

# INFERENCE FOR SPHERICAL LOCATION UNDER HIGH CONCENTRATION

BY DAVY PAINDAVEINE<sup>1,2,\*</sup> AND THOMAS VERDEBOUT<sup>1,\*\*</sup>

<sup>1</sup>Université libre de Bruxelles. \*[dpaindav@ulb.ac.be](mailto:dpaindav@ulb.ac.be); \*\*[tverdebo@ulb.ac.be](mailto:tverdebo@ulb.ac.be)

<sup>2</sup>Université Toulouse 1 Capitole

Motivated by the fact that circular or spherical data are often much concentrated around a location  $\theta$ , we consider inference about  $\theta$  under *high concentration* asymptotic scenarios for which the probability of any fixed spherical cap centered at  $\theta$  converges to one as the sample size  $n$  diverges to infinity. Rather than restricting to Fisher–von Mises–Langevin distributions, we consider a much broader, semiparametric, class of rotationally symmetric distributions indexed by the location parameter  $\theta$ , a scalar concentration parameter  $\kappa$  and a functional nuisance  $f$ . We determine the class of distributions for which high concentration is obtained as  $\kappa$  diverges to infinity. For such distributions, we then consider inference (point estimation, confidence zone estimation, hypothesis testing) on  $\theta$  in asymptotic scenarios where  $\kappa_n$  diverges to infinity at an arbitrary rate with the sample size  $n$ . Our asymptotic investigation reveals that, interestingly, optimal inference procedures on  $\theta$  show consistency rates that depend on  $f$ . Using asymptotics “à la Le Cam,” we show that the spherical mean is, at any  $f$ , a parametrically superefficient estimator of  $\theta$  and that the Watson and Wald tests for  $\mathcal{H}_0 : \theta = \theta_0$  enjoy similar, nonstandard, optimality properties. We illustrate our results through simulations and treat a real data example. On a technical point of view, our asymptotic derivations require challenging expansions of rotationally symmetric functionals for large arguments of  $f$ .

**1. Introduction.** Directional statistics is concerned with data on the unit sphere  $\mathcal{S}^{p-1} = \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x}\|^2 = \mathbf{x}'\mathbf{x} = 1\}$  of  $\mathbb{R}^p$  or more generally on Riemannian manifolds such as a torus or an infinite cylinder. Directional data are present in many fields and have attracted a lot of attention in the last decade. Recent applications include analysis of magnetic remanence through copulae on product manifolds in Jupp (2015), analysis of animal movement using angular regression in Rivest et al. (2016) or analysis of flight trajectories through principal component analysis for functional data on  $\mathcal{S}^{p-1}$  in Dai and Müller (2018), to cite only a few. For an overview of the topic, we refer to Mardia and Jupp (2000) and Ley and Verdebout (2017).

In this paper, we consider a class of distributions on  $\mathcal{S}^{p-1}$  admitting a density at  $\mathbf{x}$  that is proportional to  $f(\kappa\mathbf{x}'\theta)$ , where  $\theta \in \mathcal{S}^{p-1}$ ,  $\kappa > 0$  and  $f$  is a monotone increasing function from  $\mathbb{R}$  to  $\mathbb{R}^+$  (throughout, densities on  $\mathcal{S}^{p-1}$  will be with respect to the surface area measure). The resulting distribution on the sphere will be denoted as  $\text{Rot}_p(\theta, \kappa, f)$  to stress its *rotational symmetry*: if  $\mathbf{X} \sim \text{Rot}_p(\theta, \kappa, f)$ , then  $\mathbf{OX}$  and  $\mathbf{X}$  are equal in distribution for any  $p \times p$  orthogonal matrix  $\mathbf{O}$  such that  $\mathbf{O}\theta = \theta$ . Clearly,  $\theta$  is the modal location on the sphere, hence plays the role of a location parameter. In contrast,  $\kappa$  is a scale or *concentration* parameter. This terminology is justified by the fact that, for many functions  $f$ , the distribution  $\text{Rot}_p(\theta, \kappa, f)$  becomes arbitrarily concentrated around  $\theta$  as  $\kappa$  diverges to infinity; it is

---

Received January 2019; revised June 2019.

*MSC2020 subject classifications.* Primary 62E20, 62F30; secondary 62F05, 62F12.

*Key words and phrases.* Concentrated distributions, directional statistics, Le Cam’s asymptotic theory of statistical experiments, local asymptotic normality, superefficiency.

in particular so for the celebrated *Fisher–von Mises–Langevin* (FvML) distributions, that are obtained with  $f = \exp$ . FvML distributions play a central role in directional statistics, a role that can be compared to the one played by Gaussian distributions in classical multivariate setups. For instance, the responses of the circular/spherical regression models in Rivest (1986), Downs and Mardia (2002), SenGupta, Kim and Arnold (2013) and Rosenthal et al. (2014) are FvML with a location parameter that depends on the predictors.

In most applications, the location parameter  $\theta$  is the parameter of interest, whereas the concentration parameter  $\kappa$  and the infinite-dimensional parameter  $f$  are unspecified nuisances. The most classical estimator of  $\theta$  is the *spherical mean*, whereas the most celebrated test for  $\mathcal{H}_0 : \theta = \theta_0$ , where  $\theta_0 \in S^{p-1}$  is fixed, is the Watson test (see Sections 3 and 4, resp.). In the standard asymptotic scenario under which  $n$  diverges to infinity with  $\kappa$  fixed, the asymptotic properties of these procedures are well-known; see, for example, Mardia and Jupp (2000). In particular, the spherical mean is root- $n$  consistent, whereas the Watson test shows nontrivial asymptotic powers under sequences of local alternatives of the form  $\mathcal{H}_1^{(n)} : \theta = \theta_n$  with  $\sqrt{n}\|\theta_n - \theta_0\| \rightarrow c > 0$ .

In practice, the asymptotic results above are relevant in cases where the underlying concentration  $\kappa$  is neither too small nor too large. For small values of  $\kappa$ , the fixed- $\kappa$  asymptotic distribution of the spherical mean and the corresponding asymptotic null distribution of  $W_n$  only poorly approximate the exact distribution of these statistics, unless the sample size  $n$  at hand is extremely large. This motivates considering a double asymptotic scenario where  $\kappa = \kappa_n$  goes to zero as  $n$  diverges to infinity. The observations  $\mathbf{X}_{n1}, \dots, \mathbf{X}_{nn}$  are then assumed to form a random sample from the distribution  $\text{Rot}_p(\theta, \kappa_n, f)$ , with  $\kappa_n = o(1)$ , which makes it here strictly necessary to consider triangular arrays of observations. Such a “low-concentration double asymptotic scenario” was considered in Paindaveine and Verdebout (2017), where it was proved that the faster  $\kappa_n$  goes to zero, the poorer the consistency rates of the aforementioned inference procedures. More precisely, (i) if  $\kappa_n = o(1)$  with  $\kappa_n\sqrt{n} \rightarrow \infty$ , then  $\kappa_n\sqrt{n}(\hat{\theta}_n - \theta)$  is asymptotically normal, so that the consistency rate of the spherical mean deteriorates from  $\sqrt{n}$  (in the standard fixed- $\kappa$  case) to  $\kappa_n\sqrt{n}$  (in the present case); (ii) if  $\kappa_n = O(1/\sqrt{n})$ , then the spherical mean is not consistent anymore. Similarly, in situation (i), the Watson test shows nontrivial asymptotic powers under sequences of local alternatives of the form  $\mathcal{H}_1^{(n)} : \theta = \theta_n$  with  $\kappa_n\sqrt{n}\|\theta_n - \theta_0\| \rightarrow c > 0$ , and, in situation (ii), there is no sequence of alternatives under which this test would be consistent. These behaviors of the spherical mean and of the Watson test are nonstandard yet expected: as the concentration  $\kappa_n$  gets smaller, the distribution  $\text{Rot}_p(\theta, \kappa_n, f)$  becomes increasingly closer to the uniform distribution on  $S^{p-1}$  for which the parameter of interest  $\theta$  is not identifiable. In other words, inference on  $\theta$  is increasingly challenging as  $\kappa$  decreases to zero, which reflects in the deterioration of the consistency rates above.

The situation for large concentrations  $\kappa$  is similar yet different. On the one hand, it is still so that a standard, fixed- $\kappa$ , asymptotic analysis could in principle fail describing in a suitable way the finite-sample behaviors of the spherical mean and of the Watson test statistic under high concentration. On the other hand, inference about  $\theta$  intuitively becomes increasingly easy as the distribution gets more and more concentrated around  $\theta$ , which should make it possible to define “superefficient” estimators and tests on  $\theta$ . Inference for “concentrated” FvML distributions actually has already been quite much considered in the literature. One of the first papers tackling inference problems for the location parameter of FvML distributions under large values of  $\kappa$  is Watson (1984), where asymptotic results as  $\kappa \rightarrow \infty$  with  $n$  fixed were derived. In the same asymptotic scenario, Rivest (1986) investigated the null limiting behavior of a goodness-of-fit test for FvML distributions, whereas Rivest (1989), Downs and Mardia (2002) and Downs (2003) considered spherical regression in a concentrated FvML setup. Rosenthal et al. (2014) analyzed concentrated data using a regression model with a

FvML noise. Fujikoshi and Watamori (1992) obtained the asymptotic null distributions of various test statistics for  $\mathcal{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$  again as  $\kappa \rightarrow \infty$  with  $n$  fixed, and derived the asymptotic powers of the corresponding tests under appropriate sequences of local alternatives. Still in the framework of FvML distributions, Watamori (1996) reviewed point estimation and (one-sample and multisample) hypothesis testing in the standard asymptotic scenario where  $n \rightarrow \infty$  with  $\kappa$  fixed and in the concentrated scenario where  $\kappa \rightarrow \infty$  with  $n$  fixed. Arnold and Jupp (2013) and Arnold, Jupp and Schaeben (2018) considered estimation of “highly concentrated rotations.” Finally, Chikuse (2003a) considered inference for concentrated matrix FvML distributions, still in a setup where  $\kappa \rightarrow \infty$  with  $n$  fixed; see also Chikuse (2003b). Monographs covering inference for concentrated FvML distributions include Watson (1983) and Mardia and Jupp (2000).

This review of the literature shows that inference on  $\boldsymbol{\theta}$  under high concentration is a classical topic in directional statistics. Yet this review also reveals some important limitations in previous studies: (i) all asymptotic results available are as  $\kappa \rightarrow \infty$  with  $n$  fixed, while, parallel to the low-concentration case above, a double asymptotic scenario where  $\kappa = \kappa_n$  would go to infinity with  $n$  would be at least as natural (particularly so if  $\kappa_n$  would be allowed to diverge to infinity at an arbitrary rate as a function of  $n$ ); (ii) all results are limited to the parametric case of FvML distributions, so that the asymptotic properties of the spherical mean and of the Watson test remain unknown in the broader semiparametric class of  $\text{Rot}_p(\boldsymbol{\theta}, \kappa, f)$  distributions; (iii) for hypothesis testing, most works focused on the null hypothesis: very few results try and describe asymptotic powers under sequences of local alternatives and, more importantly, not a single optimality result, to the best of our knowledge, was obtained in the literature. In this paper, we therefore fill an important gap by deriving results that are getting rid of the limitations (i)–(iii).

The outline of the paper is as follows. In Section 2, we fix the notation, introduce the assumptions that will be used throughout and characterize the rotationally symmetric distributions that provide high concentration for arbitrarily large values of  $\kappa$ . In Section 3, we derive the asymptotic distribution of the spherical mean in a double asymptotic scenario where  $\kappa_n$  diverges to infinity at an arbitrary rate with  $n$ . Interestingly, in contrast with what happens for low concentrations, the consistency rate here depends on the nuisance function  $f$ . We also provide confidence zones for  $\boldsymbol{\theta}$  that quite naturally take the form of spherical caps centered at the spherical mean. In Section 4, we study the asymptotic behavior of the Watson and Wald tests. In Section 5, we turn to optimality issues and show that, under mild assumptions on  $f$ , the sequence of statistical experiments considered is locally asymptotically normal. We establish the Le Cam optimality of the spherical mean estimator and of the Watson and Wald tests under high concentration. Finally, a real data application is conducted in Section 6 and a wrap up is provided in Section 7. Proofs are collected in the Supplementary Material Paindaveine and Verdebout (2019).

**2. High concentration.** Throughout, we will denote as  $\mathbf{P}_{\boldsymbol{\theta}_n, \kappa_n, f}^{(n)}$  the hypothesis under which the observations  $\mathbf{X}_{n1}, \dots, \mathbf{X}_{nn}$  form a random sample from the distribution  $\text{Rot}_p(\boldsymbol{\theta}_n, \kappa_n, f)$  described in the Introduction, that is, the hypothesis under which these observations are mutually independent and share the common density

$$(2.1) \quad \mathbf{x} \mapsto \frac{c_{p, \kappa_n, f} \Gamma(\frac{p-1}{2})}{2\pi^{(p-1)/2}} f(\kappa_n \mathbf{x}' \boldsymbol{\theta}_n),$$

where  $\Gamma(\cdot)$  is the Euler Gamma function and the constant  $c_{p, \kappa, f}$  is given by

$$(2.2) \quad c_{p, \kappa, f} := 1 / \int_{-1}^1 (1 - s^2)^{(p-3)/2} f(\kappa s) ds.$$

In the sequel,  $f : \mathbb{R} \rightarrow \mathbb{R}^+$  is assumed to be monotone nondecreasing on  $(-\infty, 0]$  and monotone increasing on  $[0, \infty)$ . Under this assumption, the location parameter  $\theta_n$  is properly identified as the modal location on the sphere. One way to also make  $\kappa_n$  and  $f$  identifiable would be to further impose  $f(0) = f'(0) = 1$ . We will not impose these conditions since we also want to consider functions  $f$  that are not differentiable at zero. The resulting lack of identifiability will not be an issue in the sequel since  $\kappa_n$  and  $f$  play the role of nuisance parameters when conducting inference on  $\theta_n$ .

We will often make use of the *tangent-normal decomposition* of  $\mathbf{X}_{ni}$  with respect to  $\theta_n$ , which reads  $\mathbf{X}_{ni} = u_{ni}\theta_n + v_{ni}\mathbf{S}_{ni}$ , with

$$u_{ni} = \mathbf{X}'_{ni}\theta_n, \quad v_{ni} := \sqrt{1 - u_{ni}^2},$$

and

$$\mathbf{S}_{ni} := \frac{(\mathbf{I}_p - \theta_n\theta'_n)\mathbf{X}_{ni}}{\|(\mathbf{I}_p - \theta_n\theta'_n)\mathbf{X}_{ni}\|} = \frac{1}{v_{ni}}(\mathbf{I}_p - \theta_n\theta'_n)\mathbf{X}_{ni}.$$

The cosine  $u_{ni}$  is associated with the latitude of  $\mathbf{X}_{ni}$  with respect to the “north pole”  $\theta_n$ , whereas  $\mathbf{S}_{ni}$  determines the corresponding hyper-longitude. Under  $\mathbf{P}_{\theta_n, \kappa_n, f}^{(n)}$ ,  $u_{n1}$  and  $\mathbf{S}_{n1}$  are mutually independent,  $\mathbf{S}_{n1}$  is uniformly distributed on  $\mathcal{S}_{\theta_n}^\perp := \{\mathbf{x} \in \mathcal{S}^{p-1} : \mathbf{x}'\theta_n = 0\}$ , and  $u_{n1}$  admits the density

$$(2.3) \quad s \mapsto c_{p, \kappa_n, f} (1 - s^2)^{(p-3)/2} f(\kappa_n s) \mathbb{I}[s \in [-1, 1]],$$

where  $\mathbb{I}[A]$  stands for the indicator function of the set  $A$ . The moments of  $u_{n1}$  under  $\mathbf{P}_{\theta_n, \kappa_n, f}^{(n)}$  will play an important role below and will be denoted as  $e_{n\ell} := E[u_{n1}^\ell]$ ,  $\ell = 1, 2, \dots$ . We will also write  $\tilde{e}_{n2} = e_{n2} - e_{n1}^2$  for the corresponding variance. The function  $f$  governs (jointly with  $\kappa_n$ ) the distribution of the angle  $\arccos(u_{n1})$  between  $\mathbf{X}_{n1}$  and  $\theta_n$ , hence is sometimes referred to as an *angular function*.

The present paper is concerned with sequences of rotationally symmetric distributions that are asymptotically highly concentrated, meaning that the probability mass of any fixed spherical cap centered at  $\theta_n$  converges to one as  $n$  diverges to infinity. More precisely, we will say that the sequence of hypotheses  $\mathbf{P}_{\theta_n, \kappa_n, f}^{(n)}$  is asymptotically highly concentrated if and only if, for any sequence  $(\kappa_n)$  diverging to infinity and any  $\varepsilon \in (0, 2)$ , we have

$$(2.4) \quad \mathbf{P}_{\theta_n, \kappa_n, f}^{(n)}[\mathbf{X}'_{n1}\theta_n > 1 - \varepsilon] = c_{p, \kappa_n, f} \int_{1-\varepsilon}^1 (1 - s^2)^{(p-3)/2} f(\kappa_n s) ds \rightarrow 1,$$

that is, if and only if  $u_{n1}$  converges in probability to one as soon as  $(\kappa_n)$  diverges to infinity. Since this is clearly a property that depends on  $f$  only, we will say that  $f$  provides high concentration if and only if (2.4) holds. Not all functions  $f$  provide high concentration. The polynomial functions  $z \mapsto f(z) = z^b \mathbb{I}[z \geq 0]$ ,  $b > 0$ , are examples since, for any  $\varepsilon \in (0, 1)$ , they yield

$$c_{p, \kappa_n, f} \int_{1-\varepsilon}^1 (1 - s^2)^{(p-3)/2} f(\kappa_n s) ds = \frac{\int_{1-\varepsilon}^1 (1 - s^2)^{(p-3)/2} s^b ds}{\int_0^1 (1 - s^2)^{(p-3)/2} s^b ds} =: C < 1,$$

where  $C$  does not depend on  $n$ . It is easy to check that  $z \mapsto f(z) = \frac{\pi}{2} + \arctan(z)$  does not provide high concentration either, but that the angular FvML function  $z \mapsto f(z) = \exp(z)$  does. It is therefore desirable to characterize the functions  $f$  providing high concentration, which is the aim of the following result.

**THEOREM 2.1.** *Let  $f : \mathbb{R} \rightarrow \mathbb{R}^+$  be monotone nondecreasing on  $(-\infty, 0]$  and monotone increasing on  $[0, \infty)$ . Assume that  $f$  is differentiable in a neighborhood of  $\infty$  (in the sense that there exists  $M$  such that  $f$  is differentiable over  $(M, \infty)$ ) and put  $\varphi_f := f'/f$ , where  $f'$  is the derivative of  $f$ . Then we have the following:*

- (i) *If  $\kappa\varphi_f(\kappa) \nearrow \infty$  as  $\kappa \rightarrow \infty$ , then  $f$  provides high concentration.*
- (ii) *If  $\kappa\varphi_f(\kappa) \rightarrow c (> 0)$  as  $\kappa \rightarrow \infty$ , then  $f$  does not provide high concentration.*
- (iii) *If  $\kappa\varphi_f(\kappa) \searrow 0$  as  $\kappa \rightarrow \infty$ , then  $f$  does not provide high concentration.*

In this result,  $g(\kappa) \nearrow \infty$  (resp.,  $g(\kappa) \searrow 0$ ) as  $\kappa \rightarrow \infty$  means that (a)  $g(\kappa)$  diverges to infinity (resp., converges to zero) as  $\kappa$  diverges to infinity and that (b) there exists  $M$  such that  $\kappa \mapsto g(\kappa)$  is monotone nondecreasing (resp., monotone nonincreasing) over  $(M, \infty)$ . Essentially, Theorem 2.1 states that high concentration is obtained if  $f(z)$  diverges to infinity at least exponentially fast as  $z$  diverges to infinity. In particular, this result confirms that the polynomial and arctan functions  $f$  above do not provide high concentration, but that the FvML one does. Writing throughout  $z^b := \text{sgn}(z)|z|^b$ , it also shows that all functions  $z \mapsto f_b(z) := \exp(z^b)$ , with  $b > 0$ , do provide high concentration. These functions  $f$ , which include the FvML one, will be our main running examples below.

In the rest of the paper,  $\mathcal{F}$  will stand for the collection of functions  $f : \mathbb{R} \rightarrow \mathbb{R}^+$  that (i) are monotone nondecreasing on  $(-\infty, 0]$  and monotone increasing on  $[0, \infty)$ , (ii) are differentiable in a neighborhood of  $\infty$ , (iii) are such that  $\kappa\varphi_f(\kappa) \nearrow \infty$  as  $\kappa \rightarrow \infty$  and (iv) satisfy, for any  $\xi, \zeta > -1$ ,

$$(2.5) \quad \int_{-1}^1 g_{\xi, \zeta}(s) |e^{\log f(\kappa s) - \log f(\kappa)} - e^{(s-1)\kappa\varphi_f(\kappa)}| ds = o\left(\frac{1}{(\kappa\varphi_f(\kappa))^{\xi+1}}\right)$$

as  $\kappa \rightarrow \infty$ , with  $g_{\xi, \zeta}(s) := (1-s)^\xi(1+s)^\zeta$ . As the following result shows, our prototypical examples of angular functions  $f$  providing high concentration meet these properties.

**PROPOSITION 2.1.** *For any  $b > 0$ , the function  $z \mapsto f_b(z) = \exp(z^b)$  belongs to  $\mathcal{F}$ .*

As already mentioned, the moments of  $u_{n1} = \mathbf{X}'_{n1}\theta_n$  under  $\mathbf{P}_{\theta_n, \kappa_n, f}^{(n)}$  will play a key role in the sequel. It will actually be important to understand the asymptotic behavior of these moments under high concentration. This is the role of the following result.

**THEOREM 2.2.** *Fix an integer  $p \geq 2$  and  $f \in \mathcal{F}$ . Let  $(\kappa_n)$  be a positive real sequence that diverges to infinity. Then*

- (i)  $1 - e_{n2} = \frac{p-1}{\kappa_n\varphi_f(\kappa_n)} + o\left(\frac{1}{\kappa_n\varphi_f(\kappa_n)}\right),$
- (ii)  $\tilde{e}_{n2} = \frac{p-1}{2(\kappa_n\varphi_f(\kappa_n))^2} + o\left(\frac{1}{(\kappa_n\varphi_f(\kappa_n))^2}\right)$

and

$$(iii) \quad \mathbb{E}[v_{n1}^4] = \frac{p^2-1}{(\kappa_n\varphi_f(\kappa_n))^2} + o\left(\frac{1}{(\kappa_n\varphi_f(\kappa_n))^2}\right)$$

as  $n \rightarrow \infty$ .

As a corollary, we have

$$(2.6) \quad \frac{(1 - e_{n2})^2}{\tilde{e}_{n2}} = 2(p-1) + o(1) \quad \text{and} \quad \frac{\mathbb{E}[v_{n1}^4]}{\tilde{e}_{n2}} = 2(p+1) + o(1)$$

as  $n \rightarrow \infty$ . Also, Vitali’s theorem (see, e.g., Theorem 5.5 in [Shorack \(2000\)](#)) readily implies that, under the conditions of Theorem 2.2,  $e_{n1} = 1 + o(1)$  as  $n \rightarrow \infty$ . One could obtain an expansion of  $1 - e_{n1}$  that is similar to the one in Theorem 2.2(i), but we will not do so since this is not needed for our purposes.

**3. Point estimation.** As mentioned in the [Introduction](#), the most classical estimator of location under rotational symmetry is the spherical mean, which is given by

$$\hat{\theta}_n := \frac{\bar{\mathbf{X}}_n}{\|\bar{\mathbf{X}}_n\|},$$

with  $\bar{\mathbf{X}}_n := \frac{1}{n} \sum_{i=1}^n \mathbf{X}_{ni}$ . Under  $\mathbb{P}_{\theta, \kappa_n, f}^{(n)}$ ,  $E[\mathbf{X}_{n1}] = \lambda_{\kappa_n, f} \theta$  for some positive scalar factor  $\lambda_{\kappa_n, f}$ , so that the spherical mean is a moment-type estimator of  $\theta$ . It is easy to check that it is also the maximum likelihood estimator of  $\theta$  in the class of FvML distributions. This makes it desirable to investigate the asymptotic behavior of this estimator under high concentration. We have the following result.

**THEOREM 3.1.** *Fix an integer  $p \geq 2$ ,  $\theta \in \mathcal{S}^{p-1}$  and  $f \in \mathcal{F}$ . Let  $(\kappa_n)$  be a positive real sequence that diverges to infinity. Then, under  $\mathbb{P}_{\theta, \kappa_n, f}^{(n)}$ ,*

$$(3.1) \quad \sqrt{n\kappa_n\varphi_f(\kappa_n)}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{I}_p - \theta\theta')$$

as  $n \rightarrow \infty$ , so that, still under  $\mathbb{P}_{\theta, \kappa_n, f}^{(n)}$ ,

$$(3.2) \quad n\kappa_n\varphi_f(\kappa_n)(1 - (\theta'\hat{\theta}_n)^2) \xrightarrow{\mathcal{D}} \chi_{p-1}^2$$

as  $n \rightarrow \infty$  (throughout,  $\xrightarrow{\mathcal{D}}$  denotes convergence in distribution).

Since the sequence  $(\kappa_n\varphi_f(\kappa_n))$  diverges to infinity under high concentration, Theorem 3.1 shows that the consistency rate of the spherical mean is faster than the usual parametric root- $n$  rate. Interestingly, this consistency rate depends on the angular function  $f$ . For instance, for  $f(z) = \exp(z^b)$  with  $b > 0$ , the rate is  $n^{(b+1)/2}$ , hence can be arbitrary close to the standard root- $n$  rate for small  $b$ , but can also provide arbitrary fast polynomial convergence. Clearly, even faster rates can be achieved by considering more extreme high concentration patterns.

The asymptotic result (3.2) in principle allows constructing confidence zones for  $\theta$ . More precisely, it follows from this result that a confidence zone for  $\theta$  at asymptotic confidence level  $1 - \alpha$  is given by

$$\left\{ \theta \in \mathcal{S}^{p-1} : |\theta'\hat{\theta}_n| \geq \sqrt{1 - \frac{\chi_{p-1, 1-\alpha}^2}{n\kappa_n\varphi_f(\kappa_n)}} \right\},$$

where  $\chi_{p-1, 1-\alpha}^2$  denotes the upper  $\alpha$ -quantile of the  $\chi_{p-1}^2$  distribution. This confidence zone, however, is problematic in two respects. First, it is not connected, as it takes the form of two antipodal spherical caps centered at  $\pm\hat{\theta}_n$ , which is not natural. Second, while the  $f$ -dependent consistency rate in Theorem 3.1 is interesting, it also leads to confidence zones that cannot be used in practice since  $f$  is usually an unspecified nuisance. The first problem can be dealt with by deriving a weak limiting result for  $\theta'\hat{\theta}_n$  obtained from a second-order delta method (while Theorem 3.1 results from a classical, first-order, delta method). We have the following result.

**THEOREM 3.2.** Fix an integer  $p \geq 2$ ,  $\theta \in \mathcal{S}^{p-1}$  and  $f \in \mathcal{F}$ . Let  $(\kappa_n)$  be a positive real sequence that diverges to infinity. Then, under  $\mathbf{P}_{\theta, \kappa_n, f}^{(n)}$ ,  $2n\kappa_n\varphi_f(\kappa_n)(1 - \theta'\hat{\theta}_n) \xrightarrow{\mathcal{D}} \chi_{p-1}^2$  as  $n \rightarrow \infty$ .

This second-order result provides confidence zones at asymptotic confidence level  $1 - \alpha$  that are given by

$$(3.3) \quad \left\{ \theta \in \mathcal{S}^{p-1} : \theta'\hat{\theta}_n \geq 1 - \frac{\chi_{p-1, 1-\alpha}^2}{2n\kappa_n\varphi_f(\kappa_n)} \right\},$$

hence take, quite naturally, the form of (connected) spherical caps centered at  $\hat{\theta}_n$ . Of course, these confidence zones still cannot be used in practice since  $f$  is unspecified. Fortunately, Theorem 2.2(i) allows replacing the unknown quantity  $\kappa_n\varphi_f(\kappa_n)$  by the quantity  $(p - 1)/(1 - e_{n2}) = (p - 1)/(1 - E[(\mathbf{X}'_{n1}\theta)^2])$ , which can be naturally estimated by  $(p - 1)/(1 - \hat{e}_{n2})$ , where we let  $\hat{e}_{n2} := \frac{1}{n} \sum_{i=1}^n (\mathbf{X}'_{ni}\hat{\theta}_n)^2$ . The following result, that guarantees that this replacement has no asymptotic impact, opens the door to the construction of feasible confidence zones.

**THEOREM 3.3.** Fix an integer  $p \geq 2$ ,  $\theta \in \mathcal{S}^{p-1}$  and  $f \in \mathcal{F}$ . Let  $(\kappa_n)$  be a positive real sequence that diverges to infinity. Then, under  $\mathbf{P}_{\theta, \kappa_n, f}^{(n)}$ ,

$$\frac{\sqrt{n(p-1)}(\hat{\theta}_n - \theta)}{\sqrt{1 - \hat{e}_{n2}}} \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{I}_p - \theta\theta')$$

as  $n \rightarrow \infty$ , and, still under  $\mathbf{P}_{\theta, \kappa_n, f}^{(n)}$ ,

$$\frac{n(p-1)(1 - (\theta'\hat{\theta}_n)^2)}{1 - \hat{e}_{n2}} \xrightarrow{\mathcal{D}} \chi_{p-1}^2 \quad \text{and} \quad \frac{2n(p-1)(1 - \theta'\hat{\theta}_n)}{1 - \hat{e}_{n2}} \xrightarrow{\mathcal{D}} \chi_{p-1}^2$$

as  $n \rightarrow \infty$ , where, in all cases,  $\hat{e}_{n2} = \frac{1}{n} \sum_{i=1}^n (\mathbf{X}'_{ni}\hat{\theta}_n)^2$ .

As a direct corollary, a feasible version of the spherical cap confidence zone in (3.3) is

$$(3.4) \quad \left\{ \theta \in \mathcal{S}^{p-1} : \theta'\hat{\theta}_n \geq 1 - \frac{1 - \hat{e}_{n2}}{2n(p-1)} \chi_{p-1, 1-\alpha}^2 \right\}.$$

We conducted the following Monte Carlo exercises to check the validity of Theorems 3.2–3.3. For each combination of  $a \in \{0.5, 1\}$  and  $b \in \{0.5, 1, 1.4\}$ , we generated  $M = 10,000$  random samples of size  $n = 100$  from the rotationally symmetric distribution with location  $\theta = (1, 0, 0)' \in \mathcal{S}^2$ , concentration  $\kappa_n = n^a$  and angular function  $z \mapsto f_b(z) = \exp(z^b)$  (numerical overflows prevented us from considering larger values of  $b$ ). For each  $a$  and  $b$ , Figure 1 plots kernel density estimates of the resulting  $M$  values of  $T_n^{\text{Oracle}} := 2n\kappa_n\varphi_f(\kappa_n)(1 - \theta'\hat{\theta}_n)$  and  $T_n^{\text{Feasible}} := 2n(p-1)(1 - \theta'\hat{\theta}_n)/(1 - \hat{e}_{n2})$  (for  $a = 1$ , raw histograms are also provided). Clearly, Figure 1 supports the theoretical results above, with possibly one exception only, namely the case of  $T_n^{\text{Feasible}}$  with  $b = 0.5$ . We therefore focused on this case and repeated the same Monte Carlo exercise with  $n = 10,000$ . The results, that are shown in Figure 2, are now in perfect agreement with the theory for  $a = 1$ , whereas the fit still is not excellent for  $a = 0.5$ . A closer inspection provides the explanation: despite the large sample size  $n$  considered in Figure 2, the distribution associated with  $a = b = 0.5$  is far for being highly concentrated; see the right panel of this figure. The fit observed for  $a = 0.5$  in the left panel of Figure 2 therefore does not contradict our theoretical results, which would materialize for higher concentrations.

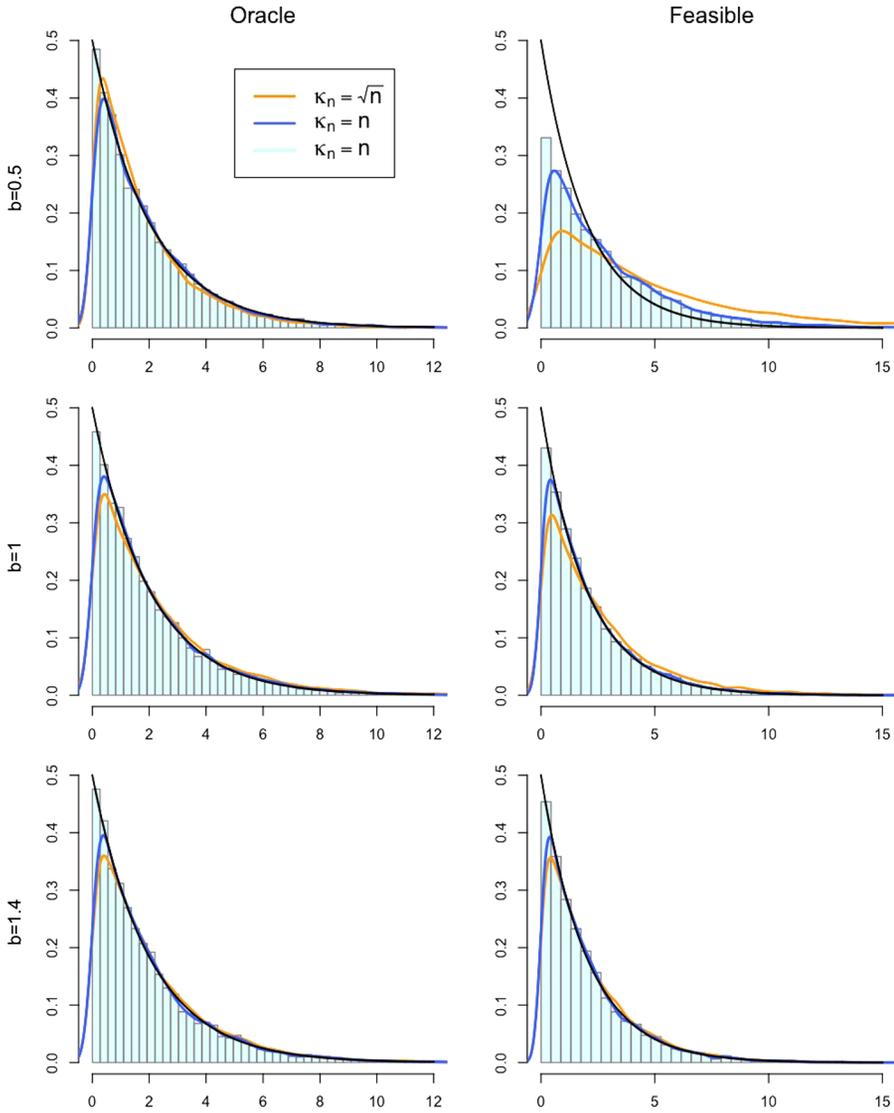


FIG. 1. Histograms of  $2n\kappa_n\varphi_f(\kappa_n)(1 - \theta' \hat{\theta}_n)$  (left) and  $2n(p - 1)(1 - \theta' \hat{\theta}_n)/(1 - \hat{e}_{n2})$  (right) computed from 10,000 random samples from  $P_{\theta, \kappa_n, f_b}^{(n)}$ , with  $p = 3$ ,  $n = 100$ ,  $\theta = (1, 0, 0)'$ ,  $\kappa_n = n$ , and  $f_b(z) = \exp(z^b)$ , for  $b = 0.5$  (top),  $b = 1$  (middle) and  $b = 1.4$  (bottom). The blue curve is the kernel density estimate resulting from the R command `density` with default parameter values. The orange curve is the corresponding kernel density estimate for random samples generated with  $\kappa_n = \sqrt{n}$ . The theoretical limiting density, namely the density of the  $\chi_2^2$  distribution, is plotted in black.

**4. Hypothesis testing.** We now turn to hypothesis testing and, more specifically, to the generic problem of testing the null hypothesis  $\mathcal{H}_0 : \theta = \theta_0$  against the alternative  $\mathcal{H}_1 : \theta \neq \theta_0$ , where  $\theta_0$  is a fixed unit  $p$ -vector. In this section, we consider the Watson test (Watson (1983), p. 140) and the Wald test (Hayakawa (1990), Hayakawa and Puri (1985)), that respectively reject the null hypothesis at asymptotic level  $\alpha$  whenever

$$(4.1) \quad W_n := \frac{n(p - 1)\bar{\mathbf{X}}_n'(\mathbf{I}_p - \theta_0\theta_0')\bar{\mathbf{X}}_n}{1 - \frac{1}{n}\sum_{i=1}^n(\mathbf{X}_{ni}'\theta_0)^2}$$

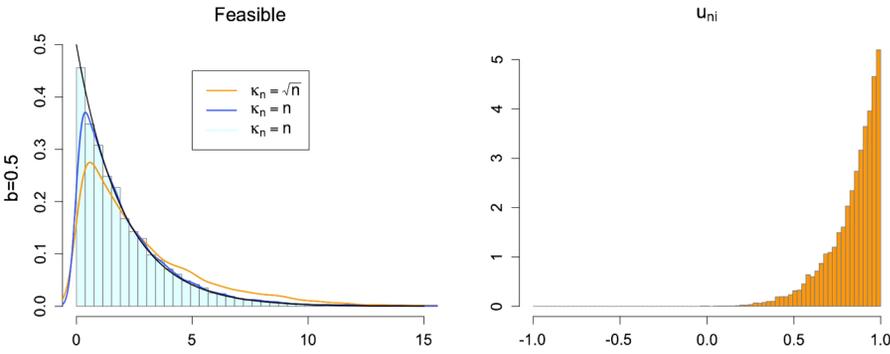


FIG. 2. (Left:) Histogram of  $2n(p-1)(1-\theta'\hat{\theta}_n)/(1-\hat{e}_{n2})$  computed from 2500 random samples from  $P_{\theta, \kappa_n, f_b}^{(n)}$ , with  $p = 3$ ,  $n = 10,000$ ,  $\theta = (1, 0, 0)'$ ,  $\kappa_n = n$ , and  $f_b(z) = \exp(z^b)$ , for  $b = 0.5$ . The blue curve is the kernel density estimate resulting from the R command `density` with default parameter values. The orange curve is the corresponding kernel density estimate for random samples generated with  $\kappa_n = \sqrt{n}$ . The theoretical limiting density is still plotted in black. (Right:) Histogram of  $u_{ni} = \mathbf{X}'_{ni}\theta$ ,  $i = 1, \dots, n$ , where the  $\mathbf{X}_{ni}$ 's form a random sample from  $P_{\theta, \kappa_n, f_b}^{(n)}$ , with  $p = 3$ ,  $n = 10,000$ ,  $\theta = (1, 0, 0)'$ ,  $\kappa_n = \sqrt{n}$ , and  $f_b(z) = \exp(z^b)$ , for  $b = 0.5$ .

and

$$(4.2) \quad S_n = \frac{n(p-1)(\bar{\mathbf{X}}'_n \theta_0)^2 \hat{\theta}'_n (\mathbf{I}_p - \theta_0 \theta'_0) \hat{\theta}_n}{1 - \frac{1}{n} \sum_{i=1}^n (\mathbf{X}'_{ni} \theta_0)^2}$$

exceed the critical value  $\chi^2_{p-1, 1-\alpha}$ . In standard asymptotic scenarios where the sample size  $n$  diverges to infinity with  $\kappa$  fixed, the Watson and Wald test statistics are asymptotically equivalent in probability under the null hypothesis, hence also under sequences of contiguous alternatives, so that these tests may be considered asymptotically equivalent. As shown in Paindaveine and Verdebout (2017), however, this asymptotic equivalence does not survive asymptotic scenarios for which  $\kappa_n = O(1/\sqrt{n})$  as  $n$  diverges to infinity. This suggests investigating the asymptotic behavior of these tests under the high concentration scenarios considered in the previous sections.

To do so, let

$$R_n := \frac{1 - \frac{1}{n} \sum_{i=1}^n (\mathbf{X}'_{ni} \theta_0)^2}{\sqrt{2(p-1)} e_{n2}^{1/2}}$$

and decompose the Watson and Wald test statistics into

$$W_n =: \frac{\tilde{W}_n}{R_n} \quad \text{and} \quad S_n =: \frac{(\bar{\mathbf{X}}'_n \theta_0)^2 \tilde{S}_n}{R_n}.$$

We then have the following lemma.

LEMMA 4.1. Fix an integer  $p \geq 2$ ,  $\theta_0 \in S^{p-1}$  and  $f \in \mathcal{F}$ . Let  $(\kappa_n)$  be a positive real sequence that diverges to infinity. Let  $(\tau_n)$  be a bounded sequence in  $\mathbb{R}^p$  such that  $\theta_n = \theta_0 + v_n \tau_n \in S^{p-1}$  for all  $n$ , with  $v_n := 1/\sqrt{n\kappa_n\varphi_f(\kappa_n)}$ . Then, under  $P_{\theta_n, \kappa_n, f}^{(n)}$ , we have  $R_n = 1 + o_p(1)$  and  $\bar{\mathbf{X}}'_n \theta_0 = 1 + o_p(1)$  as  $n \rightarrow \infty$ , so that  $W_n = \tilde{W}_n + o_p(1)$  and  $S_n = \tilde{S}_n + o_p(1)$  as  $n \rightarrow \infty$ .

This lemma ensures that, both under the sequence of null hypotheses  $P_{\theta_0, \kappa_n, f}^{(n)}$  (taking  $\tau_n \equiv \mathbf{0}$ ) and under sequences of local alternatives of the form  $P_{\theta_n, \kappa_n, f}^{(n)}$ , one may focus on  $\tilde{W}_n$  and  $\tilde{S}_n$  when studying the asymptotic behaviors of the Watson and Wald test statistics in (4.1)–(4.2). These asymptotic behaviors are provided in the following result.

**THEOREM 4.1.** Fix an integer  $p \geq 2$ ,  $\theta_0 \in \mathcal{S}^{p-1}$  and  $f \in \mathcal{F}$ . Let  $(\kappa_n)$  be a positive real sequence that diverges to infinity. Let  $(\tau_n)$  be a sequence in  $\mathbb{R}^p$  converging to  $\tau$  and such that  $\theta_n = \theta_0 + v_n \tau_n \in \mathcal{S}^{p-1}$  for all  $n$ , with  $v_n := 1/\sqrt{n\kappa_n\varphi_f(\kappa_n)}$ . Then, (i) under  $\mathbb{P}_{\theta_0, \kappa_n, f}^{(n)}$ ,

$$W_n = S_n + o_P(1) \xrightarrow{\mathcal{D}} \chi_{p-1}^2$$

as  $n \rightarrow \infty$ ; (ii) under  $\mathbb{P}_{\theta_n, \kappa_n, f}^{(n)}$ ,

$$W_n = S_n + o_P(1) \xrightarrow{\mathcal{D}} \chi_{p-1}^2(\|\tau\|^2)$$

as  $n \rightarrow \infty$ , where  $\chi_{p-1}^2(c)$  denotes the noncentral chi-square distribution with  $p - 1$  degrees of freedom and noncentrality parameter  $c$ .

This result shows that, under high concentration, the Watson and Wald test statistics remain asymptotically equivalent in probability both under the null hypothesis and under the considered sequences of local alternatives. Both tests show asymptotic size  $\alpha$  under the null hypothesis, irrespective of the angular function  $f$  and of the rate at which the concentration  $\kappa_n$  diverges to infinity. Theorem 4.1 also reveals that  $v_n$  describes the consistency rate of these tests, in the sense that the Watson and Wald tests show nontrivial asymptotic powers (i.e., asymptotic powers in  $(\alpha, 1)$ ) under sequences of local alternatives of the form  $\mathbb{P}_{\theta_n, \kappa_n, f}^{(n)}$ , with  $v_n^{-1}\|\theta_n - \theta_0\| \rightarrow c > 0$ . Like in point estimation, this rate depends on  $f$  and is faster than the standard parametric root- $n$  rate that is obtained for fixed  $\kappa$ ; that is, compared to the alternatives that can be detected in the standard fixed- $\kappa$  situation, less severe—hence, more challenging—alternatives can be detected under high concentration.

We performed the following Monte Carlo exercise to illustrate the results in Theorem 4.1. For each combination of  $a \in \{0.5, 1\}$ ,  $b \in \{0.5, 1, 1.4\}$  and  $\ell \in \{0, 1, 2, 3, 4\}$ , we generated  $M = 10,000$  random samples of size  $n = 100$  from the rotationally symmetric distribution on  $\mathcal{S}^2$  with concentration  $\kappa_n = n^a$ , angular function  $z \mapsto f_b(z) = \exp(z^b)$ , and location

$$(4.3) \quad \theta_{n\ell} = \begin{pmatrix} \cos \alpha_{n\ell} & -\sin \alpha_{n\ell} & 0 \\ \sin \alpha_{n\ell} & \cos \alpha_{n\ell} & 0 \\ 0 & 0 & 1 \end{pmatrix} \theta_0,$$

where we let  $\theta_0 = (1, 0, 0)'$  and  $\alpha_{n\ell} := 2 \arcsin(\ell v_n/2)$ , with  $v_n = 1/\sqrt{n\kappa_n\varphi_{f_b}(\kappa_n)}$ . The alternative locations  $\theta_{n\ell}$  rewrite  $\theta_0 + v_n \tau_{n\ell}$  for some  $p$ -vector  $\tau_{n\ell}$  with norm  $\ell$ . Clearly,  $\ell = 0$  refers to the null hypothesis  $\mathcal{H}_0 : \theta = \theta_0$  and  $\ell = 1, 2, 3, 4$  correspond to increasingly severe alternatives. In each sample, we performed the Watson and Wald tests at asymptotic level  $\alpha = 5\%$ . Figure 3 plots, as a function of  $\ell$ , the resulting rejection frequencies, or more precisely, the difference between these rejection frequencies and the corresponding theoretical limiting powers

$$(4.4) \quad \mathbb{P}[Y_\ell > \chi_{p-1, 1-\alpha}^2] \quad \text{with } Y_\ell \sim \chi_{p-1}^2(\ell^2);$$

see Theorem 4.1(ii). The figure also reports the results for sample size  $n = 700$ , but for the case with highest concentration (i.e., the case  $(a, b) = (1, 1.4)$ ) for which data generation led to numerical overflow. Rejection frequencies agree well with the limiting powers (note the scale of the vertical axes), particularly for  $\kappa_n = n$  which provides a higher concentration than  $\kappa_n = \sqrt{n}$ . The agreement improves as the sample size increases. In all cases but the one with lowest concentration (i.e., the case  $(a, b) = (0.5, 0.5)$ ), the asymptotic equivalence between the Watson and Wald tests materializes already for  $n = 100$ .

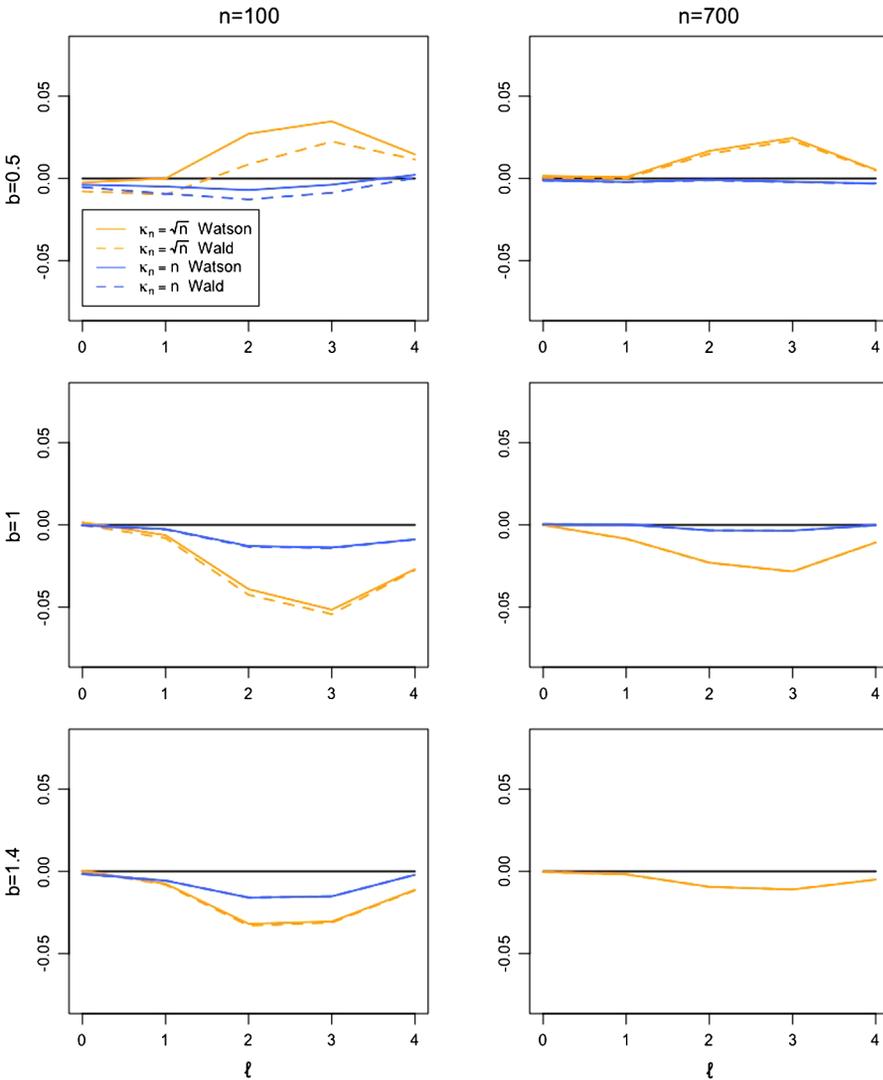


FIG. 3. Plots of the differences between (i) the rejection frequencies of the Watson and Wald tests and (ii) their theoretical limiting powers in (4.4). Rejection frequencies are obtained from a collection of  $M = 10,000$  random samples of size  $n = 100$  (left) or  $n = 700$  (right) from the rotationally symmetric distribution on  $S^2$  with location  $\theta_{n\ell}$  in (4.3), concentration  $\kappa_n = \sqrt{n}$  or  $n$ , and angular function  $z \mapsto f_b = \exp(z^b)$ , with  $b = 0.5$  (top), 1 (middle) or 1.4 (bottom); see Section 4 for details.

**5. Local asymptotic normality.** The Watson test was shown to enjoy strong optimality properties, both in the standard asymptotic scenario where the concentration  $\kappa_n$  is fixed and in the nonstandard one where the concentration goes to zero; see [Paindaveine and Verdebout \(2017\)](#). In the latter scenario, the Wald test, on the contrary, fails to be optimal. In this section, we investigate the optimality properties of the Watson and Wald tests and of the spherical mean estimator under high concentration. Optimality will be in the Le Cam sense, which requires studying the *Local Asymptotic Normality (LAN)* of the sequence of fixed- $f$  parametric submodels at hand.

To do so, we will need to reinforce our assumptions on  $f$ . Let  $p (\geq 2)$  be an integer,  $(\kappa_n)$  be a positive sequence diverging to infinity and  $(t_n)$  be a bounded positive sequence. In the sequel, we will denote as  $\mathcal{F}_{LAN}(p, \kappa_n, t_n)$  the collection of angular functions  $f \in \mathcal{F}$  such that,

as  $\kappa \rightarrow \infty$ ,

$$\frac{1}{f(\kappa)} \int_{-1}^1 (\varphi_f(\kappa s) - \varphi_f(\kappa))^2 (1 - s^2)^{(p-3)/2} f(\kappa s) ds = o\left(\frac{1}{\kappa^{(p+1)/2}(\varphi_f(\kappa))^{(p-3)/2}}\right)$$

and such that, letting  $h_n^\pm(s, w) := -\frac{1}{2}t_n^2\kappa_n v_n^2 s \pm c_n t_n \kappa_n v_n (1 - s^2)^{1/2} w^{1/2}$ , with  $v_n := 1/\sqrt{n\kappa_n\varphi_f(\kappa_n)}$  and  $c_n := (1 - \frac{1}{4}v_n^2 t_n^2)^{1/2}$ ,

$$(5.1) \quad \frac{1}{f(\kappa_n)} \int_{-1}^1 \int_0^1 |\log f(\kappa_n s + h_n^\pm(s, w)) - \log f(\kappa_n s) - h_n^\pm(s, w)\varphi_f(\kappa_n s)| f(\kappa_n s) \times (1 - s^2)^{(p-3)/2} dG_p(w) ds = o\left(\frac{1}{n(\kappa_n\varphi_f(\kappa_n))^{(p-1)/2}}\right)$$

as  $n \rightarrow \infty$ , where, for  $p \geq 3$ ,  $G_p$  is the cumulative distribution function of the Beta( $\frac{1}{2}, \frac{p-2}{2}$ ) distribution, whereas, for  $p = 2$ ,  $G_p$  is the cumulative distribution function of the Dirac distribution in 1. As shown in the next result, most angular functions  $f_b$  do satisfy these extra assumptions, sometimes under an extremely mild restriction on the rate at which the sequence  $(\kappa_n)$  diverges to infinity with  $n$ .

**PROPOSITION 5.1.** *Let  $p (\geq 2)$  be an integer,  $(\kappa_n)$  be a positive sequence diverging to infinity and  $(t_n)$  be a bounded positive sequence. Then, for any  $b \geq 1$ , the function  $z \mapsto f_b(z) = \exp(z^b)$  belongs to  $\mathcal{F}_{\text{LAN}}(p, \kappa_n, t_n)$ . Provided that there exists  $\varepsilon \in (0, 2)$  such that  $\kappa_n^b/(\log n) \geq (1 - b)/(2 - \varepsilon)$  for  $n$  large enough, the same holds for  $f_b$ , with  $b \in (\frac{1}{2}, 1)$ .*

In other words,  $f_b$ , with  $b \geq 1$ , belongs to  $\mathcal{F}_{\text{LAN}}(p, \kappa_n, t_n)$  irrespective of the sequences  $(\kappa_n)$  and  $(t_n)$ , whereas all angular functions  $f_b$ , with  $b \in (\frac{1}{2}, 1)$ , belong to  $\mathcal{F}_{\text{LAN}}(p, \kappa_n, t_n)$  in particular when  $(\kappa_n)$  diverges to infinity at least as fast as  $(\log n)^2$ , hence, for example, when  $\kappa_n = n^a$ , with  $a > 0$ . We then have the following LAN result.

**THEOREM 5.1.** *Fix an integer  $p \geq 2$  and  $\theta \in \mathcal{S}^{p-1}$ . Let  $(\kappa_n)$  be a positive real sequence that diverges to infinity. Let  $(\tau_n)$  be a bounded sequence in  $\mathbb{R}^p$  such that  $\theta_n = \theta + v_n \tau_n \in \mathcal{S}^{p-1}$  for all  $n$ , with  $v_n := 1/\sqrt{n\kappa_n\varphi_f(\kappa_n)}$ . Assume that  $f$  belongs to  $\mathcal{F}_{\text{LAN}}(p, \kappa_n, \|\tau_n\|)$ . Then, as  $n \rightarrow \infty$  under  $\mathbf{P}_{\theta, \kappa_n, f}^{(n)}$ ,*

$$\Lambda_{\theta + v_n \tau_n / \theta, \kappa_n, f} := \log \frac{d\mathbf{P}_{\theta + v_n \tau_n, \kappa_n, f}^{(n)}}{d\mathbf{P}_{\theta, \kappa_n, f}^{(n)}} = \tau_n' \Delta_{\theta, f}^{(n)} - \frac{1}{2} \tau_n' \Gamma_{\theta} \tau_n + o_{\mathbf{P}}(1),$$

where the central sequence  $\Delta_{\theta, f}^{(n)} := v_n^{-1}(\mathbf{I}_p - \theta\theta')\bar{\mathbf{X}}_n$ , still under  $\mathbf{P}_{\theta, \kappa_n, f}^{(n)}$ , is asymptotically normal with mean zero and covariance matrix  $\Gamma_{\theta} := \mathbf{I}_p - \theta\theta'$ .

This result shows that the rate  $v_n$  identified in the previous sections is actually the contiguity rate associated with the sequence of statistical experiments at hand. Remarkably, this provides one of the few semiparametric examples (if any) where the contiguity rate depends on the fixed value of the functional nuisance  $f$ . Since the contiguity rate coincides with the rate of convergence of the spherical mean (see Theorem 3.1), we conclude that the spherical mean is rate-consistent. Better: since the proof of Theorem 3.1 establishes that

$$\sqrt{n\kappa_n\varphi_f(\kappa_n)}(\hat{\theta}_n - \theta) = \Delta_{\theta, f}^{(n)} + o_{\mathbf{P}}(1)$$

as  $n \rightarrow \infty$  under  $\mathbf{P}_{\theta, \kappa_n, f}^{(n)}$ , it actually follows from Theorems 3.1 and 5.1 that the spherical mean is an asymptotically optimal estimator in the sense of the convolution theorem; see, for

example, Theorem 8.8 in *van der Vaart (1998)*. Turning to hypothesis testing, it also follows from the LAN result above that the Watson and Wald tests from the previous section are rate-consistent, since Theorem 4.1(ii) indicates that these tests show nontrivial asymptotic powers under the sequence of contiguous alternatives involved in Theorem 5.1. Actually, in the present LAN framework, an application of the Le Cam third lemma confirms these asymptotic local powers.

To show this, fix a positive real sequence  $(\kappa_n)$  that diverges to infinity and local alternatives as in Theorem 5.1. Then, under the sequence of null hypotheses  $P_{\theta_0, \kappa_n, f}^{(n)}$ ,

$$T_n^W := \frac{\sqrt{n}(p-1)^{1/4}(\bar{X}_n - e_{n1}\theta_0)}{2^{1/4}\tilde{e}_{n2}^{1/4}}$$

is asymptotically normal with mean zero and covariance matrix  $\Gamma_{\theta_0}$ ; this follows from (S.4.4) in the proof of Theorem 3.1. Now, by using Theorem 2.2(ii), we obtain that, under the same sequence of hypotheses,

$$\begin{aligned} & \text{Cov}[T_n^W, \Lambda_{\theta_0 + v_n \tau_n / \theta_0, \kappa_n, f}] \\ &= \text{Cov}[T_n^W, \Delta_{\theta_0, f}^{(n)}] \tau_n + o(1) \\ &= \frac{n\sqrt{\kappa_n \varphi_f(\kappa_n)}(p-1)^{1/4}}{2^{1/4}\tilde{e}_{n2}^{1/4}} E[(\bar{X}_n - e_{n1}\theta_0)(\bar{X}_n - e_{n1}\theta_0)'] \Gamma_{\theta_0} \tau_n + o(1) \\ &= E[T_n^W (T_n^W)'] \Gamma_{\theta_0} \tau_n + o(1) = \tau_n + o(1). \end{aligned}$$

Thus, Le Cam’s third lemma entails that, under the sequence of contiguous alternatives  $P_{\theta_n, \kappa_n, f}^{(n)}$ , with  $\theta_n = \theta_0 + v_n \tau_n$ ,  $v_n = 1/\sqrt{n\kappa_n \varphi_f(\kappa_n)}$  and  $(\tau_n) \rightarrow \tau$ ,  $T_n^W$  is asymptotically normal with mean  $\tau$  and covariance matrix  $\Gamma_{\theta_0}$ , so that, under this sequence of hypotheses,  $\tilde{W}_n = (T_n^W)' \Gamma_{\theta_0}^- T_n^W \xrightarrow{D} \chi_{p-1}^2(\|\tau\|^2)$ , where  $A^-$  stands for the Moore–Penrose inverse of  $A$ . From contiguity, we thus obtain that  $W_n = \tilde{W}_n + o_P(1) \xrightarrow{D} \chi_{p-1}^2(\|\tau\|^2)$  under the alternatives considered, which, as announced, is in agreement with Theorem 4.1(ii). As for the Wald test, the fact that  $S_n \xrightarrow{D} \chi_{p-1}^2(\|\tau\|^2)$  under the same sequence of alternatives directly follows from the result for the Watson test and from the fact that the null asymptotic equivalence  $W_n = S_n + o_P(1)$  in Theorem 4.1(i) extends, from contiguity, to the present contiguous alternatives.

Beyond this, one of the main interests of the LAN result in Theorem 5.1 is to pave the way to the construction of Le Cam optimal tests for the problem of testing, under high concentration,  $\mathcal{H}_0 : \theta = \theta_0$  versus  $\mathcal{H}_1 : \theta \neq \theta_0$  with angular function  $f$ . It directly follows from this result that, for this problem, the test rejecting the null hypothesis at asymptotic level  $\alpha$  whenever

$$Q_n := (\Delta_{\theta_0, f}^{(n)})' \Gamma_{\theta_0}^- \Delta_{\theta_0, f}^{(n)} > \chi_{p-1, 1-\alpha}^2$$

is Le Cam optimal (more precisely, locally asymptotically maximin) at asymptotic level  $\alpha$ . Since Theorem 2.2(ii) ensures that, under the null hypothesis,

$$\begin{aligned} \Delta_{\theta_0, f}^{(n)} &= \sqrt{n\kappa_n \varphi_f(\kappa_n)} (\mathbf{I}_p - \theta_0 \theta_0') (\bar{X}_n - e_{n1}\theta_0) \\ &= \frac{\sqrt{n}(p-1)^{1/4}}{2^{1/4}\tilde{e}_{n2}^{1/4}} (\mathbf{I}_p - \theta_0 \theta_0') (\bar{X}_n - e_{n1}\theta_0) + o_P(1), \end{aligned}$$

Lemma 4.1 readily entails that  $Q_n = \tilde{W}_n + o_P(1) = W_n + o_P(1)$  under the null hypothesis, hence, from contiguity, also under the sequences of local alternatives above. It follows that, under the assumptions of Theorem 5.1, the Watson test is optimal in the Le Cam sense. Since the Watson test does not depend on  $f$ , this optimality holds at any  $f$  meeting the assumptions of Theorem 5.1. From the asymptotic equivalence result in Theorem 4.1(i) and from contiguity, this extends to the Wald test.

In the high concentration framework considered, it may be intuitively appealing to linearize the problem and apply a standard Euclidean test to the data projected onto the tangent space to  $S^{p-1}$  at the null location  $\theta_0$ —or equivalently, to the data  $\mathbf{Y}_{ni} := \mathbf{P}'_{\theta_0} \mathbf{X}_{ni}$ ,  $i = 1, \dots, n$ , where  $\mathbf{P}_{\theta_0}$  is an arbitrary  $p \times (p - 1)$  matrix whose columns form an orthonormal basis of the orthogonal complement of  $\theta_0$  in  $\mathbb{R}^p$ . The null hypothesis  $\mathcal{H}_0 : \theta = \theta_0$  translates into testing that the mean of the common (under rotational symmetry about  $\theta_0$ , spherically symmetric) distribution of the  $\mathbf{Y}_{ni}$ 's is the zero vector. The Watson test can actually be seen as the (spherical) Hotelling test rejecting the null hypothesis at asymptotic level  $\alpha$  whenever  $n\tilde{\mathbf{Y}}'_n \mathbf{S}_n^{-1} \tilde{\mathbf{Y}}_n > \chi^2_{p-1, 1-\alpha}$ , with  $\tilde{\mathbf{Y}}_n := n^{-1} \sum_{i=1}^n \mathbf{Y}_{ni}$  and with a standardization matrix  $\mathbf{S}_n$  that, in line with the underlying spherical symmetry, is a multiple of the identity matrix. Quite nicely, Theorem 5.1 formally proves that this linearization provides a test that is Le Cam optimal at any  $f$ . We insist, however, that it was unclear that such a linearization would provide a test that achieves optimality in the original sequence of curved statistical experiments. Not only because the impact of linearization is difficult to control, but also because it is unknown whether or not the spherical Hotelling test is optimal in any sense under the, highly concentrated and skewed, alternatives obtained in the tangent space (to the best of our knowledge, the only optimality results for the spherical Hotelling test relate to shifted spherical Gaussian distributions; see, e.g., Hallin and Paindaveine (2002)).

**6. Real data illustration.** The real dataset we analyze here consists in measurements of magnetic remanence directions in  $n = 62$  rock specimens. The objective of Remanent magnetism or equivalently Paleomagnetism is to study the strength and the direction of the Earth's magnetic field over time. The orientation and intensity of the Earth's magnetic field can be obtained through the record of remanent magnetism preserved in rocks. The directions of remanent magnetization allow scientists to determine the position of the Earth's magnetic pole with respect to the study location at the time when the magnetization was acquired.

We consider here a well-known dataset on  $S^2$  that has already been used for inference on spherical location in Fisher, Lewis and Embleton (1987). The dataset, which is provided as Dataset A in Appendix B8 of this monograph, is showed in the left panel of Figure 4. Clearly, the data is highly concentrated. In line with this, the FvML maximum likelihood estimator of

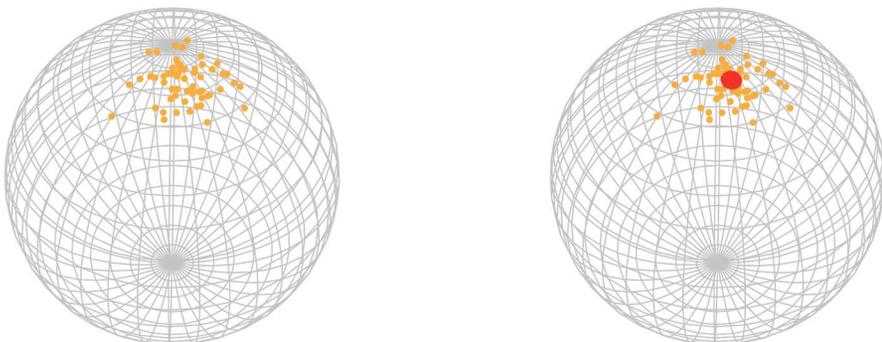


FIG. 4. (Left:) plot of the  $n = 62$  directions of magnetic remanence associated with the real dataset considered in Section 6. (Right:) plot of the same dataset with, in red, the corresponding 95% confidence cap in (3.4).

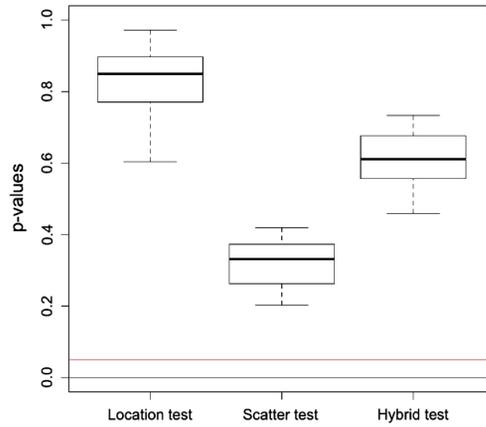


FIG. 5. Boxplots of the  $p$ -values of three tests of rotational symmetry over  $S^2$  obtained from the 62 leave-one-out samples associated with the magnetic remanence data set; see Section 6 for details. The tests considered are (the unspecified- $\theta$  versions of) the location test, scatter test and hybrid test of rotational symmetry from García-Portugués, Paindaveine and Verdebout (2019).

the concentration parameter  $\kappa$  takes value  $\hat{\kappa} = 76.12$ , which is of the same order of magnitude as the sample size  $n = 62$ . Figure 4 also suggests that rotational symmetry is a plausible assumption. To assess this, we performed the three tests of rotational symmetry on  $S^2$  that were recently proposed in García-Portugués, Paindaveine and Verdebout (2019): a location test and a scatter test, that respectively show power against location-type alternatives and scatter-type alternatives to rotational symmetry (we refer to García-Portugués, Paindaveine and Verdebout (2019) for details), as well as a hybrid test that shows power against both types of alternatives. These three tests, that are meant to test the null hypothesis of rotational symmetry about an unspecified location  $\theta$ , provided the  $p$ -values 0.844, 0.305 and 0.607, respectively, hence did not lead to rejection at any usual nominal level. To somewhat assess the robustness of this result, we performed the following analysis: on the 62 samples of size 61 obtained by leaving one of the original observations out, we performed the same three tests of rotational symmetry and provided in Figure 5 the boxplots of the 62  $p$ -values obtained for each of the three tests. Again, at any usual nominal level, none of these subsamples led any of the three tests to reject the null hypothesis of rotational symmetry.

The various statistical methods studied in this paper are therefore perfectly suitable for the present dataset. To illustrate one of these methods, we computed the 95% confidence cap for the spherical location defined in (3.4). The resulting confidence cap is showed in the right panel of Figure 4. This confidence zone is centered at the spherical mean  $\hat{\theta} = (0.210, 0.104, 0.972)'$  and, as expected in the present high concentration setup, has a very small size.

**7. Wrap up.** We discussed inference on the location parameter of rotationally symmetric distributions under high concentration. We did so by considering double asymptotic scenarios where the underlying concentration parameter  $\kappa_n$  diverges to infinity at an arbitrary rate with the sample size  $n$ . This significantly improves over the state of the art for directional inference under high concentration, since previous works not only focused on a parametric class of distributions (namely, the FvML one) but also restricted to asymptotics as  $\kappa$  diverges to infinity with  $n$  fixed. Our asymptotic results indicate that standard fixed- $\kappa$  methods are robust to high concentration, in the sense that they will remain valid in the aforementioned double asymptotic scenarios: the spherical mean remains consistent and asymptotically normal, whereas the Watson and Wald tests still asymptotically meet the nominal level constraint.

Under high concentration, however, these statistical procedures enjoy faster consistency rates than in the standard fixed- $\kappa$  asymptotic scenario. Remarkably, these consistency rates depend on the type of rotationally symmetric distributions considered, that is, they depend on the underlying angular function  $f$ ; this dependence is such that the higher the concentration, the faster the consistency rates. In contrast with all previous works on high concentration, we also considered optimality issues. We showed that, under mild assumptions on  $f$ , the aforementioned inference procedures enjoy strong, Le Cam-type, optimality properties. For some (not all) angular functions, optimality requires that  $\kappa_n$  diverges to infinity sufficiently fast as a function of  $n$ ; the corresponding restriction, as we have seen, is extremely mild for our running example associated with  $f_b(z) = \exp(z^b)$ , as optimality, for  $b \in (\frac{1}{2}, 1)$  holds in particular when  $\kappa_n$  diverges to infinity at least as fast as  $(\log n)^2$ , whereas no restriction of this sort is required for  $b \geq 1$ , hence in particular for the usual FvML case.

**Acknowledgments.** We would like to thank the Editor, Richard J. Samworth, the Associate Editor and three anonymous referees for their insightful comments and suggestions, that led to a substantial improvement of a previous version of this work.

Davy Paindaveine's research is supported by a research fellowship from the Francqui Foundation and by the Program of Concerted Research Actions (ARC) of the Université libre de Bruxelles. His affiliation at the Université libre de Bruxelles is at ECARES and the Department of Mathematics, whereas his affiliation at the Université Toulouse 1 Capitole is at the Toulouse School of Economics.

Thomas Verdebout's research is supported by the ARC Program of the Université libre de Bruxelles and by the Crédit de Recherche J.0134.18 of the FNRS (Fonds National pour la Recherche Scientifique), Communauté Française de Belgique. His affiliation at the Université libre de Bruxelles is at ECARES and the Department of Mathematics.

## SUPPLEMENTARY MATERIAL

**Supplement to “Inference for spherical location under high concentration”** (DOI: 10.1214/19-AOS1918SUPP; .pdf). In this supplement, we prove all the results of the present paper.

## REFERENCES

- ARNOLD, R. and JUPP, P. E. (2013). Statistics of orthogonal axial frames. *Biometrika* **100** 571–586. MR3094438 <https://doi.org/10.1093/biomet/ast017>
- ARNOLD, R., JUPP, P. E. and SCHAEUBEN, H. (2018). Statistics of ambiguous rotations. *J. Multivariate Anal.* **165** 73–85. MR3768753 <https://doi.org/10.1016/j.jmva.2017.10.007>
- CHIKUSE, Y. (2003a). Concentrated matrix Langevin distributions. *J. Multivariate Anal.* **85** 375–394. MR1983803 [https://doi.org/10.1016/S0047-259X\(02\)00065-9](https://doi.org/10.1016/S0047-259X(02)00065-9)
- CHIKUSE, Y. (2003b). *Statistics on Special Manifolds. Lecture Notes in Statistics* **174**. Springer, New York. MR1960435 <https://doi.org/10.1007/978-0-387-21540-2>
- DAI, X. and MÜLLER, H.-G. (2018). Principal component analysis for functional data on Riemannian manifolds and spheres. *Ann. Statist.* **46** 3334–3361. MR3852654 <https://doi.org/10.1214/17-AOS1660>
- DOWNES, T. D. (2003). Spherical regression. *Biometrika* **90** 655–668. MR2006842 <https://doi.org/10.1093/biomet/90.3.655>
- DOWNES, T. D. and MARDIA, K. V. (2002). Circular regression. *Biometrika* **89** 683–697. MR1929172 <https://doi.org/10.1093/biomet/89.3.683>
- FISHER, N. I., LEWIS, T. and EMBLETON, B. J. J. (1987). *Statistical Analysis of Spherical Data*. Cambridge Univ. Press, Cambridge. MR0899958 <https://doi.org/10.1017/CBO9780511623059>
- FUJIKOSHI, Y. and WATAMORI, Y. (1992). Tests for the mean direction of the Langevin distribution with large concentration parameter. *J. Multivariate Anal.* **42** 210–225. MR1183843 [https://doi.org/10.1016/0047-259X\(92\)90044-G](https://doi.org/10.1016/0047-259X(92)90044-G)
- GARCÍA-PORTUGUÉS, E., PAINDAVEINE, D. and VERDEBOUT, T. (2019). On optimal tests for rotational symmetry against new classes of hyperspherical distributions. *J. Amer. Statist. Assoc.*, to appear.

- HALLIN, M. and PAINDAVEINE, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.* **30** 1103–1133. MR1926170 <https://doi.org/10.1214/aos/1031689019>
- HAYAKAWA, T. (1990). On tests for the mean direction of the Langevin distribution. *Ann. Inst. Statist. Math.* **42** 359–373. MR1064794 <https://doi.org/10.1007/BF00050842>
- HAYAKAWA, T. and PURI, M. L. (1985). Asymptotic expansions of the distributions of some test statistics. *Ann. Inst. Statist. Math.* **37** 95–108. MR0790378 <https://doi.org/10.1007/BF02481083>
- JUPP, P. E. (2015). Copulae on products of compact Riemannian manifolds. *J. Multivariate Anal.* **140** 92–98. MR3372555 <https://doi.org/10.1016/j.jmva.2015.04.008>
- LEY, C. and VERDEBOUT, T. (2017). *Modern Directional Statistics. Chapman & Hall/CRC Interdisciplinary Statistics Series*. CRC Press, Boca Raton, FL. MR3752655
- MARDIA, K. V. and JUPP, P. E. (2000). *Directional Statistics. Wiley Series in Probability and Statistics*. Wiley, Chichester. Revised reprint of *Statistics of directional data* by Mardia [MR0336854 (49 #1627)]. MR1828667
- PAINDAVEINE, D. and VERDEBOUT, T. (2017). Inference on the mode of weak directional signals: A Le Cam perspective on hypothesis testing near singularities. *Ann. Statist.* **45** 800–832. MR3650401 <https://doi.org/10.1214/16-AOS1468>
- PAINDAVEINE, D. and VERDEBOUT, T. (2020). Supplement to “Inference for spherical location under high concentration.” <https://doi.org/10.1214/19-AOS1918SUPP>.
- RIVEST, L.-P. (1986). Modified Kent’s statistics for testing goodness of fit for the Fisher distribution in small concentrated samples. *Statist. Probab. Lett.* **4** 1–4. MR0822715 [https://doi.org/10.1016/0167-7152\(86\)90028-3](https://doi.org/10.1016/0167-7152(86)90028-3)
- RIVEST, L.-P. (1989). Spherical regression for concentrated Fisher–von Mises distributions. *Ann. Statist.* **17** 307–317. MR0981452 <https://doi.org/10.1214/aos/1176347018>
- RIVEST, L.-P., DUCHESNE, T., NICOSIA, A. and FORTIN, D. (2016). A general angular regression model for the analysis of data on animal movement in ecology. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **65** 445–463. MR3470586 <https://doi.org/10.1111/rssc.12124>
- ROSENTHAL, M., WU, W., KLASSEN, E. and SRIVASTAVA, A. (2014). Spherical regression models using projective linear transformations. *J. Amer. Statist. Assoc.* **109** 1615–1624. MR3293614 <https://doi.org/10.1080/01621459.2014.892881>
- SENGUPTA, A., KIM, S. and ARNOLD, B. C. (2013). Inverse circular-circular regression. *J. Multivariate Anal.* **119** 200–208. MR3061424 <https://doi.org/10.1016/j.jmva.2013.04.011>
- SHORACK, G. R. (2000). *Probability for Statisticians. Springer Texts in Statistics*. Springer, New York. MR1762415
- VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247 <https://doi.org/10.1017/CBO9780511802256>
- WATAMORI, Y. (1996). Statistical inference of Langevin distribution for directional data. *Hiroshima Math. J.* **26** 25–74. MR1380423
- WATSON, G. S. (1983). *Statistics on Spheres. University of Arkansas Lecture Notes in the Mathematical Sciences* **6**. Wiley, New York. A Wiley-Interscience Publication. MR0709262
- WATSON, G. S. (1984). The theory of concentrated Langevin distributions. *J. Multivariate Anal.* **14** 74–82. MR0734099 [https://doi.org/10.1016/0047-259X\(84\)90047-2](https://doi.org/10.1016/0047-259X(84)90047-2)