Econometrics and Statistics 27 (2023) 136–160

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

Bayesian estimation for mode and anti-mode preserving circular distributions

Toshihiro Abe[a,*], Yoichi Miyata[b], Takayuki Shiohama[c]

[a] Faculty of Economics, Hosei University, 4342 Aihara, Machida, Tokyo 194-0298, Japan
[b] Faculty of Economics, Takasaki City University of Economics, 1300 Kaminamie, Takasaki, Gunma 370-0801, Japan
[c] Department of Data Science, Nanzan University, Nagoya, 466-8673, Japan

ARTICLE INFO

Article history:
Received 10 June 2020
Revised 12 March 2021
Accepted 16 March 2021
Available online 20 April 2021

Keywords:
Approximate Bayes
Asymmetric distribution
Bayesian estimation
Circular statistics
Lindley's approximation

ABSTRACT

A Bayesian estimation is considered for unknown parameters of a unimodal skew circular distribution on the circle, where the underlying distribution has mode and anti-mode preserving properties. This distribution is obtained by using a transformation of the inverse monotone function, and the shape of the resulting density can be flat-topped or sharply peaked at its mode. With regard to Bayes estimates (BEs), the boundary-avoiding priors are assumed so that the skewness and peakedness parameters of the distribution do not lie on the boundary of the parameter space. In addition to the BEs, maximum likelihood estimations (MLEs) are conducted to compare the performances in small samples, and found that the BEs are more robust than the method of maximum likelihood. As the pairs of parameters between location and skewness and between concentration and peakedness are independent of each other, approximate BEs using Lindley's methods become rather simple. Monte Carlo simulations are performed to compare the accuracy of the BE and MLE, and some circular datasets are analyzed for illustrative purposes.

1. Introduction

In recent years, there has been increased interest in statistical data analysis where the data take values in geometrical manifolds. Examples of geometric space include a hypersphere, torus, or circle. Statistical analysis and techniques for handling these data are called directional statistics and are challenging because the usual statistical tools and methods are unsuitable. For further details of directional statistics, we refer to Mardia and Jupp (1999), Jammalamadaka and SenGupta (2001), and Pewsey et al. (2013) for comprehensive reviews.

To analyze data on a unit circle, several useful distributions are used to fit the angular data. The von Mises and wrapped Cauchy distributions are examples of symmetric circular distributions that are used frequently in practice. The Jones–Pewsey distribution provides a wide class of symmetric circular distributions, as it contains the von Mises and wrapped Cauchy distributions for special cases; see, for example, Jones and Pewsey (2005). In many cases, asymmetric or multi-modal distributions of angular data are observed, and we need statistical models to fit these data. Energy demand by time of day and arrival times associated with some events tends to skew, and the distributions of wind direction are affected by the geographical or seasonal patterns that result in skewed or multi-modal distributions. For a construction of the circular version of the skew models by Azzalini and Capitanio (2003), Umbach and Jammalamadaka (2009) adapted their distributions

* Corresponding author
E-mail address: abetosh@hosei.ac.jp (T. Abe).

https://doi.org/10.1016/j.ecosta.2021.03.004

in order to obtain a general means of skewing symmetric circular models. For a related topic for the skew models, Abe and Pewsey (2011) investigated their properties and proposed the sine-skewed Jones-Pewsey family as well as its three special cases. Miyata et al. (2020) considered the finite mixture models of skew-symmetric circular distributions to model the asymmetry and multi-modality of the data. On the other hand, a general class of transformation of scale distributions was introduced by Jones (2014). Jones and Pewsey (2012) considered its circular version, inverse Batschelet distributions, whereas the distributions induced by transforming circular variable by Papakonstantinou (1979) and Batschelet (1981) are called direct Batschelet distributions, see also Abe et al. (2009, 2013); Pewsey et al. (2011). The recent main general techniques for providing families of typically unimodal distributions was comprehensively reviewed and compared by Jones (2015).

We briefly review the testing symmetry for circular data. Pewsey (2002) provided tests for symmetry on a circle. Ley and Verdebout (2014) proposed optimal tests for circular reflective symmetry about a fixed median direction. They used the k-sine-skewed distributions of Umbach and Jammalamadaka (2009) to obtain the optimal test for symmetry and discussed the unspecified location case. Recently, Ameijeiras-Alonso et al. (2019) developed optimal tests of the null hypothesis that the distribution from which a random sample of circular data was drawn is reflectively symmetric about an unspecified location against the alternative hypothesis for a k-sine-skewed circular distribution. These topics of optimal tests were also summarized by Ley and Verdebout (2017).

Complicated multi-modal distributions can be constructed by finite mixture modeling of unimodal distributions. Examples of mixtures using the generalized von Mises distribution can be found in Yfantis and Borgman (1982) and Gatto and Jammalamadaka (2007). There exist various attempts for constructing and estimating finite mixture models in literature. Kato and Jones (2015) proposed a flexible unimodal distribution, and their four-parameter family of circular distribution has some favorable properties, such as having a simple normalizing constant and being closed under convolution and multiplication. Their density and distribution functions are tractable, with explicit moments together with a wide range of skewness and peakedness measures. Hence, we investigate flexible unimodal circular distributions that are applicable in many practical situations.

In this study, we concentrate on a general class of unimodal circular distributions using inverse monotone functions to control asymmetry and peakedness/flat-topped distribution whose mode and anti-mode are unchanged. The proposed distribution has the following preferable properties. First, the normalizing constant is easily calculated and provided explicitly. Second, from the viewpoint of parameter interpretation, our model provides an intuitive interpretation of the skewness or peakedness measures of distribution. As is often the case with usual asymmetric circular distributions, it is difficult to interpret the parameters, as the location shifts according to changes in the skewness parameter, which makes it challenging to understand the role of the parameters of location. The computational complexity of our model is a disadvantage because the scale is given by using an inverse function, which are needed to be inverted, and this inverse function is not expressed in an explicit form in general. Hence, the moments do not have closed forms, making it impossible to construct a method of moments estimation. Thus, we consider likelihood-based inferences throughout the study.

As the maximum likelihood estimate (MLE) is not robust and has a larger bias that leads to unreliable results, especially in small sample sizes, we adopt Bayesian methods and expect them to provide reliable results with prior information. In addition, the MLEs for the skewness and peakedness parameters tend to lie on the boundary of parameter space, which yields unsatisfactory confidence intervals for the parameters as the corresponding diagonal elements of the Fisher information matrix and its inverse become unstable. It is also pointed out that the Fisher information matrix for sine-skewed von Mises distributions (Abe and Pewsey, 2011) degenerates when the skewness parameter is 0 (e.g., Ley and Verdebout, 2014). To protect against influential observations in small sample settings, we use boundary-avoiding priors so that the Bayes estimate (BE) does not necessarily lie on the edge of the parameter space. As investigated in Monte Carlo simulations, more robust and smaller mean squared errors (MSEs) are obtained for BEs compared with MLEs. We also introduce Lindley's approximation for the BEs of the parameters, where the second and third derivatives of the log-likelihood functions have many elements that are close to zero, which provides the approximate posterior mean with concise expressions and calculations.

A conjugate analysis for a BE for circular distributions is available; for example, Mardia and El-Atoum (1976) considered the Bayesian analysis for a von Mises distribution. Guttorp and Lockhart (1988) studied conjugate prior distributions for the von Mises distribution in a linear model, and Nuñez-Antonio and Gutiérrez-Peña (2005) used a Bayesian analysis for a projected normal distribution. Ravindran and Ghosh (2011) used a BE for a wrapped Cauchy distribution, and Mulder and Klugkist (2017) studied the generalized linear models in a Bayesian context.

In many cases, the quantities in the Bayesian inference, which include posterior distributions, predictive density, and their ratios known as Bayes factors, are often analytically intractable and cannot be obtained in explicit form. One of the most popular methods to approximate the posterior distribution in Bayesian inference is Laplace approximation; see, for example, Tierney and Kadane (1986) and Kass et al. (1989), among others. We use the idea of Lindley (1980) to compute the approximate BEs of the unknown parameters, and the approximation works quite well, with the exception of cases wherein the estimated parameters are near the boundary of the parameter space. Lindley's approximation technique is extremely fast compared with the Markov Chain Monte Carlo (MCMC) techniques, and it provides consistent estimates for parameters. As our parameter space is restricted and careful attention is necessary for numerical computation in the Bayesian inference, we use the parameter transformation for unconstrained numerical optimization. As Tierney et al. (1989) pointed out, finding a good parameter transformation would provide better approximations for Bayesian inferences.

This study is an in-depth investigation of the distributional properties of the extended inverse transformation of scale distribution. It presents several explicit expressions of distribution functions and random number generation methods. For

parameter estimation, the limiting distributions for the MLE and Bayesian inferences are investigated. These theoretical findings are noteworthy for not only practical data analysis elucidated by simulation study but also real data analysis.

The remainder of this paper is organized as follows. Section 2 describes some basic properties of our proposed distributions with mode preserving properties. In Section 3, parameter estimations based on the MLE and BEs are introduced. Approximate BEs using Lindley's method are discussed in Section 3, and the numerical simulations are performed in Section 4. A real data analysis using three datasets is illustrated in Section 5. Finally, Section 6 concludes this study. All the proofs are given in Appendix A.

2. Transformation of scale distribution on a circle

Let $f_0(\theta)$ and $g_0(\theta)$ $(\theta \in [-\pi, \pi))$ be unimodal circular probability density functions (pdf), that is reflectively symmetric about the zero direction, and $G_0(\theta) = \int_{-\pi}^{\theta} g_0(\phi)d\phi$ is the distribution function of the latter. A weighting function w is an odd periodic function with $|w(\theta)| \leq \pi$. Then,

$$f(\theta) = 2f_0(\theta)G_0\{w(\theta)\}, \quad -\pi \leq \theta < \pi,$$

is a circular pdf; for more details, see Umbach and Jammalamadaka (2009). Jones (2014) proposed a class of skew distributions with pdf

$$f(x) = f_0(s^{-1}(x)), \quad x \in S_f,$$

where $s'(x) + s'(-x) = 2$, with s' being a derivative of function s. The function s is monotone bijective, and S_f is a support for function s. Combining these previously proposed circular densities, we consider the unimodal skew distributions on a circle given by

$$f(\theta) = f_0(s^{-1}(\theta; w)), \qquad -\pi \leq \theta < \pi,$$

which is an extension of the model in Jones and Pewsey (2012), where the function s is given by

$$s(\theta; w) = 2\int_{-\pi}^{\theta} G_0(w(t))dt - \pi.$$

The function s is needed to be inverted, and this inverse function is not expressed in an explicit form in general. This is a disadvantage in computational complexity. In this study, we propose unimodal skew distributions on a circle that are given by

$$f(\theta) = f_0(s_\lambda^{-1}(\theta)), \quad \theta \in [-\pi, \pi), \tag{1}$$

where $\lambda \in [-1, 1]$, and the function s_λ is given by

$$s_\lambda(\theta) = \theta + \lambda \sin^2(\theta).$$

The parameter λ plays the role of skewness of the distributions, such that for $\lambda = 0$, the density gives symmetric distribution. The sign of λ determines the shape of the skewness, and as $\lambda \to -1$ and $\lambda \to 1$, the distributions tend to be left and right skewed, respectively.

Hereafter, the location parameter $\mu \in [-\pi, \pi)$ is introduced by replacing $\theta \mapsto \theta - \mu$ in the above distribution. To control the peakedness around the mode, we introduce a parameter $\nu \in [-1, 1]$. Then, we also propose the following circular density function which can handle both the skewness and peakedness of the underlying circular random variables:

$$f(\theta) = \frac{1}{1 - \nu\alpha_1} f_0(t_{\lambda,\nu}^{-1}(\theta - \mu)), \quad \theta \in [-\pi, \pi), \tag{2}$$

where $\alpha_1 = E_{f_0}[\cos\Theta]$ is the first cosine moment of base density $f_0(\cdot)$, and the function $t_{\lambda,\nu}$ is given by

$$\begin{aligned} t_{\lambda,\nu}(\theta - \mu) &= s_\lambda((\theta - \mu) - \nu\sin(\theta - \mu)) \\ &= \theta - \mu - \nu\sin(\theta - \mu) + \lambda\sin^2((\theta - \mu) - \nu\sin(\theta - \mu)). \end{aligned}$$

There is no explicit expression of the inverse of $t_{\lambda,\nu}$. Therefore, numerical methods are needed to compute inverse function $t_{\lambda,\nu}^{-1}$. Let $\boldsymbol{\gamma}$ be the parameter vector of base symmetric density $f_0(\theta - \mu)$. We need the following assumption for the general inverse transformation of scale distributions.

Assumption 1. Base density function f_0 indicates a continuous circular density, which is unimodal, reflectively symmetric, and differentiable with respect to $\boldsymbol{\gamma}$. The first cosine moment of base distribution $f_0(\cdot)$ must satisfy $\alpha_1 \in (0, 1)$.

It is noted that the base symmetric density does not include the circular uniform and degenerate distributions. As our base density deals with a reflective symmetric distribution, we exclude an ℓ-fold symmetric distribution as a candidate of the base symmetric distribution. More examples of density function f_0, which satisfies Assumption 1, can be found in Mardia and Jupp (1999, Chapter 3) and Jammalamadaka and SenGupta (2001, Chapter 2).

The distribution with density (2) is sharply peaked and flat-topped as $\nu \to 1$ and $\nu \to -1$, respectively. The terms "sharply peaked and flat-topped" were introduced by Batschelet (1981) in the process of constructing Papakonstantinou's

distribution. Then the peakedness around the mode for the models was mathematically characterized using curvature in Abe et al. (2013). The density function becomes a circular pdf because its integral becomes unity. By using a change in variable $t^{-1}_{\lambda,\nu}(\theta) = y$, we observe

$$\begin{aligned}\int_{-\pi}^{\pi} f(\theta)d\theta &= \int_{-\pi}^{\pi} \frac{1}{1-\nu\alpha_1} f_0(t^{-1}_{\lambda,\nu}(\theta))d\theta = \frac{1}{1-\nu\alpha_1}\int_{-\pi}^{\pi} t'_{\lambda,\nu}(y) f_0(y)dy \\ &= \frac{1}{1-\nu\alpha_1}\int_{-\pi}^{\pi} (1-\nu\cos y)(1+\lambda\sin 2(y-\nu\sin y)) f_0(y)dy \\ &= \frac{1}{1-\nu\alpha_1}\int_{-\pi}^{\pi} (1-\nu\cos y) f_0(y)dy = 1,\end{aligned}$$

where $t'_{\lambda,\nu}$ is the derivative of function $t_{\lambda,\nu}$. Here, we assume without loss of generality that $\mu = 0$. We can see that the normalizing constant of the resulting function $f_0(t^{-1}_{\lambda,\nu}(\theta))$ becomes $1/(1-\nu\alpha_1)$ from above expression. To see whether the proposed distributions have mode and anti-mode at μ and $\mu-\pi$, we investigate the derivative of the pdf given by

$$f'(\theta) = \frac{dt^{-1}_{\lambda,\nu}(\theta)}{d\theta} f'_0(t^{-1}_{\lambda,\nu}(\theta)).$$

As the inverse function $t^{-1}_{\lambda,\nu}(\cdot)$ is monotone increasing and the base density f_0 is unimodal, the resulting pdf f is the unimodal. As $t_{\lambda,\nu}(0) = 0$ and $t_{\lambda,\nu}(\pm\pi) = \pm\pi$, the mode and anti-mode are given by

$$f(0) = \frac{1}{1-\nu\alpha_1} f_0(0)$$

and

$$f(\pm\pi) = \frac{1}{1-\nu\alpha_1} f_0(\pm\pi),$$

respectively. Therefore, the values of the density at the mode and anti-mode are invariant under transformation $t^{-1}_{\lambda,\nu}(\theta)$. From the above expression, the parameter ν controls the peakedness around the mode of the base symmetric distribution, and as ν approaches 1, the mode of the density becomes large because $\alpha_1 < 1$.

Recall that the density function of the von Mises distribution is given by

$$f_0^{(\mathrm{VM})}(\theta-\mu) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa\cos(\theta-\mu)], \quad -\pi \le \theta < \pi, \tag{3}$$

where $\mu \in [-\pi, \pi)$ denotes the location, and $\kappa \in (0, \infty)$ denotes the concentration parameter of the distributions. The function $I_p(\cdot)$ denotes the modified Bessel functions of the pth order, which is defined as

$$I_p(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} \cos p\theta\, e^{\kappa\cos\theta} d\theta = \sum_{r=0}^{\infty} \frac{1}{\Gamma(p+r+1)r!}\left(\frac{\kappa}{2}\right)^{2r+p}.$$

The density function of the wrapped Cauchy distribution is given by

$$f_0^{(\mathrm{WC})}(\theta-\mu) = \frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta-\mu)}, \quad -\pi \le \theta < \pi, \tag{4}$$

where $\mu \in [-\pi, \pi)$ and $\rho \in (0, 1)$ indicate the location and concentration parameters of the distribution, respectively. Substituting these densities for $f_0(\theta-\mu)$ in (2), the densities of the extended inverse transformation of von Mises (EIVM) and extended inverse transformation of wrapped Cauchy (EIWC) distributions are given by

$$f^{(\mathrm{EIVM})}(\theta) = \frac{1}{2\pi(I_0(\kappa) - \nu I_1(\kappa))} \exp\left[\kappa\cos\{t^{-1}_{\lambda,\nu}(\theta-\mu)\}\right], \quad -\pi \le \theta < \pi \tag{5}$$

and

$$f^{(\mathrm{EIWC})}(\theta) = \frac{1}{2\pi(1-\nu\rho)}\frac{1-\rho^2}{1+\rho^2-2\rho\cos\{t^{-1}_{\lambda,\nu}(\theta-\mu)\}}, \quad -\pi \le \theta < \pi, \tag{6}$$

respectively.

Some density functions for the EIVM and EIWC distributions are plotted in Figs. 2.1 and 2.2, respectively. Here, we choose the parameters as $\mu = 0$, $\kappa = 2$, $\nu = 0$, and $\lambda \in \{-0.9, -0.5, 0, 0.5\}$ for the EIVM distributions in Fig. 2.1. For Fig. 2.2, the parameters are chosen such that $\mu = 0, \lambda = 0.5, \rho = 0.5$, and the peakedness parameter is set at $\nu \in \{0.9, 0.5, 0, -0.5, -0.9\}$ for the EIWC distribution. These figures illustrate the effects of the parameters λ and ν of the inverse transformation function, which result in skewed and sharply peaked/flat-topped distributions.

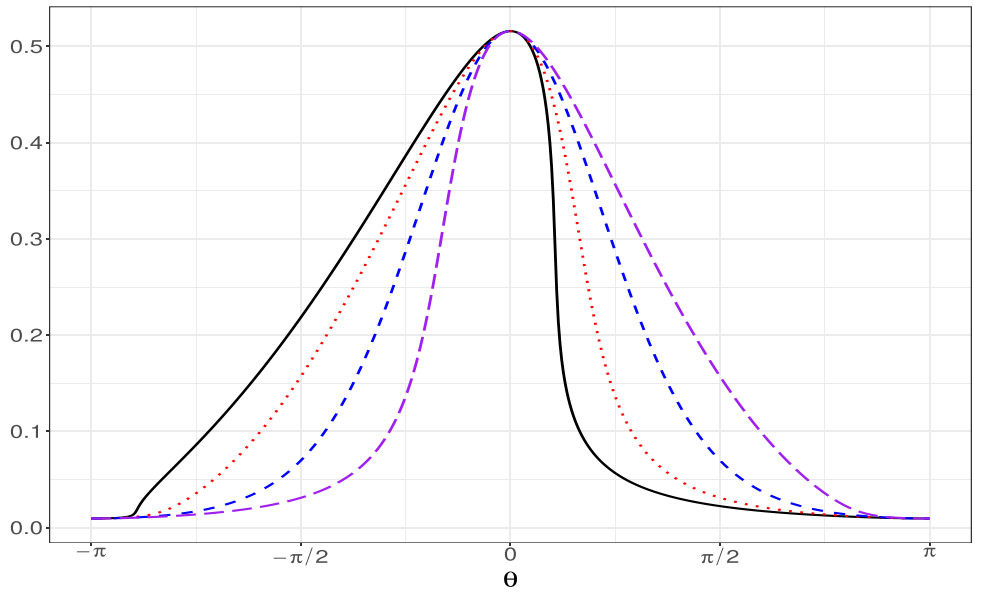

Fig. 2.1. Plots of the density functions for the extended inverse transformation of von Mises distributions with $\mu = \nu = 0$ and $\kappa = 2$. The skewness parameter λ is taken as $-0.9, -0.5, 0$, and 0.5, which correspond to the solid, dotted, dashed, and long-dashed lines, respectively.

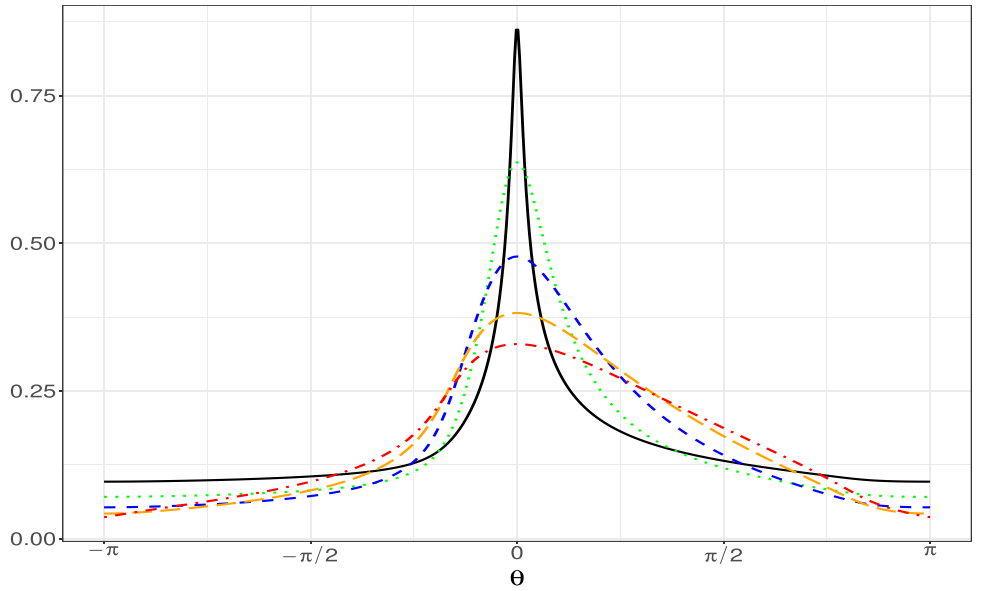

Fig. 2.2. Plots of the density functions for the extended inverse transformation of wrapped Cauchy distributions with $\mu = 0$, $\rho = 0.5$, and $\lambda = 0.5$. The peakedness parameter ν is taken as $0.9, 0.5, 0, -0.5$, and -0.9, which correspond to the solid, dotted, dashed, long-dashed, and dot-dashed lines, respectively.

The cumulative distribution function of the density (2) becomes

$$F(\theta) = \int_{-\pi}^{\theta} f(\psi)d\psi = \int_{-\pi}^{\theta} f_0(t_{0,\nu}^{-1}(s_\lambda^{-1}(\psi)))d\psi = \int_{-\pi}^{s_\lambda^{-1}(\theta)} s'_\lambda(u) f_0(t_{0,\nu}^{-1}(u))du$$
$$= \int_{-\pi}^{s_\lambda^{-1}(\theta)} f_0(t_{0,\nu}^{-1}(u))du + \int_{-\pi}^{s_\lambda^{-1}(\theta)} \lambda \sin 2u f_0(t_{0,\nu}^{-1}(u))du, \tag{7}$$

where we take $\mu = 0$ for simplicity. The explicit expression for the second term in (7) becomes rather complicated, whereas we can obtain the cumulative distribution function for the symmetric case ($\lambda = 0$) as stated in the following theorem.

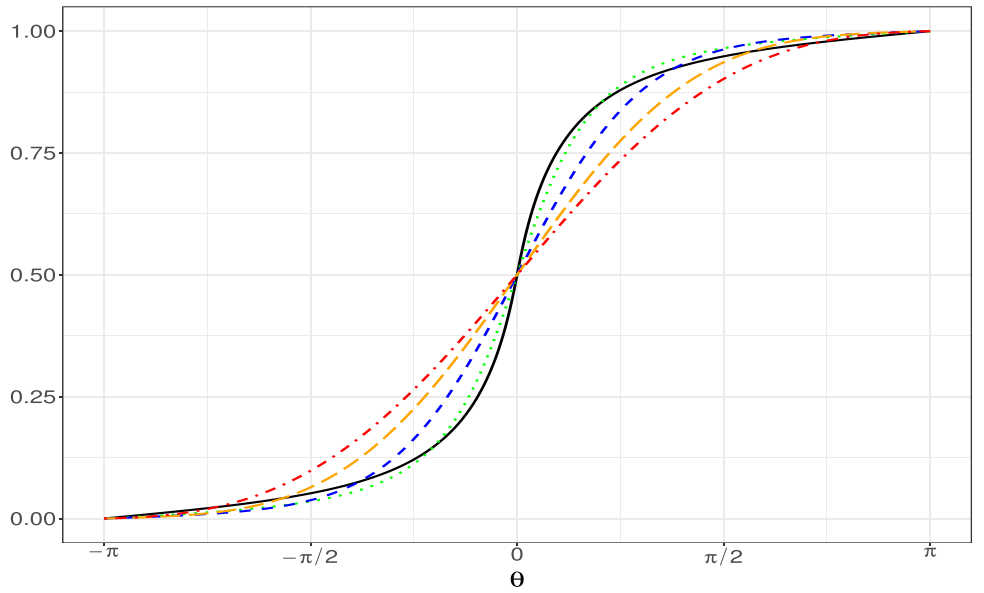

Fig. 2.3. Plots of the distribution functions for symmetric extended inverse transformation of von Mises distributions with $\mu = \lambda = 0$ and $\kappa = 2$. The peakedness parameter ν is taken as $0.9, 0.5, 0, -0.5$, and -0.9, which correspond to the solid, dotted, dashed, long-dashed, and dot-dashed lines, respectively.

Theorem 1. *For $-\pi \le \theta < \pi$ and $\lambda = 0$, the cumulative distribution functions for the EIVM and EIWC densities are given as*

$$F^{(\mathrm{EIVM})}(\theta) = \frac{t_{0,\nu}^{-1}(\theta-\mu)+\pi}{2\pi} + \frac{1}{\pi}\sum_{p=1}^{\infty}\frac{I_p(\kappa)-\nu I_p'(\kappa)}{I_0(\kappa)-\nu I_0'(\kappa)}\frac{1}{p}\sin(pt_{0,\nu}^{-1}(\theta-\mu))$$

and

$$F^{(\mathrm{EIWC})}(\theta) = \frac{1}{1-\nu\rho}\left\{\left(1-\nu\frac{1+\rho^2}{2\rho}\right)F_0^{(\mathrm{WC})}(t_{0,\nu}^{-1}(\theta-\mu)) + \nu\frac{1-\rho^2}{4\pi\rho}(t_{0,\nu}^{-1}(\theta-\mu)+\pi)\right\},$$

respectively. Here, $F_0^{(\mathrm{WC})}$ is the cumulative distribution function of the base wrapped Cauchy distribution given by

$$F_0^{(\mathrm{WC})}(\theta) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{1+\rho}{1-\rho}\tan\frac{\theta}{2}\right).$$

Some distribution functions for the symmetric EIVM and EIWC distributions are plotted in Figs. 2.3 and 2.4, respectively. Here, we choose the parameters as $\mu = 0$, $\kappa = 2$, and $\lambda = 0$ for the EIVM distribution, and set $\mu = 0, \lambda = 0$, and $\rho = 0.5$ for those with an EIWC distribution. The peakedness parameters change according to $\nu \in \{0.9, 0.5, 0, -0.5, -0.9\}$. The distribution function illustrates the effects of the peakedness parameters on the probability distribution function. As ν becomes large, the probability around anti-mode of the distribution becomes large for the EIWC distribution. The concentration around its center is more condensed for the EIWC distributions than for the EIVM distributions.

Let U be the uniform random variable on [0,1]. The following result for the random samples adheres to the extended inverse transformation of scale distributions.

Theorem 2. *Let f_0 be the arbitrary symmetric density function on the circle that satisfies Assumption 1, and let a random variable Ψ have a density function f_0. Then, the random variable Θ follows from the extended inverse transformation of scale distributions, namely,*

$$\Theta = s_\lambda(\Psi)I\left(U \le \frac{s_\lambda'(\Psi)}{2}\right) + s_\lambda(-\Psi)I\left(U > \frac{s_\lambda'(\Psi)}{2}\right) \sim f_0(s_\lambda^{-1}(\theta)),$$

where $I(\cdot)$ is an indicator function.

The proofs of Theorems 1 and 2 are given in Appendix A.

For the random number generation of the EIVM and EIWC distributions, we adopt the following two steps (see also Jones and Pewsey, 2012). The proposed random number generation method does not rely on the acceptance-rejection sampling method. First, we draw samples Ψ from the symmetric sharply peaked and flat-topped distributions for the EIVM and

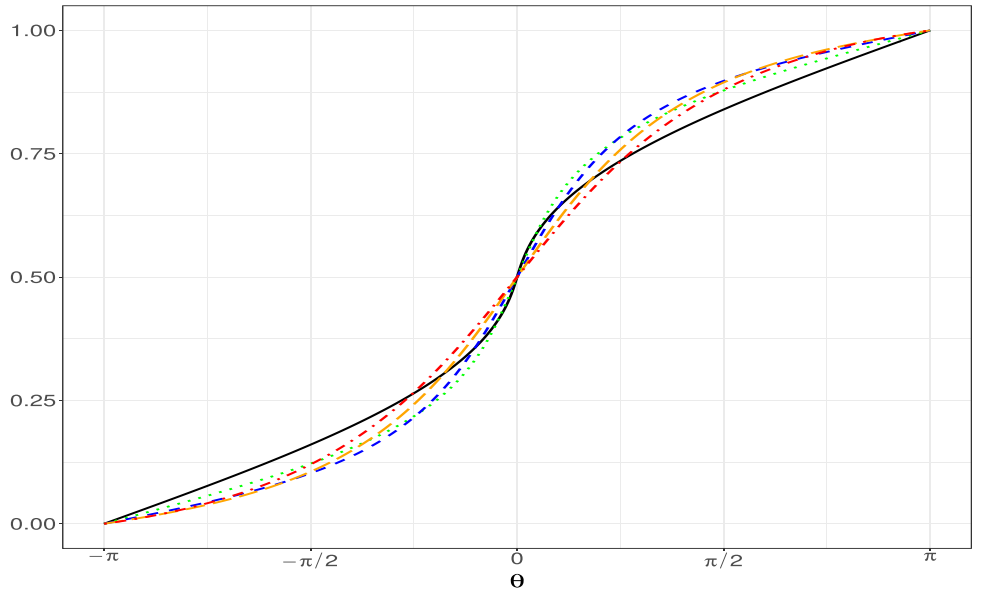

Fig. 2.4. Plots of the distribution functions for symmetric extended inverse transformation of wrapped Cauchy distributions with $\mu = \lambda = 0$ and $\rho = 0.5$. The peakedness parameter ν is taken as $0.9, 0.5, 0, -0.5$, and -0.9, which correspond to the solid, dotted, dashed, long-dashed, and dot-dashed lines, respectively.

EIWC by using the inverse transform sampling method; that is, we set $\lambda = 0$. For this step, we provide the explicit distribution function for random variable Ψ that have density function $f_0(t_{0,\nu}^{-1}(\theta - \mu))/(1 - \nu\alpha_1)$ in Theorem 1. Next, we draw independent random samples from $U[0, 1]$ and obtain Θ from the transformation

$$\Theta = \begin{cases} s_\lambda(\Psi) & \text{if } U \le s'_\lambda(\Psi)/2, \\ s_\lambda(-\Psi) & \text{if } U > s'_\lambda(\Psi)/2. \end{cases}$$

3. Parameter estimation

In this section, we describe four methods to estimate the unknown model parameters. Let $\boldsymbol{\theta}_n = (\theta_1, \theta_2, \ldots, \theta_n)$ be a random sample of size n from the extended inverse transformation of scale distribution with base density $f_0(\theta - \mu)$ given in (2). We restrict the family of base density $f_0(\theta - \mu)$ to have two parameters with location and concentration. We denote $\boldsymbol{\gamma} = (\mu, \kappa)^T$ by the parameter vector of base density function $f_0(\theta - \mu)$ and denote $\boldsymbol{\Gamma}$ by the corresponding parameter space. The first cosine moment of f_0 is determined by a single concentration parameter κ, and we rewrite α_1 as $\alpha_1(\kappa)$. Then, the model parameters are denoted by $\boldsymbol{\eta} = (\boldsymbol{\gamma}^T, \lambda, \nu)^T \in \boldsymbol{H}$, where $\boldsymbol{H}$ is the parameter space that is $\boldsymbol{H} = \boldsymbol{\Gamma} \times [-1 + \delta_\lambda, 1 - \delta_\lambda] \times [-1 + \delta_\nu, 1 - \delta_\nu]$ with $\delta_\lambda > 0$ and $\delta_\nu > 0$. For the EIVM distribution, $\boldsymbol{\eta}$ becomes $\boldsymbol{\eta}_{\text{EIVM}} = (\mu, \kappa, \lambda, \nu)^T \in \boldsymbol{H}$, where $\boldsymbol{H} = [-\pi, \pi - \delta_\mu] \times [\delta_\kappa, 1/\delta_\kappa] \times [-1 + \delta_\lambda, 1 - \delta_\lambda] \times [-1 + \delta_\nu, 1 - \delta_\nu]$ with $\delta_\mu > 0$ and $\delta_\kappa \in (0, 1)$ being constants that are small enough so that the true κ belongs to $\mathbb{R}^+$. For the EIWC distribution $\boldsymbol{\eta}_{\text{EIWC}} = (\mu, \rho, \lambda, \nu)^T \in \boldsymbol{H}$, where $\boldsymbol{H} = [-\pi, \pi - \delta_\mu] \times [\delta_\rho, 1 - \delta_\rho] \times [-1 + \delta_\lambda, 1 - \delta_\lambda] \times [-1 + \delta_\nu, 1 - \delta_\nu]$ and $\delta_\rho \in (0, 1/2)$ is a small positive number.

The log-likelihood function of the observed sample can be written as

$$\ell(\boldsymbol{\eta}) = \sum_{i=1}^{n} \log f(\theta_i; \boldsymbol{\eta}) = \sum_{i=1}^{n} \log f_0(t_{\lambda,\nu}^{-1}(\theta_i - \mu); \kappa) - n \log(1 - \nu\alpha_1(\kappa)). \tag{8}$$

Then, the MLE $\hat{\boldsymbol{\eta}}_n^{(\text{MLE})}$ is defined by

$$\hat{\boldsymbol{\eta}}_n^{(\text{MLE})} = \underset{\eta}{\text{argmax}}\ \ell(\boldsymbol{\eta}).$$

To obtain the asymptotic normality of the MLE, the following regularity conditions are required.

Assumption 2. Let $\boldsymbol{\eta}_0$ be the true parameter vector. The following regularity conditions hold:

B1: The true value of the parameter $\boldsymbol{\eta}_0$ belongs to the interior of $\boldsymbol{H}$, and $\boldsymbol{H}$ is compact

B2: For symmetric base density $f_0(\theta - \mu; \kappa)$, where $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}$ is compact, there exist constants $M > 0$ and $m > 0$ such that for any $\theta \in [-\pi, \pi)$,

$$\sup_{\gamma \in \Gamma} f_0(\theta - \mu; \kappa) < M \quad \text{and} \quad \inf_{\gamma \in \Gamma} f_0(\theta - \mu; \kappa) > m.$$

It must hold that the first cosine moment of $f_0(\theta-\mu;\kappa)$ is uniformly less than unity in κ, that is, $\sup_\kappa \alpha_1(\kappa)<1$.

B3: The first three derivatives of base density $f_0(\theta-\mu;\kappa)$ and its cosine moment $\alpha_1(\kappa)$ with respect to the parameters are continuous and bounded for all θ and $\boldsymbol{\gamma}\in\boldsymbol{\Gamma}$.

B4: A Fisher information matrix exists and is nonsingular, and it is defined by

$$I(\boldsymbol{\eta}_0)=E\left[-\frac{\partial \log f(\Theta;\boldsymbol{\eta})}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}\right]_{\eta=\eta_0}=[\iota_{ij}]_{i,j\in\{\mu,\kappa,\lambda,\nu\}}.$$

The condition for non-singularity of the Fisher information matrix is given by

$$\det\{\boldsymbol{I}(\boldsymbol{\eta}_0)\}=(\iota_{\mu\mu}\iota_{\lambda\lambda}-\iota_{\mu\lambda}^2)(\iota_{\kappa\kappa}\iota_{\nu\nu}-\iota_{\kappa\nu}^2)\neq 0.$$

B5: For model identifiability, $\boldsymbol{\eta}_1\neq\boldsymbol{\eta}_2$ indicates $f(\theta;\boldsymbol{\eta}_1)\neq f(\theta;\boldsymbol{\eta}_2)$.

We do not have closed forms of the trigonometric moments, and hence it is quite challenging to verify the identifiability of B5. One certainty of the identifiability is that the density $f(\theta;\boldsymbol{\eta})$ is identifiable with respect to the parameter μ because the mode is given by μ uniquely.

The expressions for the value $\iota_{\kappa\kappa}\iota_{\nu\nu}-\iota_{\kappa\nu}^2$ become quite complex, while we can confirm that the value approaches 0 as concentration and peakedness parameters $\kappa\to 0$ and $\nu\to\pm 1$ with numerical computation.

The first result states the consistency and asymptotic normality of $\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})}$. The proof of this theorem is outlined in Appendix A.

Theorem 3. *(Consistency and asymptotic normality) Suppose that Assumption 1 and Conditions B1, B2, and B5 hold. Then,*

$$\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})}\xrightarrow[p]{}\boldsymbol{\eta}_0,\quad \text{as}\quad n\to\infty.$$

Under Assumptions 1 and 2, the asymptotic normality holds for MLE $\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})}$ as $n\to\infty$,

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})}-\boldsymbol{\eta}_0)\xrightarrow[d]{}N(\mathbf{0},\boldsymbol{I}(\boldsymbol{\eta}_0)^{-1}).$$

Fisher information matrix $\boldsymbol{I}(\boldsymbol{\eta}_0)$ becomes

$$\boldsymbol{I}(\boldsymbol{\eta}_0)=\begin{pmatrix}\iota_{\mu\mu} & 0 & \iota_{\mu\lambda} & 0\\ 0 & \iota_{\kappa\kappa} & 0 & \iota_{\kappa\nu}\\ \iota_{\lambda\mu} & 0 & \iota_{\lambda\lambda} & 0\\ 0 & \iota_{\nu\kappa} & 0 & \iota_{\nu\nu}\end{pmatrix},\tag{9}$$

where the off-diagonal elements are given by

$$\iota_{\mu\lambda}=\iota_{\lambda\mu}=\frac{1}{1-\nu\alpha_1}\int\left\{\left(-\frac{t'^{\lambda}(y)}{t'(y)}+\frac{t^{\lambda}(y)t''(y)}{t'(y)^2}\right)(\log f_0(y))'-\frac{t^{\lambda}(y)}{t'(y)}(\log f_0(y))''\right\}f_0(y)dy$$

and

$$\iota_{\kappa\nu}=\iota_{\nu\kappa}=\frac{1}{1-\nu\alpha_1}\int(-\sin y)\left(f_0'^{\kappa}(y)-\frac{f_0'(y)f_0^{\kappa}(y)}{f_0(y)}\right)dy-\left(\frac{\partial\alpha_1}{\partial\kappa}\right)\frac{1}{(1-\nu\alpha_1)^2}.$$

The diagonal elements are given by

$$\iota_{\mu\mu}=\frac{1}{1-\nu\alpha_1}\int\left(\frac{t''(y)}{t'(y)^2}f_0'(y)-\frac{1}{t'(y)}\left(f_0''(y)-\frac{f_0'(y)^2}{f_0(y)}\right)\right)dy,$$

$$\iota_{\kappa\kappa}=-\frac{1}{1-\nu\alpha_1}\int(1-\nu\cos y)\left(f_0^{\kappa\kappa}(y)-\frac{f_0^{\kappa}(y)^2}{f_0(y)}\right)dy-\nu\left(\frac{1}{1-\nu\alpha_1}\frac{\partial^2\alpha_1}{\partial\kappa^2}+\frac{\nu}{(1-\nu\alpha_1)^2}\left(\frac{\partial\alpha_1}{\partial\kappa}\right)^2\right),$$

$$\iota_{\lambda\lambda}=\frac{1}{1-\nu\alpha_1}\int\left(\left(-2\frac{t^{\lambda}(y)t'^{\lambda}(y)}{t'(y)}+\frac{t^{\lambda}(y)^2t''(y)}{t'(y)^2}\right)f_0'(y)-\frac{t^{\lambda}(y)^2}{t'(y)}\left(f_0''(y)-\frac{f_0'(y)^2}{f_0(y)}\right)\right)dy,$$

and

$$\iota_{\nu\nu}=\frac{1}{1-\nu\alpha_1}\int\left(\left(-\frac{\sin 2y}{1-\nu\cos y}+\frac{\nu\sin^3 y}{(1-\nu\cos y)^2}\right)f_0'(y)-\frac{\sin^2 y}{1-\nu\cos y}\left(f_0''(y)-\frac{f_0'(y)^2}{f_0(y)}\right)\right)dy-\frac{\alpha_1^2}{(1-\nu\alpha_1)^2}.$$

The other elements in the Fisher information matrix become zero, that is, $\iota_{\mu\kappa}=\iota_{\kappa\mu}=\iota_{\lambda\nu}=\iota_{\nu\lambda}=0$.

The derivation of the Fisher information matrix is given in Appendix A. From this result, the pairs between (μ,λ) and (κ,ν) always provide independent relations.

The MLE of the vector of unknown parameters $\boldsymbol{\eta}=(\mu,\kappa,\lambda,\nu)^T$ is not given in a closed form, and hence numerical optimization algorithms are used for obtaining the model parameters. As some of the model parameters are restricted in

the parameter space, usual non-constrained optimization methods, such as quasi-Newton methods, do not work. To cope with this problem, we apply the following transformations of the model parameters. It is possible to obtain parameter estimates using constrained optimization algorithms, such as sequential quadratic programming; however, this requires a large amount of computational time. From the point of view of computational cost, unconstrained optimization algorithms are recommended.

Let $\boldsymbol{\phi} = (\phi_\mu, \phi_\kappa, \phi_\lambda, \phi_\nu)^T \in \mathbb{R}^4$ be the parameters in Euclidean space, and let $h : \mathbb{R}^4 \to \boldsymbol{H}$ be a suitable transformation function with continuous first derivatives. Then, $h(\boldsymbol{\phi}) = (h_1(\phi_\mu), h_2(\phi_\kappa), h_3(\phi_\lambda), h_4(\phi_\nu))^T$. As an example of function h, we parametrize the model in terms of $\boldsymbol{\eta} = (\pi \tanh(\phi_\mu), \exp(\phi_\kappa), \tanh(\phi_\lambda), \tanh(\phi_\nu))^T$ for the EIVM distributions. If we consider the EIWC distribution, $h_2(\phi_\kappa)$ could be $(\tanh(\phi_\kappa)+1)/2$, such that $h_2(\phi_\kappa) \in (0,1)$. Let $\boldsymbol{\phi}_0 = h^{-1}(\boldsymbol{\eta}_0)$ be the true parameter vector of the transformation and $\hat{\boldsymbol{\phi}}_n^{(\mathrm{MLE})} := \mathrm{argmax}_\phi\, \ell(h(\boldsymbol{\phi}))$ be the corresponding MLE. If $\hat{\boldsymbol{\phi}}_n^{(\mathrm{MLE})}$ has asymptotic normality $\sqrt{n}(\hat{\boldsymbol{\phi}}_n^{(\mathrm{MLE})} - \boldsymbol{\phi}) \to_d N(\mathbf{0}, \boldsymbol{\Sigma}_\phi)$ and $h'(\boldsymbol{\phi}_0) \neq \mathbf{0}$ holds, then by using the delta method, the MLE of $\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})} = h(\hat{\boldsymbol{\phi}}_n^{(\mathrm{MLE})})$ follows the asymptotically normal distribution as

$$\sqrt{n}(h(\hat{\boldsymbol{\phi}}_n^{(\mathrm{MLE})}) - h(\boldsymbol{\phi}_0)) \underset{d}{\to} N(\mathbf{0}, h'(\boldsymbol{\phi}_0)\boldsymbol{\Sigma}_\phi h'(\boldsymbol{\phi}_0)^T),$$

where

$$h'(\boldsymbol{\phi}) = \begin{pmatrix} \dfrac{\partial h_1}{\partial \phi_\mu} & \cdots & \dfrac{\partial h_1}{\partial \phi_\nu} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial h_4}{\partial \phi_\mu} & \cdots & \dfrac{\partial h_4}{\partial \phi_\nu} \end{pmatrix}.$$

Recall that $I(\boldsymbol{\eta}_0)^{-1} = h'(\boldsymbol{\phi}_0)\boldsymbol{\Sigma}_\phi h'(\boldsymbol{\phi}_0)^T$. Using this parametrization, we can use the standard unconstrained numerical optimization techniques to obtain the MLE $\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})}$. These techniques are implemented for estimating the BEs of the model parameters.

As the parameter space $\boldsymbol{H}$ is restricted for each parameter, estimating model parameters using Bayesian methods provides powerful statistical tools for obtaining the posterior distribution of the model parameters and asymmetric credible intervals. However, the classical MLE provides asymptotically normal distributions for the estimated model parameters. In Bayesian analyses, the choice of the prior is inevitable. If there is enough evidence on the part of the researchers, then an informative prior can be chosen; otherwise, a non-informative prior will be more appropriate. In this proposed model, a joint conjugate prior for the parameters does not exist, and hence we consider the priors for $\boldsymbol{\phi}$ as follows:

$$p_\Phi(\boldsymbol{\phi}) = N_4(\boldsymbol{\mu}_{\phi_0}, \boldsymbol{\Sigma}_{\phi_0}),$$

where $\boldsymbol{\mu}_{\phi_0}$ and $\boldsymbol{\Sigma}_{\phi_0}$ are the mean vector and covariance matrix of the four variate normal distributions, respectively. Recall that our model parameters of interest are obtained by $\boldsymbol{\eta}$. Then, the BE under a squared loss function is given by

$$\hat{\boldsymbol{\eta}}^{(\mathrm{BE})} = E(\boldsymbol{\eta}|\boldsymbol{\theta}_n) = \frac{\int_{\boldsymbol{H}} \boldsymbol{\eta} \exp\{\ell(\boldsymbol{\eta})\} p(\boldsymbol{\eta}) d\boldsymbol{\eta}}{\int_{\boldsymbol{H}} \exp\{\ell(\boldsymbol{\eta})\} p(\boldsymbol{\eta}) d\boldsymbol{\eta}}. \tag{10}$$

Then, the prior density of $\boldsymbol{\eta}$ is given by $p(\boldsymbol{\eta}) = p_\Phi(h^{-1}(\boldsymbol{\eta}))|(\partial/\partial\boldsymbol{\eta})h^{-1}(\boldsymbol{\eta})|$, where $|(\partial/\partial\boldsymbol{\eta})h^{-1}(\boldsymbol{\eta})|$ is the determinant of the Jacobian matrix of the transformation.

Since the domain of all integrals in (10) is $\boldsymbol{H}$, it is suitable to use $p(\boldsymbol{\eta})/K$ as a prior density of $\boldsymbol{\eta}$ with $K = \int_{\boldsymbol{H}} p(\boldsymbol{\eta}) d\boldsymbol{\eta}$. However, for each $i \in \{\mu, \kappa, \lambda, \nu\}$, the BE of $g^i(\boldsymbol{\eta})$ becomes

$$\frac{\int_{\boldsymbol{H}} g^i(\boldsymbol{\eta}) \exp\{\ell(\boldsymbol{\eta}) + \log(p(\boldsymbol{\eta})/K)\} d\boldsymbol{\eta}}{\int_{\boldsymbol{H}} \exp\{\ell(\boldsymbol{\eta}) + \log(p(\boldsymbol{\eta})/K)\} d\boldsymbol{\eta}} = \frac{\int_{\boldsymbol{H}} g^i(\boldsymbol{\eta}) \exp\{\ell(\boldsymbol{\eta}) + \log p(\boldsymbol{\eta})\} d\boldsymbol{\eta}}{\int_{\boldsymbol{H}} \exp\{\ell(\boldsymbol{\eta}) + \log p(\boldsymbol{\eta})\} d\boldsymbol{\eta}},$$

where $g^\mu(\boldsymbol{\eta}) = \mu$, $g^\kappa(\boldsymbol{\eta}) = \kappa$, $g^\lambda(\boldsymbol{\eta}) = \lambda$, and $g^\nu(\boldsymbol{\eta}) = \nu$. Therefore, normalizing constant K can be neglected.

As the numerator and denominator involve a four-dimensional integration to obtain the BE, we adopt Lindley's approximation for practical purposes. Instead of estimating the approximate BE, we also provide a maximum a posteriori (MAP) estimator of the posterior distributions.

For estimating the parameters in the EIVM distribution in a Bayesian context, if we set the parameters in prior distribution $p_\Phi(\boldsymbol{\phi})$ as $\boldsymbol{\Sigma}_{\phi_0} = \mathrm{diag}(\sigma_\mu^2, \sigma_\kappa^2, \sigma_\lambda^2, \sigma_\nu^2)$ and $\boldsymbol{\mu}_{\phi_0} = (m_\mu, m_\kappa, m_\lambda, m_\nu)^T$, then the joint prior becomes $p(\mu, \kappa, \lambda, \nu) = p(\mu)p(\kappa)p(\lambda)p(\nu)$, where

$$p(\mu) = \frac{1}{\sqrt{2\pi}\sigma_\mu} \exp\left\{-\frac{1}{2\sigma_\mu^2}(\tanh^{-1}(\mu/\pi) - m_\mu)^2\right\} \cdot \frac{1}{\pi(1-(\mu/\pi)^2)}, \qquad \mu \in (-\pi, \pi),$$

$$p(\kappa) = \frac{1}{\sqrt{2\pi}\sigma_\kappa \kappa} \exp\left\{-\frac{1}{2\sigma_\kappa^2}(\log(\kappa) - m_\kappa)^2\right\}, \quad \kappa \in \mathbb{R}^+,$$

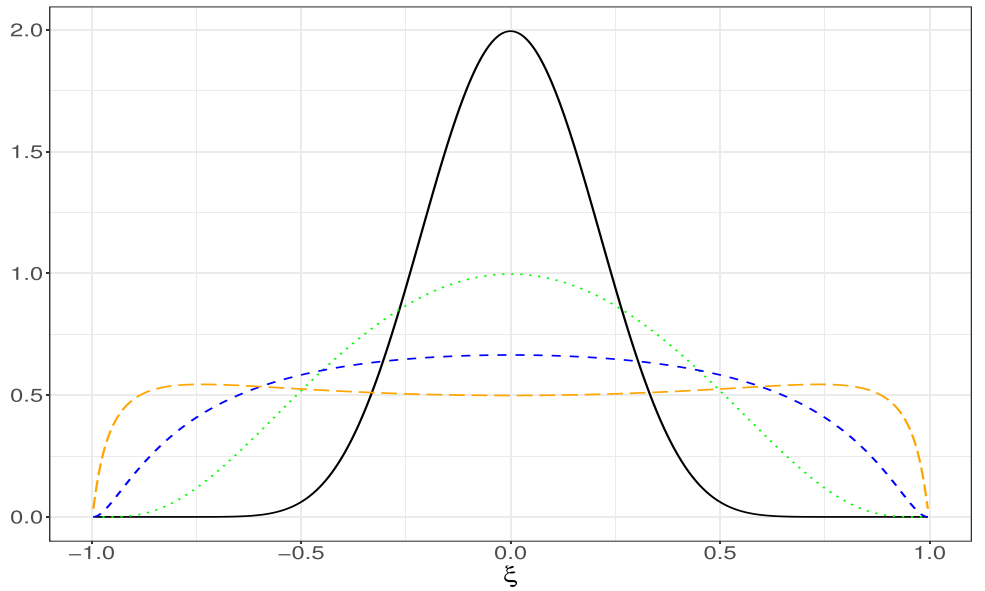

Fig. 3.1. Plots of the prior density functions for the skewness or peakedness parameter ξ, where we set $m_\xi = 0$ and $\sigma_\xi^2 \in \{0.2, 0.4, 0.6, 0.8\}$, which correspond to the solid, dotted, dashed, and long-dashed lines, respectively.

$$p(\lambda) = \frac{1}{\sqrt{2\pi}\sigma_\lambda} \exp\left\{-\frac{1}{2\sigma_\lambda^2}(\tanh^{-1}(\lambda) - m_\lambda)^2\right\} \cdot \frac{1}{(1-\lambda^2)}, \quad \lambda \in (-1, 1),$$
$$\text{and } p(\nu) = \frac{1}{\sqrt{2\pi}\sigma_\nu} \exp\left\{-\frac{1}{2\sigma_\nu^2}(\tanh^{-1}(\nu) - m_\nu)^2\right\} \cdot \frac{1}{(1-\nu^2)}, \quad \nu \in (-1, 1). \tag{11}$$

When the base density is the wrapped Cauchy density, we can use the prior for the concentration parameter ρ instead of using $p(\kappa)$ as

$$p(\rho) = \frac{1}{\sqrt{2\pi}\sigma_\rho} \exp\left\{-\frac{1}{2\sigma_\rho^2}(\tanh^{-1}(2\rho - 1) - m_\rho)^2\right\} \cdot \frac{2}{(1-(2\rho-1)^2)}, \quad \rho \in (0, 1). \tag{12}$$

We consider the BE based on weak informative prior distributions to avoid boundary estimates of the model parameters. BEs to avoid boundary estimates were investigated by Galindo-Garre et al. (2004) in log-linear models and by Galindo-Garre and Vermunt (2006) for latent class analyses.

Fig. 3.1 illustrates the prior density functions for λ and ν whose parameter space is restricted in $(-1, 1)$. Denote $\xi \in \{\lambda, \nu\}$. For the parameters in normal distribution $p(\xi) = N(m_\xi, \sigma_\xi^2)$, we choose $m_\xi = 0$ and $\sigma_\xi^2 \in \{0.2, 0.4, 0.6, 0.8\}$. As is evident in this figure, as variance σ_ξ^2 approaches 1, the density function becomes flat, which indicates that the non-informativeness together with boundary-avoiding properties are confirmed for $\sigma_\xi < 1$.

For the sequel, we introduce a MAP estimator and Lindley's approximation for the BE.

3.1. MAP estimation

All the information in Bayesian inferences is contained in the posterior distributions. It is sometimes difficult to obtain the mean or median of the posterior distributions. The MAP estimation can be used to obtain a point estimate with a shrinkage rule. Given an observation $\boldsymbol{\theta}_n$, the posterior distribution of $\boldsymbol{\eta}$ is proportional to $\exp\{\ell(\boldsymbol{\eta})\} \cdot p(\boldsymbol{\eta})$. Let $\varphi(\boldsymbol{\eta}|\boldsymbol{\theta}_n) = \ell(\boldsymbol{\eta}) + \log(p(\boldsymbol{\eta}))$, where $p(\boldsymbol{\eta})$ is the prior distribution of $\boldsymbol{\eta}$ defined by (11) and (12). Then, the MAP estimator is defined as

$$\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})} = \operatorname*{argmax}_{\eta}\left\{\varphi(\boldsymbol{\eta}|\boldsymbol{\theta}_n)\right\}.$$

For the asymptotic normality of the MAP estimator, we need the following conditions.

Assumption 3. The following regularity conditions of $p(\boldsymbol{\eta})$ hold:

C1: $p(\cdot)$ is positive almost everywhere on $\boldsymbol{H}$ and $p(\boldsymbol{\eta}_0) > 0$.
C2: $p(\cdot)$ is continuously differentiable with respect to $\boldsymbol{\eta}$, which indicates that $p(\cdot)$ is continuous on $\boldsymbol{H}$.

The prior densities given by Eqs. (11) and (12) satisfy Assumption 3. Then, we have the following result.

Theorem 4. *(Consistency and asymptotic normality) Suppose that Assumptions 1 and 3 as well as Conditions B1, B2, and B5 hold. Then, we have*

$$\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})} \underset{p}{\to} \boldsymbol{\eta}_0 \quad as \quad n \to \infty.$$

Under Assumptions B1–B3, asymptotic normality holds for MAP estimator $\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})}$ as $n \to \infty$, that is,

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})} - \boldsymbol{\eta}_0) \underset{d}{\to} N(\mathbf{0}, \boldsymbol{I}(\boldsymbol{\eta}_0)^{-1}).$$

3.2. Lindley's approximation for BEs

It may be observed from (10) that the posterior mean takes a ratio form that involves integration in both the numerator and denominator that cannot be reduced to a closed form. Let $\rho(\boldsymbol{\eta})$ be the log-prior density function as $\rho(\boldsymbol{\eta}) = \log p(\boldsymbol{\eta})$; then, the posterior expectation in (10) can be written as

$$E(\boldsymbol{\eta}|\boldsymbol{\theta}_n) = \left(\frac{\int g^{\mu}(\boldsymbol{\eta}) \exp\{\ell(\boldsymbol{\eta}) + \rho(\boldsymbol{\eta})\} d\boldsymbol{\eta}}{\int \exp\{\ell(\boldsymbol{\eta}) + \rho(\boldsymbol{\eta})\} d\boldsymbol{\eta}}, \cdots, \frac{\int g^{\nu}(\boldsymbol{\eta}) \exp\{\ell(\boldsymbol{\eta}) + \rho(\boldsymbol{\eta})\} d\boldsymbol{\eta}}{\int \exp\{\ell(\boldsymbol{\eta}) + \rho(\boldsymbol{\eta})\} d\boldsymbol{\eta}} \right)^T.$$

Once we obtain the value $\hat{\boldsymbol{\phi}}_n^{(\mathrm{MLE})}$, Lindley's approximation using $\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})} = h(\hat{\boldsymbol{\phi}}_n^{(\mathrm{MLE})})$ can be used to compute the ratio of the integrals to obtain an approximate BE of the model parameter. Recall that $\ell(\boldsymbol{\eta})$ defined by (8) is the log-likelihood function. For convenience of exposition, we write $f_{i_1 \cdots i_s}(\boldsymbol{\eta}) = (\partial^s / \partial \eta_{i_1} \cdots \partial \eta_{i_s}) f(\boldsymbol{\eta})$ for the derivative of any function $f(\boldsymbol{\eta})$ of $\boldsymbol{\eta}$, for example, $\rho_j(\boldsymbol{\eta}) = (\partial / \partial \eta_j)\rho(\boldsymbol{\eta})$, $\ell_{ijk}(\boldsymbol{\eta}) = (\partial^3 / \partial \eta_i \partial \eta_j \partial \eta_k)\ell(\boldsymbol{\eta})$. If we evaluate these functions at an estimate $\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})}$ of $\boldsymbol{\eta}$, then we omit the symbol $(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})})$, for example, $\ell_{ijk} = \ell_{ijk}(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})})$ and $g_i^s(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})}) = g_i^s$. Let σ_{ij} denote the (i, j)-th element of the inverse of the second derivative of $-\ell(\boldsymbol{\eta})$ evaluated at $\boldsymbol{\eta} = \hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})}$.

Recall that $p(\boldsymbol{\eta})$ is the prior density of $\boldsymbol{\eta}$; then, using Lindley's approximation, the BE for the sth element of $\boldsymbol{\eta}$, that is, $E(\eta_s|\boldsymbol{\theta}_n)$, can be asymptotically estimated by

$$E(\eta_s|\boldsymbol{\theta}_n) = g^s(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})}) + \frac{1}{2}\sum_{i,j}\left[(g_{ij}^s + 2g_i^s\rho_j)\sigma_{ij}\right] + \frac{1}{2}\sum_{i,j,k,l}\ell_{ijk}\sigma_{ij}\sigma_{kl}g_l^s + O_p(n^{-2}), \tag{13}$$

where $s \in \{\mu, \kappa, \lambda, \nu\}$ and the second and third terms on the right-hand side of (13) are of order $O_p(n^{-1})$. If we need to compute the exact approximation in Eq. (13), evaluating the third term, $\sum_{i,j,k,l}\ell_{ijk}\sigma_{ij}\sigma_{kl}g_l^s$, requires quite a complicated calculation. DiCiccio and Stern (1993) provided a rather simple expression for these terms, which is given by

$$\sum_{i,j,k,l}\ell_{ijk}\sigma_{ij}\sigma_{kl}g_\ell^s = \sum_{r=1}^{4}\sum_{s=1}^{4}\left\{\frac{d}{dx}g^s(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})} + x\boldsymbol{\ell}_r)\right\}_{x=0}\left\{\frac{d}{dx}\sigma_{rs}(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})} + x\boldsymbol{\ell}_s)\right\}_{x=0},$$

where a numerical differentiation is used for approximations to the two derivatives in the foregoing brackets.

To derive the second-order approximation (13), we need to apply Laplace's method for the numerator and denominator of (10) under suitable conditions. Let $B_\varepsilon(\boldsymbol{\eta}_0)$ denote the open ball of radius ε centered at $\boldsymbol{\eta}_0$. The notation *a.s.* stands for "almost surely" under the true distribution. Here, we assume the following conditions.

Assumption 4.

D1: For all $\theta \in [-\pi, \pi)$ and $\boldsymbol{\eta} \in \boldsymbol{H}$, $f(\theta; \boldsymbol{\eta}) > 0$, and for all $\theta \in [-\pi, \pi)$, $f(\theta; \boldsymbol{\eta})$ is six times continuous differentiable with respect to $\boldsymbol{\eta}$.

D2: There exists a positive constant $c_1 > 0$ such that

$$\liminf_{n\to\infty} \det\left(-\frac{1}{n}\frac{\partial^2}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}\ell(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})})\right) > c_1, \qquad \text{a.s.}$$

D3: Prior $p(\boldsymbol{\eta})$ is four-times continuously differentiable with respect to $\boldsymbol{\eta}$, and the expectation of $\boldsymbol{\eta}$, $\int_{\boldsymbol{H}} \boldsymbol{\eta} p(\boldsymbol{\eta}) d\boldsymbol{\eta}$, exists and is finite.

D4: For all $\boldsymbol{\eta}_0 \in \boldsymbol{H}$, there exist $\varepsilon > 0$ and $M < \infty$ such that $B_\varepsilon(\boldsymbol{\eta}_0) \subseteq \boldsymbol{H}$, and for all $j_1, \ldots, j_d \in \{1, 2, 3, 4\}$ with $d \le 6$,

$$\limsup_{n\to\infty}\sup_{\eta}\left\{n^{-1}|\ell_{j_1\cdots j_d}(\boldsymbol{\eta})|\boldsymbol{\eta} \in B_\varepsilon(\boldsymbol{\eta}_0)\right\} < M, \quad \text{a.s.}$$

D5: For all $\boldsymbol{\eta}_0 \in \boldsymbol{H}$ and for all $\delta > 0$,

$$\limsup_{n\to\infty}\sup_{\eta}\left\{n^{-1}[\ell(\boldsymbol{\eta}) - \ell(\boldsymbol{\eta}_0)] : \boldsymbol{\eta} \in \boldsymbol{H} - B_\delta(\boldsymbol{\eta}_0)\right\} < 0, \quad \text{a.s.}$$

Conditions D4 and D5 are the same as conditions (ii) and (iv) of Kass et al. (1990, p. 483–484). Then, under Assumption 1, Conditions B1, B2, B5, C1, D1, and D3, Lindley's approximations to the BE (10) are derived. Condition D1 corresponds to condition (i) of (Kass et al., 1990, p. 483–484) and is stronger than Condition B3. Condition D3 corresponds to condition (i) of

(Kass et al., 1990, p. 476) and is stronger than Condition C2. Condition (ii) of Kass et al. (1990) ensures the uniform boundedness of higher-order derivatives of the standardized log-likelihood functions $(1/n)\ell_{j_1\cdots j_d}(\boldsymbol{\eta})$ $(d=1,\ldots,6)$ on a neighborhood of the true parameter, $\boldsymbol{\eta}_0$. This holds from Condition D1 and the compactness of $\boldsymbol{H}$. Although B5 is a familiar consistency condition for the MLE, this also holds from the boundedness Condition B2 and identifiability Condition B5. Condition D2 ensures that the posterior density is approximated by the multivariate normal distribution with a positive definite covariance matrix $\left(-\frac{1}{n}\frac{\partial^2}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T}\ell(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})})\right)^{-1}$. Hence, if the determinant of the Hessian of $\ell(\boldsymbol{\eta})$ evaluated at $\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})}$ is close to zero numerically, then the resulting Laplace and Lindley's approximations become unstable.

From the locations of zeros in the Fisher information matrix (9), we obtain

$$\sigma_{12}=\sigma_{21}=\sigma_{14}=\sigma_{41}=\sigma_{23}=\sigma_{32}=\sigma_{34}=\sigma_{43}=O_p\left(n^{-3/2}\right).$$

We exclude these terms from Lindley's approximation for simplification and fast calculation. Owing to this result, a simplified expression for Lindley's approximations can be obtained as

$$\begin{aligned}\begin{pmatrix}\hat{\mu}^{(\mathrm{L_{MLE}})}\\ \hat{\kappa}^{(\mathrm{L_{MLE}})}\\ \hat{\lambda}^{(\mathrm{L_{MLE}})}\\ \hat{\nu}^{(\mathrm{L_{MLE}})}\end{pmatrix}&=\begin{pmatrix}g^{\mu}(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})})\\ g^{\kappa}(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})})\\ g^{\lambda}(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})})\\ g^{\nu}(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})})\end{pmatrix}+\frac{1}{2}\begin{pmatrix}\sigma_{11}(g^{\mu}_{11}+2g^{\mu}_{1}\rho_1)+2\sigma_{13}g^{\mu}_{1}\rho_3\\ \sigma_{22}(g^{\kappa}_{22}+2g^{\kappa}_{2}\rho_2)+2\sigma_{24}g^{\kappa}_{2}\rho_4\\ \sigma_{33}(g^{\lambda}_{33}+2g^{\lambda}_{3}\rho_3)+2\sigma_{31}g^{\lambda}_{3}\rho_1\\ \sigma_{44}(g^{\nu}_{44}+2g^{\nu}_{4}\rho_4)+2\sigma_{42}g^{\nu}_{4}\rho_2\end{pmatrix}\\ &\quad+\frac{1}{2}\begin{pmatrix}g^{\mu}_{1}\sum_{i,j\in\{1,3\}}\left(\sigma_{11}\sigma_{ij}\ell_{ij1}+\sigma_{31}\sigma_{ij}\ell_{ij3}\right)+g^{\mu}_{1}\sum_{i,j\in\{2,4\}}\left(\sigma_{11}\sigma_{ij}\ell_{ij1}+\sigma_{31}\sigma_{ij}\ell_{ij3}\right)\\ g^{\kappa}_{2}\sum_{i,j\in\{2,4\}}\left(\sigma_{22}\sigma_{ij}\ell_{ij2}+\sigma_{42}\sigma_{ij}\ell_{ij4}\right)+g^{\kappa}_{2}\sum_{i,j\in\{1,3\}}\left(\sigma_{22}\sigma_{ij}\ell_{ij2}+\sigma_{42}\sigma_{ij}\ell_{ij4}\right)\\ g^{\lambda}_{3}\sum_{i,j\in\{1,3\}}\left(\sigma_{13}\sigma_{ij}\ell_{ij1}+\sigma_{33}\sigma_{ij}\ell_{ij3}\right)+g^{\lambda}_{3}\sum_{i,j\in\{2,4\}}\left(\sigma_{13}\sigma_{ij}\ell_{ij1}+\sigma_{33}\sigma_{ij}\ell_{ij3}\right)\\ g^{\nu}_{4}\sum_{i,j\in\{2,4\}}\left(\sigma_{24}\sigma_{ij}\ell_{ij2}+\sigma_{44}\sigma_{ij}\ell_{ij4}\right)+g^{\nu}_{4}\sum_{i,j\in\{1,3\}}\left(\sigma_{24}\sigma_{ij}\ell_{ij2}+\sigma_{44}\sigma_{ij}\ell_{ij4}\right)\end{pmatrix},\end{aligned}$$

where

$$\sigma_{11}=\frac{\ell_{33}}{\ell_{11}\ell_{33}-\ell_{13}^2},\quad \sigma_{22}=\frac{\ell_{44}}{\ell_{22}\ell_{44}-\ell_{24}^2},\quad \sigma_{33}=\frac{\ell_{11}}{\ell_{11}\ell_{33}-\ell_{13}^2},\quad \sigma_{44}=\frac{\ell_{22}}{\ell_{22}\ell_{44}-\ell_{24}^2},$$

$$\sigma_{13}=\sigma_{31}=-\frac{\ell_{13}}{\ell_{11}\ell_{33}-\ell_{13}^2},\quad \text{and}\quad \sigma_{24}=\sigma_{42}=-\frac{\ell_{24}}{\ell_{22}\ell_{44}-\ell_{24}^2}.$$

Recall that when $g^{\mu}_{1}=g^{\kappa}_{2}=g^{\lambda}_{3}=g^{\nu}_{4}=1$ and $g^{\mu}_{11}=g^{\kappa}_{22}=g^{\lambda}_{33}=g^{\nu}_{44}=0$, the above expression becomes

$$\begin{aligned}\begin{pmatrix}\hat{\mu}^{(\mathrm{L_{MLE}})}\\ \hat{\kappa}^{(\mathrm{L_{MLE}})}\\ \hat{\lambda}^{(\mathrm{L_{MLE}})}\\ \hat{\nu}^{(\mathrm{L_{MLE}})}\end{pmatrix}&=\begin{pmatrix}\hat{\mu}^{(\mathrm{MLE})}\\ \hat{\kappa}^{(\mathrm{MLE})}\\ \hat{\lambda}^{(\mathrm{MLE})}\\ \hat{\nu}^{(\mathrm{MLE})}\end{pmatrix}+\begin{pmatrix}\sigma_{11}\rho_1+\sigma_{13}\rho_3\\ \sigma_{22}\rho_2+\sigma_{24}\rho_4\\ \sigma_{33}\rho_3+\sigma_{31}\rho_1\\ \sigma_{44}\rho_4+\sigma_{42}\rho_2\end{pmatrix}+\frac{1}{2}\begin{pmatrix}\sum_{i,j\in\{1,3\}}\left(\sigma_{11}\sigma_{ij}\ell_{ij1}+\sigma_{31}\sigma_{ij}\ell_{ij3}\right)+\sum_{i,j\in\{2,4\}}\left(\sigma_{11}\sigma_{ij}\ell_{ij1}+\sigma_{31}\sigma_{ij}\ell_{ij3}\right)\\ \sum_{i,j\in\{1,3\}}\left(\sigma_{22}\sigma_{ij}\ell_{ij2}+\sigma_{42}\sigma_{ij}\ell_{ij4}\right)+\sum_{i,j\in\{2,4\}}\left(\sigma_{22}\sigma_{ij}\ell_{ij2}+\sigma_{42}\sigma_{ij}\ell_{ij4}\right)\\ \sum_{i,j\in\{1,3\}}\left(\sigma_{13}\sigma_{ij}\ell_{ij1}+\sigma_{33}\sigma_{ij}\ell_{ij3}\right)+\sum_{i,j\in\{2,4\}}\left(\sigma_{13}\sigma_{ij}\ell_{ij1}+\sigma_{33}\sigma_{ij}\ell_{ij3}\right)\\ \sum_{i,j\in\{1,3\}}\left(\sigma_{24}\sigma_{ij}\ell_{ij2}+\sigma_{44}\sigma_{ij}\ell_{ij4}\right)+\sum_{i,j\in\{2,4\}}\left(\sigma_{24}\sigma_{ij}\ell_{ij2}+\sigma_{44}\sigma_{ij}\ell_{ij4}\right)\end{pmatrix}.\end{aligned}\tag{14}$$

Lindley's approximations based on the MAP estimates with $\ell(\boldsymbol{\eta})$ replaced by $\ell^*(\boldsymbol{\eta})=\log\ell(\boldsymbol{\eta})+\log p(\boldsymbol{\eta})$ and $\rho(\boldsymbol{\eta})=0$ are given by

$$\begin{aligned}\begin{pmatrix}\hat{\mu}^{(\mathrm{L_{MAP}})}\\ \hat{\kappa}^{(\mathrm{L_{MAP}})}\\ \hat{\lambda}^{(\mathrm{L_{MAP}})}\\ \hat{\nu}^{(\mathrm{L_{MAP}})}\end{pmatrix}&=\begin{pmatrix}\hat{\mu}^{(\mathrm{MAP})}\\ \hat{\kappa}^{(\mathrm{MAP})}\\ \hat{\lambda}^{(\mathrm{MAP})}\\ \hat{\nu}^{(\mathrm{MAP})}\end{pmatrix}+\frac{1}{2}\begin{pmatrix}\sum_{i,j\in\{1,3\}}\left(\sigma^*_{11}\sigma^*_{ij}\ell^*_{ij1}+\sigma^*_{31}\sigma^*_{ij}\ell^*_{ij3}\right)+\sum_{i,j\in\{2,4\}}\left(\sigma^*_{11}\sigma^*_{ij}\ell^*_{ij1}+\sigma^*_{31}\sigma^*_{ij}\ell^*_{ij3}\right)\\ \sum_{i,j\in\{1,3\}}\left(\sigma^*_{22}\sigma^*_{ij}\ell^*_{ij2}+\sigma^*_{42}\sigma^*_{ij}\ell^*_{ij4}\right)+\sum_{i,j\in\{2,4\}}\left(\sigma^*_{22}\sigma^*_{ij}\ell^*_{ij2}+\sigma^*_{42}\sigma^*_{ij}\ell^*_{ij4}\right)\\ \sum_{i,j\in\{1,3\}}\left(\sigma^*_{13}\sigma^*_{ij}\ell^*_{ij1}+\sigma^*_{33}\sigma^*_{ij}\ell^*_{ij3}\right)+\sum_{i,j\in\{2,4\}}\left(\sigma^*_{13}\sigma^*_{ij}\ell^*_{ij1}+\sigma^*_{33}\sigma^*_{ij}\ell^*_{ij3}\right)\\ \sum_{i,j\in\{1,3\}}\left(\sigma^*_{24}\sigma^*_{ij}\ell^*_{ij2}+\sigma^*_{44}\sigma^*_{ij}\ell^*_{ij4}\right)+\sum_{i,j\in\{2,4\}}\left(\sigma^*_{24}\sigma^*_{ij}\ell^*_{ij2}+\sigma^*_{44}\sigma^*_{ij}\ell^*_{ij4}\right)\end{pmatrix},\end{aligned}\tag{15}$$

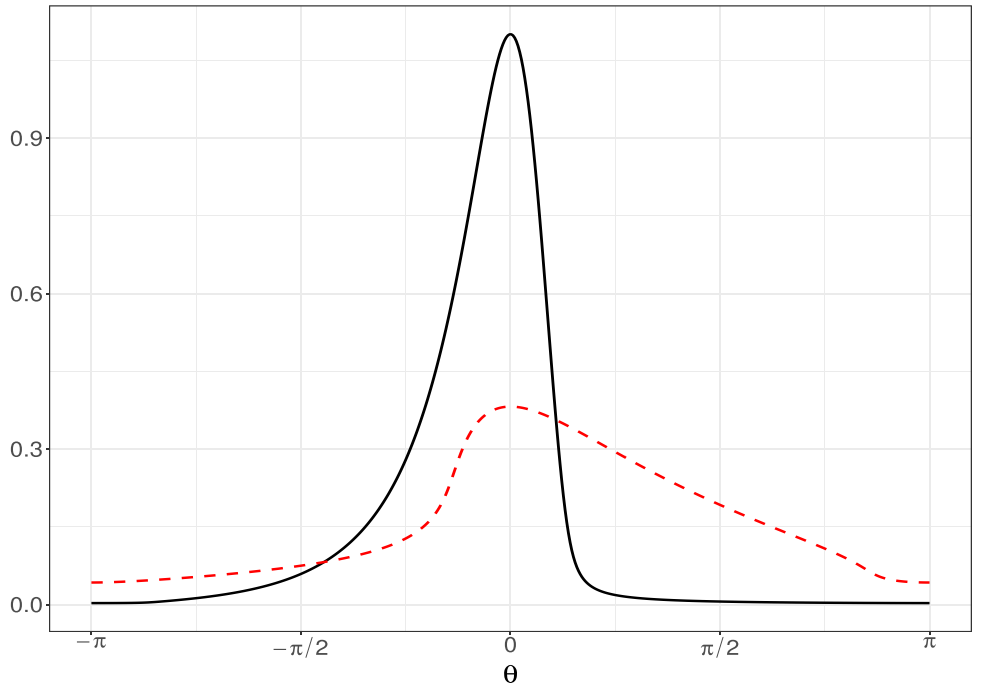

Fig. 4.1. Density plots for the extended inverse transformation of von Mises distribution with $\mu = 0, \kappa = 3, \lambda = -0.7, \nu = 0.5$ (solid line) and the EIWC distribution with $\mu = 0, \rho = 0.5, \lambda = 0.7, \nu = -0.5$ (dashed line) for simulation studies.

where

$$\sigma^*_{ij} = \left(-\frac{\partial^2}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^T} \ell^*(\boldsymbol{\eta}) \right)^{-1} \Bigg|_{\eta = \hat{\eta}_n^{(\mathrm{MAP})}} \quad \text{and} \quad \ell^*_{ijk} = -\frac{\partial^3}{\partial \eta_i \partial \eta_j \partial \eta_k} \ell^*(\boldsymbol{\eta}) \Bigg|_{\eta = \hat{\eta}_n^{(\mathrm{MAP})}}.$$

To derive approximation (15) using MAP estimator $\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})}$, we assume Condition D2 with the log-likelihood $\ell(\boldsymbol{\eta})$ and MLE $\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})}$, which are replaced by the log-posterior and MAP estimator $\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})}$, respectively. Furthermore, we need to provide a condition that corresponds with condition (iv) of Kass et al. (1990). However, this is slightly complicated and is hence omitted here to save space. For readers who are interested in the details, see condition [A4] of Miyata (2018).

For large sample asymptotics, Lindley's method provides a good approximation for the BE, that is, $\hat{\boldsymbol{\eta}}_n^{(\mathrm{BE})} = \hat{\boldsymbol{\eta}}_n^{(s)} + O_p(n^{-3/2})$ and $s \in \{\mathrm{L_{MLE}}, \mathrm{L_{MAP}}\}$, where $\hat{\boldsymbol{\eta}}^{(\mathrm{L_{MLE}})}$ and $\hat{\boldsymbol{\eta}}^{(\mathrm{L_{MAP}})}$ are defined by (14) and (15), respectively. The asymptotic distribution for the approximate BE is equivalent to that of the MLE, and as we observe from (14) and (15),

$$\sqrt{n}(\hat{\boldsymbol{\eta}}^{(\mathrm{L_{MLE}})} - \boldsymbol{\eta}_0) = \sqrt{n}(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})} - \boldsymbol{\eta}_0) + O_p(n^{-1/2})$$

and

$$\sqrt{n}(\hat{\boldsymbol{\eta}}^{(\mathrm{L_{MAP}})} - \boldsymbol{\eta}_0) = \sqrt{n}(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})} - \boldsymbol{\eta}_0) + O_p(n^{-1/2}).$$

We denote $\hat{\boldsymbol{\eta}}^{(\mathrm{L_{MLE}})}$ and $\hat{\boldsymbol{\eta}}^{(\mathrm{L_{MAP}})}$ as $\mathrm{L_{MLE}}$ and $\mathrm{L_{MAP}}$, respectively, hereafter for convenience.

4. Numerical simulations

In this section, we compare the performance of all the estimators introduced in Section 3. The random samples are drawn from the two step procedures given in Theorem 2. Simulations are conducted with 1,000 replications, with sample sizes $n = 50, 100$, and 300. We consider two conditions: (1) EIVM distribution with the parameters $\mu = 0$, $\kappa = 3$, $\lambda = -0.7$, and $\nu = 0.5$ and (2) EIWC distribution with the parameters $\mu = 0$, $\rho = 0.5$, $\lambda = 0.7$, and $\nu = -0.5$.

We use rather non-informative priors for the BEs so that the hyper parameters in Eqs. (11) are set at $(m_\mu, m_\kappa, m_\lambda, m_\nu)^T = (0, 1, 0, 0)^T$ and $(\sigma_\mu, \sigma_\kappa, \sigma_\lambda, \sigma_\nu)^T = (0.8, 1, 0.8, 0.8)^T$ for the EIVM distributions. As for the EIWC cases, we choose $(m_\rho, \sigma_\rho) = (0, 0.8)^T$, and others are the same for the EIVM distributions. The density functions of the EIVM and EIWC distributions are plotted in Fig. 4.1, which illustrates a left-skewed modestly peaked distribution and a right-skewed and flat-topped distribution, respectively.

Tables 1 and 2 summarize the simulation results of the sample averages of the estimated parameters, together with the root mean squared errors (RMSEs). As for the parameter spaces of $\lambda \in [-1 + \delta_\lambda, 1 - \delta_\lambda]$ and $\nu \in [-1 + \delta_\nu, 1 - \delta_\nu]$, we set

Table 4.1
Means and root mean square errors (RMSEs) of the maximum likelihood estimate (MLE), maximum a posteriori (MAP) estimate, L_{MLE}, and L_{MAP}, of the parameters of the extended inverse transformation of von Mises distributions, using 1,000 simulated samples of size n from the model with $\mu = 0$, $\kappa = 3$, $\lambda = -0.7$, and $\nu = 0.5$.

	MLE	MAP	L_{MLE}	L_{MAP}
$n = 50$				
mean: $\hat{\mu}$	−0.0615	−0.0077	−0.0615	0.0112
RMSE $\hat{\mu}$	0.1418	0.0780	0.1694	0.0821
mean: $\hat{\kappa}$	3.4104	2.9815	3.2487	3.0653
RMSE $\hat{\kappa}$	1.0850	0.4784	1.0241	0.6822
mean: $\hat{\lambda}$	−0.5187	−0.6956	−0.4574	−0.7307
RMSE $\hat{\lambda}$	0.4238	0.2086	0.4290	0.1835
mean: $\hat{\nu}$	0.4345	0.5363	0.4018	0.4706
RMSE $\hat{\nu}$	0.2439	0.1591	0.3217	0.2992
Effective Samples	1,000	1,000	609	835
$n = 100$				
mean: $\hat{\mu}$	−0.0235	−0.0017	−0.0116	0.0128
RMSE $\hat{\mu}$	0.0910	0.0546	0.0824	0.0576
mean: $\hat{\kappa}$	3.1457	3.0004	3.2985	3.1428
RMSE $\hat{\kappa}$	0.5109	0.4028	0.7668	0.6298
mean: $\hat{\lambda}$	−0.6255	−0.7019	−0.6340	−0.7322
RMSE $\hat{\lambda}$	0.2922	0.1553	0.2681	0.1394
mean: $\hat{\nu}$	0.4731	0.5143	0.4000	0.4435
RMSE $\hat{\nu}$	0.1485	0.1177	0.2702	0.2691
Effective Samples	1,000	1,000	728	968
$n = 300$				
mean: $\hat{\mu}$	−0.0004	0.0008	0.0081	0.0073
RMSE $\hat{\mu}$	0.0381	0.0325	0.0386	0.0316
mean: $\hat{\kappa}$	3.0595	3.0088	3.2167	3.1274
RMSE $\hat{\kappa}$	0.2507	0.2298	0.5126	0.4674
mean: $\hat{\lambda}$	−0.6989	−0.7027	−0.7239	−0.7197
RMSE $\hat{\lambda}$	0.1157	0.0953	0.1144	0.0882
mean: $\hat{\nu}$	0.4913	0.5022	0.4261	0.4514
RMSE $\hat{\nu}$	0.0758	0.0716	0.2068	0.1994
Effective Samples	1,000	1,000	936	994

$\delta_\lambda = \delta_\nu = 10^{-5}$ for sharply peaked EIVM distributions and $\delta_\lambda = \delta_\nu = 0.05$ for flat-topped EIWC distributions. This is because most of the MLEs of λ and ν for the cases in flat-topped EIWC distributions provide values on the boundary, which yields an unsatisfactory Lindley's approximation when δ_λ and δ_ν approach 0. For Lindley's methods, the resulting estimates for λ and ν may have values greater than 1 in the absolute value. We omit these solutions for computing means and RMSEs. The number of admissible solutions are denoted as the effective sample sizes in Tables 1 and 2. From these tables, the following conclusions can be obtained.

First, for the case of EIVM distribution, the ratio of unsuitable solutions for L_{MLE} and L_{MAP} are 30.1% and 16.5%, respectively, for sample size $n = 50$, whereas these values decrease as the sample size increases. On the contrary, these values greatly increase for the flat-topped EIWC distribution. This indicates that Lindley's methods are unstable, especially in small sample sizes and for the MLE-based estimation and flat-topped distributions. This is because the MLE tends to overfit the simulated data, and the resulting MLEs for λ or ν tend to have values at the edge of the parameter space. We can see that Lindley's method with the MAP estimates is more robust than that with the MLE.

Second, as the sample size increases, the bias of the estimates decreases. This confirms the consistency properties of all estimators. The MAP estimates have a smaller bias than that of other estimates for the parameters κ and ν, whereas the performance of the approximate BEs based on Lindley's method with MAP estimates is the best for the parameters μ and λ in terms of bias.

These simulation results reveal the difficulty in estimating skewness and peakedness parameters by using the MLE. It is clear from Tables 1 and 2 that the bias of MLE and the MLE-based Lindley's estimates are larger than those obtained from MAP and MAP-based Lindley's estimates for the two situations. These drawbacks are more apparent in small samples of $n = 50$. Moreover, the corresponding Lindley's method based on the MLE does not significantly improve their performances, due to the unstable Hessians and third derivatives of their log-likelihood functions. The ratios of the estimated parameters on the boundary, that is, for either $\hat{\lambda} \approx \pm 1$ or $\hat{\nu} \approx \pm 1$, are 25.8%, 10.8%, and 0.7% for sample sizes 50, 100, 300, respectively, in 1,000 simulations for EIVM distributions. These values for either $\hat{\lambda} \approx \pm 0.95$ or $\hat{\nu} \approx \pm 0.95$ are 64.9%, 48.9%, and 29.2%, respectively, based on the flat-topped EIWC distributions. However, in a Bayesian inference setting, we use transformed normal prior distributions, which aim to avoid the estimates on the boundaries, and the posterior mode and approximate BEs based on MAP estimates will never take 1 or −1.

Lindley's approximation is not applicable when the estimated parameters take values on the boundary of the parameter space, as it is doubtful that Condition D2 will hold, yielding inadmissible solutions of the estimates. However, it can be

Table 4.2
Means and root mean square errors (RMSEs) of the maximum likelihood estimate (MLE), maximum a posteriori (MAP) estimate, L_{MLE}, and L_{MAP} of the parameters of the extended inverse transformation of wrapped Cauchy distributions, using 1,000 simulated samples of size n from the model with $\mu = 0$, $\rho = 0.5$, $\lambda = 0.7$, and $\nu = -0.5$.

	MLE	MAP	L_{MLE}	L_{MAP}
$n = 50$				
mean: $\hat{\mu}$	0.1730	0.1805	0.1933	0.1774
RMSE $\hat{\mu}$	0.4919	0.4865	0.7588	0.4772
mean: $\hat{\rho}$	0.5400	0.5325	0.4661	0.4553
RMSE $\hat{\rho}$	0.0960	0.0910	0.1361	0.1394
mean: $\hat{\lambda}$	0.4216	0.4018	0.2786	0.3299
RMSE $\hat{\lambda}$	0.7367	0.7036	0.6322	0.6609
mean: $\hat{\nu}$	−0.4107	−0.3762	−0.0458	−0.2735
RMSE $\hat{\nu}$	0.6092	0.5698	0.6891	0.4776
Effective Samples	1,000	1,000	192	537
$n = 100$				
mean: $\hat{\mu}$	0.0902	0.0959	0.1687	0.0697
RMSE $\hat{\mu}$	0.3473	0.3327	0.3706	0.3154
mean: $\hat{\rho}$	0.5191	0.5133	0.4462	0.4747
RMSE $\hat{\rho}$	0.0730	0.0678	0.1107	0.0868
mean: $\hat{\lambda}$	0.5815	0.5646	0.4337	0.5428
RMSE $\hat{\lambda}$	0.4933	0.4649	0.4991	0.4573
mean: $\hat{\nu}$	−0.4797	−0.4523	−0.0997	−0.3857
RMSE $\hat{\nu}$	0.4625	0.4165	0.6314	0.3777
Effective Samples	1,000	1,000	311	611
$n = 300$				
mean: $\hat{\mu}$	0.0135	0.0216	0.0197	-0.0133
RMSE $\hat{\mu}$	0.1715	0.1621	0.1559	0.1461
mean: $\hat{\rho}$	0.5133	0.5096	0.4637	0.4921
RMSE $\hat{\rho}$	0.0545	0.0489	0.0733	0.0479
mean: $\hat{\lambda}$	0.6904	0.6744	0.6641	0.7170
RMSE $\hat{\lambda}$	0.3197	0.2771	0.3924	0.2615
mean: $\hat{\nu}$	−0.5458	−0.5237	−0.3033	−0.4953
RMSE $\hat{\nu}$	0.3197	0.2771	0.3924	0.2615
Effective Samples	1,000	1,000	615	820

seen that Lindley's approximation provides a better performance with RMSEs when the estimated parameters are within the corresponding parameter space; this is especially true for MAP-based Lindley's approximation.

5. Data analysis

To apply estimation procedures for circular data, we use the following datasets. The first dataset is the wind direction studied by Agostinelli (2007). The `circular` library of R software contains this dataset labeled as `wind` object. The data consist of 310 records of wind direction measured from the north in radians that are measured at a meteorological station in the Italian Alps from January to March 2001. The second and third datasets are the compass angles of the orientations of termite mounds of Amitermes laurensis at 14 sites in Cape York Peninsula, North Queensland. These datasets are available from Fisher (1993) and the `circular` library in R software. The data are labeled with `fisherB13`. We use the datasets numbered #1 and #8, and the sample sizes are 100 and 48, respectively, because these data had unimodal and skewed distributions based on the histograms given later in Figs. 5.1 to 5.3.

For these datasets, we compute a circular sample mean and mean resultant length for the explanatory data analysis. The wind data have circular mean of 0.2922 radians and mean resultant length of 0.6557. The compass angle of the orientations of the termite mounds has mean directions of 0.0490 and −0.0989 for datasets #1 and #8, respectively. The corresponding mean resultant length becomes 0.8826 and 0.9427 for datasets #1 and #8, respectively, which demonstrates highly concentrated data distributions. Before we fit the skewed distributions for these data, we test the null hypothesis of reflective symmetry against the alternative of a skewed distribution. The p-value of the test statistics for circular reflective symmetry of Pewsey (2002) for the wind data is less than 0.01, which indicates that there is no evidence that the underlying distribution is symmetric. On the other hand, the p-values for datasets #1 and #8 for the orientations of termite mounds data are 0.0889 and 0.9355, respectively. As the symmetry test of Pewsey (2002) is based on the asymptotic distribution theory, we cannot find strong evidence against a reflective symmetry for datasets 2 and 3 despite the histogram demonstrating somewhat asymmetric distributions. Pewsey (2002) provided a bootstrap version of the test for reflective symmetry. The test results are the similar as those obtained from the asymptotic distribution-based tests, that is, the p-values of dataset 1 are less than 0.0001, and for datasets 2 and 3, the corresponding p-values are 0.0869 and 0.9537, respectively. Here, we set the number of bootstrap replicates as $B = 9,999$.

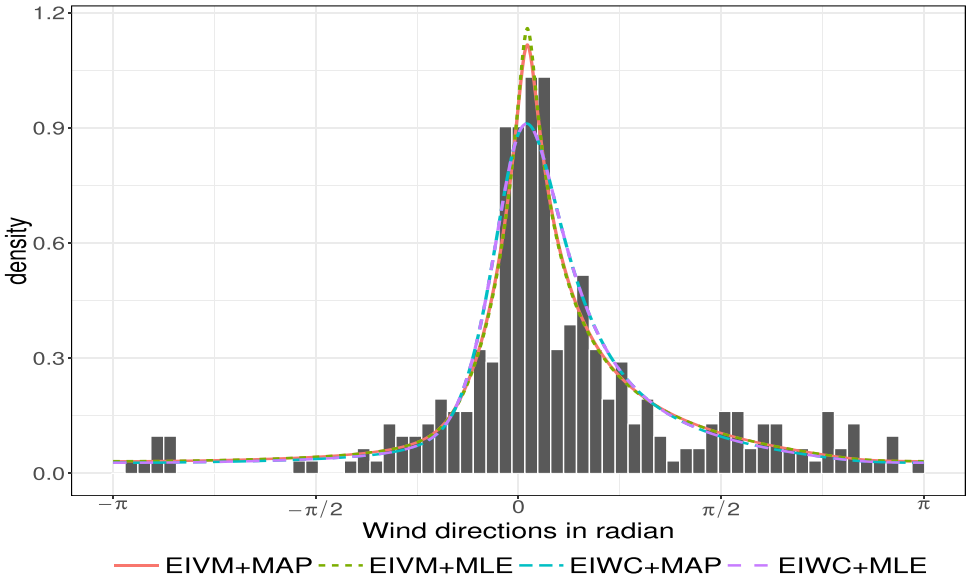

Fig. 5.1. Estimated densities with the histogram of wind direction in the Italian Alps. The abbreviations EIVM, EIWC, MLE, and MAP stand for extended inverse transformation of von Mises, extended inverse transformation of wrapped Cauchy, maximum likelihood estimate, and maximum a posteriori, respectively.

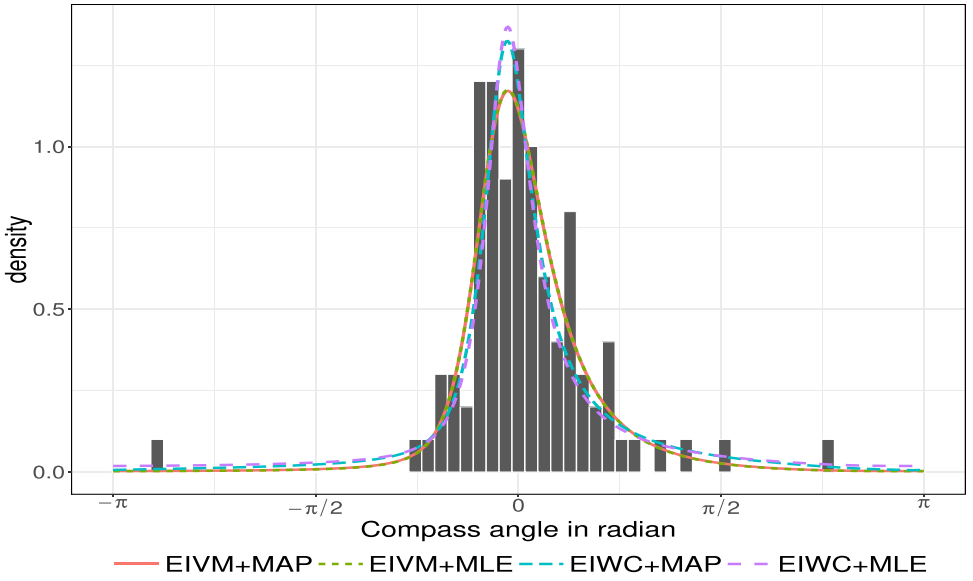

Fig. 5.2. Estimated densities with the histogram of compass orientations of termite mounds dataset #1. The abbreviations EIVM, EIWC, MLE, and MAP stand for extended inverse transformation of von Mises, extended inverse transformation of wrapped Cauchy, maximum likelihood estimate, and maximum a posteriori, respectively.

The base symmetric density function f_0 must be specified before we estimate the model parameters to fit the data. As we do not have any information about the specification of the base symmetric density, we use the Akaike information criterion (AIC) and Bayesian information criterion (BIC) to compare the performance of the model selection among the various choices of model specifications. For the choice of symmetric base density, we use the von Mises and wrapped Cauchy distributions. The best model is selected with the smallest AIC and BIC values. Table 1 summarizes the estimated model parameters using MLEs, together with AIC and BIC values.

According to the AIC values in Table 1, the four-parameter extended inverse transformation of scale distributions are selected for all three datasets, whereas BIC selects a skewed model ($\nu = 0$) of the EIWC distribution for datasets 1 and 3 and a symmetric and peakedness controlled model ($\lambda = 0$) of the EIWC distribution for dataset 2. The EIVM distributions have better fitting performances than those of the EIWC distributions with smaller AIC and BIC for all three datasets. It must

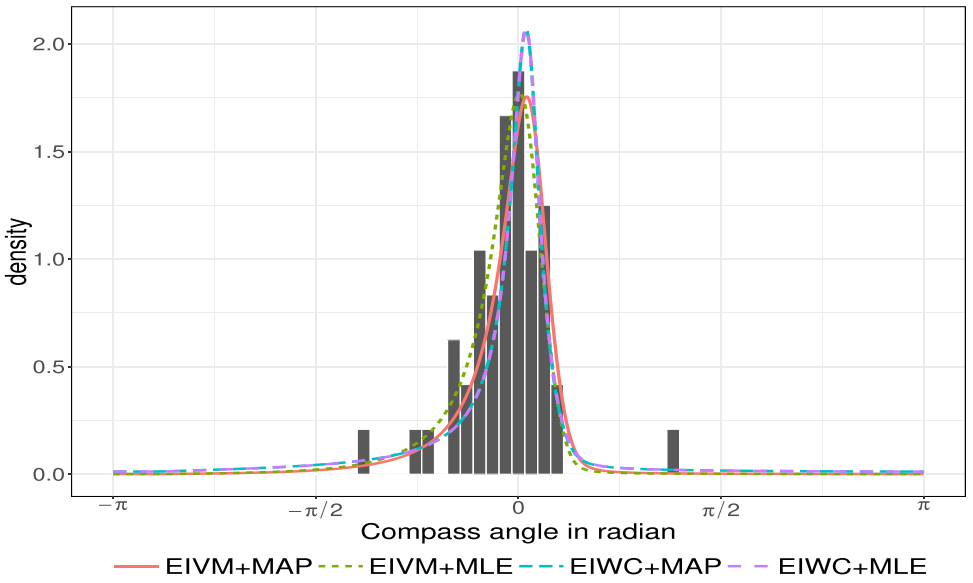

Fig. 5.3. Estimated densities with the histogram of compass orientations of termite mounds dataset #8. The abbreviations EIVM, EIWC, MLE, and MAP stand for extended inverse transformation of von Mises, extended inverse transformation of wrapped Cauchy, maximum likelihood estimate, and maximum a posteriori, respectively.

Table 5.1
Maximum likelihood estimates for the extended inverse transformation of von Mises (EIVM) and extended inverse transformation of wrapped Cauchy (EIWC) distributions with Akaike information criterion (AIC) and Bayesian information criterion (BIC) values.

	EIVM distribution					
	μ	κ	λ	ν	AIC	BIC
Dataset 1: Wind direction at Italian Alps data						
Skewed ($\nu = 0$) model	0.0584	1.8532	0.6257	–	821.7	832.9
Symmetric ($\lambda = 0$) model	0.1137	1.7854	–	0.9106	762.7	773.9
Skewed and peakedness controlled model	0.0670	1.7966	0.5020	0.8820	**751.6**	**766.5**
Dataset 2: Orientations of termite mounds No.1						
Skewed ($\nu = 0$) model	−0.1053	4.8380	0.4776	–	145.0	152.8
Symmetric ($\lambda = 0$) model	0.0062	2.9710	–	0.5426	133.9	141.7
Skewed and peakedness controlled model	−0.0852	3.0475	0.3997	0.5373	**131.2**	**141.6**
Dataset 3: Orientations of termite mounds No.8						
Skewed ($\nu = 0$) model	−0.0841	9.0281	−0.0743	–	39.5	45.1
Symmetric ($\lambda = 0$) model	−0.0711	3.8237	–	0.6488	31.3	**36.9**
Skewed and peakedness controlled model	0.0094	3.8558	−0.7136	0.6698	**30.2**	37.7
	EIWC distribution					
	μ	ρ	λ	ν	AIC	BIC
Dataset 1: Wind direction at Italian Alps data						
Skewed ($\nu = 0$) model	0.0615	0.7026	0.5044	–	756.4	**767.6**
Symmetric ($\lambda = 0$) model	0.1122	0.6629	–	0.3716	762.1	773.3
Skewed and peakedness controlled model	0.0631	0.6704	0.4808	0.3211	**754.2**	769.1
Dataset 2: Orientations of termite mounds No.1						
Skewed ($\nu = 0$) model	−0.0860	0.7916	0.3976	–	141.3	149.1
Symmetric ($\lambda = 0$) model	−0.0153	0.8787	–	−0.9989	140.3	**148.1**
Skewed and peakedness controlled model	−0.0896	0.8798	0.4177	−0.9990	**138.3**	148.7
Dataset 3: Orientations of termite mounds No.8						
Skewed ($\nu = 0$) model	0.0013	0.8572	−0.6720	–	34.4	**40.0**
Symmetric ($\lambda = 0$) model	−0.0481	0.9208	–	−1.0000	34.9	40.5
Skewed and peakedness controlled model	0.0014	0.9211	-0.6592	−1.0000	**34.2**	41.7

Note: The smallest AIC and BIC values are shown in boldface. The hyphen indicates that the corresponding parameter is set at 0.

Table 5.2
Estimated parameters of the extended inverse transformation of von Mises (EIVM) and extended inverse transformation of wrapped Cauchy (EIWC) distributions for three datasets.

	EIVM				EIWC ($\nu = 0$)		
	μ	κ	λ	ν	μ	ρ	λ
Dataset 1: Wind direction at Italian Alps data							
MLE	0.0670	1.7966	0.5020	0.8820	0.0615	0.7026	0.5044
MAP	0.0665	1.7740	0.5014	0.8625	0.0610	0.7028	0.5080
L_{MLE}	0.0645	1.7415	0.5389	0.8432	0.0593	0.7025	0.5400
L_{MAP}	0.0632	1.7838	0.5478	0.9093	0.0675	0.6986	0.4825
Dataset 2: Compass angle of the orientations of termite mounds No.1							
MLE	−0.0852	3.0475	0.3997	0.5373	−0.0860	0.7916	0.3976
MAP	−0.0848	2.9199	0.3977	0.5573	−0.0855	0.7912	0.3947
L_{MLE}	−0.0841	2.9843	0.4118	0.5390	−0.0786	0.7964	0.3868
L_{MAP}	−0.0855	2.9044	0.4180	0.5568	−0.0744	0.7971	0.3636
Dataset 3: Compass angle of the orientations of termite mounds No.8							
MLE	0.0094	3.8558	−0.7136	0.6698	0.0013	0.8572	−0.6720
MAP	0.0008	3.5118	−0.6455	0.7023	0.0027	0.8572	−0.6861
L_{MLE}	−0.0081	3.6724	−0.8225	0.6696	−0.0027	0.8718	−0.8173
L_{MAP}	−0.0014	3.4445	−0.8126	0.7169	−0.0488	0.8576	−0.5613

be noted that the MLEs for the peakedness parameter ν in the EIWC distribution have values close to -1 for datasets 2 and 3, which indicates that the peakedness around the mode is controlled by both ρ and ν, and the peakedness parameter ν is not necessary for highly concentrated data when we use the wrapped Cauchy base density function. Moreover, the corresponding Lindley's approximation does not provide reliable parameter values for these datasets.

As we compare the different estimation methods including the BEs, we use rather non-informative priors for the BEs that are $(m_\mu, m_\kappa, m_\lambda, m_\nu)^T = (0, 1, 0, 0)^T$ and $(\sigma_\mu, \sigma_\kappa, \sigma_\lambda, \sigma_\nu)^T = (0.8, 1, 0.8, 0.8)^T$ for the EIVM distribution. From the result of the model selection given in Table 1, we fit the EIWC distribution setting $\nu = 0$, and the prior parameters are chosen as $(m_\mu, m_\rho, m_\lambda)^T = (0, 0, 0)^T$ and $(\sigma_\mu, \sigma_\rho, \sigma_\lambda)^T = (0.8, 0.8, 0.8)^T$. Recall that the prior distributions are given in Eqs. (11) and (12). Table 2 summarizes the estimated parameter values for the three datasets listed above using the EIVM and EIWC distributions. All of the estimates are similar to each other among the datasets, except for Lindley's approximate BEs for λ and ν of the EIVM distribution. The estimated values for λ indicate that there is strong evidence against the null hypothesis of the reflective symmetry, which can be confirmed from Figs. 5.1 to 5.3.

For the EIWC distribution, we set $\nu = 0$ for all the datasets to provide a reliable inference of the model parameters, because Lindley's method does not provide satisfactory estimates for ν and λ. As the effects of the peakedness parameters in the EIWC distributions are absorbed by the concentration parameter ρ, a more simplified model specification is possible when assuming the wrapped Cauchy distribution for a base density function. Hence, it is important to select an appropriate base density function that provides better fitting performance for data analysis.

The estimated density functions, together with the histograms, are plotted in Figs. 5.1 to 5.3. According to these figures, we can verify that our proposed inverse transformation of scale distributions can successfully capture the skewness and sharply peaked features of underlying data distributions. The differences between the wrapped Cauchy and von Mises distributions for the base density appear in the peakedness around the mode. According to Figs. 5.2 and 5.3, the EIVM distribution has a lower peak around the mode than that with the EIWC distribution even though the peakedness parameter is larger than that of $\hat{\nu}$ in the EIWC distribution. This is because the wrapped Cauchy distribution has desirable properties that fit the sharp distribution without introducing the peakedness parameter ν.

6. Summary and conclusions

In this study, we consider the unimodal skew and sharply peaked and flat-topped distributions on a circle by using the transformation of scale by Jones (2014). The explicit expressions for the distribution functions with a base density for the von Mises and wrapped Cauchy distributions are provided. The efficient methods for random sampling of the proposed density are given, where the sampling scheme does not rely on acceptance-rejection sampling. To estimate the model parameters, we introduce the MLE, MAP estimator, and Lindley's approximations for BE. Numerical simulations are performed, and they demonstrate that the BEs perform well in small samples. The data analysis also illustrated that the proposed distributions are useful for practical purposes. For future research topics, several computational approaches and algorithms using MCMC for obtaining posterior distributions of the model parameters are worth investing.

The main limitation of the proposed extended inverse transformation of scale distribution is that the explicit forms of the trigonometric moments of the distributions are difficult to obtain, and hence we did not derive a method of moments estimation or utilize empirical characteristic function approaches for parameter estimation. It must be noted that Bayesian model selection procedures that require computation of marginal likelihoods are difficult to obtain without using Monte

Carlo simulations when the MLE occurs at the boundary of the parameter space. We leave these topics open for future research.

Acknowledgements

The authors are grateful to the editor and anonymous referees for helpful suggestions that led to the improvement of the paper. Toshihiro Abe was supported in part by JSPS KAKENHI Grant Number 19K11869 and 19KK0287. Yoichi Miyata was supported in part by JSPS KAKENHI Grant Number 19K11863 and the competitive research expenses of the Takasaki City University of Economics. Takayuki Shiohama was supported in part by JSPS KAKENHI Grant Number 18K01706.

Appendix A. Proofs of Theorems

In this Appendix, we provide proofs of theorems given in Sections 2 and 3. For the proofs of Theorems 1 and 2, we assume $\mu = 0$ without loss of generality.

Proof of Theorem 1. The cumulative distribution function for the inverse transformation of scale distribution with arbitrary base symmetric density $f_0(\theta)$ can be written as

$$\begin{aligned} F(\theta) &= \int_{-\pi}^{\theta} \frac{1}{1-\nu\alpha_1} f_0(t_{0,\nu}^{-1}(t))dt = \frac{1}{1-\nu\alpha_1}\int_{-\pi}^{t_{0,\nu}^{-1}(\theta)} (1-\nu\cos u) f_0(u)du \\ &= \frac{1}{1-\nu\alpha_1}\left\{F_0(t_{0,\nu}^{-1}(\theta)) - \nu\left(\int_{-\pi}^{t_{0,\nu}^{-1}(\theta)} \cos u f_0(u)du\right)\right\}. \end{aligned} \tag{A.1}$$

For the case with von Mises distribution as base symmetric density $f_0^{(\mathrm{VM})}(\theta)$, recall that the Fourier series for $f_0^{(\mathrm{VM})}(\theta)$ is expressed as

$$f_0^{(\mathrm{VM})}(\theta) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{p=1}^{\infty}\frac{I_p(\kappa)}{I_0(\kappa)}\cos p\theta,$$

which gives distribution function $F_0^{(\mathrm{VM})}$ as

$$F_0^{(\mathrm{VM})}(\theta) = \frac{\theta+\pi}{2\pi} + \frac{1}{\pi}\sum_{p=1}^{\infty}\frac{I_p(\kappa)}{I_0(\kappa)}\frac{1}{p}\sin p\theta.$$

As we have

$$\int_{-\pi}^{\theta}\cos t f_0^{(\mathrm{VM})}(t)dt = \frac{\theta+\pi}{2\pi}\frac{I_1(\kappa)}{I_0(\kappa)} + \frac{1}{2\pi}\sum_{p=1}^{\infty}\frac{I_{p-1}(\kappa)+I_{p+1}(\kappa)}{I_0(\kappa)}\frac{1}{p}\sin p\theta,$$

the distribution function of the extended inverse transformation of scale distribution with the base von Mises density defined by (3) and $\lambda = 0$ is given by

$$\begin{aligned} F^{(\mathrm{EIVM})}(\theta) &= \int_{-\pi}^{\theta}\frac{1}{1-\nu\alpha_1} f_0^{(\mathrm{VM})}(t_{0,\nu}^{-1}(t))dt = \frac{1}{1-\nu\alpha_1}\int_{-\pi}^{t_{0,\nu}^{-1}(\theta)}(1-\nu\cos u) f_0^{(\mathrm{VM})}(u)du \\ &= \frac{1}{1-\nu\alpha_1}\left\{F_0^{(\mathrm{VM})}(t_{0,\nu}^{-1}(\theta)) - \nu\left(\int_{-\pi}^{t_{0,\nu}^{-1}(\theta)}\cos u f_0^{(\mathrm{VM})}(u)du\right)\right\} \\ &= \frac{1}{1-\nu\alpha_1}\left\{F_0^{(\mathrm{VM})}(t_{0,\nu}^{-1}(\theta)) - \nu\left(\frac{t_{0,\nu}^{-1}(\theta)+\pi}{2\pi}\frac{I_1(\kappa)}{I_0(\kappa)} + \frac{1}{2\pi}\sum_{p=1}^{\infty}\frac{I_{p-1}(\kappa)+I_{p+1}(\kappa)}{I_0(\kappa)}\frac{1}{p}\sin(pt_{0,\nu}^{-1}(\theta))\right)\right\}, \end{aligned}$$

where $F_0^{(\mathrm{VM})}(\theta)$ is the distribution function of the base von Mises distribution. Recall that $\alpha_1 = I_1(\kappa)/I_0(\kappa)$ and the modified Bessel function has the following relations: $I_0'(\kappa) = I_1(\kappa)$ and $I_p'(\kappa) = (I_{p-1}(\kappa)+I_{p+1}(\kappa))/2$. Then, we obtain a more simplified expression as

$$F^{(\mathrm{EIVM})}(\theta) = \frac{t_{0,\nu}^{-1}(\theta)+\pi}{2\pi} + \frac{1}{\pi}\sum_{p=1}^{\infty}\frac{I_p(\kappa)-\nu I_p'(\kappa)}{I_0(\kappa)-\nu I_0'(\kappa)}\frac{1}{p}\sin(pt_{0,\nu}^{-1}(\theta)), \quad \text{when} \quad \lambda = 0.$$

For the case with symmetric wrapped Cauchy distribution for the base density, the term involved in (A.1) becomes

$$\begin{aligned} \int_{-\pi}^{\theta}\cos t f_0^{(\mathrm{WC})}(t)dt &= \frac{1+\rho^2}{2\rho}\left(\frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{1+\rho}{1-\rho}\tan\frac{\theta}{2}\right)\right) - \frac{1-\rho^2}{2\pi}\frac{\theta+\pi}{2\rho} \\ &= \frac{1+\rho^2}{2\rho}F_0^{(\mathrm{WC})}(\theta) - \frac{1-\rho^2}{4\pi\rho}(\theta+\pi). \end{aligned}$$

Therefore, the distribution function of the extended inverse transformation of scale distribution with wrapped Cauchy density defined by (4) with $\lambda = 0$ is given by

$$\begin{aligned} F^{(\mathrm{EIWC})}(\theta) &= \int_{-\pi}^{\theta} \frac{1}{1-\nu\alpha_1} f_0^{(\mathrm{WC})}(t_{0,\nu}^{-1}(t))dt \\ &= \frac{1}{1-\nu\alpha_1}\int_{-\pi}^{t_{0,\nu}^{-1}(\theta)} (1-\nu\cos u) f_0^{(\mathrm{WC})}(u)du \\ &= \frac{1}{1-\nu\rho}\left\{\left(1-\nu\frac{1+\rho^2}{2\rho}\right)F_0^{(\mathrm{WC})}(t_{0,\nu}^{-1}(\theta)) + \nu\frac{1-\rho^2}{4\pi\rho}(t_{0,\nu}^{-1}(\theta)+\pi)\right\}. \end{aligned}$$

This completes the proof of Theorem 1.

Proof of Theorem 2. We have

$$\begin{aligned} P(\Theta \le \theta) &= P\left(\Theta \le \theta, U \le \frac{s'_\lambda(\Psi)}{2}\right) + P\left(\Theta \le \theta, U > \frac{s'_\lambda(\Psi)}{2}\right) \\ &= P\left(\Psi \le s_\lambda^{-1}(\theta), U \le \frac{s'_\lambda(\Psi)}{2}\right) + P\left(\Psi > -s_\lambda^{-1}(\theta), U > \frac{s'_\lambda(\Psi)}{2}\right) \\ &= \int_{-\pi}^{s_\lambda^{-1}(\theta)} P\left(U \le \frac{s'_\lambda(\psi)}{2}\right) f_0(\psi)d\psi + \int_{-s_\lambda^{-1}(\theta)}^{\pi} P\left(U > \frac{s'_\lambda(\psi)}{2}\right) f_0(\psi)d\psi \\ &= \int_{-\pi}^{s_\lambda^{-1}(\theta)} \frac{s'_\lambda(\psi)}{2} f_0(\psi)d\psi + \int_{-s_\lambda^{-1}(\theta)}^{\pi} \left(1-\frac{s'_\lambda(\psi)}{2}\right) f_0(\psi)d\psi, \end{aligned}$$

where we have used the relations $(s'_\lambda(\psi)+s'_\lambda(-\psi))/2 = 1$ and $(s_\lambda^{-1})'(\theta) = 1/s'_\lambda(s_\lambda^{-1}(\theta))$. Then, we observe

$$\begin{aligned} f(\theta) &= \frac{d}{d\theta}P(\Theta \le \theta) \\ &= \frac{s'_\lambda(s_\lambda^{-1}(\theta))}{2} f_0(s_\lambda^{-1}(\theta))(s_\lambda^{-1})'(\theta) - \left(1-\frac{s'_\lambda(-s_\lambda^{-1}(\theta))}{2}\right) f_0(-s_\lambda^{-1}(\theta))(-s_\lambda^{-1}(\theta))' \\ &= s'_\lambda(s_\lambda^{-1}(\theta)) f_0(s_\lambda^{-1}(\theta))(s_\lambda^{-1})'(\theta) = f_0(s_\lambda^{-1}(\theta)), \end{aligned}$$

which shows that Θ has density function (2) with base symmetric density function $f_0(\theta)$.

Proof of Theorem 3. Assumption B2 implies that there exist some constants $M^* > 0$ and $m^* > 0$ such that

$$\sup_{\eta\in\boldsymbol{H}} f(\theta;\boldsymbol{\eta}) < M^* \quad \text{and} \quad \inf_{\eta\in\boldsymbol{H}} f(\theta;\boldsymbol{\eta}) > m^*,$$

which lead to $E_{\eta_0}[\log f(\Theta;\boldsymbol{\eta})] > -\infty$, and for every sufficiently small ball $U \subseteq \boldsymbol{H}$,

$$E_{\eta_0}[\sup_{\eta\in U} \log f(\Theta;\boldsymbol{\eta})] < \infty.$$

Write $\boldsymbol{H}_0$ for a set $\{\boldsymbol{\eta} \in \boldsymbol{H} | E_{\boldsymbol{\eta}_0}[\log f(\Theta|\boldsymbol{\eta}_0)] = \sup_\eta E_{\eta_0}[\log f(\Theta;\boldsymbol{\eta})]\}$ of all points at which $E_{\eta_0}[\log f(\Theta;\boldsymbol{\eta})]$ attains its global maximum. It follows from the Kullback-Leibler inequality and Assumption B5 that $\boldsymbol{H}_0 = \{\boldsymbol{\eta}_0\}$. Using Theorem 5.14 of van der Vaart (2000), for $\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})}$ such that $\ell(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})}) \ge \ell(\boldsymbol{\eta}_0)$ and for every $\epsilon > 0$, we have

$$P(||\hat{\boldsymbol{\eta}}_n^{(\mathrm{MLE})} - \boldsymbol{\eta}_0|| \ge \epsilon) \to 0 \qquad (n \to \infty).$$

Next, we prove the asymptotic normality. Assumption B3 implies that the first three derivatives of the log-density are bounded uniformly in (θ,η) on $[-\pi,\pi)\times\boldsymbol{H}$, because $\left|\frac{\partial^3}{\partial\eta_i\partial\eta_j\partial\eta_k}\log f(\theta;\eta)\right|$ is bounded above by $\sum_{\alpha=1}^4 \frac{c_\alpha}{\{t'_{\lambda,\nu}(y)\}^\alpha}\Big|_{y=t_{\lambda,\nu}^{-1}(\theta-\mu)}$ and $\inf_{\theta,\eta} t'_{\lambda,\nu}(t_{\lambda,\nu}^{-1}(\theta-\mu)) \ge c$ for some constants $c_\alpha \ge 0$ $(\alpha = 1,\dots,4)$ and $c > 0$. Hence, the Lipschitz condition with criterion function $(\partial/\partial\boldsymbol{\eta})n^{-1}\ell(\boldsymbol{\eta})$ is verified. This together with Assumption 2 and Theorem 5.21 of van der Vaart (2000) implies the results.

Finally, we derive the Fisher information matrix given in (9). Recall that the log-likelihood function is given by

$$\ell(\mu,\kappa,\lambda,\nu|\theta_1,\dots,\theta_n) = \sum_{i=1}^n \log f_0(t_{\lambda,\nu}^{-1}(\theta_i-\mu);\kappa) - n\log(1-\nu\alpha_1(\kappa)).$$

It is noted that the first cosine moment $\alpha_1(\kappa)$ depends on the concentration parameter κ of the base symmetric density. Let $\tilde{y} = \theta - \mu$ and $H_\lambda(\tilde{y})$ be even functions with $|H'_\lambda(\tilde{y})| \le 1$ and $H_\lambda(0) = 0$. Suppose that the function $s(\tilde{y})$ is given by $s(\tilde{y}) = \tilde{y} + H_\lambda(\tilde{y})$, and pdf $f_0(\theta;\kappa)$ has a concentration parameter κ. As in Jones and Pewsey (2012), write $s^\lambda = ds/d\lambda$ and $\log f(\theta) = \log f_0(t_{\lambda,\nu}^{-1}(\theta-\mu);\kappa) - \log(1-\nu\alpha_1(\kappa))$. Then, we have the following derivatives: $s'(y) = 1 + H'_\lambda(y)$, $s^\lambda(y) = H^\lambda_\lambda(y)$, $s'^\lambda(y) =$

$H_\lambda'^{\lambda}(y)$, $s''(y) = H_\lambda''(y)$, and $s^{\lambda\lambda}(y) = H_\lambda^{\lambda\lambda}(y)$. Note that $H_\lambda^{\lambda}(y)$ is an even function of y. The derivatives of function $t(y)$ in terms of λ and ν are

$$t_{\lambda,\nu}^{\lambda}(y) = \sin^2(y - \nu \sin y)$$

and

$$\begin{aligned} t_{\lambda,\nu}^{\nu}(y) &= -\sin y - \lambda \sin y \sin 2(y - \nu \sin y) \\ &= -\sin y(1 + \lambda \sin 2(y - \nu \sin y)), \end{aligned}$$

respectively. Note that $t_{\lambda,\nu}^{\lambda}$ is an even function.

Consider the case with $\nu = 0$. Write $y = s_\lambda^{-1}(\theta - \mu)$, and then we have off-diagonal elements of the Fisher information matrix:

$$\iota_{\mu\kappa}|_{\nu=0} = \iota_{\kappa\lambda}|_{\nu=0} = 0.$$

These results are also extended to the case when $\nu \neq 0$, namely, when $t_{\lambda,\nu}(\theta) = s_\lambda(t_{0,\nu}(\theta))$. Then,

$$\frac{1}{1 - \nu\alpha_1(\kappa)} f_0(t_{\lambda,\nu}^{-1}(\theta); \kappa) = \frac{1}{1 - \nu\alpha_1(\kappa)} (f_0 \circ t_{0,\nu}^{-1})(s_\lambda^{-1}(\theta); \kappa),$$

and we have

$$\iota_{\mu\kappa} = \iota_{\mu\nu} = \iota_{\kappa\lambda} = \iota_{\lambda\nu} = 0$$

by making use of the above results because $1/(1 - \nu\alpha_1)(f_0 \circ t^{-1})(\cdot)|_{\lambda=0}$ is also a symmetric density. Set $y = t_{\lambda,\nu}^{-1}(\theta - \mu)$. Then, the other off-diagonal elements of the Fisher information matrix are given by

$$\begin{aligned} \iota_{\mu\lambda} &= -\frac{1}{1 - \nu\alpha_1} \int \left\{ \frac{t'^{\lambda}(y)t'(y) - t^{\lambda}(y)t''(y)}{t'(y)^2} (\log f_0(y))' + \frac{t^{\lambda}(y)}{t'(y)} (\log f_0(y))'' \right\} f_0(y)dy \\ &= \frac{1}{1 - \nu\alpha_1} \int \left\{ \left(-\frac{t'^{\lambda}(y)}{t'(y)} + \frac{t^{\lambda}(y)t''(y)}{t'(y)^2} \right) (\log f_0(y))' - \frac{t^{\lambda}(y)}{t'(y)} (\log f_0(y))'' \right\} f_0(y)dy \end{aligned}$$

and

$$\begin{aligned} \iota_{\kappa\nu} &= -E\left[\frac{\partial^2 \log f(\theta)}{\partial\kappa\partial\nu} \right] = -E\left[\frac{\partial^2}{\partial\kappa\partial\nu} \{\log f_0(y) - \log(1 - \nu\alpha_1)\} \right] \\ &= -E\left[\frac{\partial y}{\partial\nu} \frac{\partial}{\partial\kappa} (\log f_0(y))' + \frac{\partial\alpha_1}{\partial\kappa} \frac{1}{(1 - \nu\alpha_1)^2} \right] \\ &= \int \frac{t^{\nu}(y)}{t'(y)} \frac{\partial}{\partial\kappa} (\log f_0(y))' f(y)d\theta - \frac{\partial\alpha_1}{\partial\kappa} \frac{1}{(1 - \nu\alpha_1)^2} \\ &= \frac{1}{1 - \nu\alpha_1} \int t^{\nu}(y) \left(\frac{f_0'^{\kappa}(y)}{f_0(y)} - \frac{f_0'(y)f_0^{\kappa}(y)}{f_0(y)^2} \right) f_0(y)dy - \frac{\partial\alpha_1}{\partial\kappa} \frac{1}{(1 - \nu\alpha_1)^2} \\ &= \frac{1}{1 - \nu\alpha_1} \int t^{\nu}(y) \left(f_0'^{\kappa}(y) - \frac{f_0'(y)f_0^{\kappa}(y)}{f_0(y)} \right) dy - \frac{\partial\alpha_1}{\partial\kappa} \frac{1}{(1 - \nu\alpha_1)^2} \\ &= \frac{1}{1 - \nu\alpha_1} \int (-\sin y) \left(f_0'^{\kappa}(y) - \frac{f_0'(y)f_0^{\kappa}(y)}{f_0(y)} \right) dy - \frac{\partial\alpha_1}{\partial\kappa} \frac{1}{(1 - \nu\alpha_1)^2}. \end{aligned}$$

In a similar way, it is noted that $\iota_{\kappa\lambda}$ is given by

$$\iota_{\kappa\lambda} = \int t^{\lambda}(y) \left(f'^{\kappa}(y) - \frac{f'(y)f^{\kappa}(y)}{f(y)} \right) dy.$$

The element becomes zero since $t^{\lambda}(y)$, $f^{\kappa}(y)$, and $f(y)$ are even functions, and $f'^{\kappa}(y)$ and $f'(y)$ are odd functions. The diagonal elements are given by

$$\iota_{\mu\mu} = -E\left[\frac{\partial^2 \log f(y)}{\partial\mu^2} \right] = -E\left[\frac{\partial^2 y}{\partial\mu^2} (\log f(y))' + \left(\frac{\partial y}{\partial\mu} \right)^2 (\log f(y))'' \right]$$

$$= -E\left[-\frac{1}{t'(y)}\frac{t''(y)}{t'(y)^2}(\log f(y))' + \frac{1}{t'(y)^2}(\log f(y))''\right]$$

$$= \frac{1}{1-\nu\alpha_1}\int\left(\frac{t''(y)}{t'(y)^2}(\log f_0(y))' - \frac{1}{t'(y)}(\log f_0(y))''\right)f_0(y)dy$$

$$= \frac{1}{1-\nu\alpha_1}\int\left(\frac{t''(y)}{t'(y)^2}f_0'(y) - \frac{1}{t'(y)}\left(f_0''(y) - \frac{f_0'(y)^2}{f_0(y)}\right)\right)dy,$$

$$\iota_{\kappa\kappa} = -E\left[\frac{\partial^2 \log f(y)}{\partial\kappa^2}\right] = -E\left[\frac{\partial^2}{\partial\kappa^2}\{\log f_0(y) - \log(1-\nu\alpha_1)\}\right]$$

$$= -E\left[\frac{\partial^2 \log f_0(y)}{\partial\kappa^2}\right] - \nu\left(\frac{1}{1-\nu\alpha_1}\frac{\partial^2\alpha_1}{\partial\kappa^2} + \nu\frac{1}{(1-\nu\alpha_1)^2}\left(\frac{\partial\alpha_1}{\partial\kappa}\right)^2\right)$$

$$= -\frac{1}{1-\nu\alpha_1}\int t'(y)\frac{\partial^2 \log f_0(y)}{\partial\kappa^2}f_0(y)dy - \nu\left(\frac{1}{1-\nu\alpha_1}\frac{\partial^2\alpha_1}{\partial\kappa^2} + \nu\frac{1}{(1-\nu\alpha_1)^2}\left(\frac{\partial\alpha_1}{\partial\kappa}\right)^2\right)$$

$$= -\frac{1}{1-\nu\alpha_1}\int t'(y)\left(f_0^{\kappa\kappa}(y) - \frac{f_0^{\kappa}(y)^2}{f_0(y)}\right)dy - \nu\left(\frac{1}{1-\nu\alpha_1}\frac{\partial^2\alpha_1}{\partial\kappa^2} + \frac{\nu}{(1-\nu\alpha_1)^2}\left(\frac{\partial\alpha_1}{\partial\kappa}\right)^2\right)$$

$$= -\frac{1}{1-\nu\alpha_1}\int (1-\nu\cos y)\left(f_0^{\kappa\kappa}(y) - \frac{f_0^{\kappa}(y)^2}{f_0(y)}\right)dy - \nu\left(\frac{1}{1-\nu\alpha_1}\frac{\partial^2\alpha_1}{\partial\kappa^2} + \frac{\nu}{(1-\nu\alpha_1)^2}\left(\frac{\partial\alpha_1}{\partial\kappa}\right)^2\right),$$

$$\iota_{\lambda\lambda} = -E\left[\frac{\partial^2 \log f(y)}{\partial\lambda^2}\right] = -E\left[\frac{\partial}{\partial\lambda}\left(\frac{\partial y}{\partial\lambda}(\log f(y))'\right)\right]$$

$$= E\left[\frac{\partial}{\partial\lambda}\left(\frac{t^{\lambda}(y)}{t'(y)}(\log f(y))'\right)\right]$$

$$= E\left[\left(\frac{t^{\lambda\lambda}(y)}{t'(y)} - \frac{t^{\lambda}(y)t'^{\lambda}(y)}{t'(y)^2} + \frac{\partial y}{\partial\lambda}\left(\frac{t'^{\lambda}(y)}{t'(y)} - \frac{t^{\lambda}(y)t''(y)}{t'(y)^2}\right)\right)(\log f(y))' + \left(\frac{t^{\lambda}(y)}{t'(y)}\right)^2(\log f(y))''\right]$$

$$= E\left[\left(-2\frac{t^{\lambda}(y)t'^{\lambda}(y)}{t'(y)^2} + \frac{t^{\lambda}(y)^2t''(y)}{t'(y)^3}\right)(\log f(y))' - \frac{t^{\lambda}(y)^2}{t'(y)^2}(\log f(y))''\right]$$

$$= \frac{1}{1-\nu\alpha_1}\int\left(\left(-2\frac{t^{\lambda}(y)t'^{\lambda}(y)}{t'(y)} + \frac{t^{\lambda}(y)^2t''(y)}{t'(y)^2}\right)(\log f(y))' - \frac{t^{\lambda}(y)^2}{t'(y)}(\log f(y))''\right)f_0(y)dy$$

$$= \frac{1}{1-\nu\alpha_1}\int\left(\left(-2\frac{t^{\lambda}(y)t'^{\lambda}(y)}{t'(y)} + \frac{t^{\lambda}(y)^2t''(y)}{t'(y)^2}\right)f_0'(y) - \frac{t^{\lambda}(y)^2}{t'(y)}\left(f_0''(y) - \frac{f_0'(y)^2}{f_0(y)}\right)\right)dy,$$

and

$$\iota_{\nu\nu} = -E\left[\frac{\partial^2 \log f(y)}{\partial\nu^2}\right] = -E\left[\frac{\partial^2}{\partial\nu^2}(\log f_0(y) - \log(1-\nu\alpha_1))\right]$$

$$= -E\left[\frac{\partial}{\partial\nu}\left(\frac{\partial y}{\partial\nu}(\log f_0(y))'\right)\right] - \frac{\alpha_1^2}{(1-\nu\alpha_1)^2}$$

$$
\begin{aligned}
&= E\left[\frac{\partial}{\partial \nu}\left(\frac{t^{\nu}(y)}{t'(y)}(\log f_0(y))'\right)\right] - \frac{\alpha_1^2}{(1-\nu\alpha_1)^2} \\
&= E\left[\left(\frac{t^{\nu\nu}(y)}{t'(y)} - \frac{t^{\nu}(y)t'^{\nu}(y)}{t'(y)^2} + \frac{\partial y}{\partial \nu}\left(\frac{t'^{\nu}(y)}{t'(y)} - \frac{t^{\nu}(y)t''(y)}{t'(y)^2}\right)\right)(\log f_0(y))' - \left(\frac{t^{\nu}(y)}{t'(y)}\right)^2(\log f_0(y))''\right] - \frac{\alpha_1^2}{(1-\nu\alpha_1)^2} \\
&= E\left[\left(\frac{t^{\nu\nu}(y)}{t'(y)} - 2\frac{t^{\nu}(y)t'^{\nu}(y)}{t'(y)^2} + \frac{t^{\nu}(y)^2t''(y)}{t'(y)^3}\right)(\log f_0(y))' - \frac{t^{\nu}(y)^2}{t'(y)^2}(\log f_0(y))''\right] - \frac{\alpha_1^2}{(1-\nu\alpha_1)^2} \\
&= \frac{1}{1-\nu\alpha_1}\int\left(\left(t^{\nu\nu}(y) - 2\frac{t^{\nu}(y)t'^{\nu}(y)}{t'(y)} + \frac{t^{\nu}(y)^2t''(y)}{t'(y)^2}\right)(\log f_0(y))' - \frac{t^{\nu}(y)^2}{t'(y)}(\log f_0(y))''\right)f_0(y)dy - \frac{\alpha_1^2}{(1-\nu\alpha_1)^2} \\
&= \frac{1}{1-\nu\alpha_1}\int\left(\left(t^{\nu\nu}(y) - 2\frac{t^{\nu}(y)t'^{\nu}(y)}{t'(y)} + \frac{t^{\nu}(y)^2t''(y)}{t'(y)^2}\right)f_0'(y) - \frac{t^{\nu}(y)^2}{t'(y)}\left(f_0''(y) - \frac{f_0'(y)^2}{f_0(y)}\right)\right)dy - \frac{\alpha_1^2}{(1-\nu\alpha_1)^2} \\
&= \frac{1}{1-\nu\alpha_1}\int\left(\left(-\frac{\sin 2y}{1-\nu\cos y} + \frac{\nu\sin^3 y}{(1-\nu\cos y)^2}\right)f_0'(y) - \frac{\sin^2 y}{1-\nu\cos y}\left(f_0''(y) - \frac{f_0'(y)^2}{f_0(y)}\right)\right)dy - \frac{\alpha_1^2}{(1-\nu\alpha_1)^2}.
\end{aligned}
$$

We have used the fact that $t^{\lambda\lambda}(y) = 0$. The resulting expected information matrix is given by

$$
I(\boldsymbol{\eta}_0) = \begin{pmatrix} \iota_{\mu\mu} & \iota_{\mu\kappa} & \iota_{\mu\lambda} & \iota_{\mu\nu} \\ \iota_{\mu\kappa} & \iota_{\kappa\kappa} & \iota_{\kappa\lambda} & \iota_{\kappa\nu} \\ \iota_{\mu\lambda} & \iota_{\kappa\lambda} & \iota_{\lambda\lambda} & \iota_{\lambda\nu} \\ \iota_{\mu\nu} & \iota_{\kappa\nu} & \iota_{\lambda\nu} & \iota_{\nu\nu} \end{pmatrix} = \begin{pmatrix} \iota_{\mu\mu} & 0 & \iota_{\mu\lambda} & 0 \\ 0 & \iota_{\kappa\kappa} & 0 & \iota_{\kappa\nu} \\ \iota_{\mu\lambda} & 0 & \iota_{\lambda\lambda} & 0 \\ 0 & \iota_{\kappa\nu} & 0 & \iota_{\nu\nu} \end{pmatrix},
$$

and its corresponding determinant becomes $\det\{\boldsymbol{I}(\boldsymbol{\eta}_0)\} = (\iota_{\mu\mu}\iota_{\lambda\lambda} - \iota_{\mu\lambda}^2)(\iota_{\kappa\kappa}\iota_{\nu\nu} - \iota_{\kappa\nu}^2)$.

From the above expressions for the entries of the Fisher information matrix, elements $\iota_{\kappa\nu}$, $\iota_{\kappa\kappa}$, and $\iota_{\nu\nu}$ do not depend on the skewness parameter λ. It should be noted that the pairs of the entries in the Fisher information matrix between (μ, λ) and (κ, ν) are always zero.

The following lemma is needed to prove Theorem 4. This lemma provides a consistent result for minimum contrast estimators. When we set a contrast function as $\bar{U}_n(\boldsymbol{\eta}) := -(1/n)\log\{f(\boldsymbol{\theta}_n;\boldsymbol{\eta})p(\boldsymbol{\eta})\}$, the MAP estimator $\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})}$ is viewed as the minimum contrast estimator, that is, $\bar{U}_n(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})}) \le \bar{U}_n(\boldsymbol{\eta})$ for any $\boldsymbol{\eta} \in \boldsymbol{H}$. Here, we should note that prior $p(\boldsymbol{\eta})$ can be zero in the boundary of the parameter space $\boldsymbol{H}$. Essentially, $p(\lambda)$ and $p(\nu)$, which are proposed in Eq. (11), are zeros at $\lambda = \pm 1$ and $\nu = \pm 1$, respectively. Then, because the log-prior $\log p(\boldsymbol{\eta})$ can be $-\infty$ at such points, the standard technique to prove the consistency of estimators is not available. However, the theorem below works even for such a case.

Lemma 1. *Let* $\bar{U}(\boldsymbol{\eta}) := E[-\log f(\Theta;\boldsymbol{\eta})]$. *We assume that*

E1: *For any* $\epsilon > 0$,

$$
\liminf_{n\to\infty} \inf_{||\boldsymbol{\eta}-\boldsymbol{\eta}_0||\ge\epsilon} \bar{U}_n(\boldsymbol{\eta}) \ge \inf_{||\boldsymbol{\eta}-\boldsymbol{\eta}_0||\ge\epsilon} \bar{U}(\boldsymbol{\eta}) \qquad \text{a.s.}
$$

E2: *For any* $\epsilon > 0$, $\inf_{||\boldsymbol{\eta}-\boldsymbol{\eta}_0||\ge\epsilon} \bar{U}(\boldsymbol{\eta}) > \bar{U}(\boldsymbol{\eta}_0)$.

E3: $\limsup_{n\to\infty} \bar{U}_n(\boldsymbol{\eta}_0) \le \bar{U}(\boldsymbol{\eta}_0)$ *a.s.*

Then, $\lim_{n\to\infty} ||\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})} - \boldsymbol{\eta}_0|| = 0$ *a.s., and hence* $\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})} \to_p \boldsymbol{\eta}_0$.

Condition E1 is a weaker version of the uniform strong law of large numbers, and Condition E2 means the identifiability of model $f(\theta;\boldsymbol{\eta})$.

Proof of Lemma 1. From the definition of the MAP estimator, it holds that

$$
\limsup_{n\to\infty}\left[\bar{U}_n(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})}) - \bar{U}_n(\boldsymbol{\eta}_0)\right] \le 0 \qquad \text{a.s.}
$$

Hence, combining the above equation and Condition E3, we have

$$
\begin{aligned}
\limsup_{n\to\infty} \bar{U}_n(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})}) &\le \limsup_{n\to\infty}\left\{\bar{U}_n(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})}) - \bar{U}_n(\boldsymbol{\eta}_0)\right\} + \limsup_{n\to\infty} \bar{U}_n(\boldsymbol{\eta}_0) \\
&\le \limsup_{n\to\infty} \bar{U}_n(\boldsymbol{\eta}_0)
\end{aligned}
$$

$$\leq \bar{U}(\boldsymbol{\eta}_0) \qquad \text{a.s.} \tag{A.2}$$

Let $(\bar{\Xi}, \bar{\mathcal{A}}, \bar{P})$ be an infinite Cartesian product of infinite probability spaces $(\Xi, \mathcal{A}, P)$ with $\bar{P}$ being the true probability distribution. Then, an infinite sequence $(\theta_1, \theta_2, \cdots)$ of observations becomes an element of $\bar{\Xi}$. Then, for any $\epsilon > 0$, and any element $\boldsymbol{\theta}_\infty = (\theta_1, \theta_2, \cdots) \in \bar{\Xi}$ that satisfies inequality (A.2) and Condition E1, there exists a positive constant $n_0 := n_0(\boldsymbol{\theta}, \epsilon)$ such that for every large n with $n \geq n_0$,

$$\inf_{||\boldsymbol{\eta}-\boldsymbol{\eta}_0||\geq\epsilon} \bar{U}_n(\boldsymbol{\eta}) \geq \inf_{||\boldsymbol{\eta}-\boldsymbol{\eta}_0||\geq\epsilon} \bar{U}(\boldsymbol{\eta}) > \bar{U}(\boldsymbol{\eta}_0) \geq \bar{U}_n(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})}) \qquad \text{a.s.}$$

From this, we have

$$\begin{aligned}
1 &= \bar{P}\left(\liminf_{n\to\infty} \inf_{||\boldsymbol{\eta}-\boldsymbol{\eta}_0||\geq\epsilon} \bar{U}_n(\boldsymbol{\eta}) > \bar{U}_n(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})})\right) \\
&\leq \bar{P}\left(\forall\epsilon > 0, \quad \exists n_0 := n_0(\epsilon) \quad \text{such that } n \geq n_0 \text{ implies } \inf_{||\boldsymbol{\eta}-\boldsymbol{\eta}_0||\geq\epsilon} \bar{U}_n(\boldsymbol{\eta}) > \bar{U}_n(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})})\right) \\
&\leq \bar{P}\left(\forall\epsilon > 0, \quad \exists n_0 := n_0(\epsilon) \quad \text{such that } n \geq n_0 \text{ implies } ||\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})} - \boldsymbol{\eta}_0|| < \epsilon\right) \\
&= \bar{P}\left(\lim_{n\to\infty} ||\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})} - \boldsymbol{\eta}_0|| = 0\right).
\end{aligned}$$

Recall that we have used relation $||\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})} - \boldsymbol{\eta}_0|| \geq \epsilon$, which implies $\inf_{||\boldsymbol{\eta}-\boldsymbol{\eta}_0||\geq\epsilon} \bar{U}_n(\boldsymbol{\eta}) \leq \bar{U}_n(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})})$ and its contraposition for the derivation of the above transformation.

Proof of Theorem 4. Because there exists a positive constant $M > 0$ such that $|\log f(\theta; \boldsymbol{\eta})| < M$ for any θ and $\boldsymbol{\eta}$, by the dominated convergence theorem, $\bar{U}(\boldsymbol{\eta}) = E[-\log f(\Theta; \boldsymbol{\eta})]$ is continuous with respect to $\boldsymbol{\eta}$. Combining this and the compactness of $\boldsymbol{H}$ leads to E2.

It is necessary to illustrate Conditions E1 and E3. First, we verify Condition E1. Because $|\log f(\theta; \boldsymbol{\eta})| \leq M$ by Condition B2, the boundedness of $\alpha_1(\kappa)$ as in B3, and the uniform law of large numbers, we have

$$\lim_{n\to\infty} \sup_{\boldsymbol{\eta}\in\boldsymbol{H}} \left| -\frac{1}{n}\sum_{i=1}^{n} \log f(\theta_i; \boldsymbol{\eta}) - E[-\log f(\Theta; \boldsymbol{\eta})] \right| = 0, \quad \text{a.s.} \tag{A.3}$$

from Theorem 16(a) of Ferguson (2017). Thus, we have

$$\begin{aligned}
\liminf_{n\to\infty} \inf_{||\boldsymbol{\eta}-\boldsymbol{\eta}_0||\geq\epsilon} \bar{U}_n(\boldsymbol{\eta}) &= \liminf_{n\to\infty} \inf_{||\boldsymbol{\eta}-\boldsymbol{\eta}_0||\geq\epsilon} \left\{ -\frac{1}{n}\sum_{i=1}^{n} \log f(\theta_i; \boldsymbol{\eta}) - \frac{1}{n}\log p(\boldsymbol{\eta}) + E[\log f(\Theta; \boldsymbol{\eta})] - E[\log f(\Theta; \boldsymbol{\eta})] \right\} \\
&\geq \liminf_{n\to\infty} \inf_{||\boldsymbol{\eta}-\boldsymbol{\eta}_0||\geq\epsilon} \left\{ -\frac{1}{n}\sum_{i=1}^{n} \log f(\theta_i; \boldsymbol{\eta}) - E[-\log f(\Theta; \boldsymbol{\eta})] \right\} + \liminf_{n\to\infty} \inf_{||\boldsymbol{\eta}-\boldsymbol{\eta}_0||\geq\epsilon} \left\{ -\frac{1}{n}\log p(\boldsymbol{\eta}) \right\} \\
&\quad + \inf_{||\boldsymbol{\eta}-\boldsymbol{\eta}_0||\geq\epsilon} \bar{U}(\boldsymbol{\eta}).
\end{aligned} \tag{A.4}$$

The first term in Eq. (A.4) is zero by Eq. (A.3). Moreover, $-\log p(\boldsymbol{\eta})$ is bounded from below on the neighborhood of $\boldsymbol{\eta}_0$ because $p(\boldsymbol{\eta})$ is continuous on $\boldsymbol{H}$ and $p(\boldsymbol{\eta}_0) > 0$, and hence the second term in Eq. (A.4) is also zero. Therefore, Condition E1 is verified. Next, we verify Condition E3. By the strong law of large numbers, Condition C1, and $p(\boldsymbol{\eta}_0) > 0$, we have

$$\begin{aligned}
\lim_{n\to\infty} \bar{U}_n(\boldsymbol{\eta}_0) &= \lim_{n\to\infty} \left\{ -\frac{1}{n}\sum_{i=1}^{n} \log f(\theta_i; \boldsymbol{\eta}_0) \right\} - \lim_{n\to\infty} \frac{1}{n}\log p(\boldsymbol{\eta}_0) \\
&= \bar{U}(\boldsymbol{\eta}_0) \qquad \text{a.s.},
\end{aligned}$$

which leads to Condition E3. Therefore, $\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})}$ is strongly consistent.

Finally, we show the asymptotic normality of $\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})}$. To do that, it suffices to verify the condition $n^{-1}(\partial/\partial\boldsymbol{\eta})\ell(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})}) = o_p(n^{-1/2})$ of Theorem 5.21 of van der Vaart (2000). Let $n^{-1}(\partial/\partial\boldsymbol{\eta})\ell(\boldsymbol{\eta})$ be a criterion function. Then, this immediately holds because $n^{-1}(\partial/\partial\boldsymbol{\eta})\ell(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})}) = -(1/n)(\partial/\partial\boldsymbol{\eta})\log p(\hat{\boldsymbol{\eta}}_n^{(\mathrm{MAP})}) = O_p(n^{-1})$. Hence, arguing as in the proof of Theorem 3 leads to the conclusion.

References

Abe, T., Pewsey, A., 2011. Sine-skewed circular distributions. Statistical Papers 52 (3), 683–707.
Abe, T., Pewsey, A., Shimizu, K., 2009. On Papakonstantinou's extension of the cardioid distribution. Statistics & Probability Letters 79 (20), 2138–2147.

Abe, T., Pewsey, A., Shimizu, K., 2013. Extending circular distributions through transformation of argument. Annals of the Institute of Statistical Mathematics 65 (5), 833–858.
Agostinelli, C., 2007. Robust estimation for circular data. Computational Statistics & Data Analysis 51 (12), 5867–5875.
Ameijeiras-Alonso, J., Ley, C., Pewsey, A., Verdebout, T., 2019. On optimal tests for circular reflective symmetry about an unknown central direction, *Statistical Papers*. https://doi.org/10.1007/s00362-019-01150-7
Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. Journal of the Royal Statistical Society. Series B 65 (2), 367–389.
Batschelet, E., 1981. Circular Statistics in Biology. Academic Press, New York.
DiCiccio, T.J., Stern, S.E., 1993. On Bartlett adjustments for approximate Bayesian inference. Biometrika 80 (4), 731–740.
Ferguson, T.S., 2017. A Course in Large Sample Theory. Routledge.
Fisher, N.I., 1993. Statistical Analysis of Circular Data. Cambridge University Press.
Galindo-Garre, F., Vermunt, J.K., 2006. Avoiding boundary estimates in latent class analysis by Bayesian posterior mode estimation. Behaviormetrika 33 (1), 43–59.
Galindo-Garre, F., Vermunt, J.K., Bergsma, W.P., 2004. Bayesian posterior estimation of logit parameters with small samples. Sociological Methods & Research 33 (1), 88–117.
Gatto, R., Jammalamadaka, S.R., 2007. The generalized von Mises distribution. Statistical Methodology 4 (3), 341–353.
Guttorp, P., Lockhart, R.A., 1988. Finding the location of a signal: A Bayesian analysis. Journal of the American Statistical Association 83 (402), 322–330.
Jammalamadaka, S.R., SenGupta, A., 2001. Topics in Circular Statistics. World Scientific.
Jones, M., 2015. On families of distributions with shape parameters. International Statistical Review 83 (2), 175–192.
Jones, M.C., 2014. Generating distributions by transformation of scale. Statistica Sinica 24, 749–771.
Jones, M.C., Pewsey, A., 2005. A family of symmetric distributions on the circle. Journal of the American Statistical Association 100 (472), 1422–1428.
Jones, M.C., Pewsey, A., 2012. Inverse Batschelet distributions for circular data. Biometrics 68 (1), 183–193.
Kass, R., Tierney, L., Kadane, J., 1990. The validity of posterior expansions based on Laplace's method. In: Geisser, S., Hodges, J., Press, S.J., Zellner, A. (Eds.), Bayesian and Likelihood Methods in Statistics and Econometrics. Elsevier Science Publishers BV, Amsterdam, North Holland.
Kass, R.E., Tierney, L., Kadane, J.B., 1989. Approximate methods for assessing influence and sensitivity in Bayesian analysis. Biometrika 76 (4), 663–674.
Kato, S., Jones, M., 2015. A tractable and interpretable four-parameter family of unimodal distributions on the circle. Biometrika 102 (1), 181–190.
Ley, C., Verdebout, T., 2014. Simple optimal tests for circular reflective symmetry about a specified median direction. Statistica Sinica 24, 1319–1339.
Ley, C., Verdebout, T., 2017. Modern Directional Statistics. CRC Press.
Lindley, D.V., 1980. Approximate Bayesian methods (with discussion). In: Bayesian statistics (Valencia, 1979). Univ. Press, Valencia, pp. 223–245.
Mardia, K.V., El-Atoum, S., 1976. Bayesian inference for the von Mises–Fisher distribution. Biometrika 63 (1), 203–206.
Mardia, K.V., Jupp, P.E., 1999. Directional Statistics. John Wiley & Sons.
Miyata, Y., 2018. Laplace approximations using n^{α}-consistent estimators. Journal of Statistical Planning and Inference 194, 25–31.
Miyata, Y., Shiohama, T., Abe, T., 2020. Estimation of finite mixture models of skew-symmetric circular distributions. Metrika 83, 895–922.
Mulder, K., Klugkist, I., 2017. Bayesian estimation and hypothesis tests for a circular generalized linear model. Journal of Mathematical Psychology 80, 4–14.
Nuñez-Antonio, G., Gutiérrez-Peña, E., 2005. A Bayesian analysis of directional data using the projected normal distribution. Journal of Applied Statistics 32 (10), 995–1001.
Papakonstantinou, V., 1979. Beiträge zur zirkulären Statistik. University of Zurich, Switzerland.
Pewsey, A., 2002. Testing circular symmetry. Canadian Journal of Statistics 30 (4), 591–600.
Pewsey, A., Neuhäuser, M., Ruxton, G.D., 2013. Circular Statistics in R. Oxford University Press.
Pewsey, A., Shimizu, K., de la Cruz, R., 2011. On an extension of the von Mises distribution due to Batschelet. Journal of Applied Statistics 38 (5), 1073–1085.
Ravindran, P., Ghosh, S.K., 2011. Bayesian analysis of circular data using wrapped distributions. Journal of Statistical Theory and Practice 5 (4), 547–561.
Tierney, L., Kadane, J.B., 1986. Accurate approximations for posterior moments and marginal densities. Journal of the American Statistical Association 81 (393), 82–86.
Tierney, L., Kass, R.E., Kadane, J.B., 1989. Fully exponential Laplace approximations to expectations and variances of nonpositive functions. Journal of the American Statistical Association 84 (407), 710–716.
Umbach, D., Jammalamadaka, S.R., 2009. Building asymmetry into circular distributions. Statistics & Probability Letters 79 (5), 659–663.
van der Vaart, A.W., 2000. Asymptotic Statistics. Cambridge University Press.
Yfantis, E., Borgman, L., 1982. An extension of the von Mises distribution. Communications in Statistics-Theory and Methods 11 (15), 1695–1706.