variable;
- Z_t is a $\mathcal{F}_t$-adapted one-dimensional Lévy process without a Brownian part being independent of X_0 and it satisfies that $E[Z_t] = 0$, $Var[Z_t] = t$, and $E[|Z_t|^q] < \infty$ for all $q > 0$;
- The coefficients $A : \mathbb{R} \mapsto \mathbb{R}$ and $C : \mathbb{R} \mapsto \mathbb{R}$ are measurable.

We also assume the ergodicity of X. Hereinafter we write $\pi_0(\cdot)$ and $\nu_0(\cdot)$ as an invariant measure of X and Lévy measure of Z, respectively. Under high-frequency samples $(X_{t_0}, \dots, X_{t_n})$ from (0.1) being obtained, we intend to assign a parametric one-dimensional SDE model such that

$$dX_t = a(X_t, \alpha)dt + c(X_{t-}, \gamma)dZ_t, \tag{0.2}$$

where functional forms of coefficients $a : \mathbb{R} \mapsto \mathbb{R}$ and $c : \mathbb{R} \mapsto \mathbb{R}$ are known up to finite-dimensional unknown parameter $\theta := (\alpha, \gamma)$ being an element of bounded convex domain $\Theta \subset \mathbb{R}^p$. We also suppose that $t_j := t_j^n = jh_n$ for any $j \in \{1, \dots, n\}$ with $h_n \to 0$, $nh_n \to \infty$, and $nh_n^2 \to 0$, and that true coefficients $(A, C)(\cdot)$ are not in the parametric family $\{(a, c)(\cdot, \theta) : \theta \in \Theta\}$, that is, the misspecification concerning coefficients occurs. To estimate an *optimal* value of θ in a feasible way, we utilize Gaussian quasi-likelihood approach attaching importance to the tractability and its broad applicable range; the optimality is herein determined by maximizing the limiting function of a quasi-likelihood random field which will be introduced below. The goal of this study is to verify the asymptotic behavior of Gaussian quasi-likelihood based estimators $\hat{\theta}_n$.

Our estimation scheme can be expressed as follows:

1. *Drift-free estimation of* γ. Define a maximum contrast estimator $\hat{\gamma}_n$ by

$$\hat{\gamma}_n \in \underset{\gamma\in\bar{\Theta}_\gamma}{\operatorname{argmax}}\, \mathbb{G}_{1,n}(\gamma) \left(:= -\frac{1}{nh_n} \sum_{j=1}^n \left\{ h_n \log c_{j-1}^2(\gamma) + \frac{(\Delta_j X)^2}{c_{j-1}^2(\gamma)} \right\} \right),$$

2. *Weighted least square estimation of* α. Define a maximum contrast estimator $\hat{\alpha}_n$ by

$$\hat{\alpha}_n \in \underset{\alpha\in\bar{\Theta}_\alpha}{\operatorname{argmax}}\, \mathbb{G}_{2,n}(\alpha) \left(:= -\frac{1}{nh_n} \sum_{j=1}^n \frac{(\Delta_j X - h_n a_{j-1}(\alpha))^2}{h_n c_{j-1}^2(\hat{\gamma}_n)} \right),$$

Here $\Delta_j X$ and $f_j(\cdot)$ denote $X_{t_j} - X_{t_{j-1}}$ and $f(X_{t_j}, \cdot)$, respectively (f is a $\mathbb{R}$-valued function). As was mentioned above, we define an *optimal* value $\theta^\star := (\alpha^\star, \gamma^\star)$ of θ in the following manner:

$$\gamma^\star \in \operatorname*{argmax}_{\gamma \in \bar{\Theta}_\gamma} \mathbb{G}(\gamma) \left(:= - \int_{\mathbb{R}} \left(\log c^2(x,\gamma) + \frac{C^2(x)}{c^2(x,\gamma)} \right) \pi_0(dx) \right),$$
$$\alpha^\star \in \operatorname*{argmax}_{\alpha \in \bar{\Theta}_\alpha} \mathbb{G}(\alpha) \left(:= - \int_{\mathbb{R}} c(x,\gamma^\star)^{-2} (A(x) - a(x,\alpha))^2 \pi_0(dx) \right).$$

For such estimator $\hat{\theta}_n := (\hat{\gamma}_n, \hat{\alpha}_n)$, we derived its tail probability estimates and asymptotic normality under suitable regularity and moment condition:

Theorem 0.1 (Tail probability estimates)**.** *For any $L > 0$ and $r > 0$, there exists a positive constant C_L such that*

$$\sup_{n \in \mathbb{N}} P(|\sqrt{nh_n}(\hat{\theta}_n - \theta^\star)| > r) \le \frac{C_L}{r^L}.$$

Theorem 0.2 (Asymptotic normality)**.** *We introduce $p \times p$-matrix $\Gamma := \begin{pmatrix} \Gamma_\gamma & O \\ O & \Gamma_\alpha \end{pmatrix}$ and each component is defined by:*

$$\begin{aligned}
\Gamma_\gamma &:= 2 \int_{\mathbb{R}} \frac{\partial_\gamma^{\otimes 2} c(x,\gamma^\star) c(x,\gamma^\star) - (\partial_\gamma c(x,\gamma^\star))^{\otimes 2}}{c^4(x,\gamma^\star)} (c^2(x,\gamma^\star) - C^2(x)) \pi_0(dx) \\
&\quad + 4 \int_{\mathbb{R}} \frac{(\partial_\gamma c(x,\gamma^\star))^{\otimes 2}}{c^4(x,\gamma^\star)} C^2(x) \pi_0(dx), \\
\Gamma_\alpha &:= 2 \int \frac{\partial_\alpha^{\otimes 2} a(x,\alpha^\star)}{c^2(x,\gamma^\star)} (A(x) - a(x,\alpha^\star)) \pi_0(dx) + 2 \int \frac{(\partial_\alpha a(x,\alpha^\star))^{\otimes 2}}{c^2(x,\gamma^\star)} \pi_0(dx),
\end{aligned}$$

where $\otimes 2$ denotes the tensor product. Then, there exists a nonnegative definite matrix $\Sigma \in \mathbb{R}^p \otimes \mathbb{R}^p$ such that

$$\sqrt{nh_n}(\hat{\theta}_n - \theta^\star) \xrightarrow{\mathcal{L}} N(0, \Gamma^{-1} \Sigma (\Gamma^{-1})^\top),$$

In addition, if f_1 and f_2 are twice differentiable and their first and second derivatives are of at most polynomial growth, the explicit form of $\Sigma := \begin{pmatrix} \Sigma_\gamma & \Sigma_{\alpha\gamma} \\ \Sigma_{\alpha\gamma}^\top & \Sigma_\alpha \end{pmatrix}$ is given by:

$$\begin{aligned}
\Sigma_\gamma &= 4 \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\frac{\partial_\gamma c(x,\gamma^\star)}{c^3(x,\gamma^\star)} C^2(x) z^2 - f_1(x + C(x)z) + f_1(x) \right)^{\otimes 2} \pi_0(dx) \nu_0(dz), \\
\Sigma_{\alpha\gamma} &= 4 \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\frac{\partial_\gamma c(x,\gamma^\star)}{c^3(x,\gamma^\star)} C^2(x) z^2 - f_1(x + C(x)z) + f_1(x) \right) \\
&\qquad \left(\frac{\partial_\alpha a(x,\alpha^\star)}{c^2(x,\gamma^\star)} C(x) z - f_2(x + C(x)z) + f_2(x) \right)^\top \pi_0(dx) \nu_0(dz), \\
\Sigma_\alpha &= 4 \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\frac{\partial_\alpha a(x,\alpha^\star)}{c^2(x,\gamma^\star)} C(x) z - f_2(x + C(x)z) + f_2(x) \right)^{\otimes 2} \pi_0(dx) \nu_0(dz).
\end{aligned}$$

As can be seen from above theorems, the convergence rate of $\hat{\theta}_n$ is $\sqrt{nh_n}$ and actually this is the same in correctly specified case (Masuda and Uehara [1]). This is a clear difference with the diffusion case.

References

[1] Masuda, H. and Uehara, Y. (2016). On stepwise estimation of Lévy driven stochastic differential equation., in Japanese, submitted.

Bartlett correction to the likelihood ratio test for MCAR with two-step monotone missing data

首藤　信通　(神戸大学 大学院海事科学研究科)
西山　貴弘　(専修大学 経営学部)
兵頭　　昌　(大阪府立大学 大学院工学研究科)

本報告では欠測データメカニズムの１つである Missing Completely at Random (MCAR) の成立に関する仮説検定問題を扱った. 欠測データメカニズムは, 適用可能な欠測値データ解析法を選択する上で重要な仮定である.

欠測値データに基づく尤度関数は, p 次元観測ベクトル $\boldsymbol{X} = (X_1, \ldots, X_p)'$ と p 次元欠測指標ベクトル $\boldsymbol{R} = (R_1, \ldots, R_p)'$ の同時分布に基づいて構成される. ここに, $\boldsymbol{R}$ は $\boldsymbol{X}$ における観測の有無を表す確率ベクトルであり, $i = 1, \ldots, p$ において $R_i = 1$ のとき X_i の観測, $R_i = 0$ のとき X_i の欠測を表す. このとき, 尤度関数は $\boldsymbol{X}$ の実測値に欠測値を含んでいる. また, 一般に $\boldsymbol{X}$ の分布のパラメータに対する尤度方程式は $\boldsymbol{R}$ の分布のパラメータに依存する. さらに, 欠測データメカニズムを同定することが必要となるが, 妥当なモデルを選択することは困難である. しかしながら, 欠測データメカニズムが MCAR であるとき, 特定の欠測パターンのデータセットのみを用いて $\boldsymbol{X}$ の分布のパラメータの推定を行うことができる. すなわち, MCAR が成り立つ下では Listwise deletion を適用することが可能となる (例えば, Little and Rubin (2002) を参照).

本報告では p 次元正規母集団から N 個のランダム標本 $\boldsymbol{X}_j$ $(j = 1, \ldots, N)$ が得られたとし, 欠測パターンが単調欠測である場合を仮定した下で, Little (1988) が与えた MCAR の仮説に対する尤度比検定統計量の帰無分布に対し, 漸近展開を導出した. 本報告では, 簡単のため, 欠測パターン数２の単調欠測データ (2-step 単調欠測データ) の場合, すなわち p 次元標本ベクトル $\boldsymbol{X}_j | \boldsymbol{R}_j = \mathbf{1}_p \overset{i.i.d.}{\sim} N_p(\boldsymbol{\mu}, \Sigma)$ $(j = 1, \ldots, N_1)$ と d 次元標本ベクトル

$$\boldsymbol{X}_{1j} | \boldsymbol{R}_j = \begin{pmatrix} \mathbf{1}_d \\ \mathbf{0}_{p-d} \end{pmatrix} \overset{i.i.d.}{\sim} N_d(\boldsymbol{\nu}_1, \Psi_{11}) \ (j = N_1 + 1, \ldots, N)$$

が得られた場合を考える. ここに, $d < p$, $N = N_1 + N_2$, $\mathbf{1}_p$ を 1 を成分にもつ p 次元定数ベクトル, $\boldsymbol{X}_{1j}$ は $\boldsymbol{X}_j$ の d 次元分割ベクトルである.

2-step 単調欠測データの下で, MCAR の成立に関する仮説検定問題は

$$H: \ \boldsymbol{\mu}_1 = \boldsymbol{\nu}_1, \ \Sigma_{11} = \Psi_{11} \text{ vs. } A :\neq H$$

に対する仮説検定問題として扱うことができる. ここに, $\boldsymbol{\mu}_1$, Σ_{11} はそれぞれ, $\boldsymbol{\nu}_1$, Ψ_{11} の成分に対応する $\boldsymbol{\mu}$ の分割ベクトル, Σ の分割行列である. 従って, 簡単な係数の調整を施した Little (1988) の尤度比検定統計量

$$\begin{aligned} T &= \boldsymbol{z}'\Big(\frac{1}{n}(W_{F,11} + W_{L,11} + \boldsymbol{z}\boldsymbol{z}')\Big)^{-1}\boldsymbol{z} + n\Big\{\ln\Big|\frac{1}{n}(W_{F,11} + W_{L,11} + \boldsymbol{z}\boldsymbol{z}')\Big| \\ &\qquad + \mathrm{tr}[(W_{F,11} + W_{L,11})(W_{F,11} + W_{L,11} + \boldsymbol{z}\boldsymbol{z}')^{-1}] - d\Big\} \\ &\quad -n_1 \ln\Big|\frac{1}{n_1}W_{F,11}\Big| - n_2 \ln\Big|\frac{1}{n_2}W_{L,11}\Big| \end{aligned}$$

を得る．ここに $n_g = N_g - 1\ (g = 1, 2)$, $\overline{\boldsymbol{X}}_{1F} = N_1^{-1}\sum_{j=1}^{N_1}\boldsymbol{X}_{1j}$, $\overline{\boldsymbol{X}}_{1L} = N_2^{-1}\sum_{j=N_1+1}^{N}\boldsymbol{X}_{1j}$, $\boldsymbol{z} = \sqrt{(N_1N_2)/N}(\overline{\boldsymbol{X}}_{1F} - \overline{\boldsymbol{X}}_{1L})$,

$$W_{F,11} = \sum_{j=1}^{N_1}(\boldsymbol{X}_{1j} - \overline{\boldsymbol{X}}_{1F})(\boldsymbol{X}_{1j} - \overline{\boldsymbol{X}}_{1F})', \quad W_{L,11} = \sum_{j=N_1+1}^{N}(\boldsymbol{X}_{1j} - \overline{\boldsymbol{X}}_{1L})(\boldsymbol{X}_{1j} - \overline{\boldsymbol{X}}_{1L})'$$

である．

本報告では, Nagao (1973) と同様の漸近展開の導出方針を用いて, 帰無仮説 H 及び大標本漸近枠組み $n_1, n_2 \to \infty,\ \ \gamma_g = n_g/n \to c_g \in (0, 1)\ (g = 1, 2)$ の下で, T の漸近的な特性関数

$$\varphi(t) = (1 - 2it)^{-\frac{f}{2}}\left[1 - \frac{d}{24n}g(\tilde{c})\{1 - (1 - 2it)^{-1}\}\right] + \mathrm{O}(n^{-2})$$

を導出した (詳細については, Shutoh, Nishiyama and Hyodo (2016) を参照)．ここに, $f = d(d + 3)/2$, $n = n_1 + n_2$, $g(c) = (2d^2 + 3d - 1)(c - 1) + 6d$, $\tilde{c} = \sum_g c_g^{-1}$ である．また, 上記の結果を基にして, Bartlett 修正を施した検定統計量

$$T_B = \left(1 - \frac{K}{n}\right)T, \quad K = \frac{g(\tilde{c})}{6(d + 3)}$$

を提案した．この検定統計量は $\Pr[T_B \le x] = P_f(x) + \mathrm{O}(n^{-2})$ を満たすものである．ここに, $P_f(\cdot)$ は自由度 f の χ^2-分布の分布関数である．

最後に, 数値実験を行い, T 及び T_B の自由度 f の χ^2-分布への収束の様子を考察した．具体的には

$$(p, d) = (4, 1), (4, 2), (4, 3),\ \ M = N_1 = N_2 = 10, 15, 20, 25,\ \ \alpha = 0.10, 0.05, 0.01$$

をパラメータとして与え, シミュレーションで与えられる T 及び T_B の分布の上側 $100\alpha\%$ 点と自由度 f の χ^2-分布の上側 $100\alpha\%$ 点 $\chi_f^2(\alpha)$ を比較した．数値実験の結果から, M が比較的小さい状況下においても T_B の分布の上側 $100\alpha\%$ 点は $\chi_f^2(\alpha)$ の値とほぼ等しく, 本報告の数値実験で実行したすべての場合において, T_B は T よりも正確に有意水準を保つ検定統計量であることを数値的に確認した.

参考文献

[1] Little, R. J. A., Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd eds.). John Wiley & Sons, Inc., Hoboken, New Jersey.

[2] Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *J. Amer. Statist. Assoc.*, **83**, 1198–1202.

[3] Nagao, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.*, **1**, 700–709.

[4] Shutoh, N., Nishiyama, T., Hyodo, M. (2016). Bartlett correction to the likelihood ratio test for MCAR with two-step monotone sample. *Hiroshima Statistical Research Group Technical Report No.16–15, Hiroshima University.*

Multi-group profile analysis for high-dimensional elliptical populations

大阪府立大学　田村 安徳

第 g 母集団からの無作為標本を $\boldsymbol{X}_{g1},\ldots,\boldsymbol{X}_{gn_g} \overset{i.i.d.}{\sim} C_p(\xi,\boldsymbol{\mu}_g,\Lambda_g)$ とし, $\boldsymbol{X}_{11},\ldots,\boldsymbol{X}_{1n_1},\ldots,\boldsymbol{X}_{a1},\ldots,\boldsymbol{X}_{an_a}$ は互いに独立とする．ただし, $g=1,\ldots,a$ であり, $C_p(\xi,\boldsymbol{\mu}_g,\Lambda_g)$ は楕円分布である.

プロフィール分析には, 平行性仮説, 一致性仮説, 平坦性仮説の 3 つの仮説がある．一致性仮説と平坦性仮説は, 平行性仮説が成り立った下での仮説である．標本数が次元よりも大きい大標本においては, Srivastava (1987) が, 2 つの正規母集団に対して尤度比検定を用いたプロフィール分析を提案した．また, Okamoto et al. (2006) は, 楕円母集団の下で, 検定統計量の帰無分布の漸近展開を摂動法を用いて導出した．Maruyama (2007) は, Kano (1995) で提案された手法を利用し, 非正規母集団の下で, 検定統計量の帰無分布の漸近展開を導出した．しかし, 尤度比検定は, 次元が標本数よりも大きい高次元データに対しては適用できない．そこで, Takahashi and Shutoh (2016) は等分散性を仮定した 2 つの正規母集団に対する高次元におけるプロフィール分析を提案した．また, Onozawa et al. (2014) は, 等分散性が必ずしも必要でない 2 つの正規母集団に対する高次元におけるプロフィール分析を提案した．さらに, Harra and Kong (2016) は, 等分散性が必ずしも必要でない $k(\geq 2)$ 個の正規母集団に対する高次元におけるプロフィール分析を提案した．本研究では, Harra and Kong (2016) の結果を楕円母集団へ拡張し, プロフィール分析における 3 つの仮説に対する近似検定を提案した．プロフィール分析における 3 つの仮説は以下のように表される:

(i) 平行性仮説 $H_{01}:\boldsymbol{\mu}_g-\boldsymbol{\mu}_a=\gamma_g\mathbf{1}_p$ v.s. $A_{01}:$ not H_{01}.

(ii) 一致性仮説 $H_{02}|H_{01}:\gamma_1=\cdots=\gamma_{a-1}=0$ v.s. $A_{02}|H_{01}:$ not $H_{02}|H_{01}$.

(iii) 平坦性仮説 $H_{03}|H_{01}:\mu_{g1}=\cdots=\mu_{gp}$ v.s. $A_{03}|H_{01}:$ not $H_{03}|H_{01}$.

(i), (iii) を $\boldsymbol{\mu}=(\boldsymbol{\mu}_1',\ldots,\boldsymbol{\mu}_a')'$を用いて書き換えると, 以下のように表される：

(i)' 平行性仮説 $H_{01}:\boldsymbol{\mu}'K_{AB}\boldsymbol{\mu}=0$ v.s. $A_{01}:$ not H_{01}.

(iii)' 平坦性仮説 $H_{03}|H_{01}:\boldsymbol{\mu}'K_B\boldsymbol{\mu}=0$ v.s. $A_{03}|H_{01}:$ not $H_{03}|H_{01}$.

ここで, $K_{AB}=P_a\otimes P_p, K_B=(\mathbf{1}_a\mathbf{1}_a')\otimes P_p$. ただし, $P_k=I_k-k^{-1}\mathbf{1}_k\mathbf{1}_k'$.

本研究では, この 3 つの仮説に対して, 検定統計量 T_ϕ をそれぞれ与えた. ただし, $\phi\in\{01,02,03\}$．適当な仮定の下で, T_ϕ の極限分布を導出した．具体的には, マルチンゲール差分中心極限定理を応用することで, 適当な仮定の下で T_{01} と T_{03} は漸近正規性を有する

ことがわかった．また, T_{02} は多変量中心極限定理を応用することで, 適当な仮定の下で自由度が $a-1$ の χ^2 分布に従うことがわかった．これらの結果に基づき, 近似検定方式を与え, それらの漸近的な検出力と第 1 種の過誤確率を調べた．

数値実験においては, 提案手法の有限次元・有限標本における第 1 種の過誤確率および検出力のふるまいを考察した．具体的には, 楕円分布のクラスに属する (D1) 多変量正規分布, (D2) 多変量 t-分布, (D3) 多変量ラプラス分布, (D4) 球内一様分布を母集団分布に想定し, Harra and Kong (2016) の近似検定と提案手法の比較を行った．ここでは, 特に, (i) と (iii) の検定に対して考察する．(D1) と (D4) において, 提案手法と先行研究の手法の第 1 種の過誤の確率は, ともに名目上の有意水準とさほど変わらなかった．一方, (D2) と (D3) においては, 先行研究の手法の第 1 種の過誤の確率は名目上の有意水準を大きく下回る傾向にあったのに対し, 提案手法の第 1 種の過誤の確率は名目上の有意水準とさほど変わらなかった．また, 先行研究の手法は, (D2) と (D3) において著しく検出力が低下するが, 提案手法の検出力はさほど低下しないことが確認された．

以上より, 提案手法は先行研究の手法に比べ, 母集団分布に関してロバストな手法であるということがわかった．

参考文献

[1] M. S. Srivastava, Prole analysis of several groups, *Commun.Stat. − TheoryMeth.*, **16**(3) (2006) pp.909-926

[2] N. Okamoto, N. Miura, T. Seo, On the distributions of some test statistics for profile analysis in elliptical populations, *Amer.J.Math.Manag.Sci.*, **26** (2006) pp.1-31

[3] Y. Maruyama, Asymptotic expansions of the null distributions of some test statistics for profile analysis in general conditions, *J.Statist.Plann.Inference*, **137** (2007) pp.506-526

[4] S. Takahashi, N. Shutoh, Tests for parallelism and flatness hypotheses of two mean vectors in high-dimensional settings, *J.Stat.Comput.Simul.*, **86** (2016) pp.1150-1165

[5] M. Onozawa, T. Nishiyama, T. Seo, On test statistics in profile analysis with high dimensional data, *Commun.Stat.SimuComput.*, (2014) pp.1-36

[6] S. W. Harrar, X. Kong, High-dimensional multivariate repeated measures analysis with unequal covariance matrices, *J.MultivariateAnal.*, **145** (2016) pp.1-21

[7] Y. Kano An asymptotic expansions of the distribution of Hotelling's T^2-statistic under general distributions. *Amer.J.Moth.andManage.Sci.*, **15** (1995) 317-341

Simultaneous testing of the mean vector and the covariance matrix for high-dimensional data

大阪府立大学　兵頭　昌
専修大学　　　西山　貴弘

$(\boldsymbol{X}_{g1}, \boldsymbol{X}_{g2}, \ldots, \boldsymbol{X}_{gn_g})$ を第 g 母集団 Π_g $(g = 1, 2)$ より得られる $p \times n_g$ の確率変数行列とする．各 $\boldsymbol{X}_{gi}$ に対して以下のモデルを仮定する：

$$\boldsymbol{X}_{gi} = \Sigma_g^{1/2}\boldsymbol{Z}_{gi} + \boldsymbol{\mu}_g \quad \text{for } i = 1, \ldots, n_g.$$

ただし, 各確率変数ベクトル $\boldsymbol{Z}_{gi}$ は $\mathrm{E}[\boldsymbol{Z}_{gi}] = \boldsymbol{0}$, $\mathrm{Var}[\boldsymbol{Z}_g] = I_p$ とし, $\boldsymbol{Z}_{11}, \ldots, \boldsymbol{Z}_{1n_1}, \boldsymbol{Z}_{21}, \ldots, \boldsymbol{Z}_{2n_2}$ は互いに独立であるとする．ここに, I_p は $p \times p$ の単位行列とする．

2 母集団の分布の有意差の検出のために, 平均ベクトルの同等性検定や分散共分散行列の同等性検定などが提案されている．本研究では, 平均ベクトルの同等性と分散共分散行列の同等性を同時に検定する方法を扱った．すなわち, 以下の仮説検定問題を扱った:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2,\ \Sigma_1 = \Sigma_2,\ H_1 : \text{not } H_0. \tag{1}$$

$p < \min\{n_1, n_2\}$ をデータ行列が満足し, かつ, 各母集団に正規性を仮定すればこの仮説検定問題に対する尤度比基準を構成することが可能である．具体的には,

$$\lambda = \frac{\prod_{g=1}^2 |W_g|^{n_g/2}}{|(n_1n_2)/n(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)' + \sum_{g=1}^2 W_g|^{n/2}} \frac{n^{pn/2}}{\prod_{g=1}^2 n_g^{pn_g/2}}$$

で与えられる．ただし, $n = n_1 + n_2$ であり,

$$\overline{\boldsymbol{X}}_g = \frac{1}{n_g}\sum_{i=1}^{n_g} \boldsymbol{X}_{gi},\ W_g = \sum_{i=1}^{n_g}(\boldsymbol{X}_{gi} - \overline{\boldsymbol{X}}_g)(\boldsymbol{X}_{gi} - \overline{\boldsymbol{X}}_g)'.$$

さらに, Muirhead (1982) において修正尤度比検定を与えている．

一方で, $p > \min\{n_1, n_2\}$ であるとき標本積和行列 W_1 と W_2 の少なくとも一方は特異行列となるため尤度比基準を構成することができない．また, $p < \min\{n_1, n_2\}$ であっても p,n_1,n_2 が同程度に大きい場合, Muirhead (1982) で与えられる修正尤度比検定の近似精度が悪化することが懸念される．本研究では, ノルムに基づく検定統計量を考えることで, $p < \min\{n_1, n_2\}$ や正規性の仮定を課さずとも使用可能であり, 次元が大きい場合でも近似精度が悪化しない (1) のための近似検定法を与えることができた．(1) は以下のように書き換えることができる：

$$H_0 : \|\boldsymbol{\delta}\|^2 = 0,\ \|\Delta\|_F^2 = 0,\ H_1 : \text{not } H_0,$$

ただし, $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ であり, $\Delta = \Sigma_1 - \Sigma_2$ である．ここに, $\|\cdot\|$ はユークリッドノルムであり, $\|\cdot\|_F$ はフロベニウスノルムである．$\|\boldsymbol{\delta}\|^2$ と $\|\Delta\|_F^2$ の不偏推定量はそれぞれ次のように与えられる:

$$\begin{aligned}
\widehat{\|\boldsymbol{\delta}\|^2} &= (\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)'(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2) - \frac{\mathrm{tr}S_1}{n_1} - \frac{\mathrm{tr}S_2}{n_2}, \\
\widehat{\|\Delta\|_F^2} &= \sum_{g=1}^2 \widehat{\mathrm{tr}\Sigma_g^2} - 2\,\mathrm{tr}(S_1S_2).
\end{aligned}$$

ただし,

$$\begin{aligned}
\widehat{\mathrm{tr}\Sigma_g^2} &= \frac{n_g-1}{n_g(n_g-2)(n_g-3)}\{(n_g-1)(n_g-2)\mathrm{tr}S_g^2+(\mathrm{tr}S_g)^2-n_gK_g\},\\
S_g &= \frac{1}{n_g-1}W_g,\ K_g=\frac{1}{n_g-1}\sum_{i=1}^{n_g}\{(\boldsymbol{X}_{gi}-\overline{\boldsymbol{X}}_g)'(\boldsymbol{X}_{gi}-\overline{\boldsymbol{X}}_g)\}^2.
\end{aligned}$$

不偏推定量 $\widehat{\mathrm{tr}\Sigma_g^2}$ は, Himeno and Yamada (2014) や Srivastava, Yanagihara and Kubokawa (2014) などで提案されている. これらの不偏推定量の特徴は以下のように纏められる:

(i) 正規性を仮定することなく ($\boldsymbol{Z}_{g1}$ のモーメントに関する仮定は必要), 不偏性を持つ.

(ii) $p > \min\{n_1, n_2\}$ であっても定義できる.

(iii) 適当な仮定のもとで, p, n_1, n_2 を無限大としたとき漸近正規性を持つ.

高次元の設定において, Chen and Qin (2010) は不偏推定量 $\widehat{\|\boldsymbol{\delta}\|^2}$ の漸近正規性を示し, その結果を用いた平均ベクトルの同等性検定 ($H_0 : \boldsymbol{\delta} = \boldsymbol{0}$) を与えた. さらに, Li and Chen (2012) は不偏推定量 $\widehat{\|\Delta\|_F^2}$ の漸近正規性を示し共分散行列の同等性検定 ($H_0 : \Delta = \boldsymbol{O}$) を与えた. 本研究では, $\widehat{\|\boldsymbol{\delta}\|^2}/\sigma_1$ と $\widehat{\|\Delta\|_F^2}/\sigma_2$ の同時分布の漸近的な性質を調べることで, (1) のための検定の提案を試みた. ただし, σ_i は各推定量の分散の主要な項であり, 詳細は紙面の都合上省略する.

本報告では, 適当な仮定のもとで $\widehat{\|\boldsymbol{\delta}\|^2}/\sigma_1$ と $\widehat{\|\Delta\|_F^2}/\sigma_2$ の同時分布の漸近分布が 2 次元正規分布となることを報告した. この漸近的な結果に基づき, 近似検定を与えその漸近的な第 1 種の過誤の確率および検出力を調べた. さらに, 有限次元, 有限標本における提案手法の精度をモンテカルロシミュレーションによって評価した.

参考文献

[1] S. X. Chen, Y. L. Qin, A two-sample test for highdimensional data with applications to gene-set testing. *Ann. Statist.*, **38** (2010), 808–835.

[2] T. Himeno, T. Yamada, Estimations for some functions of covariance matrix in high dimension under non-normality and its applications. *J. Multivariate Anal.*, **130** (2014) 27–44.

[3] J. Li, S. X. Chen, Two sample tests for high-dimensional covariance matrices. *Ann. Statist.*, **40** (2012) 908–940.

[4] R. J. Muirhead, *Aspects of Multivariate Statistical Theory.* John Wiley & Sons, Inc., New York (1982).

[5] M. S. Srivastava, H. Yanagihara, T. Kubokawa, Tests for covariance matrices in high dimension with less sample size. *J. Multivariate Anal.*, **130** (2014) 289–309.