ISSN 0385-2148

研究所報

No.60

Advances in Statistical Modeling and Inference: Exploring Applications on Diverse Manifolds

２０２５年 3月

法政大学
日 本 統 計 研 究 所

ISSN 0385-2148

研究所報

No.60

Advances in Statistical Modeling and Inference: Exploring Applications on Diverse Manifolds

２０２５年 3月

法政大学
日 本 統 計 研 究 所

はじめに

法政大学日本統計研究所は Angelo Efoevi Koudou 氏（University of Lorraine）と、当該分野に関連する統計学者(井本智明氏、黒田正博氏、白石博氏、宮田庸一氏、小方浩明氏)を招聘し、法政大学多摩キャンパスにおいて国際ワークショップを開催した。Koudou 氏には An independence property characterizing Kummer laws、井本氏には Bivariate distribution related to directional statistics、黒田氏には Acceleration of the EM algorithm、白石氏には Asymptotic Property for Generalized Random Forests、宮田氏には An extension of the Weibull sine-skewed von Mises distributions: advantages and limitations、小方氏には Periodicity for circular data について講演をしていただいた。本所報は、これらの報告についての概要をまとめたものであり、本国際ワークショップの研究発表と研究交流により、活発な議論が展開された。また、今後の研究の発展にも寄与することが期待される.

2025 年 3 月　日本統計研究所

目次

Bivariate distribution related to directional statistics

Tomoaki Imoto
University of Shizuoka

1 Introduction

The bivariate circular distribution, called toroidal distribution, is used for analyzing the data that is represented on a torus such as the directions of the wind at two points, and positions of the orthologs between paired circular genomes. The examples of the toroidal distributions are the bivariate von Mises distribution by Mardia (1975) and Johnson and Wehrly (1978), wrapped bivariate normal distribution by Johnson and Wehrly (1978) and Baba (1981), bivariate cardioid distribution by Wang and Shimizu (2012), bivariate wrapped Cauhcy distribution by Kato and Jones (2015) and so on. Among these, the bivariate cardioid distribution has the property that both marginal and conditional are univariate cardioid distributions. The same property can be found for the bivariate wrapped Cauhcy distribution, whose marginal and conditional are univariate wrapped Cauhcy distributions. Such a property is rare for directional statistics as explained by Mardia's comment in Pewsey and Garcìa-Porugués (2021).

In this report, a new toroidal distribution whose marginal and conditional belong to the same distribution family is proposed. Thit is considered as a bivariate version of the extended wrapped Cauchy distribution (EWC) by Kato and Jones (2013) whose density has a form

$$f(\theta) = \frac{1}{2\pi}\frac{1+(ab)^2-2ab\cos\nu}{1-(ab)^2}\frac{1-a^2}{1+a^2-2a\cos(\theta-\mu+\nu)}\frac{1-b^2}{1+b^2-2b\cos(\theta-\mu)},$$

where $a, b \in (0,1)$ and $\mu, \nu \in [-\pi, \pi)$. Hereafter, when the random variable Θ is distributed as the extended wrapped Cauhcy distribution, it is denoted by $\Theta \sim EWC(a, b, \mu, \nu)$. The marginal and conditional of the proposed distribution is proved as the EWC in Section 2. The fitting example is shown in Section 3. The concluding remarks of this report are provided in Section 4.

2 Proposition and property

Consider the function

$$g(\theta_1, \theta_2) = \frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta_1 \pm \theta_2+\nu)}\frac{1-\rho_1^2}{1+\rho_1^2-2\rho_1\cos\theta_1}\frac{1-\rho_2^2}{1+\rho_2^2-2\rho_2\cos\theta_2},$$

where $\rho, \rho_1, \rho_2 \in (0,1)$ and $\nu \in [-\pi, \pi)$. This is the product of three wrapped Cauhcy kernels. The reproductive property of the wrapped Cauhcy distribution, or

$$\int_{-\pi}^{\pi}\frac{1-a^2}{2\pi\{1+a^2-2a\cos(\theta-\psi-\mu_a)\}}\frac{1-b^2}{2\pi\{1+b^2-2b\cos(\psi-\mu_b)\}}d\psi = \frac{1-(ab)^2}{2\pi\{1+(ab)^2-2ab\cos(\theta-\mu_a-\mu_b)\}},$$

leads to the integrals of the function $g(\cdot,\cdot)$ as

$$\int_{-\pi}^{\pi} g(\theta_1,\theta_2)d\theta_2 = 2\pi\frac{1-(\rho\rho_2)^2}{1+(\rho\rho_2)^2-2\rho\rho_2\cos(\theta_1+\nu)}\frac{1-\rho_1^2}{1+\rho_1^2-2\rho_1\cos\theta_1},$$

and

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} g(\theta_1,\theta_2)d\theta_1 d\theta_2 = (2\pi)^2 \frac{1-(\rho\rho_1\rho_2)^2}{1+(\rho\rho_1\rho_2)^2-2\rho\rho_1\rho_2\cos\nu}.$$

From this, the function

$$f(\theta_1,\theta_2) = \frac{1}{(2\pi)^2}\frac{1+(\rho\rho_1\rho_2)^2-2\rho\rho_1\rho_2\cos\nu}{1-(\rho\rho_1\rho_2)^2} g(\theta_1-\mu_1,\theta_2-\mu_2)$$

is proven to be a probability density function whose marginal density is

$$f(\theta_1) = \frac{1}{2\pi}\frac{1+(\rho\rho_1\rho_2)^2-2\rho\rho_1\rho_2\cos\nu}{1-(\rho\rho_1\rho_2)^2}\frac{1-(\rho\rho_2)^2}{1+(\rho\rho_2)^2-2\rho\rho_2\cos(\theta_1-\mu_1+\nu)}\frac{1-\rho_1^2}{1+\rho_1^2-2\rho_1\cos(\theta_1-\mu_1)},$$

or density of $EWC(\rho\rho_2,\rho_1,\mu_1,\nu)$. Similarly, the marginal density $f(\theta_2)=\int_{-\pi}^{\pi} f(\theta_1,\theta_2)d\theta_1$ is proven to be $EWC(\rho\rho_1,\rho_2,\mu_2,\pm\nu)$. Let (Θ_1,Θ_2) be the random vector of $f(\theta_1,\theta_2)$. Then it is apparent that $\Theta_1 \mid (\Theta_2=\theta_2) \sim EWC(\rho,\rho_1,\pm\theta_2+\nu,\mu_1)$ and $\Theta_2 \mid (\Theta_1=\theta_1) \sim EWC(\rho,\rho_2,\pm\theta_1\pm\nu,\mu_2)$.

The examples of the plots about the proposed density $f(\cdot,\cdot)$ are displayed in Figure 1. From this figure, it can be seen that the proposed distribution models a strong correlation. The parameter ρ controls the correlation between Θ_1 and Θ_2, and the parameters ρ_1 and ρ_2 contorol the concentration of Θ_1 and Θ_2. It is noted that small ρ also makes a strong correlation when both ρ_1 and ρ_2 are small as seen in Figure 1 (e).

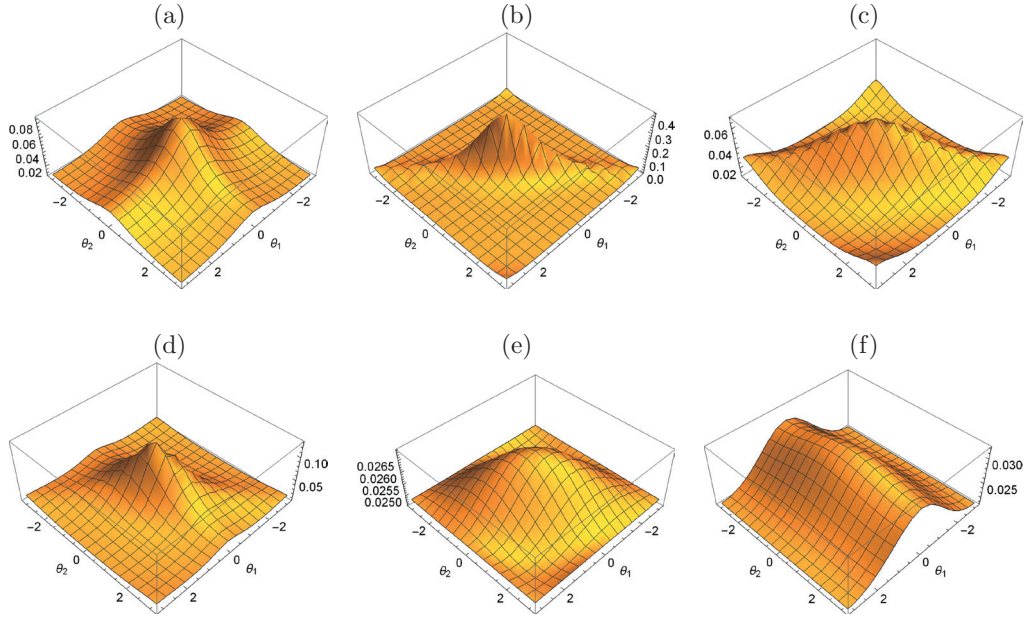


Figure 1: The plot of the proposed distribution with $\mu_1=\mu_2=\nu=0$, and (a) $\rho=0,\rho_1=\rho_2=0.3$, (b) $\rho=0.7,\rho_1=\rho_2=0.3$, (c) $\rho=0.3,\rho_1=\rho_2=0.1$, (d) $\rho=0.3,\rho_1=0.4,\rho_2=0.1$, (e) $\rho=0.01,\rho_1=\rho_2=0.01$, and (f) $\rho=0.01,\rho_1=0.1,\rho_2=0.01$.

3 Fitting example

To show the performance of the proposed distribution, this distribution is fitted to the directions of wind at 6:00 at two points ((Latitude, Longitude)= $(+29.768056^\circ, -95.220556^\circ)$ and (Latitude, Longitude)= $(+29.397778^\circ, -94.933333^\circ)$) in Texas from May 20, 2021, to July 31, 2003 which is taken from the Codiac data archive and is available a `https://data.eol.ucar.edu/dataset/85.034`. The sample size is 73. Since these two points are close to each other, the similar trends of the wind can be seen. The proposed distribution with fixing $\nu = 0$ is also fitted to this dataset

The results of the maximum likelihood estimation are shown in Table 1 and Figure 2. The proposed distribution without fixing ν gives very small estimates about ρ, ρ_1 and ρ_2, and shows a strong correlation while that with fixing ν gives moderate estimates about all parameters, and shows a strong correlation. In the sense of AIC and BIC, the case with fixing ν provides a better fitting. This might be seen from Figure 2.

Table 1: The maximum likelihood estimates of the proposed distribution (a) without fixing ν and (b) with fixing $\nu = 0$.

	ρ	ρ_1	ρ_2	μ_1	μ_2	ν	AIC	BIC
(a)	5.62×10^{-10}	1.58×10^{-25}	1.82×10^{-26}	2.23	2.91	6.10	548.66	562.40
(b)	0.71	0.52	7.17×10^{-5}	3.18	3.31	0 (fixed)	405.52	416.98

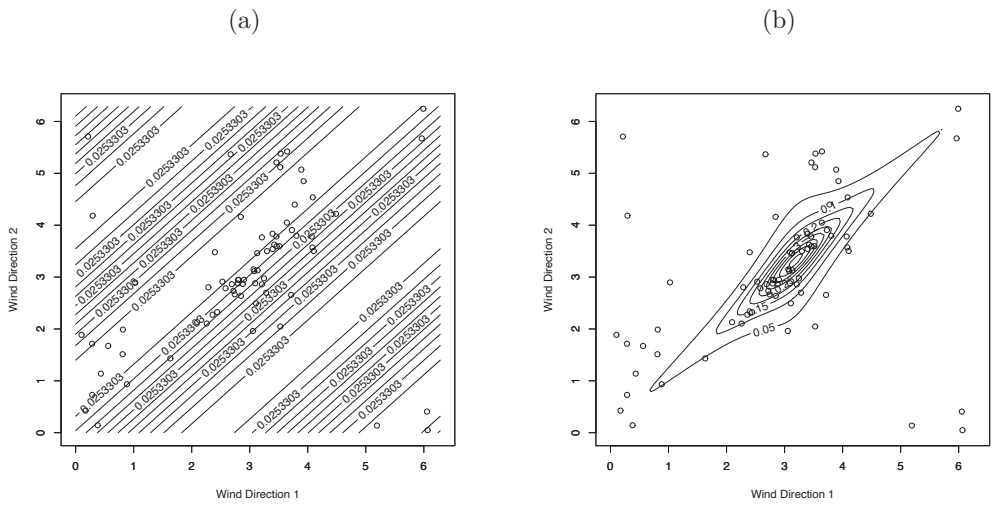


Figure 2: The contour plot of the fitted proposed distribution (a) without fixing ν and (b) with fixing $\nu = 0$.

4 Concluding remarks

In this report, a new toroidal distribution whose marginal and conditional are EWC is proposed. The distribution can model a strong correlation. However, the role of the parameters might be

difficult to interpret as explained in Section 2. More studies about the proposed distribution including the expressions of moments and Fisher information, random number generation, and multivariate extension will be valuable for the intepretation and left as future research.

Acknowledgement

Data provided by NCAR/EOL under the sponsorship of the National Science Foundation. `https://data.eol.ucar.edu/`

References

UCAR/NCAR – Earth Observing Laboratory. 2008. Texas Natural Resources Conservation Commission. Version 1.0. UCAR/NCAR – Earth Observing Laboratory. `https://data.eol.ucar.edu/dataset/85.034`.

Y. Baba. Statistics of angular data: Wrapped normal distribution model. *Proceedings of the Institute of Statistical Mathematics (in Japanese)*, 28:41–54, 1981.

R. A. Johnson and T. E. Wehrly. Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, 73:602–606, 1978.

S. Kato and M. C. Jones. An extended family of circular distributions related to wrapped Cauchy distributions via Brownian motion. *Bernoulli*, 19:154–171, 2013.

S. Kato and M. C. Jones. A tractable and interpretable four-parameter family of unimodal distributions on the circle. *Biometrika*, 102:181–190, 2015.

K. V. Mardia. Statistics of directional data. *Journal of the Royal Statistical Society:Series B*, 37:349–393, 1975.

A. Pewsey and E. Garcìa-Porugués. Rejoinder on: Recent advances in directional statistics. *Test*, 30:76–82, 2021.

M-Z. Wang and K. Shimizu. On applying Möbius transformation to cardioid random variables. *Statistical Methodology*, 9:604–614, 2012.

An independence property characterizing Kummer laws

Angelo Efoévi Koudou [1] & Jacek Wesołowski [2]

[1] *Université de Lorraine, CNRS, IECL, F-54000 Nancy, France, efoevi.koudou@univ-lorraine.fr*
[2] *Wydział Matematyki i Nauk Informacyjnych, Politechnika Warszawska, Warszawa, Poland, jacek.wesolowski@pw.edu.pl*

Abstract. We prove that if X, Y are positive, independent, non-Dirac random variables and if $\alpha, \beta \geq 0$, $\alpha \neq \beta$, then the random variables U and V defined by

$$U = Y\,\tfrac{1+\beta(X+Y)}{1+\alpha X+\beta Y} \quad \text{and} \quad V = X\,\tfrac{1+\alpha(X+Y)}{1+\alpha X+\beta Y}$$

are independent if and only if X and Y follow Kummer distributions with suitable parameters. In other words, the Kummer distributions are the only invariant measures for lattice recursion models introduced by Croydon and Sasada. This result, which fits in the literature related to independence properties of the Matsumoto-Yor type, extends earlier characterizations of Kummer and gamma laws by independence of

$$U = \tfrac{Y}{1+X} \quad \text{and} \quad V = X\left(1+\tfrac{Y}{1+X}\right),$$

which is the case of $(\alpha, \beta) = (1, 0)$. We also show that our Kummer independence property yields, as limiting cases, several well-known independence properties.

1 The Matsumoto-Yor property and related literature

Consider, for $b, c > 0$, the gamma distribution Gamma(b, c) with density proportional to

$$y^{b-1}e^{-cy}I_{(0,\infty)}(y),$$

for $p \in \mathbb{R}, a > 0, b > 0$, the generalized inverse Gaussian (GIG) distribution GIG(p, a, b) with density proportional to

$$x^{p-1}e^{-ax-b/x}I_{(0,\infty)}(x).$$

The Matsumoto-Yor property: for $p, a, b > 0$, given two independent, positive random variables X and Y such that $X \sim$ GIG$(-p, a, b)$ and $Y \sim$ Gamma(p, a), the random variables

$$U = \frac{1}{X+Y}, \quad V = \frac{1}{X} - \frac{1}{X+Y}$$

are independent (and follow the GIG$(-p, b, a)$ and Gamma(p, b) distribution, respectively).

This property was discovered by Matsumoto and Yor (2001) in the case $a = b$, while studying some functionals of exponential Brownian motion.

In Letac and Wesołowski (2000) it was noticed that it holds also when $a \neq b$ and it was proved that this independence property is in fact a characterization:
for two non-Dirac, positive and independent random variables X and Y, the random variables $\frac{1}{X+Y}$ and $\frac{1}{X} - \frac{1}{X+Y}$ are independent if and only if $X \sim \mathrm{GIG}(-p, a, b)$ while $Y \sim \mathrm{Gamma}(p, a)$ for some $p, a, b > 0$.

The detailed balance equation
Following Croydon and Sasada (2020), we say that a quadruplet of probability measures $(\mu, \nu, \tilde{\mu}, \tilde{\nu})$ on $\mathcal{U}, \mathcal{V}, \tilde{\mathcal{U}}, \tilde{\mathcal{V}}$, respectively, satisfy the *detailed balance equation* for a map $F : \mathcal{U} \times \mathcal{V} \to \tilde{\mathcal{U}} \times \tilde{\mathcal{V}}$ if

$$F(\mu \otimes \nu) = \tilde{\mu} \otimes \tilde{\nu},$$

where $F(\mu \otimes \nu)$ means $(\mu \otimes \nu) \circ F^{-1}$.

Using the terminology of Croydon & Sasada 2020, the Matsumoto-Yor property says that the quadruplet of probability measures
$\mu = \mathrm{GIG}(-p, a, b)$, $\nu = \mathrm{Gamma}(p, a)$,
$\tilde{\mu} = \mathrm{GIG}(-p, b, a)$, $\tilde{\nu} = \mathrm{Gamma}(p, b)$ satisfy the *detailed balance equation* for the map

$$F : (0, \infty)^2 \to (0, \infty)^2;\ (x, y) \overset{F}{\mapsto} \left(\frac{1}{x+y}, \frac{1}{x} - \frac{1}{x+y} \right).$$

In Koudou & Vallois 2012 we studied the question of finding decreasing and bijective functions $f : (0, \infty) \to (0, \infty)$ such that there exist a quadruplet of probability measures $(\mu, \nu, \tilde{\mu}, \tilde{\nu})$ on $(0, \infty)$ satifying the detailed balance equation for the map

$$T_f : (0, \infty)^2 \to (0, \infty)^2$$

$$(x, y) \mapsto (f(x + y),\ f(x) - f(x + y)).$$

This led, at the cost of some regularity assumptions, to other independence properties of Matsumoto-Yor type, amongst which was a property involving the Kummer distribution. For $a, c > 0$ and $b \in \mathbb{R}$, Kummer distribution $\mathcal{K}(a, b, c)$ is defined through the density proportional to

$$\frac{x^{a-1} e^{-cx}}{(1+x)^b} I_{(0,\infty)}(x).$$

More precisely, it was proved in K. & Vallois (2012) that if X and Y are independent Kummer and gamma with suitably related parameters then

$$U = X + Y \quad \text{and} \quad V = \frac{1 + (X+Y)^{-1}}{1 + X^{-1}}$$

are independent Kummer and beta random variables.

This was the starting point of a number of works on Matsumoto-Yor type characterizations of the Kummer distribution.

Firstly, starting from the latter property and looking for an involutive version of it, i.e. trying to find an involutive map $F : (X, Y) \mapsto (U, V)$ for which the Kummer distribution is

involved in a detailed balance equation, the following interesting property was discovered in Hamza and Vallois (2016): let X and Y be independent, X have the distribution $\mathcal{K}(a,b,c)$ and Y have the gamma distribution Gamma(b,c), then

$$U = \tfrac{Y}{1+X} \qquad \text{and} \qquad V = X\,\tfrac{1+X+Y}{1+X} \tag{1}$$

are independent, $U \sim \mathcal{K}(b,a,c)$ and $V \sim \text{Gamma}(a,c)$.

In Piliszec & Wesołowski (2018) this independence property was proved to give a characterization result with no assumption of existence of densities. Related characterizations were considered in Wesołowski (2015) and Piliszec & Wesołowski (2016). In Kołodziejek & Piliszek (2020), an extension to the matrix-variate case was established. In Piliszek (2022) a free probability version of the property and characterization was given. The latter needed a definition of a new distribution, a free analogue of the Kummer distribution.

A link between the detailed balance equation and a lattice recursion model as described in Croydon & Sasada (2020) :

They considered models assuming the following dynamics: For (n,t) in $\mathbb{Z}^2$, n is the spatial coordinate and t the temporal one.

For fixed $t \in \mathbb{Z}$,
$(x_n^t)_{n\in\mathbb{Z}} \in (0,\infty)^{\mathbb{Z}}$ is the configuration of the system at time t,
$(y_n^t)_{n\in\mathbb{Z}} \in (0,\infty)^{\mathbb{Z}}$ a collection of auxiliary variables through which the dynamics from t to $t+1$ are defined. Namely, (x_n^t, y_n^t) depends on (x_n^{t-1}, y_{n-1}^t) only,

$$(x_n^t, y_n^t) = G(x_n^{t-1}, y_{n-1}^t), \tag{2}$$

where for a bijection $F : \mathcal{X}\times\mathcal{Y} \to \tilde{\mathcal{X}}\times\tilde{\mathcal{Y}}$ either $G = F$, when $n+t$ is even or $G = F^{-1}$ when $n+t$ is odd.

The case when F is involutive is referred to as type I model, while the general case is referred to as type II model.

Let $x = (x_n)_{n\in\mathbb{Z}}$ be such that the above recursion with the initial condition $x_n^0 = x_n$, $n \in \mathbb{Z}$, has a unique solution $(x_n^t(x), y_n^t(x))_{n,t\in\mathbb{Z}}$. Let $\mathcal{X}^*$ denote set of all such x's.

According to Theorem 1.1 in Croydon and Sasada (2020), for a type I model, a sequence of iid random variables $X = (X_n)_{n\in\mathbb{Z}}$ with $X_1 \sim \mu$ satisfies $X \overset{d}{=} (x_n^1(X))_{n\in\mathbb{Z}}$ iff there exists a probability measure ν such that the pair (μ,ν) satisfies the detailed balance condition with respect to F. That is, $\mu\otimes\nu$ is the invariant measure for the recursion described above. In case of the type II model similar alternating invariance holds for pairs $\mu\otimes\nu$ and $\tilde{\mu}\otimes\tilde{\nu}$ depending on parity of $n+t$.

In Croydon & Sasada (2020) the authors identified four such models:

- ultra-discrete KdV (Korteweg-de Vries): type I model with $F(x,y) := (y-(x+y-J)_+ + (x+y-K)_+,\ x-(x+y-K)_+ + (x+y-J)_+)$ with μ and ν the shifted truncated exponential or shifted scaled truncated geometric laws;

- discrete KdV: type I model with

$$F(x,y) := F_{dK}^{(\alpha,\beta)}(x,y) = \left(\tfrac{y(1+\beta xy)}{1+\alpha xy},\ \tfrac{x(1+\alpha xy)}{1+\beta xy}\right)$$

  with μ the GIG law and ν the GIG (gamma) law which, when $\alpha\beta = 0$, has a direct connection with the Matsumoto-Yor property and related characterization of the GIG and gamma laws;

- ultra-discrete Toda: type II model with

$$F(x,y) := F_{udT^*}(x,y) = (x \wedge y,\ x - y)$$

  with μ, ν, $\tilde{\mu}$ the shifted exponential, $\tilde{\nu}$ asymmetric Laplace or μ, ν, $\tilde{\mu}$ shifted scaled geometric $\tilde{\nu}$ scaled discrete Laplace laws; this one is related to classical characterizations of the exponential and geometric distributions from Ferguson (1964) and Crawford (1966).

- discrete Toda: type II model with

$$F(x,y) := F_{dT^*}(x,y) = \left(x+y,\ \tfrac{x}{x+y}\right)$$

  with μ, ν, $\tilde{\mu}$ the gamma, $\tilde{\nu}$ the beta laws having a direct connection with the characterization of the gamma distribution given in Lukacs (1955).

For relations to box-ball systems and Pitman's transform one can consult Croydon, Sasada and Tsujimoto (2022) and Croydon, Kato, Sasada and Tsujimoto (2023). Recently, a connection between independence properties and Yang-Baxter equations holding for related transformations were discovered in Sasada & Uozumi (2022)

In the context of the discrete KdV model, Croydon and Sasada (2020) observed that if X and Y are independent, U and V, are independent and all four have GIG distributions with suitable parameters, then (X,Y) and (U,V) satisfy the detailed balance equation for the map $F_{dK}^{(\alpha,\beta)}$. They also conjectured that the GIG distributions are the only possible ones which let this $F_{dK}^{(\alpha,\beta)}$-detailed balance equation be satisfied.

Recently, in Letac & Wesołowski (2022) this conjecture was proved without the assumptions of existence and regularity of densities made by Bao & Noack in their proof of the same conjecture.
More precisely, Letac & Wesołowski (2022) established the following extension of the Matsumoto-Yor property: if A and B are non-degenerate, positive and independent random variables, and if α and β are two positive and distinct numbers, then the random variables

$$S = \tfrac{1}{B}\tfrac{\beta A+B}{\alpha A+B}, \quad T = \tfrac{1}{A}\tfrac{\beta A+B}{\alpha A+B}$$

are independent if and only if A and B have GIG distributions with suitable parameters.

Our work shows one more candidate for an invariant measure for a lattice recursion model. We derive the detailed balance equation for the Kummer distributions.

Specifically, we give a characterization of the Kummer laws, which is of a similar nature as in Letac & Wesołowski (2022) for the GIG laws, i.e. it says that the Kummer distributions are the only possible ones which let the detailed balance equation be satisfied for the map

$$F(x,y)=\left(y\,\tfrac{1+\beta(x+y)}{1+\alpha x+\beta y},\ x\,\tfrac{1+\alpha(x+y)}{1+\alpha x+\beta y}\right).$$

The proof uses a suitably designed "Laplace-type" transform and leads to a special second order linear differential equation for an unknown function of such form. In this sense the general methodology resembles that of the proof from Letac & Wesołowski (2022). However, at the technical level, the challenges were of quite a different nature. Interpreting this result in the context of the lattice system of recursions described above, it says that the Kummer distributions are the only relevant invariant measures for the type I model governed by the F defined above.

2 A new independence preserving property of Kummer laws

It will be convenient to introduce a scaled version of the Kummer distribution.

Let $\mathcal{K}_\alpha(a,b;c)$ for $\alpha\geq 0$, $a,c>0$ and $b\in\mathbb{R}$ be the probability distribution defined by the density

$$f(x)\propto \tfrac{x^{a-1}e^{-cx}}{(1+\alpha x)^b}\,I_{(0,\infty)}(x)$$

Remark: Note that $\mathcal{K}_0(a,b;c)=\mathcal{G}(a;c)$. Also $\mathcal{K}_\alpha(a,0;c)=\mathcal{G}(a;c)$.

Theorem 2.1 *(K., Wesołowski, 2023)*
Assume that

$$(X,Y)\sim\mathcal{K}_\alpha(a,b;c)\otimes\mathcal{K}_\beta(b,a;c)$$

for $a,b,c>0$ and $\alpha,\beta\geq 0$, $\alpha\neq\beta$. Let

$$U=Y\,\tfrac{1+\beta(X+Y)}{1+\alpha X+\beta Y}\quad\text{and}\quad V=X\,\tfrac{1+\alpha(X+Y)}{1+\alpha X+\beta Y}. \tag{3}$$

Then

$$(U,V)\sim\mathcal{K}_\alpha(b,a;c)\otimes\mathcal{K}_\beta(a,b;c).$$

Remark : Note that the above result gives a straightforward extension of one of the properties seen earlier. It suffices to take $(\alpha,\beta)=(1,0)$.

Theorem 2.2 *(K., Wesołowski, 2023)*
Let $\alpha,\beta\geq 0$, $\alpha\neq\beta$. Let X,Y be positive, independent, non-Dirac random variables and define

$$U=Y\,\tfrac{1+\beta(X+Y)}{1+\alpha X+\beta Y}\quad\text{and}\quad V=X\,\tfrac{1+\alpha(X+Y)}{1+\alpha X+\beta Y}.$$

If U and V are independent, then there exist $a, b, c > 0$ such that

$$(X, Y) \sim \mathcal{K}_\alpha(a, b; c) \otimes \mathcal{K}_\beta(b, a; c).$$

We then have

$$(U, V) \sim \mathcal{K}_\alpha(b, a; c) \otimes \mathcal{K}_\beta(a, b; c).$$

3 Recovering well-known independence properties from our result

We show that our Kummer independence property yields, as limiting cases, several well-known independence properties: the Lukacs property (Lukacs, 1955), the Kummer-Gamma property (K.& Vallois, 2012), the Matsumoto-Yor property (Matsumoto and Yor, 2001, Letac & Wesołowski, 2000) and the KdV property (Croydon and Sasada, 2020).

We will rely on the following version of Th. 5.5 from Billingsley (1968):

Theorem 3.1 *Let $X_n \xrightarrow{d} X$, with X_n and X assuming values in a separable metric space S. Let $\phi_n, \phi : S \to S$ be measurable functions such that for any $x \in S$ and any sequence $x_n \to x$ we have $\phi_n(x_n) \to \phi(x)$. Then*

$$\phi_n(X_n) \xrightarrow{d} \phi(X).$$

Proposition 3.2 *If $\alpha \to \infty$ then*

- *when $a > b$, $\mathcal{K}_\alpha(a, b, c) \xrightarrow{w} \text{Gamma}(a - b, c)$.*
- *when $b < a$, $\mathcal{K}_1(a, b, c/\alpha) \xrightarrow{w} \text{Beta}_{II}(a, b - a)$, where $\text{Beta}_{II}(p, q)$ law with $p, q > 0$ is defined by the density*
$$f(x) \propto \frac{x^{p-1}}{(1+x)^{p+q}} \mathbf{1}_{(0,\infty)}.$$
- *when $b > 0$, $\mathcal{K}_\alpha(\alpha b + a_1, \alpha b + a_2, c) \xrightarrow{w} \text{GIG}(a_1 - a_2, c, b)$.*
- *when $b > 0$, $\mathcal{K}_\alpha(a\sqrt{\alpha} + b, a\sqrt{\alpha}, c) \xrightarrow{w} \text{Gamma}(b, c)$.*
- *when $b > 0$, $\mathcal{K}_\alpha(a\alpha, a\alpha + b, c/\alpha) \xrightarrow{w} \text{InvGamma}(b, c)$, where $\text{InvGamma}(b, c)$ is defined by the density*
$$f(x) \propto x^{-b-1} e^{-c/x} \mathbf{1}_{(0,\infty)}(x).$$

3.1 The Lukacs property

Theorem 3.3 *Assume that $(X_1, Y_1) \sim \text{Gamma}(a_1, c) \otimes \text{Gamma}(b_1, c)$. Let*

$$(U_1, V_1) = \left(\frac{Y_1}{X_1}, X_1 + Y_1 \right).$$

Then

$$(U_1, V_1) \sim \text{Beta}_{II}(b_1, a_1) \otimes \text{Gamma}(a_1 + b_1, c).$$

Proof : We use our theorem with $\alpha = n$, $\beta = 0$ to see that

$$(X_1^{(n)}, Y_1) \sim \mathcal{K}_n(a_1 + b_1, b_1, c) \otimes \text{Gamma}(b_1, c)$$

implies that for

$$\phi_n(x, y) = \left(\tfrac{ny}{1+nx},\ x\tfrac{n(x+y)}{1+nx}\right)$$

we have

$$\phi_n(X_1^{(n)}, Y_1) \sim \mathcal{K}_1(b_1, a_1 + b_1, c/n) \otimes \text{Gamma}(\text{a}_1 + \text{b}_1, \text{c})$$

and we take the limit as $n \to \infty$.

3.2 The Kummer-Gamma independence property

Theorem 3.4 *Assume that* $(X_2, Y_2) \sim \mathcal{K}(a_2, a_2 + b_2, c_2) \otimes \text{Gamma}(b_2, c_2)$. *Let*

$$(U_2, V_2) = \left(X_2 + Y_2,\ \frac{1+\frac{1}{X_2+Y_2}}{1+\frac{1}{X_2}}\right).$$

Then

$$(U_2, V_2) \sim \mathcal{K}(a_2 + b_2, a_2, c_2) \otimes \text{Beta}_I(a_2, b_2),$$

where $\text{Beta}_I(p, q)$ *has the density* $f(y) \propto y^{p-1}(1-y)^{q-1}\mathbf{1}_{(0,1)}(y)$.

Proof : We use our theorem with $\alpha = 1$ and $\beta = n$ to see that

$$(X_2,\ Y_2^{(n)}) \sim \mathcal{K}_1(a_2, a_2 + b_2, c_2) \otimes \mathcal{K}_n(a_2 + b_2, a_2, c_2)$$

implies that for

$$\widetilde{\phi}_n(x, y) = \left(y\,\tfrac{1+n(x+y)}{1+x+ny},\ nx\,\tfrac{1+x+y}{1+x+ny}\right)$$

we have

$$\widetilde{\phi}_n(X_2, Y_2^{(n)}) \sim \mathcal{K}_1(a_2 + b_2, a_2, c_2) \otimes \mathcal{K}_1(a_2, a_2 + b_2, c_2/n).$$

3.3 The Matsumoto-Yor property

Theorem 3.5 *Assume that* $(X_3, Y_3) \sim \text{GIG}(-a_3, b_3, c_3) \otimes \text{Gamma}(a_3, b_3)$. *Let*

$$(U_3, V_3) = \left(\tfrac{1}{X_3+Y_3},\ \tfrac{1}{X_3} - \tfrac{1}{X_3+Y_3}\right).$$

Then

$$(U_3, V_3) \sim \text{GIG}(-a_3, c_3, b_3) \otimes \text{Gamma}(a_3, c_3).$$

Proof : We use our theorem with $\alpha = n$ and $\beta = n^2$ to see that if

$$\left(X_3^{(n)},\ Y_3^{(n)}\right) \sim \mathcal{K}_n(nc_3, nc_3 + a_3, b_3) \otimes \mathcal{K}_{n^2}(nc_3 + a_3, nc_3, b_3)$$

then with

$$\widetilde{\phi}_n(x, y) = \left(y\tfrac{1+n^2(x+y)}{1+nx+n^2y},\ nx\tfrac{1+n(x+y)}{1+nx+n^2y}\right)$$

we have

$$\widetilde{\phi}_n\left(X_3^{(n)},\ Y_3^{(n)}\right) \sim \mathcal{K}_n(nc_3 + a_3, nc_3, b_3) \otimes \mathcal{K}_n(nc_3, nc_3 + a_3, b_3/n).$$

3.4 The discrete KdV independence property

Theorem: Assume that

$$(X_4, Y_4) \sim \mathrm{GIG}(-a_4, \alpha b_4, c_4) \otimes \mathrm{GIG}(-a_4, \beta c_4, b_4).$$

Let $(U_4, V_4) = \left(Y_4 \frac{1+\alpha X_4 Y_4}{1+\beta X_4 Y_4},\ X_4 \frac{1+\beta X_4 Y_4}{1+\alpha X_4 Y_4} \right)$. Then

$$(U_4, V_4) \sim \mathrm{GIG}(-a_4, \alpha c_4, b_4) \otimes \mathrm{GIG}(-a_4, \beta b_4, c_4).$$

Proof: We use our theorem with α changed into n/α and β changed into n/β we see that if

$$\left(X_4^{(n)},\ Y_4^{(n)} \right) \sim \mathcal{K}_{n/\alpha}(nb_4 + a_4, nb_4, c_4) \otimes \mathcal{K}_{n/\beta}(nb_4, nb_4 + a_4, c_4)$$

then with

$$\widetilde{\phi}_n(x, y) = \left(y \, \frac{1+\frac{n}{\beta}(x+y)}{1+n\left(\frac{x}{\alpha}+\frac{y}{\beta}\right)},\ x \, \frac{1+\frac{n}{\alpha}(x+y)}{1+n\left(\frac{x}{\alpha}+\frac{y}{\beta}\right)} \right),$$

$$\widetilde{\phi}_n \left(X_4^{(n)},\ Y_4^{(n)} \right) \sim \mathcal{K}_{n/\alpha}(nb_4, nb_4 + a_4, c_4) \otimes \mathcal{K}_{n/\beta}(nb_4 + a_4, nb_4, c_4).$$

4 Sketch of proof of the main result

For a positive random variable W and $\gamma \geq 0$ consider an extended Laplace transform $L_W^{(\gamma)}$ of the form

$$L_W^{(\gamma)}(s, t, z) = \mathbb{E} \, \tfrac{W^s}{(1+\gamma W)^t} \, e^{-zW}.$$

We will call it the Kummer transform. Note that the Kummer tranform is well defined at least for $s, z > 0$ and $t \in \mathbb{R}$. Moreover, for any fixed $s > 0$, $t \in \mathbb{R}$, the Kummer transform as a function of $z > 0$, is just the Laplace transform of the measure $\frac{w^s}{(1+\gamma w)^t} \mathbb{P}_W(dw)$, so it uniquely determines the distribution of W.

Note also that

$$L_W^{(\gamma)}(s, t, z) + \gamma L_W^{(\gamma)}(s+1, t, z) = L_W^{(\gamma)}(s, t-1, z) \tag{4}$$

and for any $k = 1, 2, \ldots$

$$\tfrac{\partial^k \, L_W^{(\gamma)}(s,t,z)}{\partial z^k} = (-1)^k L_W^{(\gamma)}(s+k, t, z). \tag{5}$$

Proposition 4.1 *Let $X \sim \mathcal{K}_\alpha(a, b, c)$, $a, c > 0$, $b \in \mathbb{R}$. Then*

$$L_X^{(\alpha)}(s, t, z) = \frac{\Gamma(a+s) U\left(a+s, a+s-b-t+1, \frac{c+z}{\alpha}\right)}{\alpha^s \Gamma(a) U\left(a, a-b+1, \frac{c}{\alpha}\right)}, \tag{6}$$

$s > 0$, $t \in \mathbb{R}$, $z > -c$, where U is the Kummer function (see 13.2.5 in Abramowitz & Stegund, 1984) defined by

$$U(a, b, z) = \tfrac{1}{\Gamma(a)} \int_0^\infty \tfrac{x^{a-1}}{(1+x)^{a-b+1}} \, e^{-zx} \, dx, \qquad a, z > 0, \ b \in \mathbb{R}. \tag{7}$$

Proposition 4.2 *Let $b \in \mathbb{R}$, $a, c, \alpha > 0$. Assume that for some fixed $(s,t) \in (0,\infty) \times \mathbb{R}$ and all $z > 0$*

$$L_X^{(\alpha)}(s,t,z) = k(s,t) U\left(a+s, a+s-b-t+1, \tfrac{c+z}{\alpha}\right), \tag{8}$$

where $k(s,t)$ is a constant (depending also on α, a, b, c) Then $X \sim \mathcal{K}_\alpha(a,b,c)$.

Denote

$$\psi(x,y) = \left(y\,\tfrac{1+\beta(x+y)}{1+\alpha x+\beta y},\, x\,\tfrac{1+\alpha(x+y)}{1+\alpha x+\beta y}\right) =: (u,v), \qquad x,y>0.$$

Note that $\psi : (0,\infty)^2 \to (0,\infty)^2$ is an involution. Moreover, we have the following identities:

$$x + y = u + v, \tag{9}$$

$$\tfrac{x}{1+\beta y} = \tfrac{v}{1+\alpha u}, \tag{10}$$

$$\tfrac{y}{1+\alpha x} = \tfrac{u}{1+\beta v}. \tag{11}$$

Using these identities we have the following:

Lemma 4.3 *Let X and Y be independent. Assume also that U and V as defined in* (3) *are also independent. In view of* (9), (10) *and* (11) *we then have*

$$L_X^{(\alpha)}(s,t,z)\, L_Y^{(\beta)}(t,s,z) = L_U^{(\alpha)}(t,s,z)\, L_V^{(\beta)}(s,t,z), \quad (s,t,z) \in (0,\infty)\times\mathbb{R}\times(0,\infty). \tag{12}$$

Differentiating the above equation with respect to z and dividing side-wise by the same equation we get

$$\tfrac{L_X^{(\alpha)}(s+1,t,z)}{L_X^{(\alpha)}(s,t,z)} + \tfrac{L_Y^{(\beta)}(t+1,s,z)}{L_Y^{(\beta)}(t,s,z)} = \tfrac{L_U^{(\alpha)}(t+1,s,z)}{L_U^{(\alpha)}(t,s,z)} + \tfrac{L_V^{(\beta)}(s+1,t,z)}{L_Y^{(\beta)}(s,t,z)}. \tag{13}$$

Using identity (4) we obtain

$$\beta\,\tfrac{L_X^{(\alpha)}(s,t-1,z)}{L_X^{(\alpha)}(s,t,z)} + \alpha\,\tfrac{L_Y^{(\beta)}(t,s-1,z)}{L_Y^{(\beta)}(t,s,z)} = \beta\,\tfrac{L_U^{(\alpha)}(t,s-1,z)}{L_U^{(\alpha)}(t,s,z)} + \alpha\,\tfrac{L_V^{(\beta)}(s,t-1,z)}{L_Y^{(\beta)}(s,t,z)}.$$

After some calculations we get

$$\beta\, M_X(s,t,z) + \alpha\, M_Y(t,s,z) = \beta\, M_U(t,s,z) + \alpha\, M_V(s,t,z), \tag{14}$$

where

$$M_W(s,t,z) = \tfrac{L_W^{(\gamma)}(s+1,t,z) L_W^{(\gamma)}(s,t+1,z)}{L_W^{(\gamma)}(s,t,z) L_W^{(\gamma)}(s+1,t+1,z)}.$$

Note also that (12) implies

$$M_X(s,t,z) M_Y(t,s,z) = M_U(t,s,z) M_V(s,t,z). \tag{15}$$

Combining (14) with (15) we get

$$(\beta\, M_X(s,t,z) - \alpha\, M_V(s,t,z))\,(M_X(s,t,z) - M_U(t,s,z)) = 0 \tag{16}$$

$$(\beta\, M_U(t,s,z) - \alpha\, M_Y(t,s,z))\,(M_V(s,t,z) - M_Y(t,s,z)) = 0. \tag{17}$$

It follows from (16) that either $\beta\, M_X \equiv \alpha\, M_V$ or $M_X \equiv M_U$ and from (17) that either $\beta\, M_U \equiv \alpha\, M_Y$ or $M_V \equiv M_Y$.

- We prove that $\beta\, M_X \equiv \alpha\, M_V$ is impossible. It will follow by symmetry that also $\beta\, M_U \equiv \alpha\, M_Y$ is impossible.
- Then, we treat the case $M_X \equiv M_U$. The case $M_V \equiv M_Y$ will follow by the analogous approach.

We consider the equation

$$M_X(s,t,z) = M_U(t,s,z), \quad s,t \in \{0,1,\ldots\},\ z>0. \tag{18}$$

Denote

$$A(s,t,z) := \tfrac{L_X(s+1,t)L_U(t,s)}{L_X(s,t)L_U(t,s+1)}, \tag{19}$$

and

$$B(t,s,z) := \tfrac{L_U(t+1,s)L_X(s,t)}{L_U(t,s)L_X(s,t+1)}, \tag{20}$$

where we skipped the superscript (α) and the argument z in L_X and L_U.

Note that (18) implies that for all $s,t \in \mathbb{N} = \{0,1,\ldots\}$ we have

$$A(s,t,z) = A(s,t+1,z) \quad \text{and} \quad B(t,s,z) = B(t,s+1,z).$$

Consequently, for $(s,t) \in \mathbb{N}^2$ we have $A(s,t,z) = A(s,0,z) =: A(s,z)$ and $B(t,s,z) = B(t,0,z) =: B(t,z)$.

Now (18) can be written as

$$A(s,z)\,\tfrac{L_X(s,t)}{L_X(s+1,t+1)} = B(t,z)\tfrac{L_U(t,s)}{L_U(t+1,s+1)}.$$

Consider now the quotient

$$\tfrac{A(s,z)}{B(t,z)} = \tfrac{L_X(s+1,t+1)L_U(t,s)}{L_X(s,t)L_U(t+1,s+1)}.$$

We differentiate this quotient with respect to z, and after some further calculations we obtain that thIs derivative is equal to 0, and thus $\frac{A(s,z)}{B(t,z)} = \frac{a(s)}{b(t)}$, where $a(s) := A(s,0)$ and $b(t) := B(t,0)$. Consequently, we have the representations:

$$A(s,z) = f(z)a(s) \quad \text{and} \quad B(t,z) = f(z)b(t), \quad z>0,\ s,t \in \mathbb{N}, \tag{21}$$

where $f = \frac{A(0,z)}{a(0)} = \frac{B(0,z)}{b(0)}$.

We thus have

$$a(s)\,\tfrac{L_X(s,t,z)}{L_X(s+1,t+1,z)} = b(t)\tfrac{L_U(t,s,z)}{L_U(t+1,s+1,z)}. \tag{22}$$

The above equation transforms to the second order diferential equation for the function $h := L_U(t,s+1)$ as follows

$$\alpha(c+z)\,h''(z) + (\alpha b(t) - a(s)) - (c+z))\,h'(z) - b(t)\,h(z) = 0.$$

Consequently, for g defined by $g(z) = h(\alpha z - c)$ we get the Kummer equation

$$zg''(z) + (b(t) - a(s) - z)g'(z) - b(t)g(z) = 0. \tag{23}$$

It is well known that the general solution is of the form

$$g(z) = c_1 M(b(t), b(t) - a(s), z) + c_2 U(b(t), b(t) - a(s), z),$$

where (see 13.2.1 in [**?**]), for $b > a$,

$$M(a, b, z) = \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)} \int_0^1 \frac{t^{a-1}}{(1-t)^{a-b+1}} e^{zt}\, dt.$$

Recall that $M(a, b, z)$ is unbounded when $z \to \infty$ (see e.g. 13.1.4 in [**?**]) and $U(a, b, z) \to 0$ as $z \to \infty$. Since g, as a Laplace transform of a probability measure, is bounded, we necessarily have

$$g(z) = c_U(s, t) U(b(t), b(t) - a(s), z).$$

Returning to $L_U(t, s)$ (recall that g was defined through $L_U(t, s + 1)$) we get

$$L_U(t, s, z) = c_U(s, t) U(b + t, b + t - a - s + 1, \tfrac{c+z}{\alpha}),$$

with $a, b > 0$ and $c \geq 0$.

Changing the roles of L_X and L_U in the above argument we obtain

$$L_X(s, t, z) = c_X(s, t) U(a + s, a + s - b - t + 1, \tfrac{c+z}{\alpha}).$$

A technical argument yields $c > 0$ and Proposition 4.2 implies that $X \sim \mathcal{K}_\alpha(a, b, c)$ and $U \sim \mathcal{K}_\alpha(b, a, c)$. The sequel of the prove is technical and we refer to Koudou & Wesołowski (2023) for details.

References

Abramowitz, M. and Stegund, I. A. *Pocketbook of Mathematical functions.* Verlag Harri Deutsch-Thun-Frankfurt/Main, 1984.

Bao, K. V. and Noack, C. Characterizations of the generalized inverse Gaussian, asymmetric Laplace, and shifted (truncated) exponential laws via independence properties. *arXiv* **2107.01394** (2021), 1–12.

Croydon, D. A. and Sasada, M. Detailed balance and invariant mesures for discrete KdV- and Toda-type systems. *arXiv* **2007.06203** (2020), 1–49.

Hamza, M. and Vallois, P. On Kummer's distribution of type two and a generalized beta distribution. *Statist. Probab. Lett.* **118** (2016), 60-69.

Kołodziejek, B. and Piliszek, A. Independence characterization for Wishart and Kummer random matrices. *REVSTAT - Statist. J.* **18(3)** (2020), 357-373.

Koudou, A.E. and Vallois, P. Independence properties of the Matsumoto-Yor type. *Bernoulli* **18(1)** (2012), 119–136.

Koudou, A.E. and Wesołowski, J. (2023) Independence preserving property of Kummer laws. *Bernoulli*, to appear. arXiv.2212.03150 [math.PR]

Letac, G. and Wesołowski, J. An independence property for the GIG and gamma laws. *Ann. Probab.* **28(3)** (2000), 1371–1383.

Letac, G. and Wesołowski, J. About an extension of the Matsumoto-Yor property. *arXiv* **2203.05404** (2022), 1–17.

Lukacs, E. A characterization of the gamma distribution. *Ann. Statist. Math.* **26(2)** (1955), 319-324.

Matsumoto, H. and Yor, M. An analogue of Pitman's $2M-X$ theorem for exponential Wiener functionals. Part II: the role of the generalized inverse Gaussian laws. *Nagoya, Math. J.* **162** (2001), 65-86.

Piliszek, A. Regression conditions that characterize free-Poisson and free-Kummer distributions. *Rand. Matr. Theory Appl.* **11(2)** (2022), 2250019.

Piliszek, A. and Wesołowski, J. Kummer and gamma laws through independencies on trees - another parallel with the Matsumoto-Yor property. *J. Multivar. Anal.* **152** (2016) 15-27.

Piliszek, A. and Wesołowski, J. Change of measure technique in characterizations of the Kummer and gamma laws. *J. Math. Anal. Appl.* **458** (2018), 967-979.

Wesołowski, J. On the Matsumoto-Yor type regression characterization of the gamma and Kummer distributions. *Statist. Probab. Lett.* **107** (2015), 145-149.

Acceleration of the EM algorithm

Masahiro Kuroda

Okayama University of Science

1 Introduction

The expectation-maximization (EM) algorithm of Dempster et al. (1977) is a well-known iterative algorithm for finding maximum likelihood estimates from incomplete data and is used in several statistical models with latent variables and missing data. The algorithm also monotonically increases a likelihood function at each iteration and satisfies parameter constraints for convergence The popularity of the EM algorithm stems from its stable convergence, simple implementation and flexibility in interpreting data incompleteness. Despite these computational advantages, the algorithm is linear convergent, and its speed of convergence is very slow when a statistical model has many parameters, and the proportion of missing data is high.

Kuroda and Sakakihara (2006) proposed the ε-accelerated EM algorithm that accelerates the convergence of the sequence of EM iterations using the vector ε algorithm of Wynn (1962). Wang et al. (2008) proved the theoretical properties of the convergence and acceleration of the ε-accelerated EM algorithm. The merit of the ε-accelerated EM algorithm is that it requires only the sequence of EM iterations but it does not require estimates of information matrices or convergence rates during iterations.

In this paper, we introduce the ε-accelerated EM algorithm. Section 2 presents the EM algorithm. Section 3 describes the vector ε algorithm. In Section 4, we show the ε-accelerated EM algorithm.

2 The EM algorithm

Let $\mathbf{y}$ denote the incompletely observed data in a sample space $\Omega_{\mathbf{y}}$ and $\mathbf{x}$ denote the complete data augmented from $\mathbf{y}$ in a sample space $\Omega_{\mathbf{x}}$. Assume that $\mathbf{y}$ is missing at random. Let $f(\cdot|\boldsymbol{\theta})$ denote a density function with an unknown p-dimensional parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$ in a parameter space $\Omega_{\boldsymbol{\theta}} \subset \mathbb{R}^p$. Let $\ell_o(\boldsymbol{\theta}) = \ln f(\mathbf{y}|\boldsymbol{\theta})$ and $\ell_c(\boldsymbol{\theta}) = \ln f(\mathbf{x}|\boldsymbol{\theta})$ denote the log-likelihood functions for $\mathbf{y}$ and $\mathbf{x}$, respectively. Then, we have

$$\ell_o(\boldsymbol{\theta}) = \ell_c(\boldsymbol{\theta}) - \ln f(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}),$$

where $f(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}) = f(\mathbf{x}|\boldsymbol{\theta})/f(\mathbf{y}|\boldsymbol{\theta})$ is the conditional density function of $\mathbf{x}$ given $\mathbf{y}$. We define the Q function that is the conditional expectation of $\ell_c(\boldsymbol{\theta})$ given $\mathbf{y}$ and $\boldsymbol{\theta}'(\in \Omega_{\boldsymbol{\theta}})$ as

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}') = \mathrm{E}[\ell_c(\boldsymbol{\theta})|\mathbf{y}, \boldsymbol{\theta}'].$$

To seek the maximum of $\ell_o(\boldsymbol{\theta})$, the EM algorithm solves

$$\frac{\partial Q(\boldsymbol{\theta}|\boldsymbol{\theta}')}{\partial \boldsymbol{\theta}} = \mathbf{0}. \tag{1}$$

The EM algorithm finds a stationary point $\boldsymbol{\theta}^*$ of Equation (1) by alternately iterating between the expectation-step (E-step) and the maximization-step (M-step).

Let $\boldsymbol{\theta}^{(t)}$ denote the estimate of $\boldsymbol{\theta}$ at the t-th iteration of the EM algorithm. The E-step calculates $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ to obtain the conditional expectation of the missing data given $\mathbf{y}$ and $\boldsymbol{\theta}^{(t)}$. The M-step finds $\boldsymbol{\theta}^{(t+1)}$ by maximizing $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ with respect to $\boldsymbol{\theta}$. We describe the EM algorithm in Algorithm 1.

Algorithm 1 The EM algorithm

E-step: Calculate $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathrm{E}[\ell_c(\boldsymbol{\theta})|\mathbf{y}, \boldsymbol{\theta}^{(t)}]$.
M-step: Find $\boldsymbol{\theta}^{(t+1)}$ to be a value of $\boldsymbol{\theta} \in \Omega_{\boldsymbol{\theta}}$ which maximizes $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$:

$$\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta} \in \Omega_{\boldsymbol{\theta}}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}).$$

Usually, the iteration of the algorithm is terminated when the following criterion holds:

$$\ell_o(\boldsymbol{\theta}^{(t+1)}) - \ell_o(\boldsymbol{\theta}^{(t)}) < \delta \quad \text{or} \quad \|\boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^{(t)}\|^2 < \delta,$$

where $\|\cdot\|$ is the Euclidean norm and δ is the desired accuracy.

Dempster et al. (1977) showed that $\ell_o(\boldsymbol{\theta}^{(t)})$ is increased after an EM iteration, i.e.,

$$\ell_o(\boldsymbol{\theta}^{(t+1)}) \geq \ell_o(\boldsymbol{\theta}^{(t)})$$

for $t = 0, 1, \ldots$. Thus, if $\{\ell_o(\boldsymbol{\theta}^{(t)})\}_{t \geq 0}$ is bounded above, the convergence of the EM algorithm is obtained. Dempster et al. (1977) and Wu (1983) provided the convergence theorems for the sequences $\{\ell_o(\boldsymbol{\theta}^{(t)})\}_{t \geq 0}$ and $\{\boldsymbol{\theta}^{(t)}\}_{t \geq 0}$, respectively.

The EM algorithm implicitly defines a mapping M from $\Omega_{\boldsymbol{\theta}}$ to $\Omega_{\boldsymbol{\theta}}$ such that

$$\boldsymbol{\theta}^{(t+1)} = M(\boldsymbol{\theta}^{(t)}) \tag{2}$$

for $t = 0, 1, \ldots$. Suppose that $\{\boldsymbol{\theta}^{(t)}\}_{t \geq 0}$ converges to $\boldsymbol{\theta}^*$, and $M(\boldsymbol{\theta})$ is differentiable at $\boldsymbol{\theta}^*$. Then, $\boldsymbol{\theta}^*$ is a fixed point of the algorithm, i.e., $\boldsymbol{\theta}^* = M(\boldsymbol{\theta}^*)$. The Taylor series expansion of Equation (2) at $\boldsymbol{\theta}^*$ yields

$$\boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^* = DM(\boldsymbol{\theta}^*)(\boldsymbol{\theta}^{(t)} - \boldsymbol{\theta}^*) + O(||\boldsymbol{\theta}^{(t)} - \boldsymbol{\theta}^*||^2), \tag{3}$$

where

$$DM(\boldsymbol{\theta}) = \left(\frac{\partial M_i(\boldsymbol{\theta})}{\partial \theta_j} \right)$$

is the Jacobian matrix for the mapping $M(\boldsymbol{\theta}) = (M_1(\boldsymbol{\theta}), \ldots, M_p(\boldsymbol{\theta}))^\top$. Thus, in the neighborhood of $\boldsymbol{\theta}^*$, the EM algorithm is essentially a linear iteration with the iteration matrix $DM(\boldsymbol{\theta}^*)$. For a sufficiently large t, $\boldsymbol{\theta}^{(t)}$ tends to be very close to $\boldsymbol{\theta}^*$, and then, Equation (3) becomes

$$\boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^* = \lambda(\boldsymbol{\theta}^{(t)} - \boldsymbol{\theta}^*) + O(||\boldsymbol{\theta}^{(t)} - \boldsymbol{\theta}^*||^2), \tag{4}$$

where $\lambda \in [0, 1)$ is the largest eigenvalue of $DM(\boldsymbol{\theta}^*)$ (Schafer, 1997). The EM algorithm converges slowly for a large value of λ. Dempster et al. (1977) showed that the global rate of convergence of the EM algorithm is governed by the largest eigenvalue of $DM(\boldsymbol{\theta}^*)$.

3 The vector ε algorithm

The vector ε (vε) algorithm presented by Wynn (1962) is an extrapolation method and is utilized to accelerate the convergence of a slowly convergent scalar sequence. It is known that the algorithm is very effective for linearly converging sequences.

Assume that $\{\boldsymbol{\theta}^{(t)}\}_{t\geq 0}$ is a vector sequence generated from an iterative algorithm and converges to $\boldsymbol{\theta}^*$. We define the inverse $[\boldsymbol{\theta}^{(t)}]^{-1}$ of a vector $\boldsymbol{\theta}^{(t)}$ as

$$[\boldsymbol{\theta}^{(t)}]^{-1} = \frac{\boldsymbol{\theta}^{(t)}}{\|\boldsymbol{\theta}^{(t)}\|}.$$

The vε algorithm for a sequence $\{\boldsymbol{\theta}^{(t)}\}_{t\geq 0}$ starts with

$$\boldsymbol{\varepsilon}^{(t,-1)} = 0, \qquad \boldsymbol{\varepsilon}^{(t,0)} = \boldsymbol{\theta}^{(t)}$$

and generates a vector $\boldsymbol{\varepsilon}^{(t,k+1)}$ by

$$\boldsymbol{\varepsilon}^{(t,k+1)} = \boldsymbol{\varepsilon}^{(t+1,k-1)} + \left[\boldsymbol{\varepsilon}^{(t+1,k)} - \boldsymbol{\varepsilon}^{(t,k)}\right]^{-1}. \tag{5}$$

We apply the vε algorithm for $k = 1$ to accelerate the convergence of $\{\boldsymbol{\theta}^{(t)}\}_{t\geq 0}$. From Equation (5), we have

$$\begin{aligned}
\boldsymbol{\varepsilon}^{(t,1)} &= \boldsymbol{\varepsilon}^{(t+1,-1)} + \left[\boldsymbol{\varepsilon}^{(t+1,0)} - \boldsymbol{\varepsilon}^{(t,0)}\right]^{-1} = \left[\boldsymbol{\varepsilon}^{(t+1,0)} - \boldsymbol{\varepsilon}^{(t,0)}\right]^{-1}, \\
\boldsymbol{\varepsilon}^{(t,2)} &= \boldsymbol{\varepsilon}^{(t+1,1)} + \left[\boldsymbol{\varepsilon}^{(t+1,1)} - \boldsymbol{\varepsilon}^{(t,1)}\right]^{-1}.
\end{aligned}$$

Since

$$\boldsymbol{\varepsilon}^{(t,1)} = \left[\boldsymbol{\varepsilon}^{(t+1,0)} - \boldsymbol{\varepsilon}^{(t,0)}\right]^{-1}, \qquad \boldsymbol{\varepsilon}^{(t+1,1)} = \left[\boldsymbol{\varepsilon}^{(t+2,0)} - \boldsymbol{\varepsilon}^{(t+1,0)}\right]^{-1},$$

the vε algorithm is given by

$$\begin{aligned}
\boldsymbol{\varepsilon}^{(t,2)} &= \boldsymbol{\varepsilon}^{(t+1,0)} + \left[\left[\boldsymbol{\varepsilon}^{(t+2,0)} - \boldsymbol{\varepsilon}^{(t+1,0)}\right]^{-1} - \left[\boldsymbol{\varepsilon}^{(t+1,0)} - \boldsymbol{\varepsilon}^{(t,0)}\right]^{-1}\right]^{-1} \\
&= \boldsymbol{\theta}^{(t)} + \left[\left[\boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^{(t)}\right]^{-1} - \left[\boldsymbol{\theta}^{(t)} - \boldsymbol{\theta}^{(t-1)}\right]^{-1}\right]^{-1}.
\end{aligned} \tag{6}$$

Then, the sequence $\{\boldsymbol{\varepsilon}^{(t,2)}\}_{t\geq 0}$ converges to $\boldsymbol{\theta}^*$ faster than $\{\boldsymbol{\theta}^{(t)}\}_{t\geq 0}$.

4 The vector ε acceleration of the EM algorithm

Kuroda and Sakakihara (2006) proposed the ε-accelerated EM algorithm that can accelerate the convergence of the EM algorithm by using the vε algorithm. The algorithm combines the vε acceleration (6) with the EM algorithm and generates a faster convergent sequence than $\{\boldsymbol{\theta}^{(t)}\}_{t\geq 0}$. We describe the ε-accelerated EM algorithm in Algorithm 2.

The ε-acceleration step is added to accelerate the convergence of the sequence using the vector ε accelerator and does not depend on statistical models. Thus, the ε-accelerated EM algorithm accelerates the convergence without affecting its simplicity and stability.

Wang et al. (2008) provided theoretical results of the ε-accelerated EM algorithm.

Algorithm 2 The ε-accelerated EM algorithm

EM step: Find $\boldsymbol{\theta}^{(t+1)} = M(\boldsymbol{\theta}^{(t)})$.
ε-acceleration step: Compute $\boldsymbol{\psi}^{(t-1)}$ from $(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(t+1)})$ using

$$\boldsymbol{\psi}^{(t-1)} = \boldsymbol{\theta}^{(t)} + \left[\left[\boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^{(t)} \right]^{-1} - \left[\boldsymbol{\theta}^{(t)} - \boldsymbol{\theta}^{(t-1)} \right]^{-1} \right]^{-1}.$$

Repeat the two steps alternately until $\|\boldsymbol{\psi}^{(t-1)} - \boldsymbol{\psi}^{(t-2)}\|^2 < \delta$.

Theorem 1 *Suppose that the sequence of EM iterates $\{\boldsymbol{\theta}^{(t)}\}_{t\geq 0}$ converges to a stationary point $\boldsymbol{\theta}^*$. The sequence $\{\boldsymbol{\psi}^{(t)}\}_{t\geq 0}$ generated by the ε-accelerated EM algorithm converges to a stationary point $\boldsymbol{\theta}^*$ of the EM algorithm.*

We evaluate the speed of convergence of the ε-accelerated EM algorithm. For the parameter vector $\boldsymbol{\theta}$, the iterative procedure $\{\boldsymbol{\theta}^{(t)}\}_{t\geq 0}$ is said to converge linearly if

$$c = \lim_{t\to\infty} \frac{||\boldsymbol{\theta}^{(t+1)} - \boldsymbol{\theta}^*||}{||\boldsymbol{\theta}^{(t)} - \boldsymbol{\theta}^*||},$$

where c is some constant and $0 < c < 1$. The sequence of EM iterations converges linearly and the largest eigenvalue λ of $DM(\boldsymbol{\theta}^*)$ corresponds to c. To compare the convergence of the ε-accelerated EM algorithm with that of the EM algorithm, we use the following notion given by Brezinski and Redivo Zaglia (1991).

Definition 1 *Let $\{A_n\}_{n\geq 0}$ be a sequence of scalars, and $\{\hat{A}_n\}_{n\geq 0}$ be the sequence generated by applying an extrapolation method ExtM to $\{A_n\}_{n\geq 0}$, where $\hat{A}_n$ is determined from A_m, $0 \leq m \leq L_n$, for some integer L_n, $n = 0, 1, \ldots$. Assume that $\lim_{n\to\infty} A_n = \lim_{n\to\infty} \hat{A}_n = A$. Then, we say that $\{\hat{A}_n\}_{n\geq 0}$ converges more quickly than $\{A_n\}_{n\geq 0}$ if*

$$\lim_{n\to\infty} \frac{|\hat{A}_n - A|}{|A_{L_n} - A|} = 0.$$

If the above limitation holds, we also say that the extrapolation method ExtM accelerates the convergence of $\{A_n\}_{n\geq 0}$. For a sequence of vectors, $\{\mathbf{A}_n\}_{n\geq 0}$, in some general vectors space, the definition is still valid, provided we replace $|\hat{A}_n - A|$ and $|A_{L_n} - A|$ everywhere by $\|\hat{\mathbf{A}}_n - \mathbf{A}\|$ and $\|\mathbf{A}_{L_n} - \mathbf{A}\|$, respectively, where $\|\cdot\|$ is the norm in the vector space under consideration.

The following theorem shows that the convergence of the ε-accelerated EM algorithm is faster than that of the EM algorithm.

Theorem 2 *Assume that $\{\boldsymbol{\theta}^{(t)}\}_{t\geq 0}$ is the sequence of the EM iterations and the sequence $\{\boldsymbol{\psi}^{(t)}\}_{t\geq 0}$ is generated by Equation (6). Then, we have*

$$\lim_{t\to\infty} \frac{||\boldsymbol{\psi}^{(t)} - \boldsymbol{\theta}^*||}{||\boldsymbol{\theta}^{(t+2)} - \boldsymbol{\theta}^*||} = \mathbf{0}. \tag{7}$$

That is, $\{\boldsymbol{\psi}^{(t)}\}_{t\geq 0}$ converges to $\boldsymbol{\theta}^*$ more quickly than $\{\boldsymbol{\theta}^{(t)}\}_{t\geq 0}$ does.

We describe the pseudocode for the ε-accelerated EM algorithm in Algorithm 3. The algorithm can be easily implemented in R.

Algorithm 3 Pseudocode for the ε-accelerated EM algorithm

$\boldsymbol{\theta}_1 \leftarrow M(\boldsymbol{\theta}_0)$; ▷ The EM algorithm
$\boldsymbol{\psi}_0 \leftarrow \boldsymbol{\theta}_1$;
$itr \leftarrow 1$;
while $itr < maxitr$ **do** ▷ $maxitr$: the maximum number of iterations
  $\boldsymbol{\theta}_2 \leftarrow M(\boldsymbol{\theta}_1)$; ▷ The EM algorithm
  $\Delta\boldsymbol{\theta}_0 \leftarrow \boldsymbol{\theta}_1 - \boldsymbol{\theta}_0$; $\Delta\boldsymbol{\theta}_1 \leftarrow \boldsymbol{\theta}_2 - \boldsymbol{\theta}_1$;
  $\boldsymbol{\psi}_1 \leftarrow \boldsymbol{\theta}_1 + \left[\left[\Delta\boldsymbol{\theta}_1\right]^{-1} - \left[\Delta\boldsymbol{\theta}_0\right]^{-1} \right]^{-1}$; ▷ The v$\varepsilon$ acceleration
  if $\|\boldsymbol{\psi}_1 - \boldsymbol{\psi}_0\|^2 < \delta$ **then**
    Termination of iterations
  end if
  $\boldsymbol{\psi}_0 \leftarrow \boldsymbol{\psi}_1$;
  $\boldsymbol{\theta}_0 \leftarrow \boldsymbol{\theta}_1$; $\boldsymbol{\theta}_1 \leftarrow \boldsymbol{\theta}_2$;
  $itr \leftarrow itr + 1$;
end while

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number JP21K11800.

References

Brezinski, C., & Redivo Zaglia, M. (1991). *Extrapolation methods: theory and practice*. North-Holland Publishing Co., Amsterdam.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B, 39*, 1-22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Kuroda, M., & Sakakihara, M. (2006). Accelerating the convergence of the EM algorithm using the vector ε algorithm. *Comput. Statist. Data Anal., 51*, 1549-1561. https://doi.org/10.1016/j.csda.2006.05.004

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Chapman & Hall, London.

Wang, M., Kuroda, M., Sakakihara, M., & Geng, Z. (2008). Acceleration of the EM algorithm using the vector epsilon algorithm. *Comput. Statist., 23*, 469-486. https://doi.org/10.1007/s00180-007-0089-1

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *Ann. Statist., 11*, 95-103. https://doi.org/10.1214/aos/1176346060

Wynn, P. (1962). Acceleration techniques for iterated vector and matrix problems. *Math. Comp., 16*, 301-322. https://doi.org/10.2307/2004051 https://doi.org/10.1090/S0025-5718-1962-0145647-X

Asymptotic Property for Generalized Random Forests
(一般化ランダムフォレストに対する漸近的性質)

Hiroshi Shiraishi* , Tomoshige Nakamura† , Ryuta Suzuki‡

In this paper, we discuss asymptotic property of the generalized random forests (GRF) estimator introduced by Athey et.al (2019). Athey et al. derived asymptotic normality of the GRF estimator, but they do not explicitly derive the rate of convergence and asymptotic variance. Scornet (2016) also discusses asymptotic normality of the RF estimator with respect to increasing number of trees. On the other hand, we consider asymptotic normality of the GRF estimator with respect to increasing sample size n as well as number of trees.

Model Suppose that $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d. observations from the model with distribution function F, where (X_i, Y_i) takes value $\mathcal{X} \times \mathcal{Y} \subseteq \mathbb{R}^p \times \mathbb{R}$ for $i = 1, \ldots, n$, and $p \geq 1$ is a fixed integer. We are interested in a function parameter $\theta = (\theta(x))_{x \in \mathcal{X}} \in \Theta := \{\theta : \mathcal{X} \to \mathcal{Y}\}$ defined by a local estimation equation of the form

$$\mathbb{E}\left[\psi_{\theta(x)}(Y_i) | X_i = x\right] = 0 \quad \text{for all } x \in \mathcal{X} \tag{1}$$

where $\psi_\cdot(\cdot) : \mathcal{Y} \times \mathcal{X} \to \mathbb{R}$ is some scoring function. This setup encompasses several key statistical problems such as conditional means, quantiles, average partial effects, etc (see, Athey et.al (2019)). In this paper, we consider two nonparametric estimation methods, generalized random forests (GRF) estimation and Nadaraya-Watoson (NW) estimation, for estimating the θ.

GRF estimator GRF estimation introduced by Athey et.al (2019) is to first define some kind of similarity weights $\alpha_i^{\mathrm{GRF}}(x)$ that measure the relevance of the i-th training example to fitting θ at a fixed $x \in \mathcal{X}$, and then fit the target of interest via an empirical version of the estimating equation:

$$\hat{\theta}^{\mathrm{GRF}}(x) \in \arg\min_{y \in \mathcal{Y}} \left\{ \left| \sum_{i=1}^{n} \alpha_i^{\mathrm{GRF}}(x) \psi_y(Y_i) \right| \right\} \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

When the above expression has a unique root, we can simply say that $\hat{\theta}^{\mathrm{GRF}}(x)$ solves

$$\sum_{i=1}^{n} \alpha_i^{\mathrm{GRF}}(x) \psi_y(Y_i) = 0$$

with respect to $y \in \mathcal{Y}$ for a fixed $x \in \mathcal{X}$. In order to define the weights $\alpha_i(x)$, we first grow a set of B trees indexed by $b = 1, \ldots, B$ and, for each tree, define $L_b(x)$ as the set of training examples falling in the same "leaf" as x. The weights $\alpha_i^{\mathrm{GRF}}(x)$ then capture the frequency with which the i-th training example falls into the same leaf as x:

$$\alpha_i^{\mathrm{GRF}}(x) = \frac{1}{B} \sum_{b=1}^{B} \alpha_{bi}^{\mathrm{GRF}}(x), \quad \alpha_{bi}^{\mathrm{GRF}}(x) = \frac{\mathbf{1}_{\{X_i \in L_b(x)\}}}{|L_b(x)|}. \tag{3}$$

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

The main difference between GRF and other non-parametric regression techniques is their use of recursive partitioning on subsamples to generate these weights $\alpha_i^{\mathrm{GRF}}(x)$. In particular, construction of the sets $L_b(x)$ defined below is important motivated by the empirical success of regression forests across several application areas.

Before the construction of $L_b(x)$, we firest introduce a splitting rule which corresponds to that of Athey et.al (2019). Let $\mathcal{D}_n = \{(X_i, Y_i)\}_{i=1,\dots,n}$ be the sequence of i.i.d. observed data, and let $s \equiv s(n)$ be a subsample size. Define a family of the index set by

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

where the elements of $\mathcal{A}_s$ are different from each other. For any $A = \{A^{\mathcal{I}}, A^{\mathcal{J}}\} \in \mathcal{A}_s$, we define subsamples $\mathcal{I}_s$ and $\mathcal{J}_s$ of $\mathcal{D}_n$ by $\mathcal{I}_s = \mathcal{D}_{A^{\mathcal{I}}}$ and $\mathcal{J}_s = \mathcal{D}_{A^{\mathcal{J}}}$ with $\mathcal{D}_{A^{\cdot}} = \{(X_i, Y_i)\}_{i \in A^{\cdot}}$, respectively. This sampling scheme is called the double-smpling, and thanks to the double-smpling, we achieve "honesty" by dividing its training subsamples into two halves $\mathcal{I}_s$ to place the splits, and $\mathcal{J}_s$ to do within-leaf estimation.

Only using $\mathcal{J}_s$-sample, the splitting rule is defined as follows:
Given $\mathcal{J}_s$-sample, we define a sequence of partitions $\mathcal{P}_0, \mathcal{P}_1, \cdots$ by starting from $\mathcal{P}_1 = \mathcal{X}$ and then, for some $j \in \{1, \dots, p\}$, construct $\mathcal{P}_{\ell+1}$ from $\mathcal{P}_\ell$ by replacing one set (parent node) $P \in \mathcal{P}_\ell$ by (childe node) $C_1 := \{x = (x_1, \dots, x_p) \in P | x_j \le \zeta\}$ and $C_2 := \{x = (x_1, \dots, x_p) \in P | x_j > \zeta\}$, where the split direction $j \in \{1, \dots, p\}$ and the split position $\zeta = \zeta(j) \in \{X_{ij} | X_i = (X_{i1}, \dots, X_{ip}) \in P\}$ are chosen to maximize a criterion $\Delta(C_1, C_2)$. Furthermore, we impose the following specifications for the splitting rule.

Specification 1. *(S.1) (ω-Regular) Every split puts at least a fraction ω of the observations in the parent node into each child node, with $\omega \in (0, 0.2]$.*

(S.2) (Random Split) At every split, the probability that the tree splits on the j-th feature is bounded from below by some $\pi > 0$, for all $j = 1, \dots, p$.

(S.3) (PNN (Potential Nearest Neighbor) k-set) There are between k and $2k - 1$ observations in each terminal node.

(S.4) (Subsample size) Subsample size s scales $s = n^\beta$ for some $\beta \in (0, 1)$.

Remark 1. *GRF by Athey et.al (2019) is defined some criterions $\Delta(C_1, C_2)$. But in this paper, we do not specify the criterions since the asymptotic normality can be derived under the above specifications. This means our result is applicable for the "honest" CART by Athey and Imbense (2016) , or the median splitting by Breiman (2001).*

Under this splitting rule, we denote a given partition of the feature space $\mathcal{X}$ by Λ, and the subspace (leaf) of rectangular type created by the partitioning by L_ℓ ($\ell = 1, \dots, |\Lambda|$). Then,

$$\Lambda = \Lambda(\mathcal{D}_n, \xi) = \{L_1, \dots, L_{|\Lambda|}\},\quad \mathcal{X} = \bigotimes_{\ell=1}^{|\Lambda|} L_\ell,\quad L_\ell \cap L_{\ell'} = \varnothing\ (\ell \neq \ell')$$

where ξ is a random variable taking a value in $\mathcal{A}_s$ with $\mathbb{P}(\xi = A) = |\mathcal{A}_s|^{-1}$ for any $A \in \mathcal{A}_s$, and the double-sample $(\mathcal{I}_s, \mathcal{J}_s)$ is identifiled based on $(\mathcal{D}_n, \xi)$. Fix a test point $x \in \mathcal{X}$ and construct Λ in the above manner, based on a selected $\mathcal{J}_s$-sample. Now, let B be the number of trees, and suppose that random variables $\xi_1, \dots, \xi_B$ are randomly generated. Together with a fixed $\mathcal{D}_n$, we from $(\mathcal{D}_n, \xi_1), \dots, (\mathcal{D}_n, \xi_B)$. Then, generate the sequence $(\mathcal{I}_s, \mathcal{J}_s)$ of $(\mathcal{I}_{s1}, \mathcal{J}_{s1}), \dots, (\mathcal{I}_{sB}, \mathcal{J}_{sB})$. Since Λ is identified for each $(\mathcal{I}_s, \mathcal{J}_s)$, we denote $\Lambda(\mathcal{D}_n, \xi_1), \dots, \Lambda(\mathcal{D}_n, \xi_B)$. For each $\Lambda(\mathcal{D}_n, \xi_b)$ ($b = 1, \dots, B$), define $L_b(x) \in \Lambda(\mathcal{D}_n, \xi_b)$ as a leaf containing $x \in \mathcal{X}$.

NW estimator Nadaraya (1964) and Watoson (1964) have considered estimates of a regression function nonparametrically using some kernel function. In this paper, we introduce a Nadaraya-Watoson type estimator (NW estimator) of θ defined by (1) as follows:

$$\hat{\theta}^{\mathrm{NW}}(x) \in \arg\min_{y\in\mathcal{Y}} \left\{ \left| \sum_{i=1}^{n} \alpha_i^{\mathrm{NW}}(x)\psi_y(Y_i) \right| \right\} \quad \text{for all } x \in \mathcal{X}. \tag{4}$$

The difference between $\hat{\theta}^{\mathrm{GRF}}$ and $\hat{\theta}^{\mathrm{NW}}$ is just difference of the weight functions $\alpha^{\mathrm{GRF}} = \{\alpha_i^{\mathrm{GRF}}\}$ and $\alpha^{\mathrm{NW}} = \{\alpha_i^{\mathrm{NW}}\}$. Note that the α^{NW} defined below, is oracle statistics which means that the objective function $\mathbb{E}\left[\psi_{\theta(x)}(Y)|X\right]$ in (1) is unkwon, but, α^{NW} is defined by using $\mathbb{E}\left[\psi_{\theta(x)}(Y)|X\right]$. Defin a function $u : \mathcal{X}^2 \to \mathcal{Y}$ by

$$u(x, x') = \mathbb{E}\left[\psi_{\theta(x)}(Y)|X = x'\right].$$

Based on $\mathcal{D}_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, we introduce a random process $U := (U(x))_{x\in\mathcal{X}}$ by $U(x) = \{U_1(x), \ldots, U_n(x)\}$ with

$$U_i(x) = u(x, X_i) = \mathbb{E}\left[\psi_{\theta(x)}(Y_i)|X_i\right], \quad i \in \{1, \ldots, n\}.$$

In addition, we introduce $F_{n,U}^x$ as the empirical distribution function of U, that is,

$$F_{n,U}^x(z) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{U_i(x)\le z\}}$$

for all $(x, z) \in \mathcal{X} \times \mathcal{Y}$. Then, α_i^{NW} is defined by

$$\alpha_i^{\mathrm{NW}}(x) = \frac{K\left(\frac{F_{n,U}^x(U_i(x)) - F_{n,U}^x(0)}{a_n}\right)}{\sum_{j=1}^{n} K\left(\frac{F_{n,U}^x(U_j(x)) - F_{n,U}^x(0)}{a_n}\right)} \tag{5}$$

for all $x \in \mathcal{X}$, where a_n is a bandwidth, K is a Laplcase kernel function given by

$$K(z) = \frac{1}{2}\exp\left(-|z|\right), \quad z \in \mathcal{Y}.$$

Asymptotic normality for GRF and NW estimators Before the statement of theoretical result, we impose the following assumption for the model

Assumption 1. *(A.1) There exists 2nd order moment, and strictly positive, continuous p.d.f. of (X, Y) on $\mathcal{X} \times \mathcal{Y}$.*

(A.2) (Lipschitz x-signal) $M_y(x) := \mathbb{E}[\psi_y(Y)|X = x]$ is Lipschitz continuous on $\mathcal{X}$.

(A.3) (Smooth identification) M_y is twice continuously differentiable at $y = \theta(x)$ with a uniformly bounded second derivative, and that $\dot{M}_{\theta(x)}(x) := \partial_y M_y(x)|_{y=\theta(x)}$ is invertible for all $x \in \mathcal{X}$.

(A.4) (Lipschitz (θ)-variogram) $\|\mathrm{Var}\left(\psi_y(Y) - \psi_{y'}(Y)|X = \cdot\right)\|_{\mathcal{X}} \le L|y - y'|$.

(A.5) (Regularity of ψ) $\psi_y = \lambda_y + \zeta_y$ where λ_y is Lipschitz continuous in y, ζ_y is monotone and bounded function.

(A.6) (Existence of solutions) There exists $\hat{\theta}^{\mathrm{GRF}}(x)$ and $|\sum_{i=1}^{n} \alpha_i^{\mathrm{GRF}}(x)\psi_{\hat{\theta}^{\mathrm{GRF}}(x)}(Y_i)| \le C \max_i\{\alpha_i^{\mathrm{GRF}}(x)\}$ for some constant $C \ge 0$.

(A.7) (Convexity) The score function ψ_y is a negative sub-gradient of a convex function, and the expected score M_y is the negative gradient of a strong convex function.

Regarding the difference between α^{GRF} and α^{NW}, we have the following result.

Theorem 1. *Under Specification 1, Assumption 1, and* $a_n = (s/2)^{-1}$,

$$\max_{i\in\{1,\ldots,n\}} \left\|\alpha_i^{\mathrm{GRF}} - \alpha_i^{\mathrm{NW}}\right\|_{\mathcal{X}} = o_p\left(\frac{\sqrt{s}}{n}\right)$$

where $\|\cdot\|_{\mathcal{X}}$ *is an uniform norm over* $\mathcal{X}$ *(i.e.,* $\|f\|_{\mathcal{X}} = \sup_{x\in\mathcal{X}} |f(x)|$*).*

The proofs of the theorems and corollary are given in Section 5. Now we introduce

$$\Psi_n^{\mathrm{GRF}}(x,y) = \sum_{i=1}^n \alpha_i^{\mathrm{GRF}}(x)\psi_y(Y_i), \quad \Psi_n^{\mathrm{NW}}(x,y) = \sum_{i=1}^n \alpha_i^{\mathrm{NW}}(x)\psi_y(Y_i)$$

for all $(x,y) \in \mathcal{X}\times\mathcal{Y}$. Thanks to Thereom 1, the difference between Ψ_n^{GRF} and Ψ_n^{NW} are negligible with order $\sqrt{s/n}$.

Corollary 1. *Under Specification 1, Assumption 1, and* $a_n = (s/2)^{-1}$,

$$\left\|\Psi_n^{\mathrm{GRF}} - \Psi_n^{\mathrm{NW}}\right\|_{\mathcal{X}\times\mathcal{Y}} = o_p\left(\sqrt{\frac{s}{n}}\right)$$

where $\|\cdot\|_{\mathcal{X}\times\mathcal{Y}}$ *is an uniform norm over* $\mathcal{X}\times\mathcal{Y}$ *(i.e.,* $\|f\|_{\mathcal{X}\times\mathcal{Y}} = \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} |f(x,y)|$*).*

Next, we consider asymptotic normality for Ψ_n^{NW} by some modification of Schuster (1972) or Stute (1984) .

Theorem 2. *Under Assumpton 1, and* $a_n = (s/2)^{-1}$*, for any fixed* $(x,y) \in \mathcal{X}\times\mathcal{Y}$ *and* $M_y(x) = \mathbb{E}[\psi_y(Y)|X=x]$

$$\sqrt{\frac{2n}{s}}\left\{\Psi_n^{\mathrm{NW}}(x,y) - M_y(x)\right\} \xrightarrow{d} N(0, V(x,y))$$

where $V(x,y) = \int_{z\in\mathcal{Y}} K^2(z)dz \mathrm{Var}(\psi_y(Y)|X=x) = \frac{1}{4}\mathrm{Var}(\psi_y(Y)|X=x)$.

Thanks to Theorem 2 and Corollary 1, we have the followings;

Theorem 3. *Under Specification 1, Assumption 1, and* $a_n = (s/2)^{-1}$*, for any fixed* $x \in \mathcal{X}^\circ$,

$$\sqrt{\frac{2n}{s}}\left\{\hat{\theta}^{\mathrm{GRF}}(x) - \theta(x)\right\} \xrightarrow{d} N\left(0, \sigma^2(x)\right)$$

where $\sigma^2(x) = \frac{V(x,\theta(x))}{\hat{M}^2_{\theta(x)}(x)}$.

Remark 2. *The Cramér-Wold device may be applied to show that* $\sqrt{\frac{2n}{s}}\left\{\hat{\theta}^{\mathrm{GRF}}(x) - \theta(x)\right\}$ *converges jointly in distribution at finitely many points* $x_1,\ldots,x_k$ *with* $\hat{\theta}^{\mathrm{GRF}}(x_1),\ldots,\hat{\theta}^{\mathrm{GRF}}(x_k)$ *being asymptotically independent (see Stute (1984)). This result and Theorem 3 imply that we have*

$$\sqrt{\frac{2n}{s}}\left(\hat{\theta}^{\mathrm{GRF}} - \theta\right) \xrightarrow{d} G \quad \text{in } \mathcal{X}$$

where G *is a Gaussian process with mean zero and a covariance function*

$$\mathrm{Cov}(G(x), G(x')) = \mathbf{1}_{\{x=x'\}}\sigma^2(x).$$

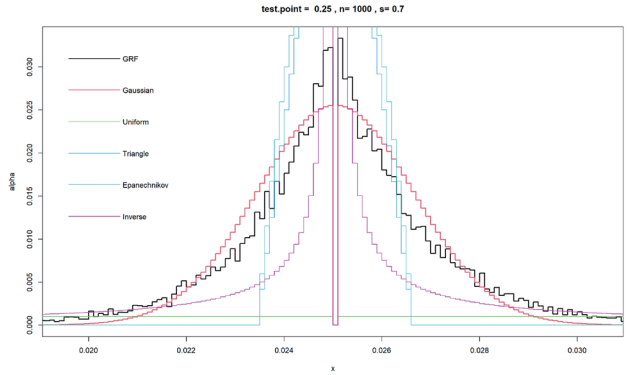


図 1: Result for fitting some kernel functions such as Gaussian, Uniform, Tiangle, Epanechinikov and Inverse when $Y_i = \sin(X_i) + N(0,1),\ X_i = i/n,\ n = 10^3,\ s = n^{0.7},\ a_n = (s/2)^{-1}$.

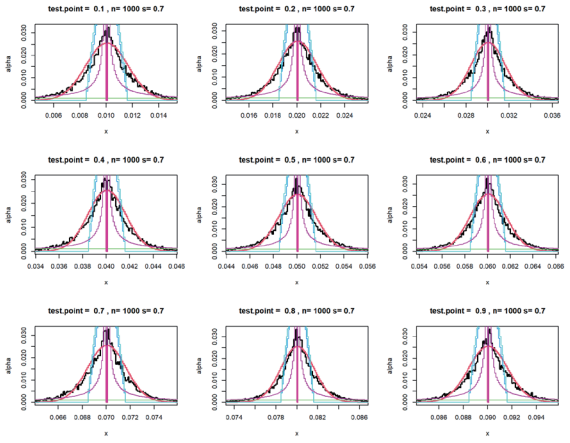


図 2: Result for test points as 0.1-quantile, 0.2-quantile, . . . ,0.8-quantile 0.9-quantile.

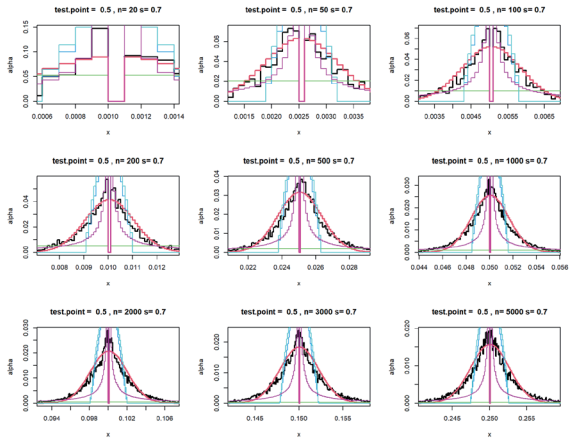


図 3: Result for sample size (n) as $20, 50, 100, 200, 500, 1000, 2000, 3000, 5000$.

参考文献

Athey, S., Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113, 7353-7360.

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

Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32.

Nadaraya, E, A. (1964). On estimating regression *Theory of Probability & Its Applications*, 9(1), 141-142.

Watson, G, S. (1964). Smooth regression analysis. *The Indian Journal of Statistics, Series A*, 359-372.

Schuster, E., F. (1972). Joint Asymptotic Distribution of the Estimated Regression Function at a finite number of distinct points. *The Annals of Mathematical Statistics*, 43(1), 84-88.

Scornet, E. (2016). On the asymptotics of random forests. *Journal of Multivariate Analysis*, 146, 72-83.

Stute, W. (1984). Asymptotic Normality of Nearest Neighbor Regression Function Estimates. *Annals of Statistics*, 12(3), 917-926.

An extension of the Weibull sine-skewed von Mises distributions: advantages and limitations

Yoichi Miyata* Takayuki Shiohama†
Toshihiro Abe‡
Advances in Statistical Modeling and Inference:
Exploring Applications on Diverse Manifolds,
Japan Statistics Research Institute in Hosei University

January 21–22, 2024

1 Introduction

Let Θ be a random variable representing an angle taking a value between $-\pi$ and π, and X be a random variable taking a non-negative or real value. The probability distribution of the pair (Θ, X) is called a statistical model on cylinders or a cylindrical distribution. Similarly, a cylindrical dataset indicates a set of values taking on $[-\pi, \pi) \times \mathbb{R}_+$ or $[-\pi, \pi) \times \mathbb{R}$. Such data can be found in various fields such as environmental studies, biology, and sports, and have been analyzed by various researchers. The relationship between wind direction and concentration of sulfur dioxide SO_2 has been studied by García-Portugués et al. (2014). The direction and speed of the currents in the Adriatic Sea have been analyzed by Lagona et al. (2015) through a hidden Markov model having the cylindrical distribution of Abe and Ley (2017) as a component.

Several statistical models on the cylinder have been proposed by Johnson and Wehrly (1978), Mardia and Sutton (1978), Kato and Shimizu (2008), Abe and Ley (2017), and Imoto et al. (2019). For a pair of random variables $(\Theta, X) \in [-\pi, \pi) \times \mathbb{R}_+$, Johnson and Wehrly (1978) propose a probability density function of the following form

$$f_{JW1}(\theta, x) = \frac{(\lambda^2 - \kappa^2)^{1/2}}{2\pi} \exp\left(-\lambda x + \kappa x \cos(\theta - \mu)\right),$$

*Faculty of Economics, Takasaki City University of Economics
†Department of Data Science, Nanzan University
‡Faculty of Economics, Hosei University

where circular location $\mu \in [-\pi, \pi)$, linear dispersion $\lambda > 0$, and $\kappa \in [0, \lambda)$ controls circular-linear dependence. On the other hand, they also propose the following density function for $(\Theta, X) \in [-\pi, \pi) \times \mathbb{R}$:

$$f_{JW2}(\theta, x) = C \frac{e^{-\kappa^2/(4\sigma^2)}}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x-\mu')^2}{2\sigma^2} + \frac{\kappa x}{\sigma^2} \cos(\theta - \mu) \right\},$$

where $C^{-1} = 2\pi \left\{ I_0\left(\frac{\kappa\mu'}{\sigma^2}\right) I_0\left(\frac{\kappa^2}{4\sigma^2}\right) + 2\sum_{i=1}^{\infty} I_i\left(\frac{\kappa^2}{4\sigma^2}\right) I_{2i}\left(\frac{\kappa\mu'}{\sigma^2}\right) \right\}$, $I_\alpha(z)$ is the modified Bessel function of the first kind and order α, $\mu', \mu \in \mathbb{R}$, $\sigma > 0$, and $\kappa \geq 0$. As can be seen from these equations, the first distribution requires the intractable constraint (i.e., $0 \leq \kappa < \lambda$) for the parameter space, while the second distribution has a normalizing constant expressed as an infinite sum of special functions. Abe and Ley (2017) proposed the tractable density function of (Θ, X) with a simple normalizing constant and no complicated parameter space constraints,

$$\begin{aligned} f_{AL}(\theta, x) = & \frac{\alpha\beta^\alpha}{2\pi\cosh(\kappa)} \left(1 + \lambda \sin(\theta - \mu)\right) x^{\alpha-1} \\ & \times \exp\left\{ -(\beta x)^\alpha \left(1 - \tanh(\kappa)\cos(\theta - \mu)\right) \right\} \quad (x \geq 0), \end{aligned} \tag{1}$$

where $\cosh(\kappa) = \{\exp(\kappa) + \exp(-\kappa)\}/2$, and $\tanh(\kappa) = \{\exp(\kappa) - \exp(-\kappa)\}/\{\exp(\kappa) + \exp(-\kappa)\}$. Furthermore, a parameter space is given by $\alpha > 0$, $\beta > 0$, $\kappa > 0$, $-\pi \leq \mu < \pi$, and $-1 \leq \lambda \leq 1$ without complicated constraints. In the distribution of Abe and Ley (2017), the marginal distribution of Θ becomes the sine-skewed wrapped Cauchy distribution with λ being a skewness parameter. However, in practice, there exist data on cylinders in cases where the circular part exhibits significant asymmetry near the mode, and cannot be fully captured by the sine-skewed wrapped Cauchy distribution.

2 Main reports

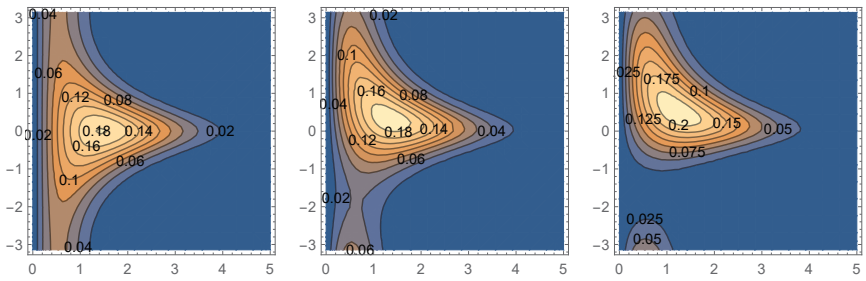
To cope with this problem, we introduce the density function proposed by Miyata et al. (2024)

$$\begin{aligned} f^{(q)}_{\mathrm{WeiESSvM}}(\theta, x) = & \frac{\alpha\beta^\alpha}{\pi\cosh(\kappa)} G_q(\lambda \sin(\theta - \mu)) x^{\alpha-1} \\ & \times \exp\left\{ -(\beta x)^\alpha \left(1 - \tanh(\kappa)\cos(\theta - \mu)\right) \right\} \quad (x \geq 0), \end{aligned} \tag{2}$$

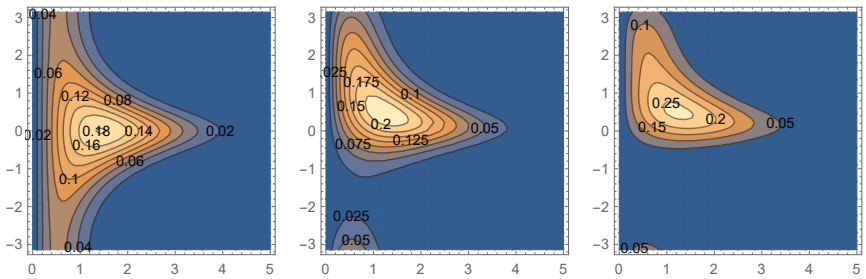
where $\boldsymbol{H} = \{(\mu, \kappa, \lambda, \alpha, \beta) | -\pi \leq \mu < \pi, \kappa > 0, -1 \leq \lambda \leq 1, \alpha > 0, \text{ and } \beta > 0\}$ is a parameter space,

$$g_q(x) = \frac{\Gamma(2(q+1))}{2^{2q+1}\Gamma(q+1)^2} (1 - x^2)^q \qquad (-1 \leq x \leq 1),$$

is a Beta density function, $\Gamma(\cdot)$ is the Gamma function, and $G_q(x) = \int_{-1}^{x} g_q(t)dt$ is its distribution function. We call the distribution (2) the Weibull Extended Sine-Skewed von Mises (WeiESSvM) distribution. $q \geq 0$ is a prespecified value, that is a hyperparameter, which we call "order". As seen from the contour plots in the upper right and lower right of Figure 1, the skewness parameter λ enhances the degree to which the distribution of Θ is skewed as order q is increased.



Order $q = 0$: $\lambda = 0, 0.5, 0.9$



Order $q = 2$: $\lambda = 0, 0.5, 0.9$

Figure 1: Contour plots of the density (2) with $\alpha = 2$, $\beta = 1$, $\mu = 0$, and $\kappa = 1$. The horizontal axis represents the variable x and the vertical axis represents the variable θ.

In this talk, we reported the following theoretical advantages and limitations.

- All the marginal and conditional distributions of Θ and those of X can be expressed in a closed form.
- Since the marginal distribution of Θ is the extended sine-skewed wrapped Cauchy distribution of order q, it can exhibit a greater degree of skewness compared to the marginal distribution of Θ in the model (1).
- The algorithm that can easily generate random numbers from the proposed model (2).

- When order q is fixed, a subfamily of the proposed distributions $\mathcal{F}_{\mathrm{cyl}}^{(q)} := \{f_{\mathrm{WeiESSvM}}^{(q)}(\theta, x; \boldsymbol{\eta}) | \boldsymbol{\eta} \in \boldsymbol{H}\}$ is identifiable for parameters. On the other hand, the family $\mathcal{F}_{\mathrm{cyl}} := \{f_{\mathrm{WeiESSvM}}^{(q)}(\theta, x; \boldsymbol{\eta}) | \boldsymbol{\eta} \in \boldsymbol{H}, q \in \{0, 1, 2, \cdots\}\}$ is not identifiable.

The details of these characteristics are provided in Miyata et al. (2024) and are omitted here. The marginal distribution of Θ in the density (2) can represent various degrees of skewness by setting the appropriate order q. To demonstrate this capability, we utilized a hidden Markov model with this distribution as a component for wind direction and wind speed data[1], consisting of $T = 426$ measurements taken at Shionomisaki by the Disaster Prevention Research Institute, Kyoto University. Using the maximum log-likelihood method and the Akaike information criterion (AIC), we confirmed that the hidden Markov model with components for the proposed distribution outperforms the model of Lagona et al. (2015). Here, the number of components for both models was estimated to be three using AICs. In the model of Lagona et al. (2015), two of the maximum likelihood estimates of skewness parameter λ in each of the three components appeared at the boundaries of the parameter space. In contrast, we reported that the maximum likelihood estimates of skewness parameter λ in the three components of the proposed model, selected via the minimum AIC criterion, are all interior points in the parameter space.

References

Abe, T. and C. Ley (2017). A tractable, parsimonious and flexible model for cylindrical data, with applications. *Econometrics and Statistics 4*, 91–104.

García-Portugués, E., A. M. Barros, R. M. Crujeiras, W. González-Manteiga, and J. Pereira (2014). A test for directional-linear independence, with applications to wildfire orientation and size. *Stochastic Environmental Research and Risk Assessment 28*(5), 1261–1275.

Imoto, T., K. Shimizu, and T. Abe (2019). A cylindrical distribution with heavy-tailed linear part. *Japanese Journal of Statistics and Data Science 2*(1), 129–154.

Johnson, R. A. and T. E. Wehrly (1978). Some angular-linear distributions and related regression models. *Journal of the American Statistical Association 73*(363), 602–606.

[1] `https://rcfcd.dpri.kyoto-u.ac.jp/frs/swel/data.html`

Kato, S. and K. Shimizu (2008). Dependent models for observations which include angular ones. *Journal of Statistical Planning and Inference 138*(11), 3538–3549.

Lagona, F., M. Picone, A. Maruotti, and S. Cosoli (2015). A hidden Markov approach to the analysis of space–time environmental data with linear and circular components. *Stochastic Environmental Research and Risk Assessment 29*(2), 397–409.

Mardia, K. V. and T. W. Sutton (1978, 12). A model for cylindrical variables with applications. *Journal of the Royal Statistical Society: Series B (Methodological) 40*(2), 229–233.

Miyata, Y., T. Shiohama, and T. Abe (2024). Cylindrical models motivated through extended sine-skewed circular distributions. *Symmetry 16*(3).

Periodicity for circular data

Hiroaki Ogata
Tokyo Metropolitan University

1 Introduction

The purpose of this work is to detect the periodicity for circular time series data. We define a spectral density for circular time series data regarding the circular time series as a complex-valued process. We also calculate the covariance matrix and spectral density for some specific circular time series models.

2 Complex-valued processes

Consider a complex-valued zero-mean process $x_t = u_t + \mathrm{i}v_t$, which is composed from the two real random process u_t and v_t for $t \in \mathbb{Z}$. The *covariance function* and *complementary covariance function* of x_t are defined by

$$r_{xx}[t,h] = E[x_t x_{t-h}^*], \qquad \tilde{r}_{xx}[t,h] = E[x_t x_{t-h}],$$

where x^* stands for the complex conjugate of x. We also introduce the notation $\underline{\boldsymbol{x}}_t = [x_t \quad x_t^*]^\top$ and define the augmented covariance matrix

$$\underline{\mathbf{R}}_{xx}[t,h] = E[\underline{\boldsymbol{x}}_t \underline{\boldsymbol{x}}_{t-h}^{\mathrm{H}}] = \begin{bmatrix} r_{xx}[t,h] & \tilde{r}_{xx}[t,h] \\ \tilde{r}_{xx}^*[t,h] & r_{xx}^*[t,h] \end{bmatrix}, \quad t \in \mathbb{Z},\ h \in \mathbb{Z},$$

where $\boldsymbol{x}^{\mathrm{H}}$ is the Hermitian transpose of $\boldsymbol{x}$. If $\underline{\mathbf{R}}_{xx}[t,h]$ is independent of t, the complex-valued process x_t is called *wide-sense stationary (WSS)*. When x_t is WSS, we can drop the t-argument from the covariance function, and complementary covariance function have the following relations:

$$r_{xx}[h] = r_{xx}^*[-h], \qquad \tilde{r}_{xx}[h] = \tilde{r}_{xx}[-h]$$

or equivalently,

$$\underline{\mathbf{R}}_{xx}[h] = \underline{\mathbf{R}}_{xx}^{\mathrm{H}}[-h]. \tag{1}$$

The Fourier transform of $\underline{\mathbf{R}}_{xx}[h]$:

$$\underline{\mathbb{P}}_{xx}(\lambda) = \begin{bmatrix} P_{xx}(\lambda) & \tilde{P}_{xx}(\lambda) \\ \tilde{P}^*_{xx}(-\lambda) & P^*_{xx}(-\lambda) \end{bmatrix} = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \underline{\mathbf{R}}_{xx}[h] \mathrm{e}^{-\mathrm{i}\lambda h}$$

is called the *augmented power spectral density (PSD) matrix*.

The augmented PSD matrix provides the spectral properties of x_t. In addition, a complex second-order random process x_t has the following spectral representation:

$$x_t = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}t\lambda} d\xi(\lambda)$$

where $\xi(\lambda)$ is a spectral process with orthogonal increments $d\xi(\lambda)$ whose second-order moments are $E|d\xi(\lambda)|^2 = P_{xx}(\lambda)d\lambda$ and $E[d\xi(\lambda)d\xi(-\lambda)] = \tilde{P}_{xx}(\lambda)d\lambda$. This representation indicates the x_t can be decomposed into periodic (rotary) components, and $P_{xx}(\lambda)$ is interpreted as the magnitude of periodic component with frequency λ.

3 Examples

In this section, we clarifies the augmented covariance and PSD matrices in two circular time series models. In this section, let $\{\Theta_t\}_{t\in\mathbb{Z}}$ be a stochastic process, taking its values on the interval $[-\pi, \pi]$. We also express a circular random variable by a unit circle on the complex plain: $x_t = \mathrm{e}^{\mathrm{i}\Theta_t}$.

3.1 Circular mixture transition density model

Ogata and Shiohama [2023] considered the circular strictly stationary pth order Markov process whose marginal distribution is set as a circular uniform and transition density is defied by

$$f(\theta_t|\theta_{t-1}, \theta_{t-2}, \ldots) = f(\theta_t|\theta_{t-1}, \ldots, \theta_{t-p}) = \sum_{k=1}^{p} a_k g(\theta_t - q_k\theta_{t-k}). \qquad (2)$$

Here $g(\cdot)$ is any circular density functions, a_k $(k = 1, \ldots, p-1) \geq 0$ and $a_p > 0$ are non negative and positive constants, respectively. The sum of mixing weights must satisfy $\sum_{k=1}^{p} a_k = 1$ and q_k's are constants which take the values -1 or 1. This model is referred as a circular mixture transition density model. Denote the characteristic function of g by

$$\phi_m = \rho_m \mathrm{e}^{\mathrm{i}\mu_m} = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}m\theta} g(\theta) d\theta \qquad (m = 0, \pm 1, \pm 2, \ldots).$$

Then, the augmented covariance matrix for h (> 0) is recursively obtained by

$$\underline{\mathbf{R}}_{xx}[h] = \sum_{k=1}^{p} \Psi_k \underline{\mathbf{R}}_{xx}[h-k] \tag{3}$$

where

$$\Psi_k = a_k \begin{bmatrix} \phi_1 & 0 \\ 0 & \phi_{-1} \end{bmatrix} \begin{bmatrix} \frac{1+q_k}{2} & \frac{1-q_k}{2} \\ \frac{1-q_k}{2} & \frac{1+q_k}{2} \end{bmatrix}. \tag{4}$$

For $h = 0$,

$$\underline{\mathbf{R}}_{xx}[0] = \begin{bmatrix} 1 & E[\mathrm{e}^{\mathrm{i}2\Theta_t}] \\ E[\mathrm{e}^{-\mathrm{i}2\Theta_t}] & 1 \end{bmatrix} = I_2 \tag{5}$$

because we assume the marginal follows the circular uniform distribution. Initial values are obtained by solving the equations

$$\begin{cases} \underline{\mathbf{R}}_{xx}[p-1] = \Psi_1 \underline{\mathbf{R}}_{xx}[p-2] + \Psi_2 \underline{\mathbf{R}}_{xx}[p-3] + \ldots + \Psi_p \underline{\mathbf{R}}_{xx}[-1] \\ \underline{\mathbf{R}}_{xx}[p-2] = \Psi_1 \underline{\mathbf{R}}_{xx}[p-3] + \Psi_2 \underline{\mathbf{R}}_{xx}[p-4] + \ldots + \Psi_p \underline{\mathbf{R}}_{xx}[-2] \\ \quad \vdots \\ \underline{\mathbf{R}}_{xx}[1] = \Psi_1 \underline{\mathbf{R}}_{xx}[0] + \Psi_2 \underline{\mathbf{R}}_{xx}[-1] + \ldots + \Psi_p \underline{\mathbf{R}}_{xx}[-p+1] \end{cases} \tag{6}$$

where $\underline{\mathbf{R}}_{xx}[h]$ with negative h is replaced by $\underline{\mathbf{R}}^{\mathrm{H}}_{xx}[-h]$.

The augmented power spectral density matrix of the circular mixture transition density model is written by

$$\underline{\mathbb{P}}_{xx}(\lambda) = \frac{1}{2\pi} \left(I_2 - \sum_{k=1}^{p} \Psi_k e^{\mathrm{i}k\lambda} \right)^{-1} \left(I_2 - \sum_{k=1}^{p} \Psi_k \underline{\mathbf{R}}^{\mathrm{H}}_{xx}(k) \right) \quad \left(I_2 - \sum_{k=1}^{p} \Psi_k^{\mathrm{H}} e^{-\mathrm{i}k\lambda} \right)^{-1}$$

where $\Psi_1, \ldots, \Psi_p$ are defined by (4) and $\underline{\mathbf{R}}_{xx}(1), \ldots, \underline{\mathbf{R}}_{xx}(p)$ are obtaind by solving the equations (6).

3.2 Wrapped Auto regressive process

Breckling [2012] introduced the wrapped autoregressive process

$$\Theta_t = Y_t \ (\mathrm{mod}\ 2\pi),$$

where $\{Y_t\}_{t\in\mathbb{Z}}$ is a stationary linear AR(p) process

$$Y_t = \sum_{k=1}^{p} \phi_k Y_{t-k} + \varepsilon_t \tag{7}$$

with ε_t independently and identically distributed as $N(0, \sigma^2)$. (7) can be rewritten into MA(∞) form

$$Y_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}$$

where

$$\sum_{k=-\infty}^{\infty} |\psi_k| < \infty, \quad \psi_0 = 1, \quad \psi_k = 0 \text{ for } k < 0.$$

In the case of wrapped AR(1) process

$$\Theta_t = Y_t \pmod{2\pi}, \quad Y_t = \phi Y_{t-1} + \varepsilon_t, \quad |\phi| < 1,$$

the process Y_t can be rewritten as

$$Y_t = \sum_{k=0}^{\infty} \phi^k \varepsilon_{t-k}.$$

Then, the augmented covariance matrix becomes

$$\underline{\mathbf{R}}_{xx}(h) = \begin{bmatrix} \mathrm{e}^{-\sigma^2 \frac{1-\phi^h}{1-\phi^2}} & \mathrm{e}^{-\sigma^2 \frac{1+\phi^h}{1-\phi^2}} \\ \mathrm{e}^{-\sigma^2 \frac{1+\phi^h}{1-\phi^2}} & \mathrm{e}^{-\sigma^2 \frac{1-\phi^h}{1-\phi^2}} \end{bmatrix} \nrightarrow O \quad (h \to \infty).$$

This implies that the covariance of wrapped autoregressive process does not tend to be zero.

References

Breckling, J. (2012). *The Analysis of Directional Time Series: Applications to Wind Speed and Direction.* Lecture Notes in Statistics. Springer New York.

Ogata, H. and T. Shiohama (2023). A mixture transition distribution modeling for higher-order circular markov processes.

研究所報(最近刊行分)

号数	タイトル	刊行年月日
36	人口センサスの現状と新展開	2007.04.01
37	統計における官学連携	2007.04.20
38	ジェンダー(男女共同参画)統計 Ⅱ	2009.02.10
39	社会生活基本調査とその利用	2010.01.15
40	地方統計の現状と課題	2010.09.15
41	Exploring Potential of Individual Statistical Records	2011.11.05
42	観光統計	2013.02.05
43	国民経済計算関連統計の新たなる展開	2014.01.30
44	タウンページデータによる事業所立地分析	2014.02.15
45	フィンランドのビジネス・レジスター	2015.03.20
46	19 世紀ドイツ営業統計史研究	2015.07.20
47	地方統計と統計 GIS	2016.01.25
48	首都圏の人口移動	2017.03.10
49	宿泊業及び飲食業の実証分析	2018.08.01
50	サービス分野の生産物分類	2019.01.31
51	全市区町村産業連関表(平成 23 年表)の推計	2019.10.15
52	商業統計調査	2021.01.31
53	産業連関表から供給・使用表へ	2021.03.31
54	統計的モデリング	2021.11.30
55	数理生態学とデータサイエンス	2022.01.31
56	様々な多様体上における統計的推測	2022.03.31
57	全市区町村産業連関表(平成 27 年表)の推計と分析	2023.02.13
58	統計的モデリングとその周辺	2023.02.13
59	英国国家統計局の年次経済調査と統合世帯調査	2024.03.31

研 究 所 報 No. 60
2025 年 3 月 31 日

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

B U L L E T I N

OF

JAPAN STATISTICS RESEARCH INSTITUTE

No.60 **March 2025**

Advances in Statistical Modeling and Inference: Exploring Applications on Diverse Manifolds

CONTENTS

Edited by
JAPAN STATISTICS RESEARCH INSTITUTE
HOSEI UNIVERSITY
TOKYO, JAPAN