Ann Inst Stat Math (2015) 67:177–193
DOI 10.1007/s10463-013-0444-7

Intrinsic means on the circle: uniqueness, locus and asymptotics

T. Hotz · S. Huckemann

Received: 24 August 2012 / Revised: 19 November 2013 / Published online: 17 January 2014

Abstract This paper gives a comprehensive treatment of local uniqueness, asymptotics and numerics for intrinsic sample means on the circle. It turns out that local uniqueness as well as rates of convergence are governed by the distribution near the antipode. If the distribution is locally less than uniform there, we have local uniqueness and asymptotic normality with a square-root rate. With increased proximity to the uniform distribution the rate can be arbitrarily slow, and in the limit, local uniqueness is lost. Further, we give general distributional conditions, e.g., unimodality, that ensure global uniqueness. Along the way, we discover that sample means can occur only at the vertices of a regular polygon which allows to compute intrinsic sample means in linear time from sorted data. This algorithm is finally applied in a simulation study demonstrating the dependence of the convergence rates on the behavior of the density at the antipode.

Keywords Circular statistics · Directional statistics · Intrinsic mean · Central limit theorem · Asymptotic normality · Convergence rate

1 Introduction

The need for statistical analysis of directional data arises in many applications, be it in the study of wind direction, animal migration or geological crack development. An

T. Hotz (✉)
Institut für Mathematik, Technische Universität Ilmenau, PF 100565, 98684 Ilmenau, Germany
e-mail: thomas.hotz@tu-ilmenau.de

S. Huckemann
Felix Bernstein Institute for Mathematical Statistics in the Biosciences,
University of Göttingen, Goldschmidtstrasse 7, 37077 Göttingen, Germany
e-mail: huckeman@math.uni-goettingen.de

overview of and introduction into this field can be found in Fisher (1993) or Mardia and Jupp (2000). Until today, nonparametric inferential techniques employing the *intrinsic mean*, i.e., the minimizer of expected squared *angular* distances, rest on the assumption that distributions underlying circular data have no mass in an entire interval opposite to an intrinsic mean. This assumption, however, is not met by many of the standard distributions for directional data, e.g., Fisher-von Mises, Bingham, wrapped normal or wrapped Cauchy distributions.

The reason for this mathematical assumption lies in the fact that the squared intrinsic distance is not differentiable at two antipodal points. However, the central limit theorem for intrinsic means on general manifolds derived by Bhattacharya and Patrangenaru (2005), cf. also Huckemann (2011), utilizes a Taylor expansion of the *intrinsic variance* by differentiating under the integral sign, i.e., by differentiating the squared intrinsic distance. In consequence, this has left the derivation of a central limit theorem along with convergence rates for circular data of most realistic scenarios and distributions an open problem until now.

Here, we fill this gap and provide for a comprehensive solution. In particular, we show that, if the distribution features a continuous density near the antipode of the intrinsic mean which stays below the density of the uniform distribution, asymptotic normality with the $n^{-1/2}$ rate well known from Euclidean statistics remains valid, the asymptotic variance, however, increases with proximity to the uniform distribution. Using different methods, this has also been observed by McKilliam et al. (2012) if the distribution features a continuous density at the antipode whose value is less than $\frac{1}{2\pi}$, though we give a different proof. When it increases further, namely if the distribution at the antipode differs from the uniform distribution only in higher order, then the asymptotic rate is accordingly lowered by this power. If and only if the distribution at the antipode is locally uniform the intrinsic mean is no longer locally unique. Moreover, at an antipode of an intrinsic mean, there may never be more mass than that of the uniform distribution. In particular there can never be a point mass. For these reasons, sample means are always locally unique and we will see that they can only occur on the vertices of a regular polygon. Hence from ordered data, sample means can be computed in linear time. These insights also allow to derive general distributional assumptions such as unimodality under which there is only one local minimizer and hence a unique intrinsic mean.

This last result extends the result of Le (1998) who guarantees uniqueness for distributions symmetric with respect to a point, being non-increasing functions of the distance to this point, strictly decreasing on a set of positive circular measure, cf. also Kaziska and Srivastava (2008), as well as the very general result of Afsari (2011) which in our case of the circle, yields uniqueness if the distribution is restricted to an open half circle.

We note that there are other estimates for a mean direction than the intrinsic mean discussed here, most notably the extrinsic mean discussed e.g., by Bhattacharya and Patrangenaru (2003, 2005). For directional data, the latter is often simply called *circular mean*. However, we focus on the mathematically more difficult case of the intrinsic mean here; the properties of the extrinsic mean, e.g., its asymptotic normality, being well established.

2 Setup

Throughout this paper, we will work with angles in the interval $[-\pi, \pi)$. On this interval, we use the distance $d(\theta, \zeta) := \left|\theta - \zeta - 2\pi\nu(\theta, \zeta)\right|$ for $\theta, \zeta \in [-\pi, \pi)$ with

$$\nu(\theta, \zeta) := \begin{cases} 1 & \text{if } \theta > 0, \zeta \in [-\pi, \theta - \pi), \\ -1 & \text{if } \theta < 0, \zeta \in (\theta + \pi, \pi), \\ 0 & \text{else.} \end{cases}$$

If one identifies every angle $\theta \in [-\pi, \pi)$ with the point $z(\theta) = (\cos\theta, \sin\theta) \in S^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ on the unit circle, then $d(\theta, \zeta)$ gives the distance between $z(\theta)$ and $z(\zeta)$ in arclength, i.e., their geodesic distance on the unit circle. We therefore call d the *intrinsic distance*. Note that d is a metric on $[-\pi, \pi)$, and for the induced topology $z : [-\pi, \pi) \to S^1$ is a homeomorphism. The advantage of working on $[-\pi, \pi)$ instead of on S^1 is that it is a subset of $\mathbb{R}$, almost a chart, so that all calculations merely concern real numbers, and all statements are to be interpreted this way. To emphasize that we do not use the standard topology on $[-\pi, \pi)$ but the one induced by the metric d, we will at times denote this set by $\mathbb{T}$. This should remind the reader of the fact that $[-\pi, \pi)$ here bears the topology of the one-dimensional *torus* $\mathbb{R}/2\pi\mathbb{Z}$, for which $[-\pi, \pi)$ is a set of representatives. An *interval* on $\mathbb{T}$ is then any interval on $\mathbb{R}$ modulo $2\pi\mathbb{Z}$.

In this paper, with the antipodal map $\phi : [-\pi, \pi) \to [-\pi, \pi)$

$$\phi(\theta) = \begin{cases} \theta - \pi & \text{if } 0 \le \theta < \pi, \\ \theta + \pi & \text{if } -\pi \le \theta \le 0, \end{cases}$$

the *antipodal set* $\widetilde{S} := \{\phi(\theta) : \theta \in S\}$ of $S \subset [-\pi, \pi) = \mathbb{T}$ will play a central role.

On $\mathbb{T} = [-\pi, \pi)$ we consider independent and identically distributed random elements $X_1, \ldots, X_n \sim X$ mapping from a given probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to $[-\pi, \pi)$, denoting the corresponding distribution of X by $\mathbb{P}_X$. Writing $\mathbb{E}$ for the usual Euclidean expectation, we are then looking for minimizers $\theta \in [-\pi, \pi)$ of the so-called Fréchet function, see Fréchet (1948),

$$V(\theta) := \mathbb{E}\left(d(\theta, X)^2\right), \text{ and } V_n(\theta) := \frac{1}{n}\sum_{j=1}^{n} d(\theta, X_j)^2. \tag{1}$$

Remark 1 Note that V and V_n, being continuous functions on the compact $\mathbb{T}$, always feature at least one global minimizer.

Definition 1 Every such global minimizer θ^* of V and θ_n of V_n is called an *intrinsic population mean* and an *intrinsic sample mean*, respectively. The values $V(\theta^*)$ and $V_n(\theta_n)$ are called *intrinsic population* and *sample variance*, respectively.

Obviously, global minimizers need not be unique, and there may be local but non-global minimizers, as the examples below will show.

Denoting the usual average by $\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$ we have that $-\pi \le \bar{X} < \pi$. Since $V(\theta) \le \mathbb{E}|\theta - X|^2$ with equality for $\theta = 0$, the latter functional being minimized by $\theta = \mathbb{E}X$, we see that $\theta^* = 0$ locally minimizing V implies $\mathbb{E}X = 0$; cf. also Kobayashi and Nomizu (1969, Section VIII.9) as well as Karcher (1977). The classical *Euclidean Central Limit Theorem* gives

$$\sqrt{n}\,\bar{X} \xrightarrow{D} \mathcal{N}(0, \sigma^2) \tag{2}$$

with the Euclidean variance $\mathrm{Var}(X) = \sigma^2$. If $\theta^* = 0$ is also a global minimizer of V, i.e., an intrinsic mean, then the Euclidean variance agrees with the intrinsic variance

$$V(0) = \sigma^2.$$

If we were interested whether any other $\theta \in \mathbb{T}$ were an intrinsic mean, we could apply a simple rotation of the circle to reduce this to the case $\theta = 0$ again. Without loss of generality we can hence restrict our attention to this special θ; recall that its antipode is $\phi(0) = -\pi$.

3 Local and global minimizers

3.1 The distribution near the antipode

Here is our first fundamental Theorem.

Theorem 1 *If $\theta^* = 0$ locally minimizes V, then:*

(i) $\mathbb{P}_X(-\pi) = 0$*, i.e., there is no point mass opposite an intrinsic mean.*

(ii) If additionally there is some $\delta > 0$ s.t. $\mathbb{P}_X$ restricted to $(-\pi, -\pi + \delta)$ features a continuous density f with respect to Lebesgue measure, then $f(-\pi+) = \lim_{\theta \downarrow -\pi} f(\theta) \le \frac{1}{2\pi}$; similarly $f(\pi-) = \lim_{\theta \uparrow \pi} f(\theta) \le \frac{1}{2\pi}$ for a continuous density f on $(\pi - \delta, \pi)$. If both $f(-\pi+) < \frac{1}{2\pi}$ and $f(\pi-) < \frac{1}{2\pi}$, then θ^* is locally unique.

(iii) In case $f(-\pi) = \frac{1}{2\pi}$, f being continuous in a neighborhood of $-\pi$, assume that there is some $\delta > 0$ s.t. f is k-times continuously differentiable on $[-\pi, -\pi + \delta)$ (i.e., with existing left limits) and $\tilde{k}$-times continuously differentiable on $(\pi - \delta, \pi]$ (i.e., with existing right limits), where k and $\tilde{k}$ are chosen minimal s.t. $f^{(k)}(-\pi+) = \lim_{\theta \downarrow -\pi} f^{(k)}(\theta) \ne 0$ as well as $f^{(\tilde{k})}(\pi-) = \lim_{\theta \uparrow \pi} f^{(\tilde{k})}(\theta) \ne 0$. Then, $f^{(k)}(-\pi+) < 0$ as well as $(-1)^{\tilde{k}} f^{(\tilde{k})}(\pi-) < 0$, θ^* is locally unique, and for $\delta > 0$ small enough

$$\mathbb{P}\{-\pi \le X \le -\pi + \delta\} = \frac{\delta}{2\pi} + \frac{\delta^{k+1}}{(k+1)!} f^{(k)}(-\pi+) + o(\delta^{k+1}),$$

as well as

$$\mathbb{P}\{\pi - \delta \le X < \pi\} = \frac{\delta}{2\pi} + \frac{\delta^{\tilde{k}+1}}{(\tilde{k}+1)!} (-1)^{\tilde{k}} f^{(\tilde{k})}(\pi-) + o(\delta^{\tilde{k}+1}).$$

Proof For any $\theta > 0$, we have

$$\begin{aligned}
& V(\theta) - V(0) \\
&= \int_{-\pi+\theta}^{\pi} (\theta - x)^2 \, \mathrm{d}\mathbb{P}_X(x) + \int_{-\pi}^{-\pi+\theta} (\theta - x - 2\pi)^2 \, \mathrm{d}\mathbb{P}_X(x) - \int_{-\pi}^{\pi} x^2 \, \mathrm{d}\mathbb{P}_X(x) \\
&= \int_{-\pi}^{\pi} \underbrace{\big((\theta - x)^2 - x^2\big)}_{=\theta^2 - 2\theta x} \mathrm{d}\mathbb{P}_X(x) + \int_{-\pi}^{-\pi+\theta} \big(-4\pi(\theta - x) + 4\pi^2\big) \, \mathrm{d}\mathbb{P}_X(x) \\
&= \theta^2 - 2\theta \underbrace{\mathbb{E}X}_{=0} - 4\pi \int_{-\pi}^{-\pi+\theta} \underbrace{(-\pi + \theta - x)}_{\geq 0} \, \mathrm{d}\mathbb{P}_X(x) \\
&\leq \theta^2 - 4\pi \, \theta \, \mathbb{P}_X(-\pi)
\end{aligned} \tag{3}$$

which for θ small enough becomes negative if $\mathbb{P}_X(-\pi) > 0$, whence (i) follows.

Now denote the (shifted) cumulative distribution function of $Y = X + \pi$ by

$$F(y) = \mathbb{P}\{0 \leq Y \leq y\} = \mathbb{P}\{-\pi \leq X \leq -\pi + y\} \tag{4}$$

for $y \geq 0$ to obtain

$$\begin{aligned}
\int_{-\pi}^{-\pi+\theta} (-\pi + \theta - x) \, \mathrm{d}\mathbb{P}_X(x) &= \int_0^{\theta} (\theta - y) F'(y) \mathrm{d}y \\
&= 0 F(\theta) - \theta \underbrace{F(0)}_{=0} + \int_0^{\theta} F(y) \mathrm{d}y,
\end{aligned}$$

F' being understood in a distributional sense. Noting that $\theta^2 = 4\pi \int_0^{\theta} \frac{y}{2\pi} \mathrm{d}y$ where $\frac{y}{2\pi}$ is the (shifted) cumulative distribution function of the uniform distribution on $[-\pi, \pi)$, in conjunction with (3), we see that

$$V(\theta) \geq V(0) \quad \text{iff} \int_0^{\theta} \left(\frac{y}{2\pi} - F(y)\right) \mathrm{d}y \geq 0, \tag{5}$$

where equality holds simultaneously, too.

Thus, under the assumptions of (ii), we may use a 2nd order Taylor expansion to obtain for $\theta \geq 0$ small enough

$$0 \leq \int_0^{\theta} \left(\frac{y}{2\pi} - F(y)\right) \mathrm{d}y = \frac{\theta^2}{4\pi} - \frac{\theta^2}{2} F'(0) + o(\theta^2),$$

hence $f(-\pi+) = F'(0+) \leq \frac{1}{2\pi}$. If this inequality is strict, $V(\theta) > V(0)$ follows for θ small enough. With the analogous argument for $\theta < 0$ this yields the assertion in (ii).

Finally, (iii) is obtained by a Taylor expansion as well, namely (for $\theta > 0$)

$$\frac{\theta}{2\pi} - F(\theta) = -\frac{\theta^{k+1}}{(k+1)!} f^{(k)}(-\pi+) + o(\theta^{k+1}),$$

which after integration, noting that uniqueness implies a sharp inequality, gives

$$0 \le \int_0^\theta \left(\frac{y}{2\pi} - F(y)\right) \mathrm{d}y = -\frac{\theta^{k+2}}{(k+2)!} f^{(k)}(-\pi+) + o(\theta^{k+2}),$$

whence $f^{(k)}(-\pi+) < 0$ follows as well as local uniqueness. The case $\theta < 0$ is again treated analogously. □

3.2 Consequences for uniqueness, loci of local minimizers and algorithms

From Theorem 1 we obtain at once a necessary and sufficient condition such that a minimizer of V is locally unique.

Corollary 1 *By (5), V is constant on some interval $S \subset \mathbb{T}$ iff the probability distribution restricted to the antipodal interval $\widetilde{S}$ has constant density $\frac{1}{2\pi}$ there. In particular, suppose that $\theta^* \in \mathbb{T}$ is a local minimizer of V. Then θ^* is locally unique iff there is no interval $\theta^* \in S \subset \mathbb{T}$ such that the distribution restricted to $\widetilde{S}$ is uniform.*

The following is a generalization of a result by Rabi Bhattacharya (personal communication from 2008):

Corollary 2 *If the distribution of X has a density with respect to the Lebesgue measure which is composed of finitely many pieces, each being analytic up to the interval boundaries, then any local minimizer of V is locally unique unless the density is constant $\frac{1}{2\pi}$ on some interval.*

Proof This follows immediately since an analytic function is constant unless one of its derivatives is non-zero. □

Proposition 1 *Consider the distribution of X, decomposed into the part λ which is absolutely continuous with respect to Lebesgue measure, with density f, and the part η singular to Lebesgue measure. Let $S_1, \dots, S_k$ be the distinct open arcs on which $f < \frac{1}{2\pi}$, assume they are all disjoint from $\operatorname{supp} \eta$, and that $\{x \in \mathbb{T} : f(x) = \frac{1}{2\pi}\}$ is a Lebesgue null-set. Then X has at most k intrinsic means and every $\widetilde{S_j}$ contains at most one candidate.*

Proof Suppose that $\theta^* = 0$ is a local minimizer of V. By hypothesis and virtue of Corollary 1 there is an open arc $S_1 \ni -\pi$ followed (going into positive x-direction) by a closed arc T_1 such that $f(x) < 1/(2\pi)$ for $x \in S_1 \backslash \{-\pi\}$ and $f(x) > 1/(2\pi)$ for $x \in T_1$ a.e.. Let $\widetilde{S_1} = (-\delta', \delta)$ and $\widetilde{T_1} = [\delta, \delta'']$ for some $0 < \delta', \delta$ and $\delta'' > \delta$. It suffices to show that no $0 < \theta < \delta''$ can be another minimizer of V.

To this end with F from (4) consider

$$G(\theta) = \frac{\theta}{2\pi} - F(\theta) \quad \text{and} \quad H(\theta) = \int_0^\theta G(x)\mathrm{d}x = \int_0^\theta \left(\frac{x}{2\pi} - F(x)\right)\mathrm{d}x,$$

cf. (5) and recall that θ is a minimizer of V iff $H(\theta) \geq 0$. By construction, G' is positive on $(0, \delta)$ and it is negative on $[\delta, \delta']$ a.e. Hence G is continuous and $G(0) = 0 < G(\theta)$ for small $\theta > 0$. Hence, $H(\theta) > 0$ for all $0 < \theta \leq \delta$. Arguing again with Corollary 1 that there cannot be a minimizer of V if its antipode carries more density than the uniform density, gives that $\delta < \theta < \delta''$ cannot be a minimizer either, completing the proof. □

We note two straightforward consequences.

Corollary 3 *(i) Every population mean of a unimodal distribution is globally unique.*
(ii) If the distribution of X is composed of $k < \infty$ point masses at distinct locations, then V has at most k local minimizers, each being locally unique; for each interval formed by two neighboring point masses there is at most one local minimizer in the interior of the antipodal interval.
(iii) In particular, any intrinsic sample mean is locally unique.

Curiously, candidates for intrinsic sample means are very easy to obtain from one another.

Corollary 4 *For a sample $X_1, \ldots, X_n$ the candidates for minimizers of V_n described in Proposition 1 and Corollary 3 form the vertices of a regular n-polgyon. If $(X_1, \ldots, X_n)$ is continuously distributed on $\mathbb{T}^n$, then there is almost surely one and only one intrinsic sample mean.*

Proof W.l.o.g. assume that the sample is ordered, i.e., $-\pi \leq X_1 \leq \ldots \leq X_n < \pi$. We consider for $\theta \geq 0$ the case that $X_i < \theta - \pi < X_{i+1}$ for $1 \leq i \leq n-1$ or $\theta - \pi < X_1$ for $i = 0$. Note that equalities $X_i = \theta - \pi$ etc. are excluded by Theorem 1(i); also, observe that $\theta - \pi > X_n$ for $i = n$ cannot lead to a local minimum at θ. Then, setting the first sum to zero in case $i = 0$,

$$\begin{aligned} V_n(\theta) &= \frac{1}{n}\left(\sum_{j=1}^{i}(X_j - \theta + 2\pi)^2 + \sum_{j=i+1}^{n}(X_j - \theta)^2\right) \\ &= \frac{1}{n}\sum_{j=1}^{n}(X_j - \theta)^2 - \frac{4\pi}{n}\sum_{j=1}^{i}(\theta - X_j - \pi). \end{aligned} \tag{6}$$

This is minimal for $\theta^{(i)} = \bar{X} + \frac{2\pi i}{n}$. However, only if $X_i + \pi < \theta^{(i)} < X_{i+1} + \pi$, this minimum corresponds to a local minimum of V_n. Similarly, for $\theta < 0$ and $X_i < \theta + \pi < X_{i+1}$ for $1 \leq i \leq n-1$ or $X_n < \theta + \pi$ for $i = n$, we get

$$
\begin{aligned}
V_n(\theta) &= \frac{1}{n}\left(\sum_{j=1}^{i}(X_j-\theta)^2+\sum_{j=i+1}^{n}(X_j-\theta-2\pi)^2\right)\\
&= \frac{1}{n}\sum_{j=1}^{n}(X_j-\theta)^2-\frac{4\pi}{n}\sum_{j=i+1}^{n}(X_j-\theta-\pi),
\end{aligned}
\tag{7}
$$

which is minimal for $\theta^{(i)}=\bar{X}-\frac{2\pi(n-i)}{n}$. Again, note that $X_1>\theta+\pi$ for $n=0$ cannot lead to a local minimum at θ.

With $\bar{X}_i=\frac{1}{i}\sum_{j=1}^{i}X_j$ and $\underline{X}_i=\frac{1}{n-i}\sum_{j=i+1}^{n}X_j$, we obtain for a local minimizer $\theta^{(i)}$ that $V_n(\theta^{(i)})=v_{n,i}$ where

$$
v_{n,i}=\frac{1}{n}\sum_{j=1}^{n}(X_j-\bar{X})^2+\begin{cases}-\left(\frac{2\pi i}{n}\right)^2+\frac{4\pi i}{n}\left(\pi+\bar{X}_i-\bar{X}\right), & \theta^{(i)}\geq 0,\\ -\left(\frac{2\pi(n-i)}{n}\right)^2+\frac{4\pi(n-i)}{n}\left(\pi-\underline{X}_i+\bar{X}\right), & \theta^{(i)}<0,\end{cases}
\tag{8}
$$

whence $V_n(\theta^{(i)})=V_n(\theta^{(j)})$ for $i\neq j$ implies that there are $a,b,c,d\in\mathbb{Z}$, with $a\neq 0$ or $b\neq 0$ s.t. $a\bar{X}_i+b\bar{X}_j+c\bar{X}+d\pi=0$, the probability of which is zero for continuously distributed data. □

Remark 2 The fact that intrinsic sample means of continuous distributions are almost surely globally unique has also been observed by Bhattacharya and Patrangenaru (2003, Remark 2.6).

Remark 3 Corollary 4 has application for algorithmically determining an intrinsic sample mean: to do so, determine the minimizers of $V_n(\theta^{(i)})$ for $\theta^{(i)}\equiv\bar{X}+\frac{2\pi i}{n}$ mod 2π, $i=1,\dots,n$; this requires as many steps as there are data points in the sample. In fact, (8) easily leads to an implementation requiring $O(n)$ time for computing the intrinsic mean(s) of a sorted sample. An example is shown in Fig. 1; note that not all points on the polygon form candidates $\theta^{(i)}$: if $\bar{X}>\frac{2\pi i}{n}$ then $\bar{X}-\frac{2\pi i}{n}$ cannot be a minimizer. Also note that $V_n(\theta^{(i)})$ and $v_{n,i}$ only agree if $\theta^{(i)}$ is a local minimizer. A conceptually different algorithm for computing the intrinsic sample mean in linear time using combinatorial optimization has recently been introduced by McKilliam et al. (2012).

For the computation of population means, Proposition 1 allows for a similar procedure: compute in each interval with density less than $1/(2\pi)$ the unique minimizer.

Finally, we give an illustration to Corollary 1.

Example 1 Suppose that X is uniformly distributed on $[-\pi,-\pi+\delta\pi]\cup[\pi-\delta\pi,\pi)$ with $0\leq\delta\leq\frac{1}{2}$ and total weight $0\leq\alpha\delta\leq 1$, i.e., $0\leq\alpha\leq\delta^{-1}$ giving a density of $\frac{\alpha}{2\pi}$ near $\pm\pi$, and with a point mass of weight $1-\alpha\delta$ at 0. Then in case of $\alpha=1$, by Corollary 1, $V(\theta)$ is constant for $-\delta\pi\leq\theta\leq\delta\pi$, and $[-\delta\pi,\delta\pi]$ is precisely the set of intrinsic means. Moreover for $\alpha>1$, $\{-\alpha\delta\pi,\alpha\delta\pi\}$ is the set of intrinsic means whereas for $\alpha<1$, 0 is the unique intrinsic mean.

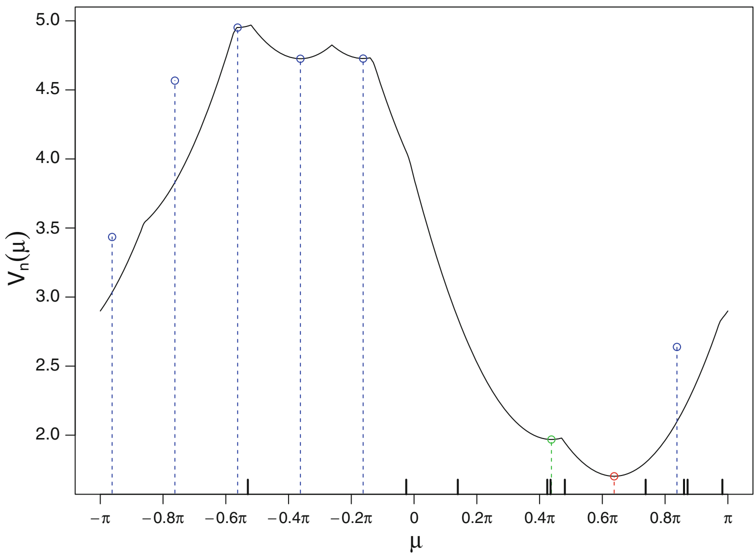

Fig. 1 Numerical example for determining a sample's intrinsic mean; see Remark 3: $n = 10$ observations $X_1, \ldots, X_{10}$ from a wrapped normal distribution are indicated as *vertical strokes at the bottom*; the curve depicts their $V_n(\theta)$; *vertical dashed lines* give the $\theta^{(i)}$ lying on a regular polygon from the proof of Corollary 4, the corresponding values $v_{n,i}$ from (8) are indicated by *circles*; the average $\bar{X}$ of the data is shown in *green*, the intrinsic mean in *red* (color figure online)

Proof Indeed, for $0 \le \theta \le \delta\pi$ we have

$$
\begin{aligned}
V(\theta) &= (1-\alpha\delta)\,\theta^2 + \frac{\alpha}{2\pi}\left(\int_{\pi-\delta\pi}^{\pi+\theta}(x-\theta)^2\,\mathrm{d}x + \int_{-\pi+\theta}^{-\pi+\delta\pi}(x-\theta)^2\,\mathrm{d}x\right)\\
&= (1-\alpha\delta)\,\theta^2 + \frac{\alpha}{6\pi}\left(\underbrace{\pi^3 - ((1-\delta)\pi-\theta)^3 + (-(1-\delta)\pi-\theta)^3 + \pi^3}_{=2\pi^3-2\,(1-\delta)^3\pi^3\,-2.3\,(1-\delta)\pi\,\theta^2}\right)\\
&= \frac{\alpha}{3}(3\delta - 3\delta^2 + \delta^3)\pi^2 + (1-\alpha)\,\theta^2
\end{aligned}
$$

which is constant in θ for $\alpha = 1$, minimal for $\theta = 0$ in case of $\alpha < 1$ and minimal for $\theta = \delta\pi$ in case of $\alpha > 1$. On the other hand for $\delta\pi \le \theta \le \pi$ we have

$$
\begin{aligned}
V(\theta) &= (1-\alpha\delta)\,\theta^2 + \frac{\alpha}{2\pi}\int_{\pi-\delta\pi}^{\pi+\delta\pi}(x-\theta)^2\,\mathrm{d}x\\
&= (1-\alpha\delta)\,\theta^2 + \frac{\alpha}{6\pi}\left(\underbrace{((1+\delta)\pi-\theta)^3 - ((1-\delta)\pi-\theta)^3}_{=((1+\delta)^3-(1-\delta)^3)\pi^3-3((1+\delta)^2-(1-\delta)^2)\pi^2\,\theta+2.3\,\delta\pi\,\theta^2}\right)\\
&= \frac{\alpha}{3}(3\delta + \delta^3)\pi^2 - 2\alpha\delta\pi\,\theta + \theta^2\\
&= V(0) + \delta^2\alpha(1-\alpha)\pi^2 + (\theta - \alpha\delta\pi)^2
\end{aligned}
$$

which is minimal for $\theta = \alpha\delta\pi$. In case of $\alpha = 1$ this minimum agrees with $V(0)$, in case of $\alpha < 1$ it is larger than $V(0)$, and in case of $\alpha > 1$ it is smaller than $V(\delta\pi)$. □

4 Asymptotics

The strong law of large numbers established by Ziezold (1977) for minimizers of squared quasi-metrical distances applied to the circle with its intrinsic metric, which renders it a compact space for which the sequence of $\theta_n \in \mathbb{T} = [-\pi, \pi)$ necessarily features an accumulation point, gives the following theorem; cf. also Bhattacharya and Patrangenaru (2003, Theorem 2.3(b)).

Theorem 2 *If θ^* is the unique minimizer of V and $(\theta_n)_{n\in\mathbb{N}}$ a measurable choice of minimizers of V_n, then $\theta_n \to \theta^*$ almost surely.*

More generally, if E_n denotes the set of intrinsic sample means, and E the set of intrinsic population means, then

$$\bigcap_{n=1}^{\infty} \overline{\bigcup_{k=n}^{\infty} E_n} \subset E \text{ almost surely.} \tag{9}$$

We now characterize the asymptotic distribution of $\theta_n \in \mathbb{T} = [-\pi, \pi)$ under similar assumptions as in Theorem 1, though additionally requiring that the locally unique intrinsic mean p^* is in fact globally unique, say $p^* = 0$. Then, $p_n - p^* = p_n$ might fulfill a central limit theorem, i.e., $\sqrt{n}\,p_n$ may converge to a normal distribution. Note that by the choice $p^* = 0$, we have $d(p_n, p^*) = p_n$, so that we may naturally identify p_n with the corresponding tangent vector at p^* with length p_n and that orientation. Then, $\sqrt{n}\,p_n \in \mathbb{R}$ would in fact converge in this tangent space, as is to be expected from the general results on manifolds, cf. e.g., Bhattacharya and Patrangenaru (2003, 2005). Part (i) of the following theorem, for which we give a new proof, is due to McKilliam et al. (2012).

Theorem 3 *Assume that the distribution of X restricted to some neighborhood of $-\pi \in \mathbb{T}$ features a continuous density f, has Euclidean variance σ^2 and that $\theta^* = 0$ is its unique intrinsic mean. Then the following assertions hold for the intrinsic sample mean $\theta_n \in \mathbb{T} = [-\pi, \pi)$ of independent and identically distributed $X_1, \ldots, X_n \sim X$:*

(i) If $f(-\pi) < \frac{1}{2\pi}$ then

$$\sqrt{n}\,\theta_n \xrightarrow{D} \mathcal{N}\left(0, \frac{\sigma^2}{\left(1 - 2\pi f(-\pi)\right)^2}\right).$$

(ii) If $f(-\pi) = \frac{1}{2\pi}$, and if f is $(k-1)$-times continuously differentiable in a neighborhood U of $-\pi$ with these $k-1$ derivatives vanishing at $-\pi$ while f is even k-times continously differentiable in $U \backslash \{-\pi\}$ with

$$\begin{aligned}0 < (-1)^{k+1} f^{(k)}(\pi-) &= (-1)^{k+1} \lim_{\theta\uparrow\pi} f^{(k)}(\theta)\\ &= -\lim_{\theta\downarrow-\pi} f^{(k)}(\theta) = -f^{(k)}(-\pi+) < \infty\end{aligned}$$

then

$$\sqrt{n}\ \mathrm{sign}(\theta_n)\,|\theta_n|^{k+1} \xrightarrow{D} \mathcal{N}\left(0, \frac{\sigma^2\big((k+1)!\big)^2}{\big(2\pi f^{(k)}(-\pi+)\big)^2}\right).$$

Proof With the indicator function

$$\chi_A(X) = \begin{cases} 1 & \text{if } X \in A\\ 0 & \text{if } X \notin A\end{cases}$$

we have that

$$\nu(X,\theta) = \begin{cases} \chi_{[-\pi,\theta-\pi)}(X_j) & \text{if } \theta > 0,\\ -\chi_{(\theta+\pi,\pi)}(X_j) & \text{if } \theta < 0.\end{cases}$$

In consequence of Theorem 1, if θ_n is an intrinsic sample mean, almost surely none of the X_j ($j = 1,\dots,n$) can be opposite of θ_n. Since the sample mean θ_n minimizes $V_n(\theta)$, we have hence with the well- defined derivative $^{\mathrm{d}}\mathrm{d}\theta V_n(\theta_n)$ that

$$0 = \frac{1}{2}\,\frac{\mathrm{d}}{}\mathrm{d}\theta V_n(\theta_n) = \begin{cases} \theta_n - \bar{X} - 2\pi\frac{1}{n}\sum_{j=1}^n \chi_{[-\pi,\theta_n-\pi)}(X_j) & \text{for } \theta_n \geq 0,\\ \theta_n - \bar{X} + 2\pi\frac{1}{n}\sum_{j=1}^n \chi_{(\theta_n+\pi,\pi)}(X_j) & \text{for } \theta_n < 0,\end{cases} \tag{10}$$

cf. (6) and (7). Under the assumptions of (i) above let us now compute

$$\mathbb{E}\left(\chi_{[-\pi,\theta-\pi)}(X)\right) = \int_{-\pi}^{\theta-\pi} f(x)\,\mathrm{d}x = \theta f(-\pi) + o(\theta)$$

in case of $\theta > 0$ and similarly

$$\mathbb{E}\left(\chi_{(\theta+\pi,\pi)}(X)\right) = -\theta f(-\pi) + o(\theta)$$

in case of $\theta < 0$. In consequence, using that the variance of these Bernoulli variables is less or equal than their expectation, we get

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^n \chi_{[-\pi,\theta-\pi)}(X_i) &= \theta f(-\pi) + O_P\left(\frac{\sqrt{\theta}}{\sqrt{n}}\right) + o(\theta),\\ \frac{1}{n}\sum_{j=1}^n \chi_{(\theta+\pi,\pi)}(X_i) &= -\theta f(-\pi) + O_P\left(\frac{\sqrt{-\theta}}{\sqrt{n}}\right) + o(\theta)\end{aligned}$$

for $\theta > 0$ and $\theta < 0$, respectively, the bounds for $O_P(\sqrt{\theta/n})$ being uniform in θ. In conjunction with (10), and using the strong law of large numbers for θ_n (Theorem 2), i.e., $\theta_n = o_P(1)$, we obtain

$$\sqrt{n}\left((1 - 2\pi f(-\pi))\,\theta_n - \bar{X}\right) = o_P(1).$$

This gives assertion (i).

Under the assumptions of (ii) we get, noting that by Theorem 1(iii) $\mathrm{sign}(f^{(k)}(-\pi+)) = -1$ while $\mathrm{sign}(f^{(k)}(\pi-)) = (-1)^{k+1}$,

$$\left.\begin{array}{l}\mathbb{E}\left(\chi_{[-\pi,\theta-\pi)}(X)\right)\\ \mathbb{E}\left(\chi_{(\theta+\pi,\pi)}(X)\right)\end{array}\right\} = \frac{|\theta|}{2\pi} + \frac{|\theta|^{k+1}}{(k+1)!} f^{(k)}(-\pi+) + o(\theta^{k+1}).$$

In consequence, as above, we infer from (10) that

$$\sqrt{n}\left(2\pi\,\frac{\mathrm{sign}(\theta_n)|\theta_n|^{k+1}}{(k+1)!}\,f^{(k)}(-\pi+) + \bar{X}\right) = o_P(1),$$

which gives assertion (ii). □

Remark 4 We note that under the assumptions in Theorem 3, namely that f differs from the uniform distribution at $-\pi$ for the first time in its k-th derivative there, then the convergence rate of θ_n is precisely $n^{-\frac{1}{2(k+1)}}$.

Comparing with (2), we see that the asymptotic distribution of $\bar{X}$ is more concentrated than the one of the intrinsic mean unless $f(-\pi) = 0$, the intrinsic mean exhibiting slower convergence rates than $\bar{X}$ if $f(-\pi) = \frac{1}{2\pi}$.

Concerning the distributions typically occurring in practice, we obtain by combining Proposition 1 and Theorem 3(i):

Corollary 5 *Assume that the distribution of X features a continuous density f with respect to Lebesgue measure such that the set where f is strictly less than the uniform density is connected. Then it features a unique intrinsic mean θ^*, and $\sqrt{n}(\theta_n - \theta^*)$ is asymptotically normally distributed where θ_n is the almost surely unique sample mean of independent and identically distributed $X_1, \ldots, X_n \sim X$.*

Thus, for the classically used distributions, e.g., wrapped normal, Fisher, or von Mises, as well as for suitable mixtures thereof, the intrinsic mean behaves qualitatively as its Euclidean counterpart.

5 Simulation

For illustration of the theoretical results we consider here examples exhibiting different convergence rates: we generalize the density from Example 1 to behave like a polynomial of order k near $\pm\pi$. To be precise, we assume that the distribution of X is composed of a point mass at 0 with weight $1 - \alpha\delta$ with $0 \le \alpha \le \delta^{-1}$, $0 \le \delta \le \frac{1}{2}$, and of a part absolute continuous with respect to Lebesgue measure with density g where $g(-\pi + x) = g(\pi - x) = f(x)$ for $0 \le x \le \pi$, and

Table 1 Parameters for the simulations, and respective colors used in Figs. 2 and 4

	α	k	δ	Color
Case 0a	0.9	0	0.4	Blue
Case 0b	1	0	0.4	Red
Case 1a	0.9	1	0.4	Green
Case 1b	1	1	0.4	Brown
Case 2	1	2	0.4	Violet
Case 3	1	3	0.4	Purple

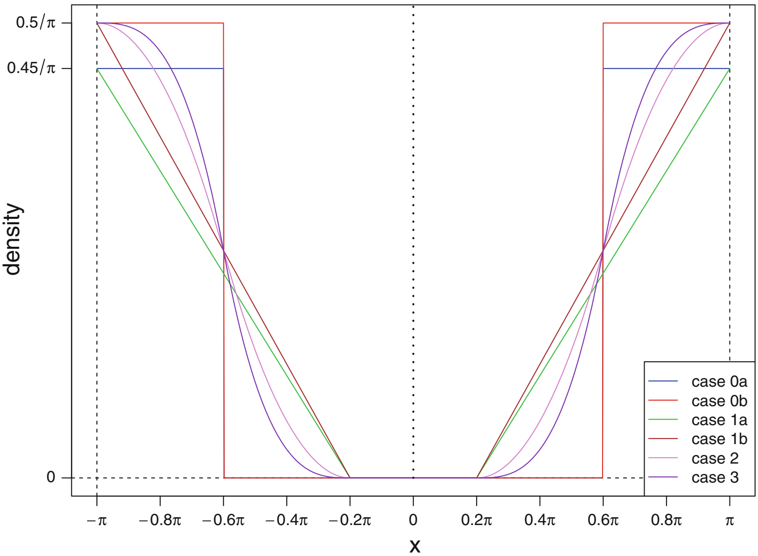

Fig. 2 Densities g of the simulated distributions with parameters in Table 1; the *dotted vertical line in the center* indicates the point mass at 0 (color figure online)

$$f(x) = \begin{cases} \frac{\alpha}{2\pi} & \text{for } 0 \le |x| \le \delta\pi, \\ 0 & \text{else} \end{cases}$$

for $k = 0$, while for $k > 0$

$$f(x) = \begin{cases} \frac{\alpha}{2\pi} - \frac{\alpha}{4\pi}\left(\frac{|x|}{\delta\pi}\right)^k & \text{for } 0 \le |x| \le \delta\pi, \\ \frac{\alpha}{4\pi}\left(2 - \frac{|x|}{\delta\pi}\right)^k & \text{for } \delta\pi \le |x| \le 2\delta\pi, \\ 0 & \text{else.} \end{cases}$$

Note that $f(-\pi) = \frac{\alpha}{2\pi}$ while $\int_{-\pi}^{\pi} f(x)\mathrm{d}x = \alpha\delta$. We simulated several examples with parameters given in Table 1, the corresponding densities are shown in Fig. 2.

Example 1 is the special case for which $k = 0$; in particular, for case 0b, we computed 10,000 intrinsic means, each of which was based on $n = 10{,}000$ independent and identically distributed observations, a histogram of these intrinsic means is shown

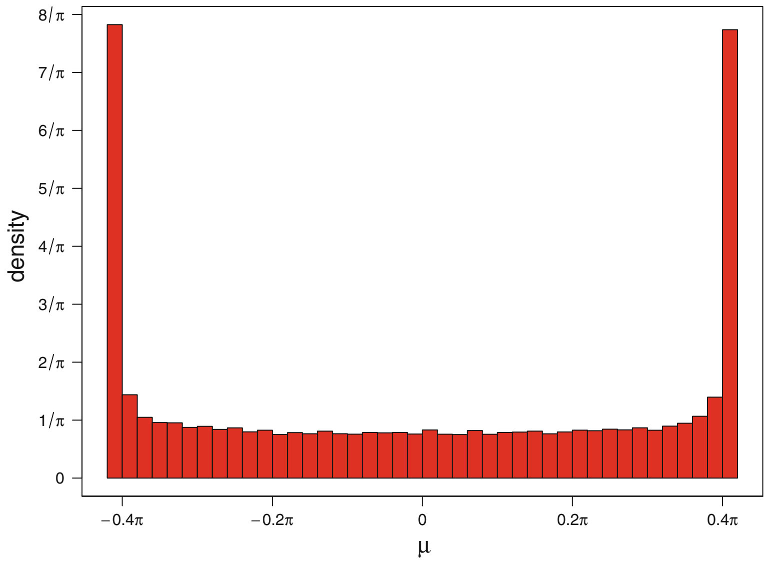

Fig. 3 Histogram of 10,000 intrinsic means each based on n = 10,000 draws from case 0b in Table 1

in Fig. 3. There, the distribution of the intrinsic sample mean for case 0b appears to be composed of two parts: an essentially constant density over $[-\delta\pi, \delta\pi]$, the set of the intrinsic population mean, and two peaks with their modes located close to the interval's endpoints. Their presence can be explained as follows: approximately with probability one half, we observe less than $(1-\delta)n$ zeros, whence there is too little mass at 0 to keep the intrinsic sample mean in the interval $[-\delta\pi, \delta\pi]$ but for n large enough there will with large probability still be many zeros so that the intrinsic sample mean cannot move far away from that interval. According to Theorem 2, these peaks' locations converge to the interval's boundary when $n \to \infty$. In fact, our simulations suggest that we have $\cap_{n=1}^{\infty} \overline{\cup_{k=n}^{\infty} E_n} \cap A \neq \emptyset$ with positive probabiliy for any $A \subset [-\delta\pi, \delta\pi]$ having non-zero Lebesgue measure.

We also determined the median absolute deviation of the intrinsic sample means for the different cases in Table 1, and compared them to the median absolute deviation predicted from the asymptotic distribution given in Theorem 3 (except for case 0b where it does not apply), see Fig. 4. For this, one easily computes

$$
\begin{aligned}
\sigma^2 &= 2\int_0^{2\delta\pi} (\pi - x)^2 f(x)\mathrm{d}x = 2\int_0^{\delta\pi} (\pi - x)^2 f(x)\mathrm{d}x + 2\int_{\delta\pi}^{2\delta\pi} (\pi - x)^2 f(x)\mathrm{d}x \\
&= -\frac{2\alpha}{6\pi}\left[(\pi - x)^3\right]_0^{\delta\pi} - \frac{\alpha}{2\pi(\delta\pi)^k}\left[\frac{\pi^2}{k+1}x^{k+1} - \frac{2\pi}{k+2}x^{k+2} + \frac{1}{k+3}x^{k+3}\right]_0^{\delta\pi} \\
&\quad + \frac{\alpha}{2\pi(\delta\pi)^k}\left[\frac{\pi^2(1-2\delta)^2}{k+1}x^{k+1} + \frac{2\pi(1-2\delta)}{k+2}x^{k+2} + \frac{1}{k+3}x^{k+3}\right]_0^{\delta\pi} \\
&= \frac{\alpha\pi^2}{3}\left(1 - (1-\delta)^3\right) - \frac{\alpha\delta\pi^2}{2(k+1)}\left(1 - (1-2\delta)^2\right) + \frac{\alpha(\delta\pi)^2}{k+2}(2 - 2\delta),
\end{aligned}
$$

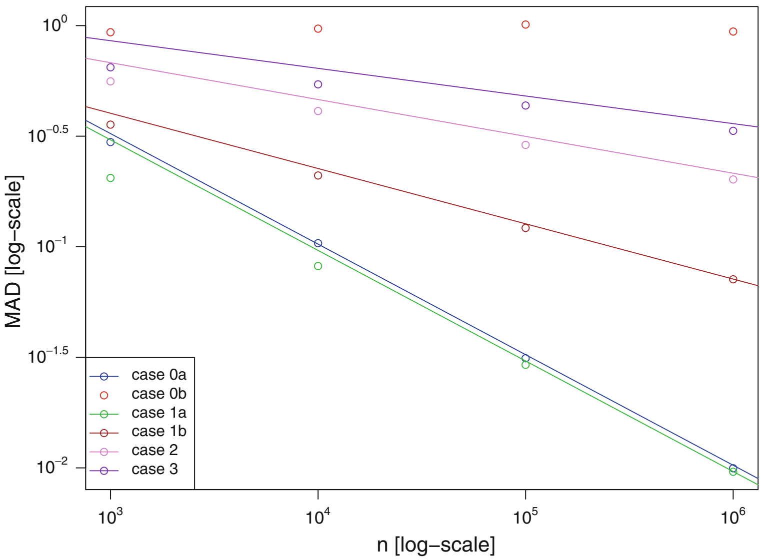

Fig. 4 Median absolute deviations of the intrinsic mean of n draws from the simulated distributions with parameters in Table 1, based on 1,000 repetitions; *lines* give the values predicted using the asymptotic distribution (color figure online)

as well as $f^{(k)}(0+) = -\alpha \frac{k!}{4\pi(\delta\pi)^k}$. We chose the median absolute deviation as it commutes with the power transforms in Theorem 3(ii), as opposed to the standard deviation, and is therefore easier to compute exactly. We then found the rates predicted in Remark 4, namely $n^{-\frac{1}{2}}$ if $\alpha < 1$ and $n^{-\frac{1}{2(k+1)}}$ if $\alpha = 1$ and $k > 0$, to match the observed median absolute deviation in Fig. 4 well.

Furthermore, for case 1a, Fig. 5 shows normal q–q plots for the intrinsic sample means, transformed and standardized according to their asymptotic distribution, i.e., for $\sqrt{n}\ \mathrm{sign}(\theta_n)\,|\theta_n|^{k+1}\frac{-2\pi f^{(k)}(0+)}{(k+1)!\sigma)}$. Note that there appears to be a peak at 0, visible from the curve getting almost horizontal there, which decreases with increasing n.

6 Discussion

Let us conclude with a discussion of our rather comprehensive results on locus, uniqueness, asymptotics and numerics for intrinsic circular means. In the past, there has been a fundamental mismatch between distributional and asymptotic theory on non-Euclidean manifolds. While a great variety of distributions for circular data had been developed which very well reflect the non-Euclidean topology, e.g., nowhere vanishing densities, the central limit theorem had only been available for distributions essentially restricted to a subset of Euclidean topology. On the circle we have eliminated this mismatch. In particular our results state that

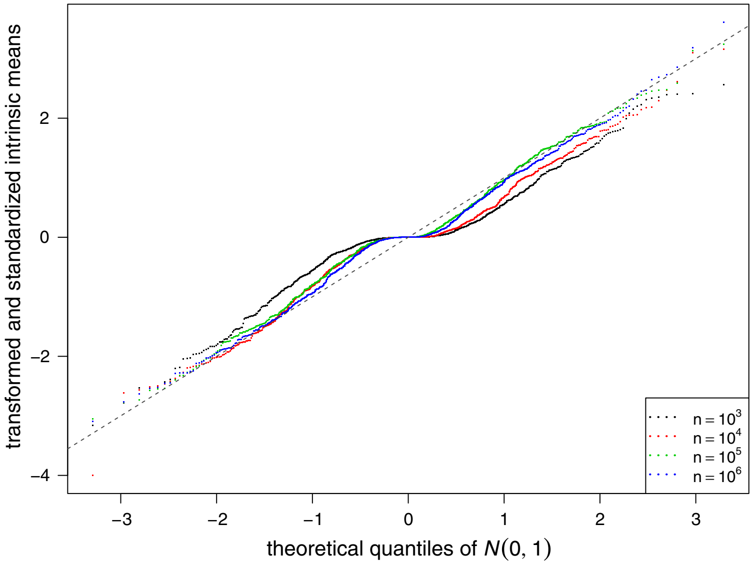

Fig. 5 *Q–q* plots of 1,000 intrinsic means, transformed and standardized according to the asymptotic distribution, each based on *n* draws from case 1b in Table 1; the *dashed line* depicts the identity line (color figure online)

the more the non-Euclidean topology is reflected by a probability distribution, i.e., the closer the distribution near the antipode is to the uniform distribution, the larger the deviation from Euclidean asymptotics.

We expect that similar results are valid for general manifolds where the antipode needs to be replaced by the *cut locus* $C(\theta)$. Unless $C(\theta)$ carries positive mass, $V(\theta)$ is still differentiable at θ, see Pennec (2006). From what we observed for the circle, we conjecture that $C(\theta)$ cannot carry mass if θ is a local minimizer, and that an analogue of Corollary 5 will hold. However, generalizing all results obtained here to arbitrary Riemannian manifolds is the subject of future research, but note that our results apply componentwise to tori, they being direct products of circles. The asymptotic distribution of each component of the intrinsic sample mean on a torus can then be used in image analysis, since the projective shape space of finite configurations of k landmarks on a line, with the first three landmarks being distinct, is the torus $\mathbb{T}^{k-3}$, cf. Mardia and Patrangenaru (2005).

One may compare our results on the circle to those known for the *extrinsic* or *circular mean*. It is also obtained by minimizing the square of a distance, namely of the so-called *extrinsic* or *chordal* distance given by $2\sin\big(d(\theta,\zeta)/2\big)$, cf. Sect. 2. It is easy to see that the extrinsic population mean is (essentially) given by $\arctan(\mathbb{E}(\cos(X))/\mathbb{E}(\sin(X)))$, and it fulfills a law of large numbers as well as a central limit theorem if the extrinsic population mean is unique, see e.g., Mardia and Jupp (2000). However, no condition on the latter's antipode enters. This is because the squared extrinsic distance is everywhere differentiable. We thus conjecture that a central limit theorem holds for any distance but that the rate depends on the density at points where the squared distance to the mean is not differentiable. However, the explicit computations, and in particular the result on local minimizers in Corollary 4, will not necessarily carry over easily.

Acknowledgments The authors thank the two anonymous reviewers as well as the editorial committee for their comments which helped in improving this manuscript. T. Hotz acknowledges support by DFG CRC 803, S. Huckemann by DFG HU 1575/2-1 and the Niedersachsen Vorab of the Volkswagen Foundation. Both authors are more than grateful for support by SAMSI whose 2010 workshop on AOOD and the successive research stays there considerably enabled this research.

References

Afsari, B. (2011). Riemannian L^p center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, *139*, 655–773.

Bhattacharya, R.N., Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds I. *The Annals of Statistics*, *31*(1), 1–29.

Bhattacharya, R.N., Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds II. *The Annals of Statistics*, *33*(3), 1225–1259.

Fisher, N. (1993). *Statistical analysis of circular data*. Cambridge: Cambridge University Press.

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'institut Henri Poincaré*, *10*(4), 215–310.

Huckemann, S. (2011). Inference on 3D Procrustes means: tree boles growth, rank-deficient diffusion tensors and perturbation models. *Scandinavian Journal of Statistics*, *38*(3), 424–446.

Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, *XXX*, 509–541.

Kaziska, D., Srivastava, A. (2008). The Karcher mean of a class of symmetric distributions on the circle. *Statistics and Probability Letters*, *78*(11), 1314–1316.

Kobayashi, S., Nomizu, K. (1969). *Foundations of differential geometry* (vol. II). Chichester: Wiley.

Le, H. (1998). On the consistency of Procrustean mean shapes. *Advances of Applied Probability (SGSA)*, *30*(1), 53–63.

Mardia, K., Patrangenaru, V. (2005). Directions and projective shapes. *The Annals of Statistics*, *33*, 1666–1699.

Mardia, K. V., Jupp, P. E. (2000). *Directional statistics*. New York: Wiley.

McKilliam, R.G., Quinn, B.G., Clarkson, I.V.L. (2012). Direction estimation by minimum squared arc length. *IEEE Transactions on Signal Processing*, *60*(5), 2115–2124.

Pennec, X. (2006). Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, *25*(1), 127–154.

Ziezold, H. (1977). Expected figures and a strong law of large numbers for random elements in quasi-metric spaces. *Transaction of the 7th Prague Conference on Information Theory, Statistical Decision Function and Random Processes*, *A*, 591–602.