ISSN　0385-2148

日本統計研究所報

No.63

Frontiers in Directional Statistics Theory, Methods, and Applications

２０２６年 6月

法政大学
日 本 統 計 研 究 所

ISSN　0385-2148

日本統計研究所報

No.63

Frontiers in Directional Statistics Theory, Methods, and Applications

２０２６年 6月

法政大学
日　本　統　計　研　究　所

目次

Order Selection for Extended Sine-Skewed Circular Distributions

高崎経済大学・経済学部 [*]　宮田 庸一

南山大学・理工学部 [†]　塩濱 敬之

法政大学・経済学部 [‡]　阿部 俊弘

Frontiers in Directional Statistics Theory, Methods, and Applications
法政大学 多摩キャンパス,
2025 年 11 月 29, 30 日

1　はじめに

Abe and Pewsey (2011) により提案された正弦関数摂動分布族は, 円周上の非対称な確率分布族である．この正弦関数摂動分布族は，von Mises 分布や巻き込みコーシー分布のようなよく知られた円周上の対称な密度関数に歪み関数 (Skewing function) を乗ずることで，非対称な確率分布を導いている．この分布族は，三角モーメントが陽に表すことができ, 乱数の生成が容易であり, なおかつ分布族の単峰性の確認が容易であるなど，いくつかの理論的に望ましい性質を持つことが知られている．三角モーメントは, パラメーターに関する識別可能性を明らかにすることやモーメント推定量を導出できることができるため，役に立つ量となる．一方で, 正弦関数摂動分布族はモードの周辺で強い非対性を持つ分布を表すのにやや制限がある．Miyata et al. (2024) では, より大きな歪みを与えることができる歪み関数を提案し, それに基づいて, 拡張された正弦関数摂動分布族を提案した．この歪み関数の形状を決めるためには, オーダーと呼ばれるハイパーパラメーターを定める必要がある．本報告では, 拡張された正弦関数摂動分布族の紹介, および, このオーダーを定めるためのいくつかの手法を考え, そのパフォーマンスをシミュレーションを通じて検証した．

[*] 〒370-0801 群馬県高崎市上並榎町 1300

[†] 〒466-8673 愛知県名古屋市昭和区山里町 18

[‡] 〒194-0298 東京都町田市相原町 4342

2 提案する確率モデル

まず初めに, ここで提案する歪み関数を定義する. ここでは, 閉区間 $[-1,1]$ 上の確率密度関数

$$g_m(x) = \frac{\Gamma(2(m+1))}{2^{2m+1}\Gamma(m+1)^2}(1-x^2)^m \qquad (-1 \le x \le 1),$$

を考え, その分布関数を $G_m(x) = \displaystyle\int_{-1}^{x} g_m(t)dt$ により定義する. ただし $m > -1$ はハイパーパラメーターとし, ここではオーダーと呼ぶことにする. 関数 G_m は, ベータ分布の分布関数を $G_{\mathrm{Be}}(x,\alpha,\beta) = \int_0^x \{\Gamma(\alpha+\beta)/\Gamma(\alpha)\Gamma(\beta)\} t^{\alpha-1}(1-t)^{\beta-1}dt$ $(\alpha > 0, \beta > 0)$ とおくと, $G_m(x) = G_{\mathrm{Be}}(x, m+1, m+1)$ $(m > -1)$ と表すことができるため, R のようなプログラムにおいては, 容易に計算することができる. このとき, オーダー m の拡張された正弦関数摂動分布 (Exnended sine-skewed circular distribution, ESS) の密度関数を以下の形で定義する:

$$f_{\mathrm{ESS}}^{(m)}(\theta;\mu,\boldsymbol{\rho},\lambda) = 2f_0(\theta-\mu;\boldsymbol{\rho})G_m\left(\lambda\sin(\theta-\mu)\right), \tag{1}$$

ただし $f_0(\theta;\boldsymbol{\rho})$ は対称なベース密度とし, $\mu \in [-\pi,\pi)$ を位置パラメーター, $\lambda \quad \in [-1,1]$ を非対称パラメーターとする. パラメーター $\boldsymbol{\rho}$ は, 分布の集中度など, 位置以外の形状を制御するパラメーターとする. 式 (1) は, $\lambda = 0$ のとき, $G_m(0) = 1/2$ となるため, μ に関して対称な分布となる.

密度関数 (1) は任意の $m > -1$ に対して定義できるが, 非負の整数 (すなわち, $m = 0,1,2,\cdots$) に限定した場合は

$$\begin{aligned} G_m(x) &= \int_{-1}^{x} g_m(t)dt \\ &= C_m \sum_{\ell=0}^{m} \binom{m}{\ell} \frac{1}{2\ell+1}(-1)^{\ell}\left(x^{2\ell+1}+1\right) \end{aligned}$$

のように多項式を用いて表現ができる.

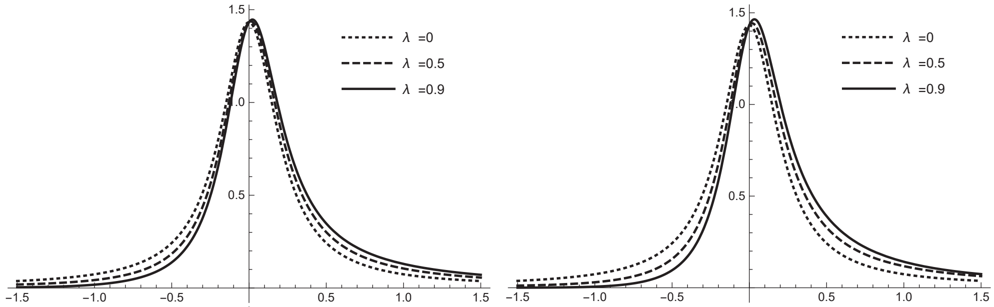
図 1 では, 強めの集中度 $\rho = 0.8$ を持つ巻き込みコーシー分布をベース密度に持つ場合の, 関数 (1) のグラフを示す. この図からもわかるように, 集中度が高いベース密度を歪ませるためには, 大きなオーダー m を選ぶ必要がある.

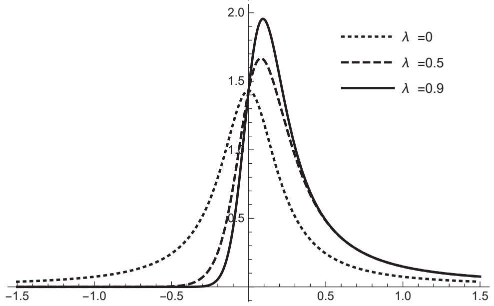
3 オーダー m の選択方法

確率密度関数 (1) において, $\lambda = 0$ の場合, 任意の m に対して, $G_m(0) = 1/2$ となるため, 拡張された正弦関数摂動分布は, オーダー m に関して識別不可能であることがわかる. しかし, 実用上は m を何らかの方法で選ぶ必要があるため, 本報告では, 以下の 3 つの方法を試みた.

1. 最大対数尤度アプローチ: $\ell_m(\boldsymbol{\gamma}) = (1/n)\sum_{i=1}^{n}\log f_{\mathrm{ESS}}^{(m)}(\theta_i;\mu,\boldsymbol{\rho},\lambda)$ を確率密度関数 (1) に基づく対数尤度関数とするとき, オーダー m を $\hat{m} = \mathrm{argmax}_{m\in\mathcal{M}}\ell_m(\hat{\boldsymbol{\gamma}}_n)$ により推定す

図 1　ベース密度を巻き込みコーシー分布としたときの正弦関数摂動分布族





ベース密度: $\mu = 0$, $\rho = 0.8$ としたときの, 巻き込みコーシー分布左上:オーダー $m = 0$
右上:オーダー $m = 1$
左下:オーダー $m = 50$
これらの, 密度関数の定義域は $[-\pi, \pi)$ であるが, 歪み具合を明らかにするために, $[-1.5, 1.5]$ に領域を制限する形で与えている.

る方法である. ただし, $\mathcal{M} = \{f_{\mathrm{ESS}}^{(m)}(\theta; \mu, \boldsymbol{\rho}, \lambda) | m = m_1, \ldots, m_M\}$ は候補となる確率密度関数 (1) の集合とする.

2. 何らかの情報量規準を最適化する m を選ぶ. ここでは, 竹内情報量規準 (Takeuchi, 1976), およびベイジアンアプローチである対数周辺尤度関数を考える.
3. マッチング法: 既存の円周上の非対称分布 (例えば, Umbach and Jammalamadaka, 2009) に最も近くなるような m を選ぶ.

n 個の観測値 $\theta_1, \ldots, \theta_n$ が与えられたとき, 対数周辺尤度は以下の形で定義される:

$$\log q(\boldsymbol{\theta}) = \log \int \prod_{i=1}^{n} f_{\mathrm{ESS}}^{(m)}(\theta_i; \mu, \boldsymbol{\rho}, \lambda) p(\mu, \boldsymbol{\rho}, \lambda) d\mu d\boldsymbol{\rho} d\lambda,$$

ただし $p(\mu, \boldsymbol{\rho}, \lambda)$ はパラメーター $(\mu, \boldsymbol{\rho}^\top, \lambda)^\top$ に対する事前分布とする. この積分は一般的には陽に解くことができないため, モンテカルロ法もしくは漸近展開により近似をする. 次に, 竹内情報量規準 (Takeuchi information criterion, TIC) を与える. まず始めに, 以下の 2 つの行列を定義する.

$$\boldsymbol{J}_n(\boldsymbol{\gamma}) := -\frac{\partial^2}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}^\top} \ell_m(\boldsymbol{\gamma}),$$
$$\boldsymbol{I}_n(\boldsymbol{\gamma}) := \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\gamma}} \log f_{\mathrm{ESS}}^{(m)}(\theta_i; \boldsymbol{\gamma}) \frac{\partial}{\partial \boldsymbol{\gamma}^\top} \log f_{\mathrm{ESS}}^{(m)}(\theta_i; \boldsymbol{\gamma})$$

このとき, TIC は

$$\mathrm{TIC} = -2n\ell_m(\hat{\boldsymbol{\gamma}}_n) + 2\mathrm{tr}\{\boldsymbol{J}_n(\hat{\boldsymbol{\gamma}}_n)^{-1} \boldsymbol{I}_n(\hat{\boldsymbol{\gamma}}_n)\},$$

と定義される．次にマッチング法について解説する．f_0, g を巻き込みコーシー分布の密度関数とする，すなわち $f_0(\theta) = \dfrac{1-\rho^2}{2\pi(1+\rho^2-2\rho\cos\theta)}$, $f_0(\theta) = g(\theta)$ とする．ここで, $w(\theta) = \lambda\sin(\theta)$, $G(\theta)$ を $g(\theta)$ の分布関数とする．このとき, Umbach and Jammalamadaka (2009) により, 以下のような確率密度関数をえることができる:

$$f_{UJ}(\theta) = 2f_0(\theta)G(w(\theta)) \tag{2}$$

この確率密度関数は, $\lambda = 0$ のときには対称な分布となり, $\lambda \neq 0$ のときには歪んだ分布となる．密度関数 f_0 および g に対して, 同じ確率分布を与える方法は, Azzalini (1985) による方法と同じ考え方になる．このとき, この分布の歪みの大きさを測る簡便的な尺度として, 以下のようなものが考えられる:

$$\begin{aligned} f_{UJ}(\theta) - f_{UJ}(-\theta) &= 2f_0(\theta)\{G(w(\theta)) - G(w(-\theta))\} \\ &\approx 4w(\theta)f_0(\theta)G'(0) \\ &= 4\lambda\pi\sin\theta f_0(\theta)\frac{1}{2\pi}\frac{1-\rho}{1+\rho}. \end{aligned} \tag{3}$$

その一方で, 確率密度関数 (1) において, ベース密度を巻き込みコーシー分布の密度関数としたときの, 歪みの大きさを測る簡便的な尺度は, 以下の式で与えられる．

$$\begin{aligned} f^{(m)}_{\mathrm{ESSWC}}(\theta;\boldsymbol{\gamma}) - f^{(m)}_{\mathrm{ESSWC}}(-\theta;\boldsymbol{\gamma}) &\approx 4w(\theta)f_0(\theta)G'_m(0) \\ &= 4\lambda\sin\theta f_0(\theta)\frac{\Gamma(2(m+1))}{2^{2m+1}\Gamma(m+1)^2}, \end{aligned} \tag{4}$$

ただし $G'_m(x) = (d/dx)G_m(x)$ とする．ここで, 式 (3) と式 (4) が何らかの意味で, 最も近い値になるように m を選ぶ方法となる．最も素朴な方法としては, 式 (3) から式 (4) を割った量が, 最も 1 に近くなるように m を求めることである．

4　結果の報告

上記の 1. 最大対数尤度アプローチおよび, 2. の竹内情報量規準および対数周辺尤度アプローチにおいては, オーダー m を正しく選ぶことができるかをシミュレーションを通じて検証を行った．残念ながら, 今回行ったシミュレーションの下では, サンプルサイズを大きくしても, 正しいオーダー m を特定できなかった．この結果に対して, 参加者の方から, いくつかのコメントをもらった．

参考文献

Abe, T. and A. Pewsey (2011). Sine-skewed circular distributions. *Statist. Papers 52*(3), 683–707.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, 171–178.

Miyata, Y., T. Shiohama, and T. Abe (2024). An extension of sine-skewed circular distributions. *arXiv preprint arXiv:2402.09788*.

Takeuchi, K. (1976). The distribution of information statistics and the criterion of goodness of fit of models (in japanese). *Mathematical Science 153*, 12–18.

Umbach, D. and S. R. Jammalamadaka (2009). Building asymmetry into circular distributions. *Statist. Probab. Lett. 79*(5), 659–663.

Topics in Complex-Valued Time Series Modeling: Estimation and Applications

Takayuki Shiohama[a]

[a]*Department of Data Science, Nanzan University, 18 Yamazato-cho, Showa, Nagoya, 466-8673, Aichi, Japan*

Abstract

Complex-valued time series modeling is useful for fitting spiral sequences that often occur in the natural and environmental sciences. To incorporate time-dependent structures in the observed real and complex parts of the complex-valued process, we apply a vector autoregressive model in the wide-sense stationary framework. An extension of the complex-valued Yule-Walker estimators is introduced, and its asymptotic properties are investigated. Testing the null hypothesis of the underlying observed process is considered proper.

Keywords: Complex-valued time series, Yule-Walker estimators, wide-sense stationary, hypothesis testing

1. Introduction

Statistical analysis of complex-valued time series is a concept within spectral analysis and has a long tradition in time series analysis, as in (Hannan, 1970; Brockwell and Davis, 1998; Brillinger, 1981). In particular, complex-valued representations naturally arise in applications involving oscillatory and wave-like phenomena. Frequency-domain analysis of a time series provides an alternative to time-domain analysis, offering distinct advantages for identifying unknown signal frequencies associated with complex numbers. Such an approach enables a clearer characterization of periodic structures and phase relationships in the data. Since the 1960s, theoretical developments in signal processing have advanced rapidly, leading to the emergence of machine-learning methods. These developments have further expanded the scope of frequency-domain techniques in modern data analysis.

The defining characteristic of a complex-valued process lies in its rotational, or oscillatory, behavior in the complex plane. Such behavior makes complex-valued models particularly suitable for representing signals with phase and amplitude components. A complex-valued process is typically represented in terms of its real and imaginary parts, which are real-valued random variables. If these two components have equal variances and are uncorrelated, the process is said to be proper (or circular). In contrast, the process is termed improper when the two components are correlated or have unequal variances. This distinction plays a central role in statistical modeling and inference for complex-valued data.

For autoregressive processes with complex-valued coefficients, Le Breton (1988) studied maximum likelihood estimation in the proper case and discussed its extension to the improper case; see also Sykulski et al. (2016) for further developments. For a comprehensive treatment of proper complex-valued processes, we refer the reader to the book by Miller (1974).

The merits of modeling variables in polar coordinates, rather than employing a bivariate normal distribution on the Cartesian plane, lie in the ability to explicitly capture angular and radial dependence, particularly for center-biased, circular, or rotational data. Such a representation is especially advantageous when the underlying phenomenon exhibits periodicity or directional symmetry. By separating distance and direction, this framework facilitates more flexible modeling of nonlinear structures and asymmetric patterns. The analysis of complex-valued processes has therefore attracted increasing attention across various fields of data science, where phase-amplitude interactions play a fundamental role.

In the context of directional statistics, circular or cylindrical time series models defined on specific geometric manifolds–such as the unit circle or the cylinder–are closely related to complex-valued processes expressed in polar coordinates. These models provide a natural probabilistic framework for data constrained to non-Euclidean spaces. Various approaches to time series modeling have been proposed in the field of directional statistics; see, for example, Taniguchi et al. (2020), Shiohama (2023), Ogata and Shiohama (2024), Abe et al. (2025), and Gatto (2022).

In this short note, we will consider the wide-sense stationary process defined in the subsequent section, and the asymptotic properties of standard Yule-Walker estimators are investigated.

2. Models and Assumptions

A complex-valued time series is a sequence of complex-valued random variables Z_t. For a stationary complex-valued process $\{Z_t\}_{t\in\mathbb{Z}}$, the autocovariance function at lag $h \in \mathbb{Z}$ is defined by $\gamma_Z(h) = \mathrm{cov}(Z_t, Z_{t-h}) = E[Z_t Z_{t-h}^*]$ for a zero mean sequence of $\{Z_t\}_{t\in\mathbb{Z}}$. Recall that the autocovariance function is a conjugate symmetric such that $\gamma_Z(h) = \gamma_Z^*(-h)$, where $\gamma_Z^*(-h)$ denotes the complex conjugate of $\gamma_Z(-h)$. The autocorrelation function is defined by $\rho_Z(h) = \gamma_Z(h)/\gamma_Z(0)$. The sample autocovariance function at lag $h \geq 0$ is defined by

$$\hat{\gamma}_Z(h) = \frac{1}{n-h}\sum_{t=h+1}^{n} Z_t Z_{t-h}^*.$$

For $h < 0$, it is defined as $\hat{\gamma}_Z(h) = \hat{\gamma}_Z^*(-h)$. The sample autocorrelation function at lag h is defined by $\hat{\rho}_Z(h) = \hat{\gamma}_Z(h)/\hat{\gamma}_Z(0)$.

For the zero mean random variables X and Y with correlation coefficient ρ_{XY}, and let the complex-valued random variable Z be $X = X + \mathrm{i}Y$. Then, the variance becomes

$$\mathrm{Var}(Z) = E|Z|^2 = E[(X + \mathrm{i}Y)(X - \mathrm{i}Y)] = \sigma_X^2 + \sigma_Y^2,$$

where σ_X^2 and σ_Y^2 are the variance of X and Y, respectively. Since this $\mathrm{Var}(Z)$ does not carry information about the correlation coefficient between the pairs (X, Y), we need another complex second-order moment as

$$\widetilde{\mathrm{Var}}(Z) = E(Z^2) = E[(X + \mathrm{i}Y)^2] = \sigma_X^2 - \sigma_Y^2 + 2\mathrm{i}\rho_{XY}\sigma_X\sigma_Y,$$

which we call $\widetilde{\mathrm{Var}}(Z)$ as the complementary variance or pseudo-variance, see (Neeser and Massey, 1993; Schreier and Scharf, 2010). If $\sigma_X^2 = \sigma_Y^2$ and $\rho_{XY} = 0$, the complementary variance becomes zero. This is the so-called proper or circular case; all others are called improper cases. For the improper case, the complementary autocovariance function at lag h is defined by $\tilde{\gamma}_Z(h) = E(Z_t Z_{t-h})$. In order to investigate the complementary autocovariance structure of the process, we need an augmented process as $\underline{\boldsymbol{Z}}_t = (Z_t, Z_t^*)^T$, which is defined by

$$\begin{aligned}
Z_t =& X_t + \mathrm{i}Y_t, \quad X_t, Y_t \in \mathbb{R}, \quad t \in \mathbb{Z}, \quad E[Z_t] = 0,\\
\underline{\boldsymbol{Z}}_t =& \begin{bmatrix} Z_t \\ Z_t^* \end{bmatrix} = \begin{bmatrix} X_t + \mathrm{i}Y_t, \\ X_t - \mathrm{i}Y_t, \end{bmatrix}
\end{aligned}$$

Then the augmented covariance matrix becomes

$$\underline{\boldsymbol{R}}_Z(h) = E[\underline{\boldsymbol{Z}}_t \underline{\boldsymbol{Z}}_{t-h}^H] = \begin{bmatrix} \gamma_Z(h) & \tilde{\gamma}_Z(h) \\ \tilde{\gamma}_Z^*(h) & \gamma_Z^*(h) \end{bmatrix},$$

where $\boldsymbol{x}^H$ denotes complex conjugate of a vector or a matrix $\boldsymbol{x}$. Because the augmented covariance matrix depends only on lag h, the bivariate process is called wide-sense stationary (WSS) or widely stationary.

Hereafter, we consider the complex-valued AR process as follows. The following bivariate process $\{\underline{\boldsymbol{Z}}_t\}_{t\in\mathbb{Z}}$ is called complex-valued process of order p (CV-VAR(p)):

$$\underline{\boldsymbol{Z}}_t = \sum_{j=1}^{p} \underline{\boldsymbol{A}}_i \underline{\boldsymbol{Z}}_{t-j} + \boldsymbol{\varepsilon}_t, \quad \text{or} \quad \begin{bmatrix} Z_t \\ Z_t^* \end{bmatrix} = \sum_{j=1}^{p} \begin{bmatrix} \phi_j & \psi_j \\ \psi_j^* & \phi_j^* \end{bmatrix} \begin{bmatrix} Z_{t-j} \\ Z_{t-j}^* \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ \varepsilon_t^* \end{bmatrix},$$

where $\{\varepsilon_t = u_t + \mathrm{i}v_t\}$ is a complex-valued white noize with $E[\varepsilon_t] = E[u_t] + \mathrm{i}E[v_t] = 0$, $E[\varepsilon_t\varepsilon_t^*] = \sigma_u^2 + \sigma_v^2$, and $E[\varepsilon_t\varepsilon_t] = \sigma_u^2 - \sigma_v^2 + 2\mathrm{i}\rho_{uv}\sigma_u\sigma_v$. Hence, the error term ε_t is, in general, an improper complex-valued white noise process. Recall that $p = 1$, $\psi_1 = 0$, and ε_t is the proper process, the CV-VAR(1) reduces wrapped Cauchy process considered by Shiohama (2023).

As with the case for the real-valued VAR(p) processes, CV-VAR(p) process has the following Yule-Walker equations

$$\begin{bmatrix} \gamma_Z(h) & \tilde{\gamma}_Z(h) \\ \tilde{\gamma}_Z^*(h) & \gamma_Z^*(h) \end{bmatrix} = \sum_{j=1}^{p} \begin{bmatrix} \phi_j & \psi_j \\ \psi_j^* & \phi_j^* \end{bmatrix} \begin{bmatrix} \gamma_Z(h-j) & \tilde{\gamma}_Z(h-j) \\ \tilde{\gamma}_Z^*(h-j) & \gamma_Z^*(h-j) \end{bmatrix}$$

The corresponding matrix form expression becomes:

$$\underline{\boldsymbol{R}}_Z(h) = \sum_{j=1}^{p} \underline{\boldsymbol{A}}_j \underline{\boldsymbol{R}}_Z(h-j)$$

The variance of the process $\underline{\boldsymbol{Z}}_t$ is given by

$$\underline{\boldsymbol{R}}_Z(0) = \sum_{j=1}^{p} \underline{\boldsymbol{A}}_j \underline{\boldsymbol{R}}_Z(-j) + \underline{\boldsymbol{R}}_\varepsilon,$$

where $\underline{\boldsymbol{R}}_\varepsilon$ is

$$\underline{\boldsymbol{R}}_\varepsilon = E \begin{bmatrix} \varepsilon_t \\ \varepsilon_t^* \end{bmatrix} \begin{bmatrix} \varepsilon_t^* & \varepsilon_t \end{bmatrix} = \begin{bmatrix} R_\varepsilon & \tilde{R}_\varepsilon \\ \tilde{R}_\varepsilon & R_\varepsilon \end{bmatrix} = \boldsymbol{T}_u \boldsymbol{R}_{uv} \boldsymbol{T}_u^H$$

with

$$\boldsymbol{R}_{uv} = E\begin{bmatrix} u_t \\ v_t \end{bmatrix}\begin{bmatrix} u_t & v_t \end{bmatrix} = \begin{bmatrix} \sigma_u^2 & \rho_{uv}\sigma_u\sigma_v \\ \rho_{uv}\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{T}_u = \begin{bmatrix} 1 & \mathrm{i} \\ 1 & -\mathrm{i} \end{bmatrix}.$$

Using this expression, we can recursively obtain the theoretical augmented auto-covariance and auto-correlation functions for the widely stationary process $\{\underline{\boldsymbol{Z}}_t\}_{t\in\mathbb{Z}}$.

To obtain the asymptotic distribution of the estimators, we introduce the following complex Gaussian distribution. Recall that, the complex augmented vector $\underline{\boldsymbol{\varepsilon}_t}$ is related to the real composite vector $\boldsymbol{u}_t = (u_t, v_t)^T$ as

$$\underline{\boldsymbol{\varepsilon}}_t = \boldsymbol{T}_u \boldsymbol{u}_t \Longleftrightarrow \boldsymbol{u}_t = \frac{1}{2}\boldsymbol{T}_u^H \underline{\boldsymbol{\varepsilon}}_t,$$

and the probability density function of a complex biivariate Gaussian distribution is defined as

$$f(\boldsymbol{u}_t) = \frac{1}{(2\pi)\det \boldsymbol{R}_{uv}} \exp\left\{-\frac{1}{2}\boldsymbol{u}_t^T \boldsymbol{R}_{uv}^{-1}\boldsymbol{u}_t\right\},$$

where

$$\boldsymbol{R}_{uv}^{-1} = \boldsymbol{T}^H \underline{\boldsymbol{R}}_{\varepsilon}^{-1}\boldsymbol{T}, \quad \text{and} \quad \det \boldsymbol{R}_{zz} = 4\det \boldsymbol{R}_{\varepsilon}$$

3. Main Results and Some Related Tesing Problems

3.1. Yule-Walker Estimators

To define the well-known Yule-Walker estimators for the CV-VAR(p) process, we introduce the following notations.

For the sequence of VAR coefficient matrices $\{\underline{\boldsymbol{A}}_i\}_{i=1,2,\ldots,k}$, let $\underline{\boldsymbol{A}}_{1:k} = \begin{bmatrix} \underline{\boldsymbol{A}}_1 & \underline{\boldsymbol{A}}_2 & \cdots & \underline{\boldsymbol{A}}_k \end{bmatrix}$ and

$$\boldsymbol{R}_{0:k-1}^{(Toe)} = \begin{bmatrix} \underline{\boldsymbol{R}}_Z(0) & \underline{\boldsymbol{R}}_Z(1) & \cdots & \underline{\boldsymbol{R}}_Z(k-1) \\ \underline{\boldsymbol{R}}_Z^H(1) & \underline{\boldsymbol{R}}_Z(0) & \cdots & \underline{\boldsymbol{R}}_Z(k-1) \\ \vdots & \vdots & \ddots & \vdots \\ \underline{\boldsymbol{R}}_Z^H(k-1) & \underline{\boldsymbol{R}}_Z^H(k-2) & \cdots & \underline{\boldsymbol{R}}_Z(0) \end{bmatrix}.$$

For the sequence of $\{\underline{\boldsymbol{R}}_Z(j)\}_{j=1,2,\ldots,k}$, let

$$\underline{\boldsymbol{R}}_{1:k} = \begin{bmatrix} \underline{\boldsymbol{R}}_Z(1) & \underline{\boldsymbol{R}}_Z(2) & \cdots & \underline{\boldsymbol{R}}_Z(k) \end{bmatrix}$$

Here we assume the following stable conditions on the CV-VAR(p) process,

Assumption 1. The VAR representation of a CV-VAR(p) process is stable, such that the roots of the characteristic polynomial

$$\det(\boldsymbol{I}_2 - \underline{\boldsymbol{A}}_1 z - \ldots - \underline{\boldsymbol{A}}_p z^p) = 0$$

has no roots in and on the complex unit circle.

The matrix version of the Yule-Walker equation becomes

$$\underline{\boldsymbol{A}}_{1:p} \boldsymbol{R}_{0:p-1}^{(Toe)} = \underline{\boldsymbol{R}}_{1:p}.$$

Then, we obtain

$$\underline{\boldsymbol{A}}_{1:p} = \underline{\boldsymbol{R}}_{1:p} \left(\boldsymbol{R}_{0:p-1}^{(Toe)} \right)^{-1}$$

and the Yule-Walker estimator is defined by

$$\widehat{\underline{\boldsymbol{A}}}_{1:p}^{(YW)} = \widehat{\underline{\boldsymbol{R}}}_{1:p} \left(\widehat{\boldsymbol{R}}_{0:p-1}^{(Toe)} \right)^{-1}$$

where $\widehat{\underline{\boldsymbol{R}}}_{1:p}$ and $\widehat{\boldsymbol{R}}_{0:p-1}^{(Toe)}$ are block matrices whose elements $\underline{\boldsymbol{R}}(k)$'s $(k = 1, 2, \ldots, p)$ are replaced by

$$\widehat{\underline{\boldsymbol{R}}}_Z(k) = \frac{1}{n-k} \sum_{t=k+1}^{n} \underline{\boldsymbol{Z}}_t \underline{\boldsymbol{Z}}_{t-k}^H.$$

To derive the asymptotic distribution of the YW estimators, the following notations are introduced.

Recall that the matrix of the parameters is of the form:

$$\underline{\boldsymbol{A}}_{1:p} = \begin{pmatrix} \underline{\boldsymbol{A}}_1 & \cdots & \underline{\boldsymbol{A}}_p \end{pmatrix} = \begin{bmatrix} \phi_1 & \psi_1 & \cdots & \phi_p & \psi_p \\ \psi_1^* & \phi_1^* & \cdots & \psi_p^* & \phi_p^* \end{bmatrix}.$$

As this matrix consists of the entries of the complex-valued parameters, we need to arrange this as the real valued matrix:

$$\boldsymbol{A}_{1:p} = (\boldsymbol{A}_1 \quad \cdots \quad \boldsymbol{A}_p) = \begin{bmatrix} \mathrm{Re}(\phi_1) & \mathrm{Re}(\psi_1) & \cdots & \mathrm{Re}(\phi_p) & \mathrm{Re}(\psi_p) \\ \mathrm{Im}(\phi_1) & \mathrm{Im}(\psi_1) & \cdots & \mathrm{Im}(\phi_p) & \mathrm{Im}(\psi_p) \end{bmatrix}$$

This arrangement can be done by introducing the following matrix $\boldsymbol{T}_p$ as defined below. Let $\boldsymbol{T}_p$ be a block diagonal matrix defined by $\boldsymbol{T}_p := \mathrm{diag}(\boldsymbol{T}, \ldots, \boldsymbol{T})$. The matrix $\boldsymbol{T}_p$ is unitary

$$\boldsymbol{T}_p \boldsymbol{T}_p^H = \boldsymbol{T}_p^H \boldsymbol{T}_p = 2I_{4p}$$

and

$$\boldsymbol{T}_p = \underbrace{\begin{bmatrix} \boldsymbol{T} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{T} \end{bmatrix}}_{(4p)\times(4p)} \quad \text{with} \quad \boldsymbol{T} = \begin{bmatrix} 1 & 0 & 0 & \mathrm{i} \\ 0 & 1 & \mathrm{i} & 0 \\ 0 & 1 & -\mathrm{i} & 0 \\ 1 & 0 & 0 & -\mathrm{i} \end{bmatrix}$$

Then, we obtain the asymptotic distributions of the complex-valued YW estimators.

Theorem 1. *(Asymptotics of the CV-YW Estimators). Let the process $\{\underline{\boldsymbol{Z}}_t\}$ be defined by the CV-VAR(p) processes that satisfy Assumptions A. Then, we have as $n \to \infty$*

$$\widehat{\boldsymbol{A}}_{1:p}^{(YW)} \underset{p}{\to} \boldsymbol{A}_{1:p}$$

and

$$\begin{aligned} &\sqrt{n}(vec(\widehat{\boldsymbol{A}}_{1:p}^{(YW)}) - vec(\boldsymbol{A}_{1:p})) \\ =&\sqrt{n}(\widehat{\boldsymbol{\mathcal{A}}}_p^{(YW)} - \boldsymbol{\mathcal{A}}_p) \to_d N(\boldsymbol{0}, \boldsymbol{T}_p^H \left(\boldsymbol{R}_{0:p-1}^{(Toe)-1} \otimes \underline{\boldsymbol{R}}_\varepsilon\right) \boldsymbol{T}_p) \end{aligned}$$

where

$$\underline{\boldsymbol{R}}_\varepsilon = E[\underline{\boldsymbol{\varepsilon}}_t \underline{\boldsymbol{\varepsilon}}_t^H] = E\left[\left(\underline{\boldsymbol{Z}}_t - \sum_{j=1}^{p} \boldsymbol{A}_j \underline{\boldsymbol{Z}}_{t-j}\right)\left(\underline{\boldsymbol{Z}}_t - \sum_{j=1}^{p} \boldsymbol{A}_j \underline{\boldsymbol{Z}}_{t-j}\right)^H\right]$$

.

Remark 1. If we consider the CV-VAR(1) model, the CV-YW estimators becomes

$$\begin{aligned} &\widehat{\underline{\boldsymbol{A}}}_1 = \widehat{\underline{\boldsymbol{R}}_Z}(1)\left(\widehat{\underline{\boldsymbol{R}}_Z}(0)\right)^{-1} \\ \Longleftrightarrow &\begin{bmatrix} \hat{\phi}_1 & \hat{\psi}_1 \\ \hat{\psi}_1^* & \hat{\phi}_1^* \end{bmatrix} = \begin{bmatrix} \hat{\gamma}_Z(1) & \widehat{\tilde{\gamma}}_Z(1) \\ \widehat{\tilde{\gamma}}_Z^*(1) & \hat{\gamma}_Z^*(1) \end{bmatrix} \begin{bmatrix} \hat{\gamma}_Z(0) & \widehat{\tilde{\gamma}}_Z(0) \\ \widehat{\tilde{\gamma}}_Z^*(0) & \hat{\gamma}_Z^*(0) \end{bmatrix}^{-1}, \end{aligned}$$

and its asymptotic distribution becomes

$$\sqrt{n}\begin{pmatrix}\mathrm{Re}(\widehat{\phi_1})-\mathrm{Re}(\phi_1)\\ \mathrm{Im}(\widehat{\phi_1})-\mathrm{Im}(\phi_1)\\ \mathrm{Re}(\widehat{\psi_1})-\mathrm{Re}(\psi_1)\\ \mathrm{Im}(\widehat{\psi_1})-\mathrm{Im}(\psi_1)\end{pmatrix} \to_d N\left(\mathbf{0}, \boldsymbol{T}^H\begin{bmatrix}\underline{R}_Z^{11}(0)\underline{\boldsymbol{R}}_\varepsilon & \underline{R}_Z^{12}(0)\underline{\boldsymbol{R}}_\varepsilon\\ \underline{R}_Z^{21}(0)\underline{\boldsymbol{R}}_\varepsilon & \underline{R}_Z^{22}(0)\underline{\boldsymbol{R}}_\varepsilon\end{bmatrix}\boldsymbol{T}\right)$$

The asymptotic distribution of the YW estimators enables us to test the null hypothesis of circular time series. If the processes are circular, we observe that $\psi_1 = \cdots = \psi_p = 0$ which coinsides $\psi_1^* = \cdots = \psi_p^* = 0$.For real-valued VAR($p$) processes, these conditions are known as the test for the null hypothesis of no Granger-causality. Consider the testing problem of the null hypothesis that $H_0 : \psi_1 = \cdots = \psi_p = \psi_1^* = \cdots = \psi_p^* = 0$ against the alternative $H_1 : \psi_j \neq 0$ and $\psi_j^* \neq 0$ for at least one $j \in \{1, \ldots p\}$.

Under the null hypothesis, the CV-VAR(p) coefficient matrix becomes

$$\boldsymbol{A}_{1:p} = (\boldsymbol{A}_1 \quad \cdots \quad \boldsymbol{A}_p) = \begin{bmatrix}\mathrm{Re}(\phi_1) & 0 & \cdots & \mathrm{Re}(\phi_p) & 0\\ \mathrm{Im}(\phi_1) & 0 & \cdots & \mathrm{Im}(\phi_p) & 0\end{bmatrix}$$

Let $\boldsymbol{C}$ be the $(2p \times 4p)$ matrix with the form

$$\boldsymbol{C} = \begin{bmatrix}0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 1\end{bmatrix}$$

Using this notation, the null and alternative hypotheses become $H_0 : \boldsymbol{C}\boldsymbol{\mathcal{A}}_p = \mathbf{0}$ and $H_1 : \boldsymbol{C}\boldsymbol{\mathcal{A}}_p \neq \mathbf{0}$, respectively. We consider the Wald test statistics of the form

$$W = n(\boldsymbol{C}\widehat{\boldsymbol{\mathcal{A}}_p})^T[\boldsymbol{C}\boldsymbol{T}_p^H\left(\widehat{\boldsymbol{R}_{0:p-1}^{(Toe)-1}} \otimes \widehat{\underline{\boldsymbol{R}}_\varepsilon}\right)\boldsymbol{T}_p\boldsymbol{C}^T]^{-1}(\boldsymbol{C}\widehat{\boldsymbol{\mathcal{A}}_p}).$$

Theorem 2. *(Asymptotic distribution of the Wald Statistics.) Let assumption 1 hold. Then, Under the null hypothesis, we have*

$$W \to_d \chi_{2p}^2,$$

where χ_{2p}^2 denotes the chi-squared distribution with $2p$ degrees of freedom.

To test the independency of the residual terms, we apply the following result.

$$\frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \hat{\boldsymbol{u}}_t \to_d N(\boldsymbol{0}, \boldsymbol{R}_\varepsilon)$$

Under the null hypothesis of the circular residual processes, this leads to the test statistics

$$V = \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \begin{pmatrix} \hat{u}_t^2 - \hat{v}_t^2 \\ \hat{u}_t \hat{v}_y \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \begin{bmatrix} \mathrm{Re}(\hat{\varepsilon}_t^2) \\ \mathrm{Im}(\hat{\varepsilon}_t^2) \end{bmatrix} \to_d N\left(\begin{pmatrix} \sigma_u^2 - \sigma_v^2 \\ \rho_{uv}\sigma_u\sigma_v \end{pmatrix}, \boldsymbol{R}_\varepsilon \right)$$

Theorem 3. *Let assumption 1 hold. Then, Under the null hypothesis of circluar residual processes, we have*

$$nV^T(\widehat{\boldsymbol{R}_\varepsilon})^{-1}V \to_d \chi_2^2,$$

where χ_2^2 denotes the chi-squared distribution with 2 degrees of freedom.

Acknowledgements

This research was supported by JSPS KAKENHI Grant Numbers 22K11944, and Nanzan University Pache Research I-A-2 for the 2025 academic year.

References

Abe, T., Imoto, T., Shiohama, T., Miyata, Y., 2025. Modeling joint cylindrical distributions and related markov processes, in: Directional and Multivariate Statistics: A Volume in Honour of Ashis SenGupta. Springer, pp. 103–129.

Brillinger, D.R., 1981. Time Series: Data Analysis and Theory. SIAM.

Brockwell, P.J., Davis, R.A., 1998. Time Series: Theory and Methods. Springer Science & Business Media.

Gatto, R., 2022. Information theoretic results for stationary time series and the gaussian-generalized von mises time series, in: Directional Statistics for Innovative Applications: A Bicentennial Tribute to Florence Nightingale. Springer, pp. 187–209.

Hannan, E.J., 1970. Multiple time series. John Wiley & Sons.

Le Breton, A., 1988. A note on maximum likelihood estimation for the complex-valued first-order autoregressive process. Statistics & probability letters 7, 171–173.

Miller, K.S., 1974. Complex stochastic processes: an introduction to theory and application. (No Title) .

Neeser, F.D., Massey, J.L., 1993. Proper complex random processes with applications to information theory. IEEE Transactions on Information Theory 39, 1293–1302.

Ogata, H., Shiohama, T., 2024. A mixture transition distribution modeling for higher-order circular markov processes. Journal of Time Series Analysis .

Schreier, P.J., Scharf, L.L., 2003. Second-order analysis of improper complex random vectors and processes. IEEE Transactions on Signal Processing 51, 714–725.

Schreier, P.J., Scharf, L.L., 2010. Statistical signal processing of complex-valued data: the theory of improper and noncircular signals. Cambridge University Press.

Shiohama, T., 2023. Complex-valued time series models and their relations to directional statistics, in: Research Papers in Statistical Inference for Time Series and Related Models: Essays in Honor of Masanobu Taniguchi. Springer, pp. 475–496.

Sykulski, A.M., Olhede, S.C., Lilly, J.M., 2016. A widely linear complex autoregressive process of order one. IEEE Transactions on Signal Processing 64, 6200–6210.

Taniguchi, M., Kato, S., Ogata, H., Pewsey, A., 2020. Models for circular data from time series spectra. Journal of Time Series Analysis 41, 808–829.

Distributions for directional data

Imoto Tomoaki
University of Shiozuoka

1 Circular distribution

An observation with periodicity, such as direction and clock time, is represented as a point on the unit circle and is thus called a circular observation. Owing to the periodicity, the origin of the circular observation can be set as an arbitrary point c, and then, observation will take the value on $B_c := [c, c + 2\pi)$ when the period is 2π. Circular distributions are an essential tool for analyzing phenomena related to circular observations. The probability density function (pdf) $f(\cdot)$ of the circular distribution is characterized by nonnegativity on $\mathbb{R}$, integration on B_c to be one, and 2π-periodicity,

$$f(\theta) = f(\theta + 2k\pi) \quad \text{for} \quad k \in \mathbb{Z}.$$

Hereafter, we put $c = -\pi$ for convenience. Such distributions are used to analyze wind direction at an observation point, the disease onset day of the year, the time of crime occurrence in a day, and so on.

2 Simple Construction of Toroidal Distribution

One of the bivariate extensions of the circular distribution is the toroidal distribution, each of whose random variables has a periodicity. Toroidal distributions have recently drawn much attention for analyzing the dihedral angles of 3D protein structure, consecutive animal movements, and wind directions at two observation points, and so on.

Imoto and Abe [2] propose the simple construction of toroidal distributions, and Abe et al. [1] consider its generalization and provide more flexible distributions. The following is the more generalized method. Let $G(\cdot)$ be the nonegative function which satisfies $G(z) = 1 - G(-z)$ for $z \in \mathbb{R}$, and $w(\cdot)$ is the bivariate function which satisfies $w(\theta_1, \theta_2) = -w(-\theta_1, \theta_2) = -w(\theta_1, -\theta_2) = w(\theta_1 + 2k\pi, \theta_2 + 2l\pi)$ for $\theta_1, \theta_2 \in \mathbb{R}$ and $k, l \in \mathbb{Z}$. Put $f_1(\cdot)$ and $f_2(\cdot)$ are the circular pdfs which are symmetric around 0. Then the function

$$f_T(\theta_1, \theta_2) = 2G(w(\theta_1, \theta_2))f_1(\theta_1)f_2(\theta_2), \quad \theta_1, \theta_2 \in [-\pi, \pi) \tag{1}$$

is the toroidal pdf whose marginal densities are $f_1(\theta_1)$ and $f_2(\theta_2)$. The function (1) reduces to that by Abe et al. [1] when $w(\theta_1, \theta_2) = \lambda \sin\theta_1 \sin\theta_2$, and reduces to that by Imoto and Abe [2] when $G(z) = (1 + z)/2$ for $z \in [-1, 1]$.

Interestingly, when the trivariate distribution with pdf

$$f(\theta_1, \theta_2, x) = 2g(x - w(\theta_1, \theta_2))f_1(\theta_1)f_2(\theta_2), \quad \theta_1, \theta_2 \in [-\pi, \pi), \quad x \in (0, \infty), \tag{2}$$

is considered, the marginal pdf of the pair (θ_1, θ_2) coresponds to (1). Using this property, we can implement the EM algorithm with simple update functions for the maximum likelihood estimation.

3 Application to Hidden Markov Model

The hidden Markov model is a type of mixture model widely used for time series data. In the model, each observation is assumed to be generated by one of K distributions that is determined by the latent state $k = 1, 2, \ldots K$, and the latent state follows a Markov process. For $t = 1, 2, \ldots, T$, let $\boldsymbol{z}_t = (z_{t1}, z_{t2}, \ldots, z_{tK})$ be the latent variable whose components indicate the state at time t, or $z_{tk} = 1$ and $z_{tl} = 0$ for $l \neq k$ if the state is k at time t. Assume that the transition probability from state k to state l is A_{kl} and independent of time t, and the observation $\boldsymbol{\theta}_t = (\theta_{1t}, \theta_{2t})$ is generated from the distribution with pdf

$$f_k(\theta_{1t}, \theta_{2t}) = 2G(w_k(\theta_{1t}, \theta_{2t}))f_{1k}(\theta_{1t})f_{2k}(\theta_{2t}),$$

correspoindg to the family (1) when the state is k. Then, if $\boldsymbol{z}_t$ is obseved, the log-likelihood function is represented as

$$\begin{aligned}\ell =& T\log 2 + \sum_{t=1}^{T}\sum_{k=1}^{K} z_{tk}G(w_k(\theta_{1t}, \theta_{2t})) + \sum_{t=1}^{T}\sum_{k=1}^{K} z_{tk}f_{1k}(\theta_{1t}) + \sum_{t=1}^{T}\sum_{k=1}^{K} z_{tk}f_{2k}(\theta_{2t}) \\ &+ \sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{l=1}^{K} z_{t-1k}z_{tl}\log A_{jk} + \sum_{k=1}^{K} z_{1k}\log a_k,\end{aligned}$$

where a_k is the probability of $z_{1k} = 1$.

Using the expression (2) and regarding x_t as a latent variable at time t, more tractable log-likelihhod function

$$\begin{aligned}\ell =& T\log 2 + \sum_{t=1}^{T}\sum_{k=1}^{K} z_{tk}g(x_t - w(x_t, \theta_1, \theta_2)) + \sum_{t=1}^{T}\sum_{k=1}^{K} z_{tk}f_{1k}(\theta_{1t}) + \sum_{t=1}^{T}\sum_{k=1}^{K} z_{tk}f_{2k}(\theta_{2t}) \\ &+ \sum_{t=2}^{T}\sum_{k=1}^{K}\sum_{l=1}^{K} z_{t-1k}z_{tl}\log A_{jk} + \sum_{k=1}^{K} z_{1k}\log a_k\end{aligned}$$

is obtained. When $g(\cdot)$ is a member of the exponential family, this expression is useful to obtain simple update functions in the EM algorithm.

The examples of the update functions and application to a real dataset are found in the main talk in the workshop.

References

[1] Abe, T., Imoto, T., Shiohama, T., and Miyata Y. (2022). On some flexible models for circular, toroidal, and cylindrical data. In Directional Statistics for Innovative Applications: *A Bicentennial Tribute to Florence Nightingale*, pp. 229–243. Springer.

[2] Imoto, T., and Abe, T. (2021). Simple construction of a toroidal distribution from independent circular distributions *Journal of Multivariate Analysis*, 186, 104799.

The second order generalization of Hájek-Le Cam asymptotic minimax theorem

Junichi Hirukawa, Masanobu Taniguchi and Marc Hallin
Nanzan University, Waseda University and Université libre de Bruxelles

The basic results concerning with the asymptotic theory of estimation and testing, Le Cam [2] introduced so-called locally asymptotically normal (LAN) family of distributions. The convolution theorem for LAN case is obtained by Hájek [1]. The convolution result was extended by Le Cam [3] to more general situations than that of LAN case. These results sometimes called the Hájek-Le Cam asymptotic minimax theorem. In this talk we derive the second order generalization of Hájek's convolution theorem. Furthermore, as a application of the second order Hájek's convolution theorem, we lead to the second order Hájek-Le Cam asymptotic minimax theorem. It automatically provides the conditions that the second order asymptotic efficient estimators should satisfy.

First, we define the class of the second order regular estimators, which do not depend on parameter h up to $O\left(c_T^{-1}\right)$, as follows:

Assumption 1. *The class $\mathcal{R}$ of statistics $S = c_T\left(\widehat{\theta}_T - \theta_0\right)$, $\left(\overline{S}_T = c_T\left(\widehat{\theta}_T - \theta_T\right) = S - h\right)$ is called the class of the second order regular estimators if*

$$\mathcal{R} = \left[\{S\} \ : \ \mathcal{L}\left(\widehat{\theta}_T - \theta_T \mid P_{\theta_T,T}\right) = L_{\theta_0}(\cdot) + O\left(c_T^{-2}\right)\right],$$

where the distribution functions of L_{θ_0} do not depend on the parameter h up to order c_T^{-1}.

To clarify the concrete structure of the statistic S belonging to the class $\mathcal{R}$ of the second order regular estimators, we introduce the following class of estimators:

The statistics S in the class $\mathcal{S}$ has the stochastic expansion

$$\mathcal{S} = \left[S \mid S = a_1 W_{1,T}(\theta_0) + c_T^{-1} Q\left(W_{1,T}(\theta_0), W_{2,T}(\theta_0)\right) + \zeta_T\right], \tag{1}$$

where $Q\left(W_{1,T}(\theta_0), W_{2,T}(\theta_0)\right) = O_{P_{\theta_0,T}}(1)$, a_1 is non random constants and $\zeta_T = O_{P_{\theta_0,T}}\left(c_T^{-2}\right)$.

We set $\omega_1 := E_{\theta_0}\left\{Q\left(W_{1,T}(\theta_0), W_{2,T}(\theta_0)\right)\right\}$. Let the statistics S belong to the some sufficiently large class $\mathcal{D}$ of the estimators and satisfy the condition for the first order AMU.

Now, we can describe the following statment regarding to the class $\mathcal{R}$ of the second order regular estimators.

Lemma 1. *Under regularity conditions, it holds that*

$$P_{\theta_T,T}\left\{\Gamma^{1/2}(\theta_0)\, c_T\left(\widehat{\theta}_T - \theta_T\right) \le x\right\} = \Phi(x) - \phi(x)\, c_T^{-1}\left[\Gamma^{1/2}(\theta_0)\,\omega_1 + \Gamma^{3/2}(\theta_0)\frac{\kappa_{111}}{6}\left(x^2 - 1\right)\right] + o\left(c_T^{-1}\right). \tag{2}$$

Note that the right hand side of (2) does not depend on h. That is, the statistics S also belongs to the class $\mathcal{R}$ of the second order regular estimators in Assumption 1.

In order to clarify the properties of the statistics S belonging to the class $\mathcal{R}$ of regular statistics more precisely, we consider the statistics S in (1) as the following concrete form. If $S \in \mathcal{D}$, it is shown

$$Q(w_1, w_2) := \lambda_1 w_1^2 + \lambda_2 w_1 w_2 + \lambda_3.$$

Then, we obtain

$$\omega_1 = E_{\theta_0}\left\{Q\left(W_{1,T}(\theta_0), W_{2,T}(\theta_0)\right)\right\} = \lambda_1 + \lambda_3,\ \kappa_{112} = h\left\{4\lambda_1 + \Gamma^{-2}(\theta_0)\, K(\theta_0)\right\},\ \kappa_{111} = 6\Gamma^{-1}(\theta_0)\,\lambda_1 + \Gamma^{-3}(\theta_0)\, K(\theta_0).$$

The condition for the statistics S belongs to the class $\mathcal{R}$ of the second order regular estimators is given by $\lambda_1 = -\frac{K(\theta_0)}{4\Gamma^2(\theta_0)}$ (and $\omega_1 = \lambda_3 - \frac{K(\theta_0)}{4\Gamma^2(\theta_0)}$).

Now, it is interesting to compare the class of regular statistics $\mathcal{R}$ with other classes of statistics.

We consider the comparison with other cases: (i) the second order unbiasedness and (ii) the second order AMU. Taniguchi [4] gave different definition of the class of the second order asymptotically unbiased estimators from (i). Therefore, we also consider the comparison with (iii) the second order unbiasedness in the sense of Taniguchi [4].

The case (i) is corresponding to $\omega_1 = \lambda_1 + \lambda_3 = 0$. Moreover, if we take $\lambda_3 = -\lambda_1 = \frac{K(\theta_0)}{4\Gamma^2(\theta_0)}$, the statistics S also belong to the class $\mathcal{R}$ of the second order regular estimator and are identical up to order c_T^{-1}.

For the case (ii), if we assume the second order AMU of $\widehat{\theta}_T$, we obtain $\omega_1 = \lambda_1 + \lambda_3 = \Gamma(\theta_0)\frac{\kappa_{111}}{6} = -\frac{2K(\theta_0)+3J(\theta_0)}{6\Gamma^2(\theta_0)}$. Moreover, if we take $\lambda_3 = -\lambda_1 - \frac{2K(\theta_0)+3J(\theta_0)}{6\Gamma^2(\theta_0)} = -\frac{K(\theta_0)+6J(\theta_0)}{12\Gamma^2(\theta_0)}$, the statistics S also belong to the class $\mathcal{R}$ of the second order regular estimator and are identical up to order c_T^{-1}.

The case (iii) corresponds to $\lambda_1\,(= a_1\Gamma(\theta_0) + a_2 J(\theta_0)$ (in Taniguchi [4])$) = -\frac{K(\theta_0)}{6\Gamma^{3/2}(\theta_0)}$, where $\omega_1 = \lambda_3 - \frac{K(\theta_0)}{6\Gamma^{3/2}(\theta_0)}$, which implies that the statistics S do not belong to the class $\mathcal{R}$ of the second order regular estimator. On the other hand, if we take, $\lambda_3 = -\lambda_1$, then $\omega_1 = 0$ and the statistics S are the second order unbiasedness and identical up to order c_T^{-1}. Similarly, if we take, $\lambda_3 = -\lambda_1 - \frac{2K(\theta_0)+3J(\theta_0)}{6\Gamma^2(\theta_0)}$, then $\omega_1 = -\frac{2K(\theta_0)+3J(\theta_0)}{6\Gamma^2(\theta_0)}$ and the statistics S are the second order AMU and identical up to order c_T^{-1}.

(i) The condition $\lambda_1 = -\frac{K(\theta_0)}{4\Gamma^2(\theta_0)}$ implies that

$$\psi_{\overline{S}_T}(t,h) = e^{\left\{\frac{(it)^2\Gamma^{-1}(\theta_0)}{2}\right\}}\left[1 + c_T^{-1}\left\{it\omega_1 - \frac{(it)^3 K(\theta_0)}{12\Gamma^3(\theta_0)}\right\}\right] + O\left(c_T^{-2}\right)$$

and the estimator $\overline{S}_T$ belongs to the class of the second order regular estimators $\mathcal{R}$ in Assumption 1 and therefore, it becomes the second order optimal estimator in the class $\mathcal{D}$.

(ii) Moreover, in the cases (i) the second order unbiasedness, and (ii) the second order AMU, we can choose λ_3 appropriately so that the estimators $\overline{S}_T$ also belongs to the class $\mathcal{R}$ of the second order regular estimators. On the other hand, under (iii) the second order unbiasedness in the sense of Taniguchi [4], we see that the estimators $\overline{S}_T$ do not belong to the class of the second order regular estimators $\mathcal{R}$.

References

[1] Hájek, J. : A characterization of limiting distributions of regular estimates. Zeitschrift für Wahrscheinlichkeits-theorie und Verwandte Gebiete**14**, 323–330 (1970)

[2] Le Cam, L. : Locally asymptotically normal families of distributions. University of California Press (1960)

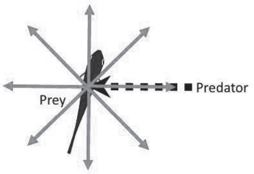
[3] Le Cam, L. : Limits of experiments. Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability Vol. I, 245–261 (1972)

[4] Taniguchi, M. : An automatic formula for the second-order approximation of the distributions of test statistics under contiguous alternatives. International Statistical Review **60**, 211–225 (1992)

Kawabata's bimodal circular distribution: optimal escape direction of prey animals

Ichiro Ken Shimatani
The Institute of Statistical Mathematics

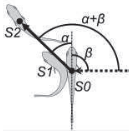
Kawabata et al. 2023

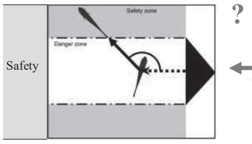


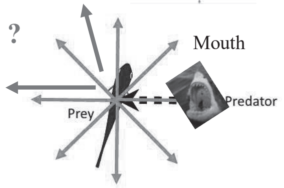
本発表はこの論文の内容の紹介が中心ですが、若干、新しい図表なども加えています。

When a predator animal is approaching to a prey animal, to which direction should the prey escape?

If a prey first bends a body, then, dashes, and this requires a few milliseconds depending on the turning angle (larger turn needs a longer period, impossible for very small turn), the prey needs the period of turning and dashing before entering into the safety zone. The optimal direction depends on initial body directions (and various biological factor specific to the prey and the predator).

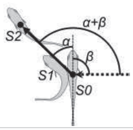


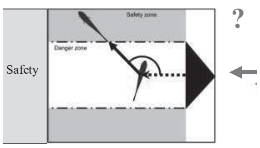


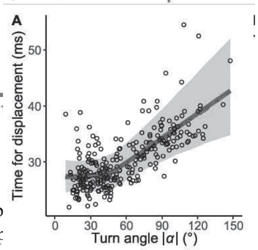


When a predator animal is approaching to a prey animal, to which direction should the prey escape?

If a prey first bends a body, then, dashes, and this requires a few milliseconds depending on the turning angle (larger turn needs a longer period, impossible for very small turn), the prey needs the period of turning and dashing before entering into the safety zone. The optimal direction depends on initial body directions (and various biological factor specific to the prey and the predator).

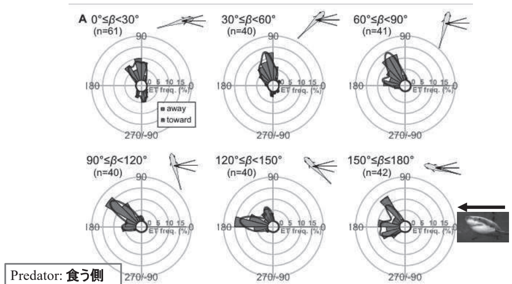






When a predator animal is app
to which direction should the p

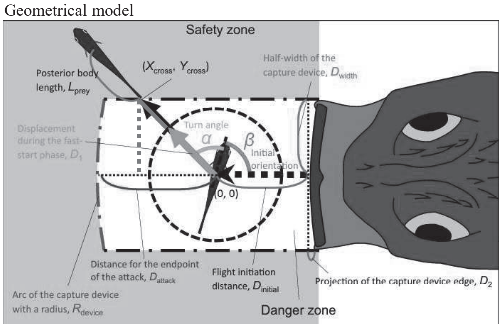
Some experimental results showed bimodality.
Objective: Explore biological mechanisms that cause the bimodal patterns of escape directions using mathematical models.



Predator: 食う側
Prey: 食われる側

Kawabata et al. 2023

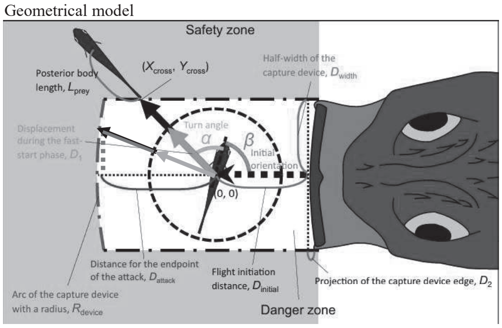
Geometrical model



D_{attack} : distance the predator gives up chasing because the prey quickly escapes away

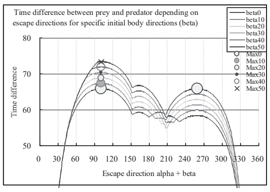
Kawabata et al. 2023

Geometrical model

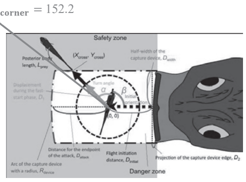


D_{attack}: distance the predator gives up chasing because the prey quickly escapes away

Kawabata et al. 2023

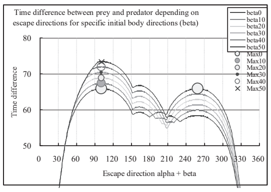


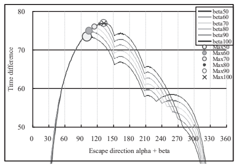
Example of the time differences for 6 prey's initial body direction (beta) for 0 – 360 escape directions.
When the initial direction is small, the prey escapes to the side safety zone.
If the direction is 0, T_{diff} is symmetric.
The parameters were set according to the experimental design,
e.g. D_{width} = mouth size of the movie
$D_{initial}$ = distance between the prey and the screen



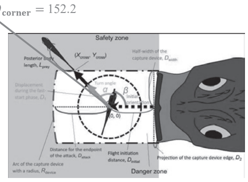
Predator: 食う側
Prey: 食われる側

$$T_{diff}' = \begin{cases} \frac{D_2}{U_{pred}} - \frac{X_{cross}}{U_{pred}} - T_1(|\alpha|) - \frac{\sqrt{X_{cross}^2 + Y_{cross}^2}}{U_{prey}}, \alpha + \beta < \theta_{corner} \\ \frac{D_{attack}}{U_{pred}} - T_1(|\alpha|) - \frac{\sqrt{X_{cross}^2 + Y_{cross}^2}}{U_{prey}}, \alpha + \beta \geq \theta_{corner} \end{cases}$$

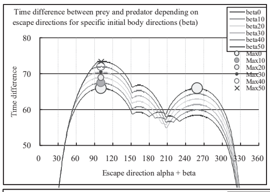
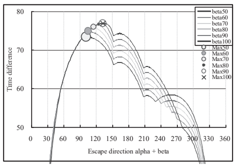
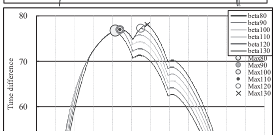




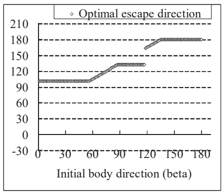
When the initial direction becomes larger, the prey escapes toward the corner of the safety zone.

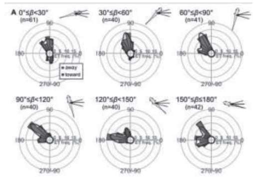
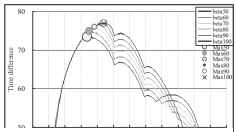


$$T_{diff}' = \begin{cases} \frac{D_2}{U_{pred}} - \frac{X_{cross}}{U_{pred}} - T_1(|\alpha|) - \frac{\sqrt{X_{cross}^2 + Y_{cross}^2}}{U_{prey}}, \alpha + \beta < \theta_{corner} \\ \frac{D_{attack}}{U_{pred}} - T_1(|\alpha|) - \frac{\sqrt{X_{cross}^2 + Y_{cross}^2}}{U_{prey}}, \alpha + \beta \geq \theta_{corner} \end{cases}$$

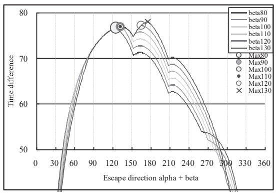
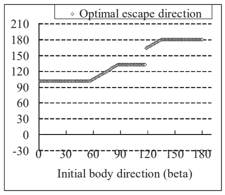




When the initial direction is closed to 180, the prey selects the opposite direction (180) against the predator.

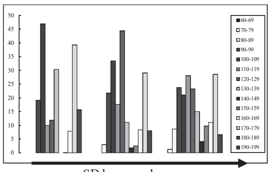


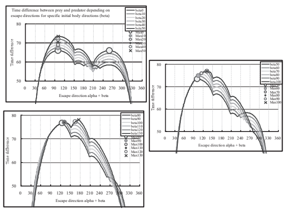
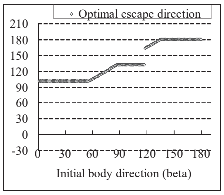
Basically, there are three choices, and preys seem to select one of them.

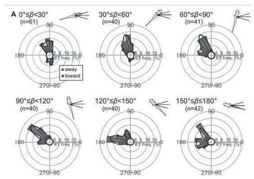



If the try's initial directions are uniformly distributed over 0 – 180, the escape directions show a tri-modal pattern.
If the observed escape directions contain random noise due to a measurement error or/and prey's motion, escape directions show bi-modality if the variance of random noise is not small.

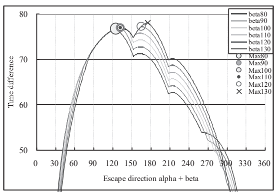
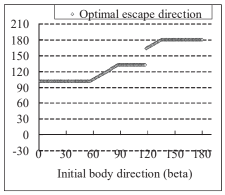


SD becomes larger

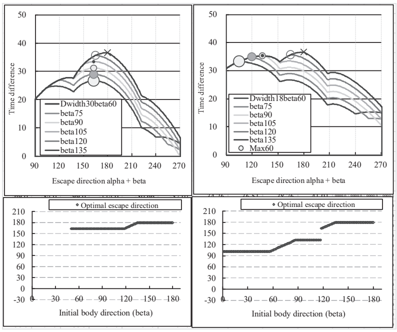


Occasionally, preys seem to select the suboptimal escape direction.
This may be because the prey wrongly estimate some quantity(ies).
D_{width}: predator's mouth size
D_{attack}: How long will predator chase?
U_{pred}: Speed of the predator
…

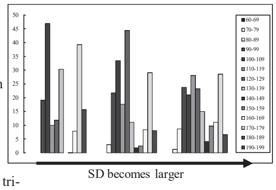



If the prey is assumed to wrongly estimate D_{width} as 30 (the true size of the experiment was 18), the optimal escape directions are changed to more horizontal (close to 180).

Side escape becomes not optimal.

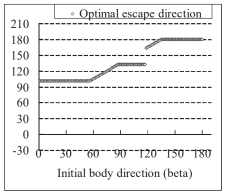


If the prey's initial directions are uniformly distributed over 0 – 180, the escape directions show a tri-modal pattern.
If the observed escape directions contain random noise due to a measurement error or/and prey's motion, escape directions show bi-modality if the variance of random noise is not small.



Let me emphasize here that this bi- and tri-modal circular distribution is naturally derived from the biological principles, not a mathematical artifact.

Historically, important probability distribution has been found from real observations (e.g. the normal distribution in astronomy).
It is expected that interesting circular distributions will be found in biological observations in future, as Dr. Kawabata succeeded in it.



Bias correction for kernel density estimation with spherical data

Yasuhito Tsuruta

2025/11/30

Table of contents

1. Introduction
2. Proposed estimators
3. Numerical experiment
4. Summary

Directional data

- Assume that directional observations $X_1, \dots, X_n$ form a random sample from an unknown directional density $f(x)$ defined on a unit sphere $\mathbb{S}_d := \{x \in \mathbb{R}^{d+1} : ||x|| = 1\}$.
- Let w_d be the surface Lebesgue measure on $\mathbb{S}^d$.
  Then the density f satisfies

$$\int_{\mathbb{S}^d} f(x)\, w_d(dx) = 1.$$

- The area on sphere $\mathbb{S}_d$ is $w_d = 2\pi^{(d+1)/2}/\Gamma((d+1)/2)$, where $\Gamma(y) := \int_0^\infty e^{-r} r^{y-1} dr$ is the Gamma function.

Kernel density estimation

- Nonparametric methods such as kernel density estimation allow flexible estimations of directional data without assuming a specific parametric form.

Kernel density estimator

A kernel density estimator is defined as

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

- The kernel function $K_\kappa : \mathbb{S}_d \to \mathrm{R}$ is a rotationally symmetric.
- κ is a smoothing parameter, and corresponds to $\kappa = h^{-2}$ for a bandwidth $h > 0$.
- The kernel takes form $K_\kappa(x^{\mathrm{T}}X_i) := C_\kappa(L)^{-1}L(\kappa(1 - x^{\mathrm{T}}X_i))$, where
  - Bounded function $L(r) : [0, \infty) \to \mathbb{R}$
  - Normalizing constant $C_\kappa(L) := \int_{\mathbb{S}_d} L(\kappa(1 - x^{\mathrm{T}}X_i))w_d(dx)$.

5 / 29

Motivation (1)

- The performance of a kernel density estimator on the sphere is evaluated by the mean integrated squared error (MISE):

$$\mathrm{MISE}[\hat{f}_\kappa(x)] := \int_{\mathbb{S}_d} \mathrm{E}\left[\{\hat{f}_\kappa(x) - f(x)\}^2\right] w_d(dx).$$

This decomposes into $\int_{\mathbb{S}_d} \mathrm{Bias}^2 + \int_{\mathbb{S}_d} \mathrm{Var}$.
  - Assume that smoothing parameter $\kappa := \kappa_n$ ($\kappa > 0$) satisfies that $\lim_{n\to\infty} \kappa = \infty$ and $\lim_{n\to\infty} n^{-1}\kappa^{1/2} = 0$.
- For standard kernel density estimators (e.g., using the von Mises kernel):
  - $\mathrm{Bias}[\hat{f}_\kappa(x)] = O(\kappa^{-1})$.
  - $\mathrm{Var}[\hat{f}_\kappa(x)] = O(n^{-1}\kappa^{1/2})$.
  - Resulting MISE: $\mathrm{MISE}[\hat{f}_\kappa(x)] = O(n^{-4/(4+d)})$ (Hall et al., 1987; García-Portugués, 2013).
  - The slow convergence rate is caused by the dominant bias term.
- In circular data, higher-order bias correction exists such as *generalized jackknifing method* (Di Marzio et al., 2011; Tsuruta and Sagae, 2017; Bedouhene and Zougab, 2022).

6 / 29

Motivation (2)

- Why circular bias correction cannot be directly extended;
  - On sphere, the asymptotic bias is expressed using a Taylor expansion with directinal derivatives.
  - The higher-order expression become complicated.
  - However, the (Geodesic-based) directional derivatives defined by Klemelä(2000) yield a simpler and tractable formulation.
- To improve the convergence rate of MISE, we need estimators that:
  - eliminate the leading bias term,
  - maintain the convergence rate of variance,
  - can be constructed from practical kernels (e.g., vM).

7 / 29

Contributions

- Extension of generalized jackknifing to spherical kernel density estimation
  - Derive a bias expansion using geodesic directional derivatives.
- Two new bias-corrected estimators;
  - Jones and Foster kernel density estimator constructs a higher-order kernel from a second-order kernel (standard kernel e.g., vM).
  - Additive kernel density estimator eliminates the second-order bias term via a linear combination.
- Asymptotic theory:
  - Obtain bias, variance, and AMISE for both estimators.
  - Achieve the improved MISE rate: $O\left(n^{-8/(8+d)}\right)$.
- Numerical experiments show consistent improvement over the VM kernel density estimator.

8 / 29

Smoothness of directional densities

- Higher-order smoothness conditions are required to derive the higher-order bias expansion of $\mathrm{Bias}[\hat{f}_\kappa(x)]$.
- A commonly smoothness definition is the directional derivatives of f (Hall et al, 1987):

Directional derivatives

- Let function $g : \mathbb{R}^{d+1} \to \mathbb{R}$ and $x, \xi \in \mathbb{R}^{d+1}$ with $x \perp \xi$.
- The derivative of g at x in the direction of ξ is
  - $D_\xi g(x) = \lim_{h \to 0} h^{-1}\{g(x + h\xi) - g(x)\}$;
  - $D_\xi^s g(x) := D_\xi D_\xi^{s-1} g(x)$ for an integer $s \geq 2$.
- To apply this to $f : \mathbb{S}^d \to \mathbb{R}$, we extend $f : \mathbb{S}_d \to \mathbb{R}$ to $\bar{f}(x) := f(x/||x||) : \mathbb{R}^{d+1} \setminus \{0\} \to \mathbb{R}$.

Smoothness via tangent–normal decomposition (Klemelä, 2000)

- Klemelä (2000) introduced another smoothness definition based on the tangent–normal decomposition at a fixed point $y \in \mathbb{S}^d$.
- Every $X \in \mathbb{S}^d$ can be written as

$$X = ty + (1 - t^2)^{1/2}\xi,$$

where $t = X^{\mathrm{T}}y$ and $\xi := (X - ty)/||X - ty||$ is a unit vector orthogonal to y in space $T_y := \{\xi \in \mathbb{S}_d | \xi \perp y\}$.

- Equivalently, the parametrization $\phi_y : T_y \times (0, \pi) \to \mathbb{S}^d \setminus \{y, -y\}$ is given by

$$\phi_y^{-1}(\xi, \theta) = \cos(\theta)y + \sin(\theta)\xi. \tag{1}$$

- Each parameterization corresponds to an integration formula; that is, (1) has

$$\int_{\mathbb{S}_d} f(x) w_d(dx) = \int_0^\pi d\theta \sin(\theta)^{d-1} \int_{T_y} f(\phi_y^{-1}(\xi, \theta)) w_{d-1}(d\xi).$$

(Geodesic-based) directional derivatives

- The derivative of order s is $D^s f : \mathbb{S}_d \to \mathbb{R}$, defined by

$$D^s f(x) = \frac{1}{w_{d-1}} \int_{T_x} D_\xi^s \bar{f}(x) w_{d-1}(d\xi),$$

where $T_x = \{\xi \in \mathbb{S}_d | \xi \perp x\}$.

Lemma 1 (Klemelä, 2000)

Let $f : \mathbb{S}_d \to \mathbb{R}$ and $x, \xi \in \mathbb{S}_d$, with $\xi \perp x$. Set $\phi_x^{-1}(\xi, \theta) = x\cos(\theta) + \xi\sin(\theta)$ with $\theta \in \mathbb{R}$. Define $\phi_x^{-1}(\xi, \theta) = x\cos\theta + \xi\sin\theta$. Then for any integer $s \geq 0$,

$$D_\xi^s \bar{f}(x) = \frac{\partial^s}{\partial\theta^s} f(\phi_x^{-1}(\xi, \theta))\Big|_{\theta=0},$$

We obtain $D_{-\xi}^s \bar{f}(x) = (-1)^s D_\xi^s \bar{f}(x)$.

- all odd-order derivatives satisfy $D^s f(x) = 0$ for all $x \in \mathbb{S}_d$.

Assumption of true density

Assumption A1

A1 Partial derivative $\partial^s/\partial\theta^s f(\phi_x^{-1}(\xi,\theta))$ is continuous as a function of $\theta \in \mathbb{R}$ for $s \in \{1,\cdots,m\}$. $D^s f(x)$ exists and is bounded by $s \in \{1,\cdots,m\}$.

Under Assumption A1, f admits Taylor expansion around x:

$$f(\phi_x^{-1}(\xi,\theta)) = f(x) + \sum_{s=1}^{m-1} \frac{\theta^s}{s!} f_\theta^{(s)}(\phi_x^{-1}(\xi,\theta))\Big|_{\theta=0} + O(\theta^m),$$

where $f_\theta^{(s)}(\phi_x^{-1}(\xi,\theta)) := \frac{\partial^s}{\partial\theta^s} f(\phi_x^{-1}(\xi,\theta))$.

Conditions of Kernel function to derive bias and variance

- For even integers $\ell \geq 0$, define the ℓ-th moment as $\mu_\ell(L) := \int_0^\infty L(r) r^{(\ell+d-2)/2} dr$.
- Let $v \geq 4$ be an even integer.

Conditions on (bounded function) L in Kernel function

We assume that the following conditions on L hold:

L1 Derivative $L'(r) := dL(r)/dr$ is continuous;

L2 $L(r) r^{(v+d)/2} \to 0,\ r \to \infty$;

L3 For $\ell \in \{0,2,\ldots,v\}$, $\mu_\ell(L)$ is bounded and $\mu_\ell(L) = \int_0^\kappa L(r) r^{(\ell+d-2)/2} dr + O(\kappa^{-(v+2)/2})$;

L4 For $\ell \in \{0,2,4\}$, $\mu_\ell(|L|^m)$ is bounded and $\mu_\ell(|L|^m) = \int_0^\kappa |L(r)|^m r^{(\ell+d-2)/2} dr + O(\kappa^{-(v+2)/2})$, $m \in \{2,4,6,8\}$.

- Conditions L1 and L2 are required to provide the partial integral $\int_0^\infty L'(r) r^{(v+d)/2} dr = -(v+d)\mu_v(L)/2$.
- Condition L3 is used to derive the bias.
- Condition L4 is required to ensure the boundedness of the variance.

Derive aymptotic bias

- Combining Lemma 1 with the Taylor expansion gives

$$\mathrm{E}[\hat{f}_\kappa(x)] = f(x) + \sum_{t=1}^{p/2} \beta_{2t}(K_\kappa) \frac{D^{2t} f(x)}{(2t)!} + O(\beta_{p+2}(K_\kappa)), \qquad (2)$$

where $\beta_{2t}(K_\kappa) := w_{d-1} \int_0^\pi \theta^{2t} \sin(\theta)^{d-1} K_\kappa(\cos(\theta)) d\theta$. Note that $\beta_0(L) = 1$.

- Using the kernel representation, $\beta_{2t}(K_\kappa)$ has the expansion:

$$\beta_{2t}(K_\kappa) \simeq \frac{w_{d-1}}{C_\kappa(L)} \kappa^{-d/2} \left\{ \sum_{q=t}^{z/2} \sum_{m=0}^{z/2-q} \frac{\alpha_{q,z,t}}{\kappa^{(q+m)} m!} \mu_{2(q+m)}(L) + O(\kappa^{-(z+2)/2}) \right\}.$$

- If we use a kernel whose lower-order moments vanish, the leading terms in $\beta_{2t}(K_\kappa)$ improve significantly.

p-th order kernel function

A kernel K_κ is of p-th order (even $p \geq 2$) if L1–L4 hold with $v \geq p+2$ and

$$\mu_0(L) \neq 0, \quad \mu_\ell(L) = 0, \ell \in \{2,4,6,\ldots,p-2\}, \quad \mu_p(L) \neq 0.$$

- the normal constant is $C_\kappa(L) = w_{d-1} \kappa^{-d/2} \int_0^{2\kappa} L(r) r^{(d-2)/2} dr$.
- For $d = 2$, this simplifies to $C_\kappa(L) = w_{d-1} \kappa^{-d/2} \int_0^{2\kappa} L(r) dr$.
- Its asymptotic form is

$$C_\kappa(L) = \begin{cases} w_1 \kappa^{-1} \{\mu_0(L) + O(\kappa^{-(v+2)/2})\}, & \text{if } d = 2, \\ 2^{(d-2)/2} w_{d-1} \kappa^{-d/2} \mu_0(L) \{1 + O(\kappa^{-p/2})\}, & \text{if } d \neq 2. \end{cases}$$

- For a p-th order kernel, combining this with the earlier expansion yields

$$\beta_{2t}(K_\kappa) = \kappa^{-p/2} \mu_0(L)^{-1} \mu_p(L) b_{p,2t} + O(\kappa^{-(p+2)/2}), \qquad (3)$$

for $t \in \{0,1,\ldots,p/2\}$, where $b_{p,2t}$ is constant.

- By substituting (3) into the right-hand side of (2), the bias:

$$\mathrm{Bias}[\hat{f}_\kappa(x)] = \kappa^{-p/2} \mu_0(L)^{-1} \mu_p(L) \sum_{t=1}^{p/2} b_{p,2t} \frac{D^{2t} f(x)}{(2t)!} + O(\kappa^{-(p+2)/2}), \qquad (4)$$

- the variance of a p-th order kernel estimator remains $\mathrm{Var}[\hat{f}_\kappa(x)] = O(n^{-1}\kappa^{1/2})$ regardless of p.

Asymptotic MISE of p-th order kernel density estimator

Theorem 1 (Tsuruta 2024)

Assume that A1 with $m = p + 2$ and A2 hold. When a pth-order kernel is employed, we have

$$\mathrm{MISE}[\hat{f}_\kappa(x)] = \mathrm{AMISE}[\hat{f}_\kappa(x)] + O(\kappa^{-(p+1)} + n^{-1}\kappa^{(d-2)/2} + n^{-1}),$$

$$\mathrm{AMISE}[\hat{f}_\kappa(x)] = \kappa^{-p}\frac{\mu_p(L)^2}{\mu_0(L)^2} R\left(\sum_{t=1}^{p/2} \frac{b_{p,2t} D^{2t} f(x)}{(2t)!}\right) + \frac{\kappa^{d/2}}{n}\gamma(L), \qquad (5)$$

where $\gamma(L) := 2^{-(d-2)/2}\mu_0(L^2)[w_{d-1}\mu_0(L)^2]^{-1}$ and $R(g) := \int_{\mathbb{S}_d} g(x)^2 w_d(dx)$. The minimizer of the right-hand side (RHS) of (5) is

$$\kappa_* = \left\{\frac{2^{d/2} p w_{d-1}\mu_p(L)^2 R\left(\sum_{t=1}^{p/2} \frac{b_{p,2t}D^{2t}f(x)}{(2t)!}\right)}{d\mu_0(L^2)}\right\}^{2/(2p+d)} n^{2/(2p+d)}.$$

Therefore, the optimal MISE of $\hat{f}_\kappa$ is $O(n^{-2p/(2p+d)})$.

Additive bias correction

we discuss two additive bias correction methods (Jones and Foster, 1993) employing a linear combination:

- generating higher-order kernels;
  - Jones and Foster (JF) kernel.
- eliminating the first term of the bias;
  - Additive (ADD) kernel density estimator

Generating higher-order kernels

- Let a pth-order kernel function denote $K_{\kappa,[p]}(x^{\mathrm{T}}\eta) := C_\kappa^{-1}(L_{[p]})L_{[p]}(\kappa(1 - x^{\mathrm{T}}\eta))$ for the fixed direction $\eta \in \mathbb{S}_d$.

Definition of JF kernel

For a JF kernel with the order of $p + 2$: $K^{\mathrm{JF}}_{\kappa,[p+2]}$, its main term $L^{\mathrm{JF}}_{[p+2]}(r)$ is generated from $L_{[p]}(r)$, which satisfies L1 to L4 with $v \geq p + 4$ as follows:

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

- JF kernel generates a higher-order kernel from a second kernel.
- The popular second-order kernel is the von Mises kernel with the function $L_{\mathrm{VM}}(r) = e^{-r}$ having the ℓ-th moment $\mu_\ell(L_{\mathrm{VM}}) = \Gamma((\ell + d)/2)$.
- Generate the following fourth-order JF kernel based on the von Mises kernel:

$$L^{\mathrm{JFVM}}_{[4]}(r) = \left(\frac{2+d}{2} - r\right) e^{-r},$$

where $r = \kappa(1 - x^{\mathrm{T}}\eta)$.
- The higher-order kernel needs to be a negative value to eliminate the moments above the second order.

eliminating the first term of the bias

- Theorem 1 implies that the bias with a second-order kernel $K_{a\kappa}(x^{\mathrm{T}}X_i)$ for $a \in (0, 1)$ is $\mathrm{Bias}[\hat{f}_{a\kappa}(x)] = a^{-1}\kappa^{-1}b_{2,2}D^2 f(x)/2 + O(\kappa^{-2})$.
- A linear combination of $K_\kappa(x^{\mathrm{T}}X_i)$ and $K_{a\kappa}(x^{\mathrm{T}}X_i)$ can eliminate the term of order κ^{-1} in the bias.

Definition of ADD kernel density estimator

$$\hat{f}^{\mathrm{ADD}}_{\kappa,a}(x) := \frac{1}{1-a}\hat{f}_\kappa(x) - \frac{a}{1-a}\hat{f}_{a\kappa}(x), \quad a \in (0, 1).$$

- we add an additional condition of L as follows:
  - L5 For $\ell \in \{0, 2, 4\}$, $\mu_\ell(L; a) := \int_0^\infty L(r)L(ar)r^{(\ell+d-2)/2}dr$ is bounded and $\mu_\ell(L; a) = \int_0^\kappa L(r)L(ar)r^{(\ell+d-2)/2}dr + O(\kappa^{-(v+2)/2})$.
- The bias of the ADD kernel density estimator is

$$\mathrm{Bias}[\bar{f}_{\kappa,a}(x)] \simeq -a^{-1}\kappa^{-2}\left\{B_4(L)\sum_{t=1}^{2} b_{4,2t}\frac{D^{2t}f(x)}{(2t)!} - B(L; d)\frac{D^2 f(x)}{2}\right\},$$

for $a \in (a, 1)$, where $B_4(L) := \mu_0(L)^{-1}\mu_4(L)$ and

$$B(L; d) := \begin{cases} 0, & d = 2, \\ 2^{-1}(d-2)\mu_2(L)^2\mu_0(L)^{-2}, & \text{otherwise.} \end{cases}$$

Asymptotic MISE of ADD kernel density estimator

Theorem 2 (Tsuruta, 2024)

Assume that A1 with $m = 6$ and A2 hold. We employ a second-order kernel that satisfies L1 to L5 with $v \geq 6$. Then, we have the asymptotic MISE:

$$\mathrm{AMISE}[\hat{f}^{\mathrm{ADD}}_{\kappa,a}(x)] = \kappa^{-4}\frac{\bar{R}}{a^2} + \frac{\kappa^{d/2}}{n}\bar{\gamma}(L;a), \tag{6}$$

where $\bar{R} = B_4(L)^2R_1 - 2B_4(L)B(L;d)R_2 + B(L;d)^2R_3$ with R_1, R_2, and $R_3 := R(D^2f(x)/2)$ are squaed integrera of derivatives of f, and

$$\bar{\gamma}(L;a) := \frac{(1+a^{2+d/2})\mu_0(L^2) - 2a^{1+d/2}\mu_0(L;a)}{2^{(d-2)/2}w_{d-1}\mu_0(L)^2(1-a)^2},$$

The minimizer of the RHS of (6) is given by

$$\kappa_{\mathrm{ADD}} = \left[\frac{8\bar{R}}{da^2\bar{\gamma}(L;a)}\right]^{2/(8+d)} n^{2/(8+d)}.$$

Therefore, the optimal MISE of $\bar{f}_{\kappa,a}$ is $O(n^{-8/(8+d)})$.

Exact MISE of proposed estimators

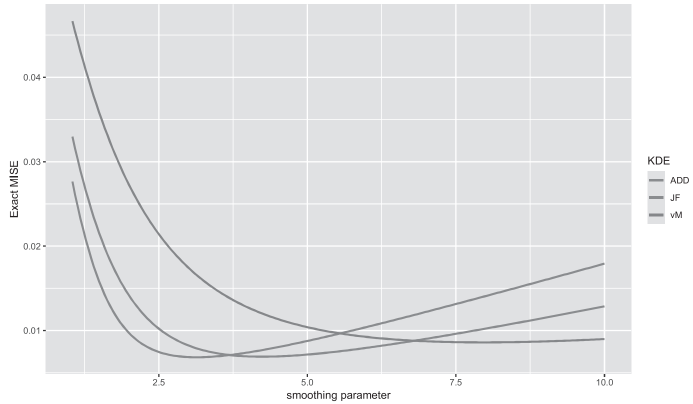


Fig. 1: Exact MISE (numerically calculated) over smoothing parameter κ. Samples from the vM distribution ($d = 2$, $n = 100$). Red, green, and blue lines: ADD-KDE (based on VM, $a = 0.5$), JF-KDE(based on VM), and vM-KDE, respectively.

Setting

- To compare proposed estimators (JF and ADD kernel density estimations based on the vM kernel) and previous estimator (vM kernel density estimators):
  - Sample size: $n = 1000$;
  - Repetitions: $N = 1000$;
  - Scenarios: seven mixtures of vM densities ($d = 2$).
- Smoothing parameter selector:
  - Least squares cross-validation (LSCV);
  - Exact mixtures (EMI) selector (for only vM) (García-Portugués, 2013).

Results

Table 1: Each value is MISE(SD)

	vM		JF	ADD	
				a=0.5	a=0.9
Scenario	LSCV	EMI	LSCV	LSCV	LSCV
1(vM)	21.2(0.2)	19.9(0.2)	14.2(0.2)	14.4(0.2)	14.3(0.2)
2(2vM)	23.5(0.2)	22.4(0.2)	17.8(0.2)	18.0(0.2)	17.8(0.2)
3(5VM)	71.0(0.5)	71.7(0.5)	62.1(0.5)	62.4(0.5)	62.1(0.5)
4(3vM)	76.8(0.5)	74.5(0.5)	55.2(0.5)	55.9(0.5)	55.2(0.5)
5(10vM)	131.7(0.8)	130.3(0.8)	118.2(0.7)	118.6(0.7)	118.2(0.7)
6(4vM)	403.3(3.2)	401.2(3.0)	288.2(2.8)	292(2.8)	288.3(2.8)
7(20vM)	78.6(0.5)	79.4(0.4)	74.3(0.4)	74.3(0.4)	74.3(0.4)

- The proposed estimators outperform the previous estimators.
- JF kernel density estimator performs slightly better than the ADD kernel density estimator with $a = 0.5$.

Summary

- We extended the higher-order kernel and generalized jackknifing to the spherical setting.
  - JF kernel (construct a higher-order kernel from a second kernel)
  - ADD kernel density estimator (eliminates the second term in the bias)
- The proposed estimators achieve a faster rate of MISE than standard kernel density estimators.
  - Proposed estimators: $O(n^{-8/(8+d)})$.
  - Previous estimators: $O(n^{-4/(4+d)})$.
- Numerical experiments illustrate the theoretical results.

Limitations

- The proposed estimators may take negative values because they cancel higher-order moments in the bias.
- A nonnegative version can be obtained by $\bar{f}_\kappa(\theta) := \max(\hat{f}_\kappa(\theta), 0)$.
- To obtain a proper density, an additional normalization step can be applied to ensure that the estimator integrates to one.

References

- Bedouhene, K. and Zougab, N. (2022) Commun. Stat. Simul. Comput., 51, 774–792.
- Di Marzio, M., Panzera, A., and Taylor, C. C. (2011). J. Stat. Plan. Inference., 141, 2156-2173.
- García-portugués, E. (2013). Electron. J. Stat., 7, 1655–1685.
- Hall, P., Watson, G. S., and Cabrera, J. (1987). Biometrika, 74, 751–762.
- Jones, M.C. and Foster, P.J. (1993) J. Nonparametric Stat., 3, 81–94.
- Klemelä, J. (2000) J. Multivar. Anal., 73, 18–40.
- Tsuruta, Y. (2024) J. Multivar. Anal., 203, 105338.
- Tsuruta, Y., and Sagae, M. (2017). Stat. Probab. Lett., 131, 46-50.

日本統計研究所報（最近刊行分）

日本統計研究所報　No. 63
2026 年 6 月 30 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

B U L L E T I N
OF
JAPAN STATISTICS RESEARCH INSTITUTE

No.63 **June 2026**

Frontiers in Directional Statistics Theory, Methods, and Applications

CONTENTS

Edited by
JAPAN STATISTICS RESEARCH INSTITUTE
HOSEI UNIVERSITY
TOKYO, JAPAN