

On geometric probability distributions on the torus with applications to molecular biology

Alessandro Selvitella

*Department of Mathematics and Statistics, McMaster University
and*

*Department of Mathematics and Statistics, University of Ottawa
and*

*Department of Psychiatry and Department of Computing Science, University of Alberta
AMII Lab, Computing Science Center 2-05, Edmonton, AB, Canada T6G 2E8
E-mail: e-mail: selvitel@ualberta.ca*

Abstract: In this paper, we study a family of probability distributions, alternative to the *von Mises* family, called *Inverse Stereographic Normal Distributions*. These distributions are counterparts of the *Gaussian Distribution* on \mathbb{S}^1 (univariate) and \mathbb{T}^n (multivariate). We discuss some key properties of the models, such as unimodality and closure with respect to marginalizing and conditioning. We compare this family of distributions to the *von Mises*' family and the *Wrapped Normal Distribution*. Then, we discuss some inferential problems, introduce a notion of moments which is natural for inverse stereographic distributions and revisit a version of the CLT in this context. We construct point estimators, confidence intervals and hypothesis tests and discuss briefly sampling methods. Finally, we conclude with some applications to molecular biology and some illustrative examples. This study is motivated by the *Protein Folding Problem* and by the fact that a large number of proteins involved in the DNA-metabolism assume a toroidal shape with some amorphous regions.

MSC 2010 subject classifications: Primary 62E15, 62H10, 62H11; secondary 62E10, 62E20, 62F03, 62F10, 62F12, 62H15.

Keywords and phrases: Distributions on manifolds, von Mises distributions, inverse stereographic distributions, molecular biology, protein folding problem, DNA-metabolism.

Received June 2017.

Contents

1	Introduction	2718
2	Notation and preliminary results	2722
2.1	Inverse stereographic projected distributions	2722
2.2	Marginals and conditionals of MISND	2723
2.3	Unimodality and multimodality conditions	2724
3	Comparison theorems: ISND vs Von Mises and Wrapped Normal	2726
4	ISMs, ISMGF and a version of the Central Limit Theorem	2727
4.1	The inverse stereographic moments	2728
4.2	A CLT and the ISMGF	2730

5	Inference	2732
6	Applications and numerical examples	2734
6.1	Plots	2734
6.2	Comparison between the VM and the ISND	2734
6.3	Applications to molecular biology	2739
7	Discussion	2744
8	Conclusion	2746
A	Proofs	2747
B	Supplementary material	2752
	Acknowledgements	2761
	References	2761

1. Introduction

In the recent years, statisticians paid increasing attention to random variables taking values on manifolds. Amongst the most important statistical problems on manifolds, those arising from the analysis of circular data, spherical data and toroidal data play a fundamental role (see [15] and [21] for some references and [29]). Motivated by important applications in molecular biology, we will concentrate on circular and toroidal data.

The motivation behind this paper relates to the *Protein Folding Problem* (PFP), which is one of the major open problems in biochemistry. Protein folding is the physical process that a protein open chain undertakes before reaching its final three dimensional shape (*conformation*). The shape of the protein ensures that the protein does its job properly, while a misfolding can be the cause of several neurodegenerative diseases and other types of diseases as well [37].

Since the physical process underlying protein folding is complicated, there has been limited success in predicting the final folded structure of a protein from its amino acid sequence. A better understanding of this would definitely have clinical impact and result in the design of efficient drug molecules for the cure of several diseases such as the ones mentioned above.

The PFP can be divided into two parts: the static problem and the dynamic problem. As explained in [6], the static problem concerns the prediction of the active conformation of the protein given only its amino acid sequence. The dynamic problem, instead, consists in rationalizing the folding process as a complex physical interaction. Both these problems have a precise mathematical formulation [6].

Some regions of the conformation of a protein may look amorphous and so require random models and probability distributions to be described properly. As a matter of fact, a large number of proteins involved in DNA metabolism with different evolutionary origin and catalyzing different reactions assume a toroidal shape [10], [11]. It has been argued that the preference towards the toroidal form may be due to special advantages in the DNA-binding [10], [11]. A long list of proteins share the toroidal form [10], [11]. The importance of the

PPF and the DNA-binding process motivated statisticians to find appropriate statistical models that describe these phenomena.

The most famous circular distribution is the *von Mises distribution*. The interest on the theory as well as practical applications stimulated researchers to look for a good higher dimensional analogue of the one dimensional *von Mises Distribution*. The *Bivariate von Mises Distribution* (BVM) is a probability distribution describing a two dimensional random vector, taking values on a torus $\mathbb{T}^2 := \mathbb{S}^1 \times \mathbb{S}^1$. It aims at representing an analogue on the torus of the *Bivariate Normal Distribution* (BVN). The *Full Bivariate von Mises Distribution* (FBVM) was first proposed by Mardia [17]. Some of its variants are currently used in the field of *bioinformatics* to formulate probabilistic models of protein structure.

The FBVM seems to be over-parametrized for being the *toroidal counterpart* of the BVN. In fact, the FBVM depends on eight parameters, while the BVN possesses only five parameters. This situation becomes even clearer in the high-concentration limit ($\kappa \rightarrow +\infty$) [26]. For this reason, several submodels have been proposed. Four commonly used variants of the bivariate von Mises distribution have been originally proposed by Mardia [18] and then revisited by Rivest [35] and also by Singh-Hnidzo-Demchuk [39]. They are models with a reduced number of parameters and are derived by setting to zero the off-diagonal elements of the interaction matrix. The following models are submodels of the FBVM which have been discussed in the literature: the *Cosine Model with Positive Interaction* [26], the *Cosine Model with Negative Interaction* [26], the *Sine Model* [39] and the *Hybrid Model* [13].

The use of these distributions has pros and cons. The pros: the *von Mises* distributions resemble the BVN in the high-concentration limit, they are closed with respect to conditioning, it is relatively easy to give unimodality conditions, and the parameters have easy interpretability, even when they do not exactly match the ones of the BVN. The cons: the family is not closed under marginalization (but in the case of high-concentration limit), inference and prediction are not trivial and require numerical methods. For example, MLEs cannot be computed explicitly, but just via optimization algorithms. To overcome this problem, more advanced procedures, like pseudo-likelihood estimators, have been suggested [21].

It is a common belief that the geometry of the torus implies that a completely natural counterpart of the BVN on the torus is not available (see, for example, the discussion in [13]). Therefore, despite *von Mises*' models have been proven to be successful, there has not been a definite answer yet to which of the models proposed is the "best" candidate to represent the *toroidal counterpart* of the BVN.

In this paper, we aim at giving an alternative to *von Mises* models both in univariate and multivariate case, which maintain the good properties of the *von Mises* distributions, like the asymptotic normality, but also possess some extra properties, such as the simplicity in the estimation that the BVN possesses in the Euclidean case, but the *von Mises* distributions do not.

The candidate distribution in the univariate case of \mathbb{S}^1 is called *Inverse Stere-*

ographic Normal Distribution (ISND). The name comes from the fact that the *Stereographic Projection* of this distribution is the Normal Distribution on \mathbb{R} .

Definition 1.1. We say that a random variable $\Theta \in \mathbb{S}^1$ has an *Inverse Stereographic Normal Distribution*, denote by $\mathcal{ISN}(\mu, \sigma^2)$, if and only if its pdf is given by

$$f_{\Theta}(\theta) := \sqrt{\frac{1}{2\pi}} \frac{1}{(1 + \cos \theta)\sigma} e^{-\frac{1}{2\sigma^2} \left(\frac{\sin \theta}{1 + \cos \theta} - \mu \right)^2},$$

for some $\mu \in \mathbb{R}$, $\sigma > 0$ and for any $\theta \in [-\pi, +\pi)$. The distribution $\mathcal{ISN}(0, 1)$ is called *Inverse Stereographic Standard Normal Distribution*.

In a similar manner, we can define on $\mathbb{T}^n := [-\pi, +\pi)^n$ the *Multivariate Inverse Stereographic Normal Distribution* (MISND).

Definition 1.2. We say that a random vector $\Theta := \{\Theta_1, \dots, \Theta_n\} \in \mathbb{T}^n$ has a *Multivariate Inverse Stereographic Normal Distribution*, denoted by $\mathcal{MISN}(\mu, \Sigma)$, if and only if its pdf is given by

$$f_{\Theta}(\theta) := (2\pi)^{-n/2} |\Sigma|^{-1/2} \prod_{i=1}^n \frac{1}{1 + \cos \theta_i} \times e^{-\frac{1}{2} \left(\frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1, \dots, \frac{\sin \theta_n}{1 + \cos \theta_n} - \mu_n \right)^T \Sigma^{-1} \left(\frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1, \dots, \frac{\sin \theta_n}{1 + \cos \theta_n} - \mu_n \right)},$$

for some $\mu := (\mu_1, \dots, \mu_n) \in \mathbb{R}^n$, $\Sigma \in \text{Sym}_{n \times n}^+$ with $|\Sigma| := \det(\Sigma)$, and for any $\theta = (\theta_1, \dots, \theta_n) \in \mathbb{T}^n$. The distribution $\mathcal{MISN}(0, Id)$ is called *Multivariate Inverse Stereographic Standard Normal Distribution*.

In the following, we concentrate on the case $n = 2$, namely the *Bivariate Inverse Normal Distribution* (BISND), since it is more relevant to the applications to molecular biology (See Subsection 6.3).

Remark 1.3. Note that this is not the only way in which one can introduce a location parameter in the ISND. Previously, the location parameter for the ISND has been introduced as the location of the axis of symmetry of the distribution on the circle (see for example [1]). In our case μ does not play that role. Instead, μ corresponds to the mean value of the corresponding projected distribution on the real line. We refer to Section 4 for more comments on this.

There are several advantages of this approach, listed below not necessarily in order of importance:

- The *Inverse Stereographic Projection* suggests a natural way to construct distributions on manifolds, by transforming distributions on the Euclidean Space (see Section 2.1);
- The number of parameters of the BISND matches the number of parameters of the BVN, without imposing any further constraint and there is a natural interpretation for the parameters of the ISND in terms of the parameters of the BVN (see Section 4);

- The BVM and BISND resemble each other in the case of high-concentration limit for certain ranges of parameters. Analogously for the ISND and *Wrapped Normal Distribution* (WN). Moreover, the ISND approximates the BVN in the case of high concentration limit (see Section 3);
- The *Stereographic Projection* suggests a more geometric counterpart of the *Euclidean Moments* which differs from the *Circular Moments* (see Section 4);
- The definition of moments using the *Stereographic Projection* helps address some interpretation problems in the parameters of directional variables (for further comments, we refer to Section 7);
- The MLE estimates for the ISND do not need numerical methods, since the estimates of the parameters are the transformed estimates of the parameters of the BVN (see Section 5);
- The *Stereographic Projection* allows to transfer some test statistics, confidence intervals and hypothesis testing from the Euclidean space to manifolds (see Section 5).

The rest of this paper is organized as follows.

In Section 2, we collect some notation and describe some preliminary results. We describe the construction of general inverse stereographic projected distributions; we give results about the marginals of the ISND which are still ISND (Theorem 2.3), and conditionals which are ISND as well (Theorem 2.4); we give unimodality conditions for this family of distributions (Theorem 2.6).

In Section 3, we compare the ISND with the VM and the WN. In particular, we prove approximation results in the case of high-concentration limit which connect ISND, VM and WN for some subsets of the parameter space (Theorem 3.1, Theorem 3.2 and Theorem 3.3). Note that WN and VM are close also for moderate values of the concentration parameter κ , which in the past has been an argument for using the VM for circular data, due to its compact form.

In Section 4, we introduce *Inverse Stereographic Moments* (ISMs) and *Inverse Stereographic Moment Generating Functions* (ISMGFs) which lead naturally to a corresponding “intrinsic” CLT on the torus (Theorem 4.12). We compare ISMs with the classical *Circular Moments* and with the moments of Euclidean random variables.

In Section 5, we pass to *Statistical Inference*. We propose a way to do point estimation, confidence intervals and hypothesis testing for the model parameters following the lines of Euclidean cases.

In Section 6, we present some numerical examples and applications. In Subsection 6.1, we produce some plots of the ISND for some choices of the parameters, both in the unimodal and the multimodal cases. In Subsection 6.2, we compare numerically the BVM and the ISND, using the results in Section 3 for the corresponding ranges of the parameters. In Subsection 6.3, we give an application to problems in Molecular Biology, in particular to the problem of the distribution of *dihedral angles* on the torus.

We complete our analysis with a discussion (Section 7), the conclusion (Section 8) and the technical proofs (Appendix A: Proofs). In the supplementary

material, we collect some of the codes (Appendix B: Supplementary Material) used in the applications.

2. Notation and preliminary results

We denote with $Mat_{n \times n}$ the set of square matrices of dimension n , and with $Sym_{n \times n}^+$ the set of positive symmetric matrices of dimension n .

In Subsection 2.1, we briefly discuss the construction of *Inverse Stereographic Projected Distributions*. In Subsection 2.2, we present results about the marginals of ISND which are still ISND (Theorem 2.3), and conditionals of the ISND which are again ISND (Theorem 2.4). In Subsection 2.3, we give unimodality conditions for ISND (Theorem 2.6).

2.1. Inverse stereographic projected distributions

Suppose $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$ is a pdf on \mathbb{R}^n . The *Stereographic Projection* is defined as

$$(\Theta_1, \dots, \Theta_n) \mapsto (X_1, \dots, X_n) := \left(\frac{\sin(\Theta_1)}{1 + \cos(\Theta_1)}, \dots, \frac{\sin(\Theta_n)}{1 + \cos(\Theta_n)} \right),$$

for $(\Theta_1, \dots, \Theta_n) \in \mathbb{T}^n$ and so with $\Theta_i \in [-\pi, +\pi)$ for every $i = 1, \dots, n$. Therefore,

$$\begin{aligned} f_{\Theta_1, \dots, \Theta_n}(\theta_1, \dots, \theta_n) d\theta_1 \dots d\theta_n = \\ f_{X_1, \dots, X_n} \left(\frac{\sin(\theta_1)}{1 + \cos(\theta_1)}, \dots, \frac{\sin(\theta_n)}{1 + \cos(\theta_n)} \right) \prod_{i=1}^n \frac{d\theta_i}{1 + \cos(\theta_i)}. \end{aligned}$$

Viceversa, suppose $f_{\Theta_1, \dots, \Theta_n}(\theta_1, \dots, \theta_n)$ is a pdf on \mathbb{T}^n . The *Inverse Stereographic Projection* is defined as

$$(X_1, \dots, X_n) \mapsto (\Theta_1, \dots, \Theta_n) := (2 \arctan(X_1), \dots, 2 \arctan(X_n)),$$

for every $(X_1, \dots, X_n) \in \mathbb{R}^n$. Therefore,

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n = \\ 2^n \prod_{i=1}^n \frac{dx_i}{1 + x_i^2} f_{\Theta_1, \dots, \Theta_n}(2 \arctan(x_1), \dots, 2 \arctan(x_n)). \end{aligned}$$

As a consequence, a natural measure on the torus when the probability distribution is obtained through *Inverse Stereographic Projection* P is the pull-back measure of the Lebesgue measure on \mathbb{R} :

$$P^* dx_1 \dots dx_n = \prod_{i=1}^n \frac{d\theta_i}{1 + \cos \theta_i}.$$

Remark 2.1. There are other possible multidimensional generalizations of the inverse stereographic projection. In the case of the sphere, the authors of [1] note that the Jacobian of the stereographic projection is non-trivial and they use a projection which produces a slightly simpler Jacobian to the expenses of a slightly more complicated construction (see also [41]). When reduced to $p = 2$, namely \mathbb{S}^1 , the projections here and in [1] coincide.

Remark 2.2. The *Stereographic Projection* is a *conformal transformation* and so it does not change angles. In our context, this is a very useful property as it means that the *Stereographic Projection* maps *elliptical contours* to *elliptical contours*.

2.2. Marginals and conditionals of MISND

In strong contrast with the *Multivariate von Mises Distribution*, the ISND is closed under marginalization.

Theorem 2.3. *Suppose a random variable*

$$\Theta := \{\Theta_1, \dots, \Theta_n\} \sim \text{MISN}(\mu, \Sigma).$$

Then, for every $i = 1, \dots, n$, we have

$$\Theta_i \sim \text{ISN}(\mu_i, \Sigma_{ii}).$$

Proof. See Appendix A. □

The ISND is closed under conditioning too.

Theorem 2.4. *Suppose a random variable*

$$\Theta := \{\Theta_1, \dots, \Theta_n\} \sim \text{MISN}(\mu, \Sigma)$$

and we have the partition $\Theta = (\Theta_1, \Theta_2)$, $\mu = (\mu_1, \mu_2)$,

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

where Θ_1 and μ_1 are $k \times 1$ vectors and $\Sigma_{11} \in \text{Mat}_{k \times k}$. Let Σ_{22}^- be the generalized inverse of Σ_{22} . Then, the conditional distribution of Θ_1 , given Θ_2 is given by

$$\Theta_1 \mid \Theta_2 \sim \text{MISN}(\mu_1 - \Sigma_{12}\Sigma_{22}^-(\Theta_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^-\Sigma_{21}).$$

Proof. The proof follows the same lines as in the proof of Theorem 2.3 and the techniques used in the Euclidean case (see for example Theorem 1.2.11 page 12 in [28]). □

Remark 2.5. As far as we know, these results are new at least in the case $\mu \neq 0$.

2.3. Unimodality and multimodality conditions

The ISND can be both unimodal or multimodal depending on the parameters μ and σ^2 (see the plots in Subsection 6.1). In this subsection, we give necessary and sufficient conditions for the ISND to be unimodal. Recall that in the case $n = 1$, $\Theta \sim \mathcal{ISN}(\mu, \sigma^2)$ has a pdf given by

$$f_{\Theta}(\theta) := \sqrt{\frac{1}{2\pi}} \frac{1}{(1 + \cos \theta)\sigma} e^{-\frac{1}{2\sigma^2} \left(\frac{\sin \theta}{1 + \cos \theta} - \mu \right)^2},$$

for some $\mu \in \mathbb{R}$, $\sigma > 0$ and for any $\theta \in [-\pi, +\pi)$. Note that, for $\theta = -\pi$, the density has a removable singularity and it can be extended to a smooth function by choosing $f_{\Theta}(-\pi) = 0$. This makes $\theta = -\pi$ a critical point of the density and hence the global minimum. Since \mathbb{S}^1 is compact and $f_{\Theta}(\theta)$ is not constant, it admits a maximum and a minimum distinct from the maximum. To ensure unimodality, we need to find μ and σ^2 such that $f_{\Theta}(\theta)$ admits only one further critical point apart from $\theta = -\pi$.

Note that $f_{\Theta}(\theta)$ is differentiable for every $\theta \in (-\pi, +\pi)$. By direct computation, the first derivative of $f_{\Theta}(\theta)$ is given by

$$\begin{aligned} \frac{d}{d\theta} f_{\Theta}(\theta) &= \sqrt{\frac{1}{2\pi\sigma^2}} \frac{1}{(1 + \cos \theta)^2\sigma^2} e^{-\frac{1}{2\sigma^2} \left(\frac{\sin \theta}{1 + \cos \theta} - \mu \right)^2} \times \\ &\quad \left\{ \sigma^2 \sin \theta - \left(\frac{\sin \theta}{1 + \cos \theta} - \mu \right) \right\}. \end{aligned}$$

Therefore, we need to find conditions for which

$$\sigma^2 \sin \theta - \left(\frac{\sin \theta}{1 + \cos \theta} - \mu \right) = 0$$

only once for $\theta \in (-\pi, +\pi)$.

The case $\mu = 0$ is simpler and has been treated in [1], where authors proved that $f_{\Theta}(\theta|\mu, \sigma^2)$ is unimodal if and only if $\sigma^2 \leq \frac{1}{2}$. In this case, the maximum is achieved at $\theta = 0$ and the minimum at $\theta = -\pi$. The interpretation is that when the mass of the density is too spread out on the real line, when mapped to the circle, it tends to accumulate on itself and produces more than one peak.

The case $\mu \neq 0$ is more complicated and it is not present in the literature. Note that one solution always exist because the functions $f_1(\theta; \mu, \sigma^2) := \sigma^2 \sin \theta$ and $f_2(\theta; \mu, \sigma^2) := \frac{\sin \theta}{1 + \cos \theta} - \mu$ cross each others. In fact, $f_1(\theta; \mu, \sigma^2)$ is bounded in $(-\pi, \pi)$ while $f_2(\theta; \mu, \sigma^2)$ goes from $-\infty$ to $+\infty$, while θ goes from $-\pi$ to $+\pi$. We have the following theorem.

Theorem 2.6. *Consider the random variable $\Theta \sim \mathcal{ISN}(\mu, \sigma^2)$ with pdf $f_{\Theta}(\theta|\mu, \sigma^2)$ and parameters $\mu \neq 0$ and $\sigma^2 > 0$. Then, $f_{\Theta}(\theta|\mu, \sigma^2)$ is unimodal if and only if $\Delta(\mu, \sigma^2) < 0$, where $\Delta(\mu, \sigma^2)$ is given by*

$$\Delta(\mu, \sigma^2) = 18\mu^2(1 - 2\sigma^2) - \mu^4 + \mu^2(1 - 2\sigma^2)^2 - 4(1 - 2\sigma^2)^3 - 27\mu^2.$$

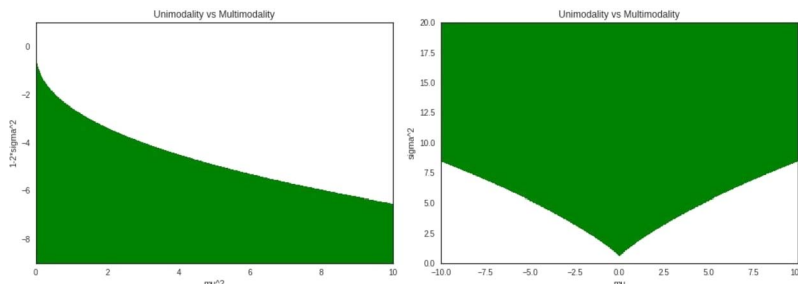


FIG 1. Unimodality region $\Delta(\mu, \sigma^2) < 0$, coloured in white. Left plot: $x = \mu^2$ and $y = 1 - 2\sigma^2$. Right plot: $x = \mu$ and $y = \sigma^2$. In both plots: $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$. Note that the value of μ does not represent the position of the mean, mode or median of the $ISND(\mu, \sigma^2)$, even if it does represent the position of the mean, mode and median of the corresponding $N(\mu, \sigma^2)$.

Proof. See Appendix A. □

Remark 2.7. Note that the condition on $\Delta(\mu, \sigma^2) < 0$ can be extremely simplified in the variables $x := \mu^2$ and $y := 1 - 2\sigma^2$. Indeed, $\tilde{\Delta}(x, y) := \Delta(\mu, \sigma^2) < 0$ becomes

$$18xy - x^2 + xy^2 - 4y^3 - 27x < 0.$$

with $x > 0$ and $y < 1$. This inequality is quadratic in x and so it can be solved easily by

$$\max \left\{ 0, \frac{1}{2} \left(y^2 + 18y - 27 - \sqrt{y^4 + 20y^3 + 270y^2 - 972y + 729} \right) \right\} < x < \frac{1}{2} \left(y^2 + 18y - 27 + \sqrt{y^4 + 20y^3 + 270y^2 - 972y + 729} \right),$$

for $y < 1$. Note that for $y < 1$, the argument of the square root is always positive. We refer to Figure 1 for a picture of the region $\tilde{\Delta}(x, y) = \Delta(\mu, \sigma^2) < 0$.

Remark 2.8. Note that for $\mu = 0$, the discriminant $\Delta(0, \sigma^2) = 4(2\sigma^2 - 1)^3$ is negative if and only if $0 \leq \sigma^2 < \frac{1}{2}$. When $\Delta(0, \sigma^2) = 0$, namely $\sigma^2 = \frac{1}{2}$, we have three identical solutions. Therefore, Theorem 2.6 in the case $\mu = 0$ is in agreement with the results of [1] (see also [7]).

It is important to recall that the inverse stereographic projection does not necessarily send critical points to critical points and so we cannot study the critical points of f_Θ by studying the critical points f_X and then transform them. This makes the analysis of unimodality non-trivial. It is interesting to notice that, in the symmetric case, the inverse stereographic projection does not change the parity of the set of critical points. Indeed, we have the following lemma.

Lemma 2.9. *Suppose f_X is a pdf on \mathbb{R} and f_Θ is a pdf on $[-\pi, +\pi)$ such that $\Theta = 2 \arctan(X)$ and such that both have a finite number of critical points. Moreover, suppose that $f_\Theta(\theta) = f_\Theta(-\theta)$ for every $\theta \in [-\pi, +\pi)$. Then, both $f_\Theta(\theta)$ and $f_X(x)$ have an odd number of critical points.*

Proof. Direct computation or direct consequence of the fact that the stereographic projection is a conformal transformation. \square

Example 2.1. For a random variable $\Theta \sim ISN(0, \sigma^2)$, the density $f_\Theta(\theta)$ has three critical points $0, \pm\pi$ for $\sigma^2 < 1/2$ and five $0, \pm\pi, \pm \arccos(\frac{1}{\sigma^2} - 1) \in [-\pi, +\pi)$ for $\sigma^2 \geq 1/2$. On the other hand, the corresponding random variable $X = \tan \frac{\Theta}{2} \sim N(0, \sigma^2)$ has always three critical points $-\infty, 0, +\infty$, with $x = -\infty, +\infty$ corresponding to $\theta = -\pi, +\pi$ and $x = 0$ corresponding to $\theta = 0$.

3. Comparison theorems: ISND vs Von Mises and Wrapped Normal

In this subsection, we compare the ISND with the VM and the WN. A classical argument to promote the use of the VM as a natural circular counterpart of the *Normal Distribution* is that, in the case of high-concentration limit ($\kappa \rightarrow +\infty$), the two distributions resemble each others

$$f_{VM}(\theta | 0, \kappa) = \frac{e^{\kappa \cos(\theta)}}{2\pi I_0(\kappa)} \simeq \sqrt{\frac{\kappa}{2\pi}} e^{\kappa(\cos \theta - 1)} \simeq \sqrt{\frac{\kappa}{2\pi}} e^{-\kappa\theta^2/2},$$

for $0 < \theta \ll 1$ or for $\kappa \gg 1$. This approximation is valid in every $L^p([-\pi, +\pi])$ -norm for $p \in [1, +\infty]$ and, as mentioned, does not uniquely identify the VM; in fact it is valid for several other distributions like the truncated normal and, indeed, it is valid for the ISND as well.

Theorem 3.1. *Consider the two distributions*

$$f_{VM}(\theta | 0, \kappa) = \frac{e^{\kappa \cos(\theta)}}{2\pi I_0(\kappa)}, \quad f_{ISN}(\theta | 0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{1 + \cos \theta} e^{-\frac{1}{2\sigma^2} \left(\frac{\sin \theta}{1 + \cos \theta}\right)^2},$$

with $\theta \in \mathbb{R}$ and $\kappa^{-1} = 4\sigma^2$. Then

$$\|f_{VM}(\theta | 0, \kappa) - f_{ISN}(\theta | 0, \sigma^2)\|_{L^\infty([-\pi, +\pi])} \rightarrow 0, \quad \text{as } \kappa \rightarrow +\infty.$$

Proof. See Appendix A. \square

The bivariate case for certain ranges of the parameters follows easily.

Theorem 3.2. *Consider the following BVM:*

$$f_{VM}(\phi, \psi) \propto \exp[\kappa_1[\cos(\phi) - 1] + \kappa_2[\cos(\psi) - 1] + (\cos(\phi) - 1, \sin(\phi))\mathbf{A}(\cos(\psi) - 1, \sin(\psi))^T],$$

defined for $\phi \in [-\pi, +\pi)$ and $\psi \in [-\pi, +\pi)$. Here, κ_1 and κ_2 are the concentration parameters, and the matrix $\mathbf{A} = [a_{11}, a; a, a_{22}] \in Mat_{2 \times 2}$ is the interaction matrix. Further, consider the BISND:

$$f_{BISN}(\phi, \psi) := \frac{1}{2\pi|\Sigma|^{1/2}} \frac{1}{1 + \cos \phi} \frac{1}{1 + \cos \psi} \times e^{-\frac{1}{2} \left(\frac{\sin \phi}{1 + \cos \phi}, \frac{\sin \psi}{1 + \cos \psi}\right)^T \Sigma^{-1} \left(\frac{\sin \phi}{1 + \cos \phi}, \frac{\sin \psi}{1 + \cos \psi}\right)},$$

defined for $\phi \in [-\pi, +\pi)$ and $\psi \in [-\pi, +\pi)$. Here, $\Sigma \in \text{Sym}_{2 \times 2}^+$ with $\Sigma^{-1} = [b_{11}, b; b, b_{22}]$. In the regime where $\kappa_1 = \frac{b_{11}}{4}$, $\kappa_2 = \frac{b_{22}}{4}$ and $a_{22} = -\frac{b}{4}$, the following holds:

$$\|f_{BISN}(\phi, \psi) - f_{VM}(\phi, \psi)\|_{L^\infty([-\pi, +\pi])} \rightarrow 0, \quad \kappa_1, \kappa_2 \rightarrow +\infty.$$

Proof. In the case where $b = 0$, the variables ϕ and ψ decouple and so the proof is a direct consequence of Theorem 3.1. The general case follows from a small modification of the proof of Theorem 3.1, that you can find in Appendix A. \square

Another consequence of Theorem 3.1 is that the ISND is close to the WN. This distribution is particularly relevant, as it is the unique solution to the heat equation on the circle and it is the only distribution preserving causality when used for kernel smoothing [12].

Theorem 3.3. Consider the two distributions

$$f_{WN}(\theta | 0, \sigma_{WN}^2) = \frac{1}{\sigma_{WN}\sqrt{2\pi}} \sum_{k=-\infty}^{+\infty} e^{-\frac{(\theta - 2\pi k)^2}{2\sigma_{WN}^2}}$$

and

$$f_{ISN}(\theta | 0, \sigma_{ISN}^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{1 + \cos \theta} e^{-\frac{1}{2\sigma_{ISN}^2} \left(\frac{\sin \theta}{1 + \cos \theta}\right)^2},$$

with $\theta \in \mathbb{R}$, $\kappa^{-1} = 4\sigma_{ISN}^2$ and $\sigma_{WN}^2 = A(\kappa)$, where $A(\kappa) := \frac{I_1(\kappa)}{I_0(\kappa)}$ and

$$I_p(\kappa) := \frac{1}{2\pi} \int_0^{2\pi} \cos(p\theta) e^{\kappa \cos \theta} d\theta$$

is the modified Bessel function of the first kind and order p . Then

$$\|f_{WN}(\theta | 0, \sigma_{WN}^2) - f_{ISN}(\theta | 0, \sigma_{ISN}^2)\|_{L^\infty([-\pi, +\pi])} \rightarrow 0, \quad \text{as } \kappa \rightarrow +\infty.$$

Proof. We have

$$\begin{aligned} & \|f_{WN}(\theta | 0, \sigma_{WN}^2) - f_{ISN}(\theta | 0, \sigma_{ISN}^2)\|_{L^\infty([-\pi, +\pi])} \leq \\ & \|f_{WN}(\theta | 0, \sigma_{WN}^2) - f_{VM}(\theta | 0, \kappa)\|_{L^\infty([-\pi, +\pi])} \\ & + \|f_{VM}(\theta | 0, \kappa) - f_{ISN}(\theta | 0, \sigma_{ISN}^2)\|_{L^\infty([-\pi, +\pi])}. \end{aligned}$$

As $\kappa \rightarrow +\infty$, the first summand goes to zero because of equations (3.5.23) and (3.5.24) on page 58 of [15], while the second summand goes to zero because of Theorem 3.1. \square

4. ISMs, ISMGF and a version of the Central Limit Theorem

Classical circular (trigonometric) moments (CMs) have been studied in the literature in the context of inverse stereographic distributions (see [30], [31], [33] and the references therein). There, the authors do not seem to consider the inverse

map and deal with classical CMs. As the results in [30], [31], [33] underline, CMs seem not natural. In fact the formulas for CMs derived in those papers are complicated functions of the parameters of their models. As a consequence, when related to the parameters, CMs lack in interpretability for those distributions. This motivates the search for a notion of moments which suits better inverse stereographic distributions.

We introduce the *Inverse Stereographic Moments* (ISMs) and compare them with CMs in Subsection 4.1. In Subsection 4.2, we introduce the *Inverse Stereographic Moment Generating Function* (ISMGF) and use ISMs to rephrase the Central Limit Theorem (CLT) in the context of \mathbb{T}^n for $n \geq 1$.

4.1. The inverse stereographic moments

The classical way in which *Moments on the Circle* are defined is the following.

Definition 4.1. Consider a random variable Θ defined on \mathbb{S}^1 with pdf given by $f_\Theta(\theta)$, for every $\theta \in [-\pi, +\pi)$. Then, the *Circular Moment (CM) of Θ of order k* with $k \in \mathbb{N}$ is defined as

$$m_k[\Theta] := E_\Theta[e^{ik\theta}] = \int_{[-\pi, +\pi)} f_\Theta(\theta)e^{ik\theta} d\theta.$$

This procedure has no connection with the extra structure coming from the *Stereographic Projection*, since the *Stereographic Projection* does not send the *Lebesgue Measure* on $[-\pi, +\pi)$ to the *Lebesgue Measure* on \mathbb{R} and also does not send polynomials defined on $\mathbb{C} \simeq \mathbb{R}^2$ to their restriction to \mathbb{S}^1 .

We propose to compute *Moments* in a way which is consistent with the *Stereographic Projection*, so that it is particularly suitable for inverse stereographic projected random variables defined on \mathbb{S}^1 . This procedure identifies in a geometrically natural way the circular counterpart of the moments defined in the *Euclidean Space* with respect to the *Lebesgue Measure*. A corresponding construction works in higher dimensions too.

Definition 4.2. Consider a random variable Θ defined on \mathbb{S}^1 with pdf $f_\Theta(\theta)$. Then, we define the *Inverse Stereographic Moment (ISM) of Θ of order k* with $k \in \mathbb{N}$ as

$$m_k^{\mathbb{S}^1}[\Theta] := E_\Theta \left[\left(\frac{\sin \theta}{1 + \cos \theta} \right)^k \right] := \int_{-\pi}^{+\pi} \left(\frac{\sin \theta}{1 + \cos \theta} \right)^k f_\Theta(\theta) d\theta.$$

Remark 4.3. Note that, by applying the *Stereographic Projection*, we get

$$m_k^{\mathbb{S}^1}[\Theta] = \int_{-\infty}^{+\infty} x^k f_\Theta(2 \arctan(x)) \frac{2dx}{1+x^2}.$$

Therefore, if $f_\Theta(\theta)$ comes from an *Inverse Stereographic Projection*, namely $\Theta = 2 \arctan(X)$, then

$$m_k^{\mathbb{S}^1}[\Theta] = \int_{-\infty}^{+\infty} x^k f_X(x) dx.$$

ISMs are definitely different from CMs, in particular in how they relate to the parameters of the underlying circular distribution. We give a small proposition on the comparison between CMs and ISMs for the ISND, but definitely the relation between CMs and ISMs deserve a deeper and more general investigation.

Proposition 4.4. *Consider a random variable $\Theta \sim ISN(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma^2 \geq 0$. We have:*

$$E_{\Theta}[\sin \theta] = E_{\Theta} \left[\frac{\sin \theta}{1 + \sin \theta} \right] \Leftrightarrow \mu = 0 \text{ or } \sigma^2 = 0.$$

Proof. We have that

$$E_{\Theta} \left[\frac{\sin \theta}{1 + \sin \theta} \right] = \mu \text{ and } E_{\Theta}[\sin \theta] = \mu \sqrt{\frac{\pi}{2\sigma^2}} e^{-\frac{1}{2\sigma^2}} \left(1 - \Phi(\sqrt{1/2\sigma^2}) \right),$$

where Φ is the error function (cdf of $N(0, 1)$). Therefore, the two moments are equal if and only if $\mu = 0$ or $\sigma^2 = 0$. \square

Note that $\mu = 0$ is an obvious solution of $E_{\Theta}[\sin \theta] = E_{\Theta}[\frac{\sin \theta}{1 + \sin \theta}]$, due to oddness of the integrands. The case $\sigma^2 = 0$ represents the degenerate case where $\Theta = c \in \mathbb{R}$ a.s wrt to the Lebesgue measure on \mathbb{R} .

A consequence of Proposition 4.4 is that a large ISM of order 2 affects a CM of order 1, but does not affect the ISM of order 1. This can be interpreted as some sort of orthogonality between the first and second ISMs and mimics the orthogonality of \bar{X} and S^2 for the MVN.

Corollary 4.5. *Consider a random variable $\Theta \sim ISN(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma \geq 0$. If*

$$Var^{S^1}(\Theta) := E_{\Theta} \left[\left(\frac{\sin \theta}{1 + \sin \theta} \right)^2 \right] - E_{\Theta}^2 \left[\frac{\sin \theta}{1 + \sin \theta} \right] = \sigma^2 \rightarrow +\infty,$$

then

$$E_{\Theta} \left[\frac{\sin \theta}{1 + \sin \theta} \right] = \mu,$$

independently of σ^2 , while

$$E_{\Theta}[\sin \theta] \rightarrow 0.$$

Proof. Direct computation plus the computations in the proof of Proposition 4.4. \square

Not all random variables defined on S^1 admit finite ISMs. For example, the *Uniform Distribution* on S^1 ($\Theta \sim Unif(-\pi, +\pi)$), which corresponds to the *Cauchy Distribution on the Real Line* through *Inverse Stereographic Projection*, does not admit any finite ISM, while all its CMs are finite. We have the following result.

Proposition 4.6. Consider a random variable $\Theta \in \mathbb{S}^1$ whose pdf $f_{\Theta}(\theta)$ is continuous and such that $f_{\Theta}(\theta) \sim (\theta - \pi)^{\alpha}$ for some $\alpha > 0$ when $\theta \sim +\pi$ and $f_{\Theta}(\theta) \sim (\theta + \pi)^{\beta}$ for some $\beta > 0$ when $\theta \sim -\pi$. Then

$$m_k^{\mathbb{S}^1}[\Theta] < +\infty \Leftrightarrow k \leq \min\{\alpha, \beta\}.$$

Proof. It is a direct consequence of the Taylor expansion of $\frac{\sin \theta}{1 + \cos \theta}$ at $+\pi$, where $\frac{\sin \theta}{1 + \cos \theta} \sim -\frac{2}{\theta - \pi}$, and at $-\pi$, where $\frac{\sin \theta}{1 + \cos \theta} \sim -\frac{2}{\theta + \pi}$. \square

The definition of ISMs can be easily extended to random variables defined on \mathbb{T}^n .

Definition 4.7. Consider a random variable $\Theta = (\Theta_1, \dots, \Theta_n) \in \mathbb{T}^n$ with pdf given by $f_{\Theta}(\theta)$ and $\theta := (\theta_1, \dots, \theta_n)$. Then, we define the *Inverse Stereographic Moment (ISM)* of Θ of order k with $k = (k_1, \dots, k_n) \in \mathbb{N}^n$ as

$$\begin{aligned} m_{(k_1, \dots, k_n)}^{\mathbb{T}^n}[\Theta] &:= E_{\Theta} \left[\left(\frac{\sin \theta_1}{1 + \sin \theta_1} \right)^{k_1} \times \dots \times \left(\frac{\sin \theta_n}{1 + \sin \theta_n} \right)^{k_n} \right] \\ &= \int_{[-\pi, +\pi]^n} \prod_{i=1}^n d\theta_i \left[\left(\frac{\sin \theta_1}{1 + \sin \theta_1} \right)^{k_1} \times \dots \times \left(\frac{\sin \theta_n}{1 + \sin \theta_n} \right)^{k_n} \right] f_{\Theta}(\theta). \end{aligned}$$

Example 4.1. With this definition, it is easy to see that the ISMs of a random variable distributed as $\mathcal{MLSN}(\mu, \Sigma)$ coincide with the standard moments of the MVN in \mathbb{R}^n and so

$$E^{\mathbb{T}^n}[\mathcal{MLSN}(\mu, \Sigma)] = \mu, \quad \text{Var}^{\mathbb{T}^n}[\mathcal{MLSN}(\mu, \Sigma)] = \Sigma.$$

Note that the interpretation is a little tricky because μ and Σ are not scale and location parameter for the variable Θ , but only for the variable X . In particular, it is easy to see that \mathcal{MLSN} 's family is not closed under re-location and re-scaling (affine transformations).

Remark 4.8. An important consequence of this definition of moments is that we can compute moments of distributions coming from *Inverse Stereographic Projections* analytically and by performing simple integrations on \mathbb{R}^n . This seems a big simplification for estimation purposes, since the estimation of parameters would not need numerical optimization algorithms, in general. Note that the simplification in the computation is directly related to how simple the computation of the moments is in the corresponding random variable on the Euclidean space.

4.2. A CLT and the ISMGF

The ISMGF is defined in the following way.

Definition 4.9. Consider a random variable $\Theta \in \mathbb{S}^1$ with pdf $f_\Theta(\theta)$. The *Inverse Stereographic Moment Generating Function* (ISMGF) is defined as

$$M_\Theta^{\mathbb{S}^1}(t) := E_\Theta \left[e^{t \tan \frac{\theta}{2}} \right] = \int_{-\pi}^{+\pi} e^{t \tan \frac{\theta}{2}} f_\Theta(\theta) d\theta,$$

for every $t \in I$, with $0 \in I \subset \mathbb{R}$.

Remark 4.10. Applying the *Stereographic Projection*, we get

$$M_\Theta^{\mathbb{S}^1}(t) = \int_{-\infty}^{+\infty} e^{tx} f_\Theta(2 \arctan(x)) \frac{2dx}{1+x^2}.$$

Therefore, if $f_\Theta(\theta)$ comes from an *Inverse Stereographic Projection*, namely $\Theta = 2 \arctan(X)$, then

$$M_\Theta^{\mathbb{S}^1}(t) = \int_{-\infty}^{+\infty} e^{tx} f_X(x) dx = M_X(t).$$

This makes the ISMGF a natural choice. We can construct an ISMGFs in higher dimensions as well. We concentrate on the case of \mathbb{T}^n .

Definition 4.11. Consider a random vector $\Theta = (\Theta_1, \dots, \Theta_n) \in \mathbb{T}^n$ with pdf given by $f_\Theta(\theta)$ and $\theta := (\theta_1, \dots, \theta_n)$. Then, we define the *Inverse Stereographic Moment Generating Function (ISMGF) on the Torus* as

$$M_\Theta^{\mathbb{T}^n}(t_1, \dots, t_n) := E_{\mathbb{T}^n} \left[e^{\sum_{i=1}^n t_i \tan \frac{\theta_i}{2}} \right] := \int_{[-\pi, +\pi]^n} \prod_{i=1}^n d\theta_i e^{\sum_{i=1}^n t_i \tan \frac{\theta_i}{2}} f_\Theta(\theta).$$

Here, $t_1, \dots, t_n \in I$, with $0 \in I \subset \mathbb{R}$.

We have the following theorem, which represents an “intrinsic” version of the CLT on \mathbb{S}^1 .

Theorem 4.12. Suppose $\{\Theta_i\}_{i \in \mathbb{N}} \subset \mathbb{S}^1$ is a sequence of iid random variables such that $\Theta_i = 2 \arctan(X_i)$ for every $i \in \mathbb{N}$ and some $\{X_i\}_{i \in \mathbb{N}} \subset \mathbb{R}$. Suppose $M_{\Theta_i}^{\mathbb{S}^1}(t)$ exists in a neighbourhood of $t = 0$ for every $i \in \mathbb{N}$. Further, suppose that $E^{\mathbb{S}^1}[\Theta_i] = \mu$ and $Var[\Theta_i] = \sigma^2 > 0$. Define

$$\Phi_n := P^{-1} \left(\frac{\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n P(\Theta_i) - \mu \right)}{\sigma} \right),$$

with P being the Stereographic Projection. Then,

$$\Phi_n \rightarrow \mathcal{ISN}(0, 1) \text{ as } n \rightarrow +\infty,$$

in the sense of the ISMGF and so in distribution.

This is not the most general version of the CLT possible and the theorem works in higher dimensions as well.

Proof. See Appendix A. □

Remark 4.13. The standard ISND is multimodal. However, we have that

$$\Phi_n^\tau := P^{-1} \left(\tau \frac{\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n P(\Theta_i) - \mu \right)}{\sigma} \right) \rightarrow \mathcal{ISN}(0, \tau^2).$$

Therefore, we can tune the parameter τ^2 so that $\tau^2 < 1/2$ and so the convergence in distribution is towards a unimodal ISND. This possibly surprising fact is a consequence of the fact that the stereographic projection (and in general any continuous function) does not preserve the number of modes. This connects to our previous discussion in Lemma 2.9 and Example 2.1.

5. Inference

In this section, we discuss some key inferential issues such as point estimation, confidence intervals and hypothesis testing and we briefly comment on some sampling methods for inverse stereographic probability distributions.

The first statistical problem that we address is parameter estimation.

Definition 5.1. Consider a *statistic* T_X on the Euclidean Space. Then, we call *Inverse Stereographic Statistic*, the following real-valued or vector-valued function: $T_\Theta := P^{-1} \circ T_X \circ P$, where P is the *Stereographic Projection*.

These estimators are natural for probability distributions defined through *Inverse Stereographic Projection*. Suppose we have a random sample $\Theta_i \sim \mathcal{ISN}(\mu, \sigma^2)$, for $i = 1, \dots, n$. We can define the *Inverse Stereographic Sample Mean* as

$$\bar{\Theta}^S := P^{-1} \left(\sum_{i=1}^n \frac{P(\Theta_i)}{n} \right),$$

where $\Theta_i \sim \mathcal{ISN}(\mu, \sigma^2)$, for $i = 1, \dots, n$. By the definition of P and since $\Theta_i \sim \mathcal{ISN}(\mu, \sigma^2)$, for $i = 1, \dots, n$, we have $P(\Theta_i) \sim N(\mu, \sigma^2)$, for $i = 1, \dots, n$, and so $\sum_{i=1}^n \frac{P(\Theta_i)}{n} \sim N(\mu, \sigma^2/n)$. Again, by the definition of P , we have $\bar{\Theta}^S \sim \mathcal{ISN}(\mu, \sigma^2/n)$. Note that similar considerations can be done for the *Sample Variance* and other estimators. $P(\bar{\Theta}^S)$ is a point estimator of μ . The MLEs can be computed very easily, because the extra term with respect to the likelihood of the MVN, appearing because of the pull-back measure, is parameter independent.

Theorem 5.2. Suppose $\Theta_1, \dots, \Theta_n$ is a random sample drawn from a random variable $\Theta \sim \mathcal{ISN}(\mu, \sigma^2)$ with μ and σ^2 unknown parameters. The MLEs of μ and σ^2 are given by:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \tan \left(\frac{\theta_i}{2} \right), \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \left[\tan \left(\frac{\theta_i}{2} \right) - \hat{\mu} \right]^2.$$

Proof. Direct computation. □

We can develop *Interval Estimation* for random variables distributed as an *Inverse Projected Distributions* as well.

Example 5.1. Suppose $\Theta \sim \mathcal{ISN}(\mu, \sigma^2)$ is a random variable defined on \mathbb{S}^1 with parameters μ unknown and σ^2 known. Let P be the *Stereographic Projection*. A $(1 - \alpha)$ -*Confidence Interval* for μ can be constructed as follows. We have

$$\begin{aligned} 1 - \alpha &= Pr(c \leq P(\bar{\Theta}^S) - \mu \leq d) = Pr(c + \mu \leq P(\bar{\Theta}^S) \leq d + \mu) \\ &= Pr\left(c + \mu \leq \frac{1}{n} \sum_{i=1}^n P(\Theta_i) \leq d + \mu\right) = Pr\left(\frac{c}{\sigma/\sqrt{n}} \leq Z \leq \frac{d}{\sigma/\sqrt{n}}\right) \\ &= \Phi(d\sqrt{n}/\sigma) - \Phi(c\sqrt{n}/\sigma), \end{aligned}$$

with $Z \sim N(0, 1)$ and Φ the error function. A possible choice is then $c = -d = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ with $z_{\alpha/2}$ the corresponding normal quantiles. Therefore,

$$C(\bar{\theta}^S) = C(\theta_1, \dots, \theta_n) := \left\{ \mu : P(\bar{\theta}^S) + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq P(\bar{\theta}^S) - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right\}$$

becomes an *Inverse Stereographic Interval Estimator* of μ . Since

$$P(C(\theta_1, \dots, \theta_n)) = 1 - \alpha$$

for some $\alpha \in [0, 1]$, then $C(\theta_1, \dots, \theta_n)$ is a $(1 - \alpha)$ -*Inverse Stereographic Confidence Interval* of μ .

Remark 5.3. Note that $\bar{\Theta}^S \sim \mathcal{ISN}(0, \sigma^2)$ if and only if $P(\bar{\Theta}^S) \sim N(0, \sigma^2)$ and so $C(\bar{\theta}^S)$ does not depend on P . This is an important property, because makes $C(\bar{\theta}^S)$ “intrinsic” of \mathbb{S}^1 , namely independent of the charts used on \mathbb{S}^1 .

Remark 5.4. There are two fundamental advantages in using this perspective: a theoretical and a practical one. From a theoretical point of view, everything is geometrically consistent and fully “intrinsic”. From a practical point of view, the estimators are explicit and their distributions can be computed as explicitly as they can be computed for the corresponding projected distribution. Therefore, there is no extra need for numerical optimization. The main negative feature of these estimators derives from the fact that the way in which they approach the true values of the parameters is counter-intuitive in the Θ variables, because it is measured with P^*dx . We will discuss this in a particular example in Section 6, but definitely it deserves further investigation. See Figure 2 below, where it is shown how confidence bands become much worse close to the cut-point.

We turn the discussion to *Hypothesis Tests*.

Example 5.2. Suppose $\Theta \sim \mathcal{ISN}(\mu, \sigma^2)$ is a random variable defined on \mathbb{S}^1 with μ unknown and σ^2 known. Let P be the *Stereographic Projection*. We test the following hypothesis:

$$H_0 : \mu = \mu_0 \text{ vs } H_a : \mu \neq \mu_0.$$

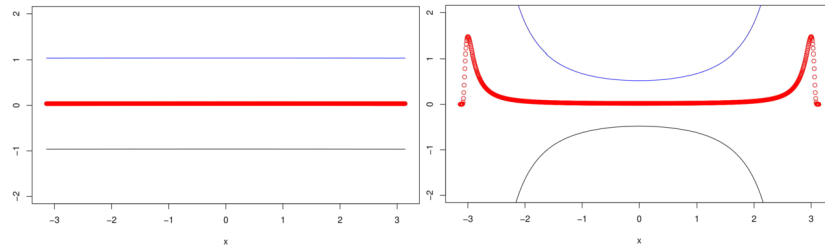


FIG 2. The plots illustrate how good confidence bands for the normal distribution (left plot) might become very bad for the ISND (right plot) when we approach the cut-point. In these plots, $\mu = 0$ and $\sigma^2 = 10$. This is actually a problem for the interpretability of the estimators, especially if peaks are concentrated close to the cut-point.

Consider a random sample $\Theta_1, \dots, \Theta_n \sim \mathcal{ISN}(\mu, \sigma^2)$. We can use the inversion theorem and build a Rejection Region from the Confidence Intervals described above. For a fixed level α , the *Most Powerful Unbiased Test* (see [4]) has rejection region

$$R(\theta) := \{\theta = (\theta_1, \dots, \theta_n) : |P(\bar{\theta}^S) - \mu| > z_{\alpha/2}\sigma/\sqrt{n}\}.$$

This test has size α if $P(H_0 \text{ is rejected} \mid \mu = \mu_0) = \alpha$.

It is worth mentioning that it is easy to sample from *Inverse Stereographic Projected Distributions*. The strategy is to sample from the corresponding Euclidean Distributions and then inverse stereographic project the sample to the circle or the torus. To accomplish this, we can simply use *Box-Muller Sampling* or other classical algorithms (see [4]).

6. Applications and numerical examples

We give some numerical examples and applications of the theory developed in the previous sections.

6.1. Plots

We collect some plots in order to visualize the ISND and BISND and the dependency on the parameters of the location, scale and shape of the distribution. The parameter μ has a mixed role: it behaves more as a location parameter for small values, but it affects skewness and concentration a lot more for higher values. We refer to Figure 3 up to Figure 9 for the plots and some more comments.

6.2. Comparison between the VM and the ISND

We use Theorem 3.1 to illustrate that for $\kappa \rightarrow +\infty$ the VM distribution is well approximated by the ISND. For $\kappa = 10$, the densities of the $ISN(0, (4\kappa)^{-1})$ and the $VM(0, \kappa)$ overlap (see Figure 10).

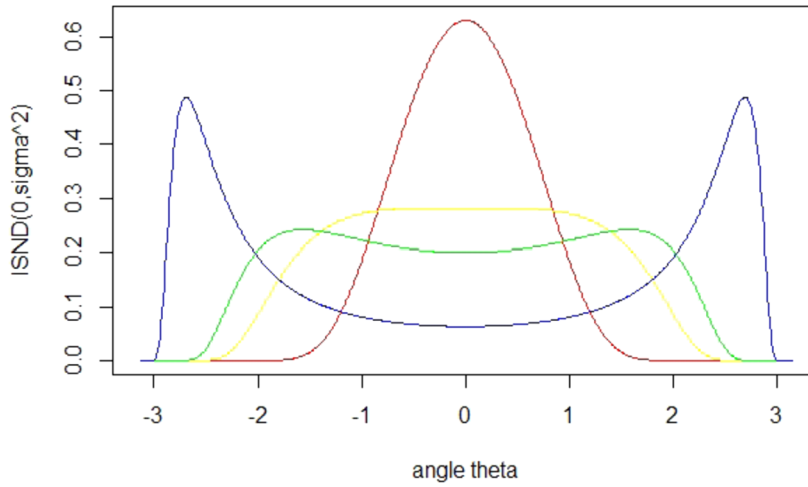


FIG 3. The plots show the $ISN(0, \sigma^2)$ coloured with levels of concentration in decreasing order: red $\sigma^2 = 0.1$, yellow $\sigma^2 = 0.5$, green $\sigma^2 = 1$ and blue $\sigma^2 = 10$). As demonstrated by the plots, the parameter $\mu = 0$ forces the distribution to be symmetric.

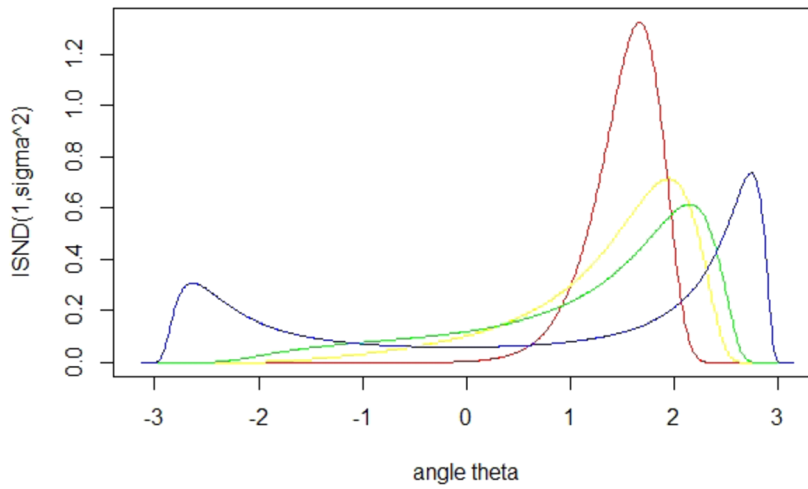


FIG 4. The plots show the $ISN(1, \sigma^2)$ coloured with levels of concentration in decreasing order: red $\sigma^2 = 0.1$, yellow $\sigma^2 = 0.5$, green $\sigma^2 = 1$ and blue $\sigma^2 = 10$. The introduction of the parameter $\mu \neq 0$ breaks the symmetry, increases the flexibility of the model and produces a higher variety of shapes. Note that when $\sigma^2 = 1$, the ISND was bimodal for $\mu = 0$, but it is unimodal for $\mu = 1$ in agreement with Theorem 2.6.

A similar comparison can be done in the bivariate case, using Theorem 3.2. If we analyze the contour plots of the $BISN(0, (4\kappa)^{-1})$ and the $BVM(0, \kappa)$ for $\kappa = 10$, the lines overlap (see Figure 11). Note that the VM is unimodal independently of its parameters μ and κ , while there are regimes where the

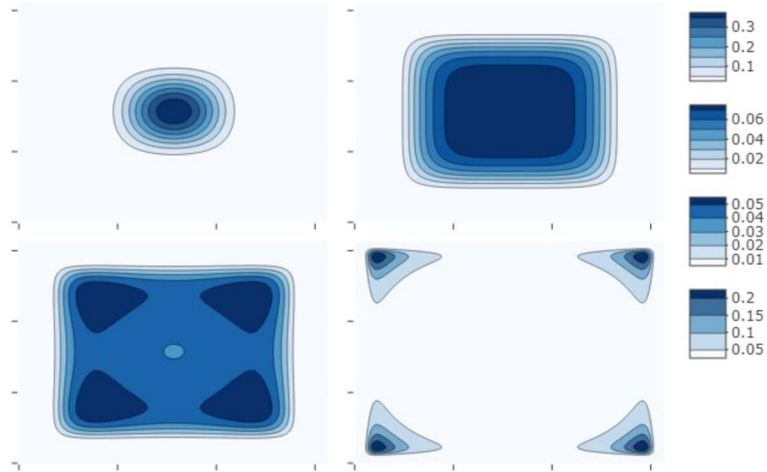


FIG 5. The plots show the uncorrelated $BLSN(0, \Sigma)$ with decreasing level of concentration. From the top, left to right, $\Sigma = 0.1 * Id$, $\Sigma = Id$, $\Sigma = 2 * Id$ and $\Sigma = 10 * Id$. We can see that the number of modes increases for less concentrated distributions and peaks move towards the cut-point. This is somehow a limitation of the model, in the sense that skewness and location are ruled by one single parameter: there is gain in flexibility of skewness with respect to the VM, but the change in location cannot be achieved without increasing skewness as it happens for the VM.

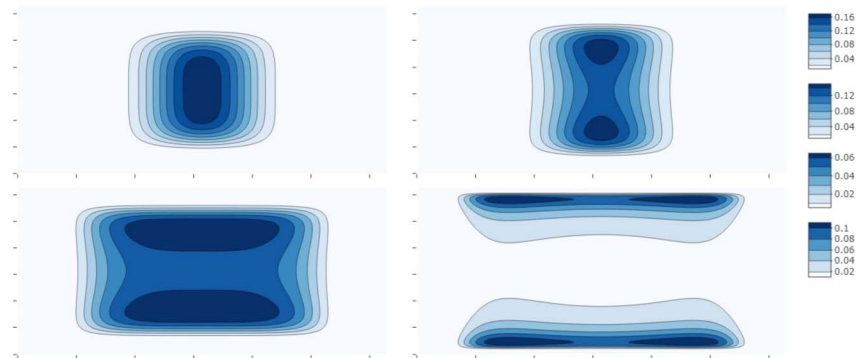


FIG 6. The plots show the uncorrelated $BLSN(0, \Sigma)$ with different marginal variances. From the top, left to right, $\Sigma = [0.1, 0; 0, 0.5]$, $\Sigma = [0.1, 0; 0, 1]$, $\Sigma = [0.5, 0; 0, 1]$ and $\Sigma = [1, 0; 0, 10]$.

ISND is bimodal (See Theorem 2.6). This of course happens in the case in which concentration is not high in agreement with the results of Section 3.

To overcome this lack of flexibility in the number of modes of the VM, people use mixtures of VM (see for example [26]). It is useful to know in which cases it is better to use a mixture of VM with the expenses of extra parameters and when it is enough to use a ISND, which is more parsimonious. Note that a limited number of parameters is often associated to a better fit of the model

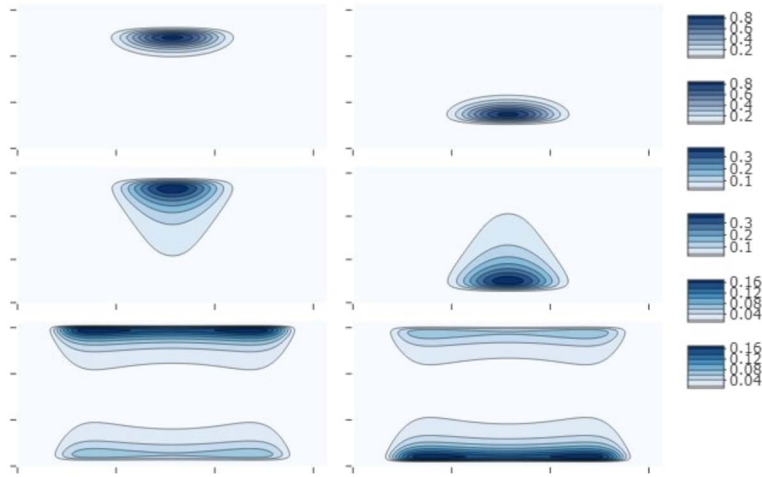


FIG 7. The plots show the uncorrelated $BLSN(\mu, \Sigma)$ with decreasing level of concentration and with parameter $\mu \neq 0$. From the top, left to right, $\Sigma = [0.1, 0; 0, 0.1]$ and $\mu = (0, 1)$, $\Sigma = [0.1, 0; 0, 0.1]$ and $\mu = (0, -1)$, $\Sigma = [0.1, 0; 0, 1]$ and $\mu = (0, 1)$, $\Sigma = [0.1, 0; 0, 1]$ and $\mu = (0, -1)$, $\Sigma = [1, 0; 0, 10]$ and $\mu = (0, 1)$, and $\Sigma = [1, 0; 0, 10]$ and $\mu = (0, -1)$.

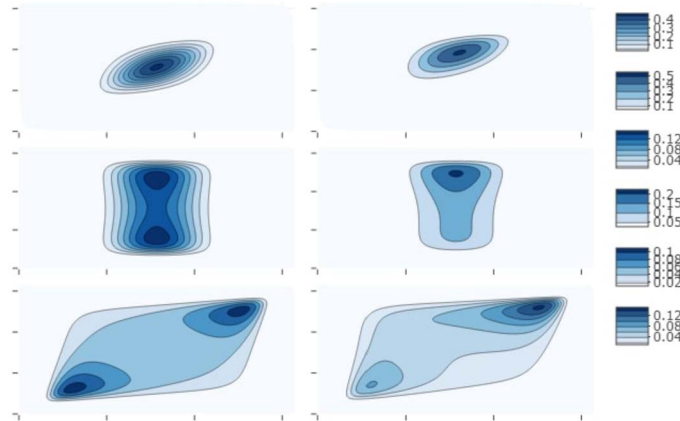


FIG 8. The plots show several cases of correlated and not correlated $BLSN(\mu, \Sigma)$ with different values of μ . The left column is centred at $\mu = (0, 0)$, while the right one is centred at $\mu = (0, 0.3)$. From top to bottom the variance matrices are $\Sigma = [0.1, 0.05; 0.05, 0.1]$, $\Sigma = [0.1, 0; 0, 0.1]$ and $\Sigma = [1, 0.5; 0.5, 1]$.

and a better understanding of the phenomenon under study. From Figure 12 and Figure 13, we note that there are some differences in the shape of the bimodal ISND and the mixture of VM. In the case in which peaks are far away, the mixture of VM keeps its symmetry around peaks, while the ISND becomes skewed towards the zero angle. In the case in which peaks are close by, the mixture of VM separates peaks more distinctly, while the ISND seems more

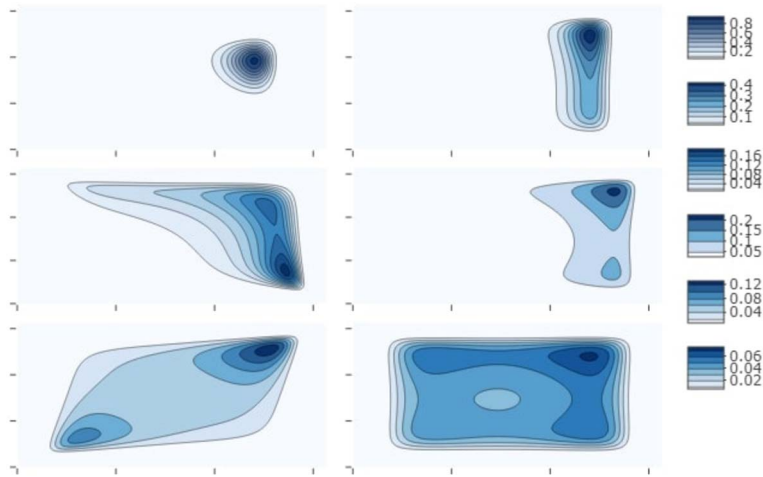


FIG 9. The plots show different correlated $BLSN(\mu, \Sigma)$ with varying values of μ and Σ . The different combinations produce a high variety of shapes and more complicated mode structures. With small μ , we see a translation of the peaks almost without skewness. With higher μ skewness becomes significant. Left column, top to bottom: $\Sigma = [0.1, 0; 0, 0.1]$ and $\mu = (1, 0.3)$, $\Sigma = [1, 0.5; 0.5, 1]$ and $\mu = (1, 0.3)$, and $\Sigma = [1, 0.5; 0.5, 1]$ and $\mu = (0.1, 0.1)$, Right column: $\Sigma = [0.1, 0; 0, 1]$ and $\mu = (1, 0.3)$, $\Sigma = [1, 0.05; 0.05, 1.5]$ and $\mu = (1, 0.3)$, and $\Sigma = [1, 0; 0, 1]$ and $\mu = (0.1, 0.1)$.

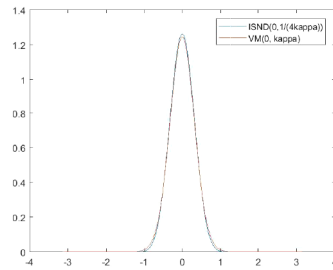


FIG 10. We plot the $ISN(0, (4\kappa)^{-1})$ and the $VM(0, \kappa)$ for $\kappa = 10$. The two curves are basically indistinguishable.

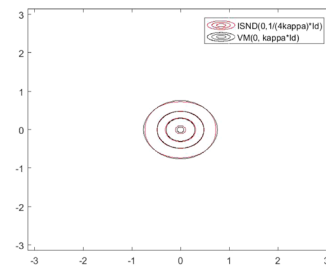
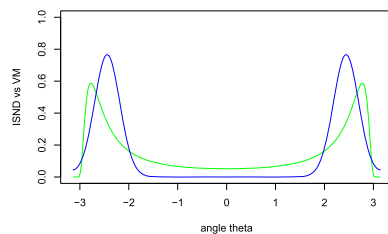
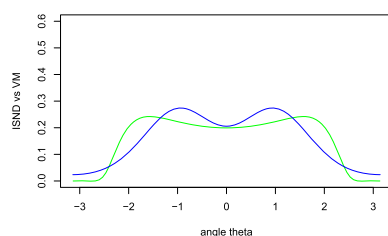


FIG 11. We plot the $BISN(0, (4\kappa)^{-1} * Id)$ and the $BVM(0, \kappa * Id)$ for $\kappa = 10$. The two level curves are basically indistinguishable.

FIG 12. *Bimodal ISND vs mixture of VM distribution with far away modes.*FIG 13. *Bimodal ISND vs mixture of VM distribution with nearby modes.*

suitable in the more complicated cases of almost indistinguishable peaks.

6.3. Applications to molecular biology

In this subsection, we consider some applications of our results to molecular biology.

To illustrate the distribution of the *dihedral angles* ϕ and θ (sometimes called *conformational angles* or *torsional angles*) in the protein main chain, people use the so called *Ramachandran map*. The *Ramachandran map* identifies a point in the protein main chain with a point on a flat square of the Euclidean plane \mathbb{R}^2 with opposite sides identified. From the mathematical point of view, the *Ramachandran map* represents the embedding of the *Flat Torus* into the Euclidean space \mathbb{R}^4 . It turns out to be a very useful starting point in the inference process. We analyze the Ramachandran Plots of three proteins and model the distributions of the dihedral angles with inverse stereographic distributions.

Myoglobin PDB=1MNB

The first dataset we consider contains the Dihedral Angles of Myoglobin from the Protein Database [32]. Myoglobin is a globular protein whose function is to store molecular oxygen in muscles. Its secondary structure consists mostly of α -helices, so it is expected that the Ramachandran plot shows a highly concentrated region.

To fit the ISND model, we did not use the MLE, because it was giving unintuitive results with lack of interpretability, possibly due to the fact that the

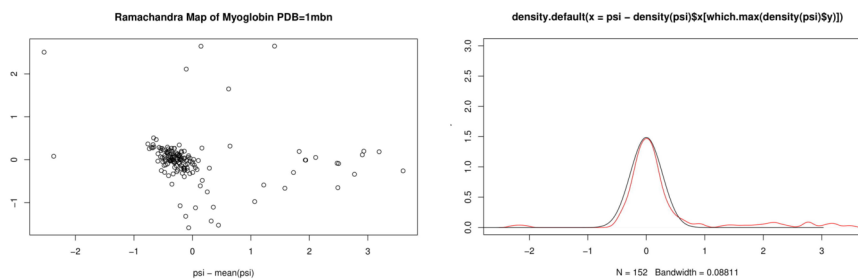


FIG 14. Ramachandran Plot of Myoglobin and associated density estimate of the marginal distribution of the dihedral angle Φ . The red line represent the empirical distribution, the black the ISND fit.

convergence are phrased in norms given in the variables $\tan(\Theta/2)$ and not Θ (See also Remark 5.4 and Figure 2). The estimation problem definitely needs further investigation. We decided to attack the problem with a semi-parametric approach. We centred the dihedral angle ψ so we could apply our Comparison Theorems from Section 3 and fit the parameter σ^2 with the value which was giving the ISND pdf closer to the empirical density in L^∞ -norm (which implies closeness in distribution), in agreement with the convergence results of Section 3 (See also Figure 2). We did also the following further simplification in the code. Reading through the lines of the proofs of the Comparison Theorems, one can see that outside the mode (the distribution of the ψ angle for 1MND is unimodal), there is no need to match the parameters of the distributions that we are comparing since both distributions approach zero heavily. Where the parameters matter the most is in the asymptotic shape of the mode. Therefore, it is reasonable to think that the fit is most likely good if the modes are close. Indeed, using this criterion, the fit seems pretty good as you can verify from Figure 14. The distance between the fitted distribution and the kernel density estimate of the empirical distribution is 0.01775444. This value has been found by a simple search over the parameter σ^2 with step size 10^{-3} . In terms of the mode structure, the model is also plausible.

Important Comment

The estimation in the Myoglobin PDB=1MNB example uncovers a very significant issue: since the MLE approach basically fits a normal distribution on the real line, outliers, which are projected very far out by the inverse stereographic map, lead to serious problems. A possible remedy could be found in trimmed estimators which are known to be more robust in such cases.

Gomesin PDB=1KFP (Marginals and Bivariate ISND)

The second dataset we consider contains the Dihedral Angles of Gomesin from the Protein Database [32]. Gomesin is a peptide isolated from the blood cells of a spider. The interest in analyzing the dataset is that the Ramachandran Plot of the dihedral angles of this protein show some skewness. Fitting the ISND as in

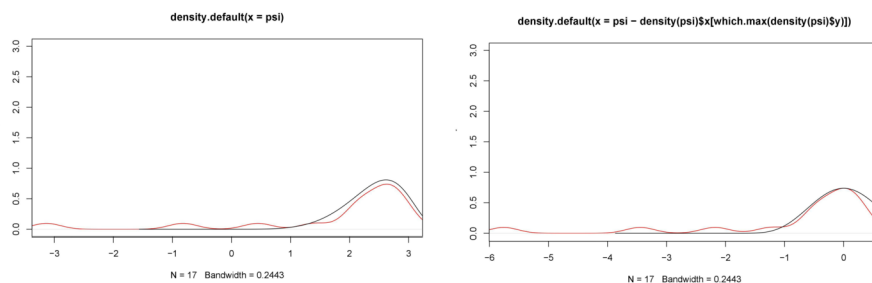


FIG 15. Comparison of the two parameters and one parameter fit of the marginal density estimate of the marginal distribution of the dihedral angle Ψ . The red line represent the empirical distribution, the black the ISND fit.

the previous example, give us the plot on the right in Figure 15 and a distance of 0.0006968093. However, we can see that the right tail is a little off and does not show a really good fit. To prevent this to happen, we searched for the best couple of parameters μ and σ^2 jointly. The plot on the left in Figure 15 confirms a better result and the distance in norm decreases to 2.409174×10^{-6} . This example shows definitely an improvement in the fit, when using ISND instead of VM in cases where the distribution of dihedral angles is significantly skewed.

For what concerns the joint distribution of the dihedral angles for the Gomesin, we fitted a BISND. The MLE for the various parameters with corresponding 95% confidence intervals are given by $\hat{\mu}_\phi = -1.096495$ ($-1.427, -0.766$), $\hat{\mu}_\psi = -0.946447$ ($-2.197, 0.304$), $\hat{\sigma}_\phi^2 = 4.289921$ ($3.467962, 5.444849$), $\hat{\sigma}_\psi^2 = 61.29062$ ($49.54718, 77.79121$) and $\hat{\sigma}_{\phi,\psi} = -0.946447$ ($-2.675504, 2.464707$). The original fit did not seem optimal, and heuristics in the choice of the cut-point did not seem to work. Therefore, we performed a “grid-search” over both angles in steps of $\pi/16$ and we obtained the fit in Figure 16 with cut-point at $P = (-9\pi/16, -14\pi/16)$. The BISND catches the skewness in the distribution of the dihedral angles, but seems confused by presence of an extra observation, to which it assigns an extra mode. This example shows both the the pros and the cons in the use of the BISND to fit data on \mathbb{T}^2 , namely the extra flexibility vs the problem of the choice of the cut-point. We refer to Section 7 for more comments on the choice of the cut-point. Note that as shown in Figure 16, the chain seems not long enough (153 *dihedral angles*) for a non-parametric method, like kernel density estimation, to take over. We fitted also the BISND using the criterion we used to fit the marginals. The distribution obtained fit better close to the mode, but missed completely to catch the dihedral angles in the zone $-\pi < \phi < 0$, $0 < \psi < \pi$, which is not desirable, since it is not a forbidden region in the Ramachandran Plot. Considering the sensitivity of the fit to the cut-point, we tried to estimate the parameters using robust statistics, like the median, the interquartile range and the MCD by Rouseeuw and Van Driessen [36]. The estimates for the various parameters give $\hat{\mu}_\phi = -1.134703$, $\hat{\mu}_\psi = -0.6737781$, $\hat{\sigma}_\phi = 0.1638183$, $\hat{\sigma}_\psi = 0.2821952$ and $\hat{\sigma}_{\phi,\psi} = 0.007492636$.

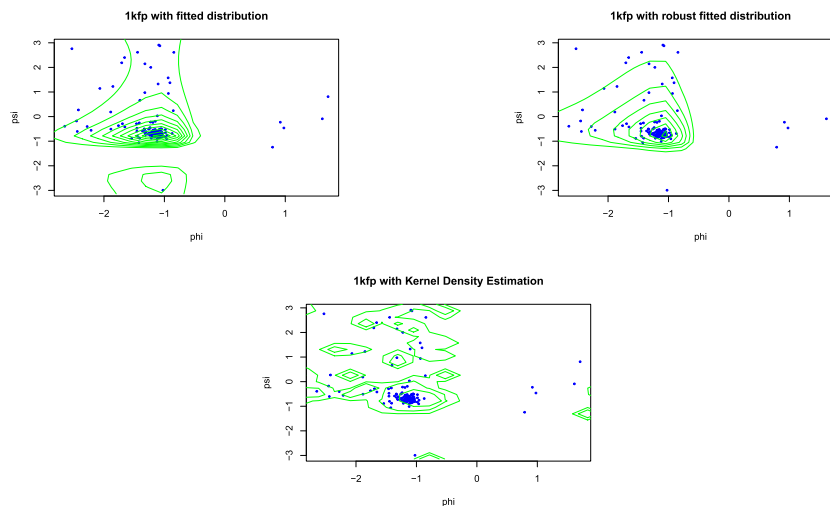


FIG 16. Two fitted BISND for the Gomesin 1KFP vs Kernel Density Estimation. BISND seems doing a better job in capturing some pattern, in particular if using robust statistics, while KDE shows overfitting.

Note that the bigger variation in the estimate is given by the marginal variances, which now, since more robust, are responsible of the unimodality of the fit (see Figure 16). We believe this last method gives the fit more representative of the distribution of the dihedral angles, even if the best goodness of fit test is produced by the non parsimonious KDE. Further investigation needs to be done on this issue.

Important Comment

In general situations, it is not easy to provide a good parametric model for distributions on manifolds, especially because the models available are much less than the ones available on the Euclidean space. For example, it is often hard to find a correct model if the VM, WN and ISND are not good models. However, with the next example, we outline a procedure which might be useful in such cases. Namely, we use the stereographic projection to send our data to the Euclidean space, we build a model there and we project it back to the manifold.

1208 Protein Structures from CDB

We consider a data set from the open access *Conformational Angles DataBase* (see [5]). It consists of 8190 *Conformation Angles* from 1208 PDB structures in 25% non-redundant protein chains. See Figure 17. The experiment method taken into consideration is the NMR (*Nuclear Magnetic Resonance*). The *Ramachandran Plot* for the *residue ALA* is shown in Figure 17. The *Shapiro-Wilk Normality Test* rejects beyond every reasonable doubt the hypothesis of normality on the projected data set and so we cannot conclude that these data are

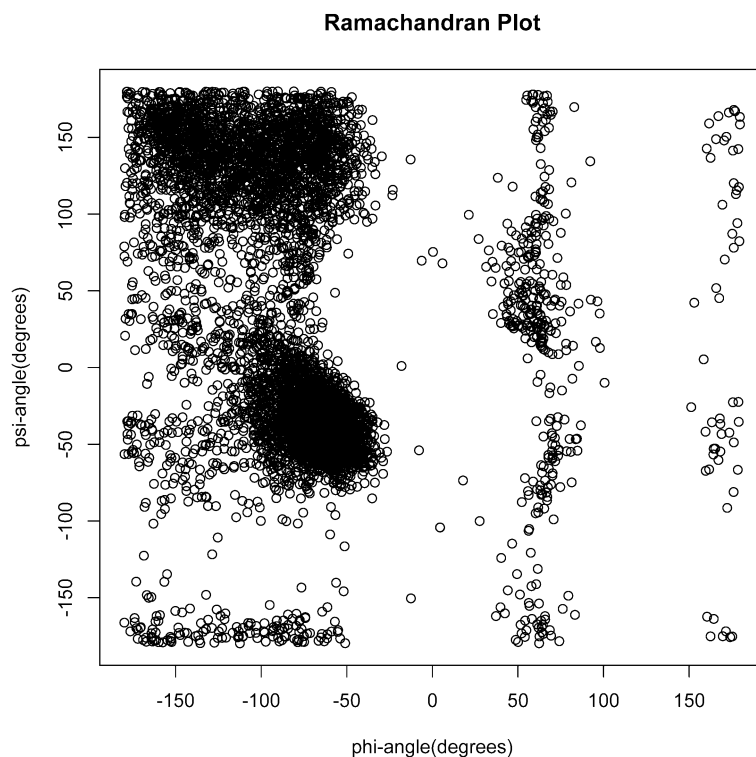


FIG 17. *Ramachandran Plot of the 8190 Conformation Angles from 1208 PDB structures in 25% non-redundant protein chains.*

distributed as an ISND (see Appendix B: Supplementary Material).

At this point, without taking advantage of the projected data set, it seems not easy to find a reasonable model which fits the data well. Let us see how the projected marginals behave under *Stereographic Projection* (see Figure 10). It seems reasonable to test if $-X \sim \exp(\lambda)$ and if $Y \sim \exp(\mu)$ for some $\lambda > 0$ and $\mu > 0$. To do this, we perform a *Kolmogorov-Smirnov Test*.

Remark 6.1. Although this hypothesis has been formulated after looking at the data, here we are not interested in the result of the test, but in proposing test procedures. A similar point of view has already been taken in [23] (See Section 5.3 of that paper).

From the *Kolmogorov-Smirnov Test* (see Appendix B), we cannot reject the *Null Hypothesis*, which does support our model. Note that the rejection region has a more intuitive shape in the projected variables.

From the *Ramachandran Plot*, it is clear that there are two peaks and that the data are clustered along two main major circles of the torus. Our model, even if it fits well, does not see these differences, which translates to some lack of power of our test. This is also unsatisfactory from a biological point of view,

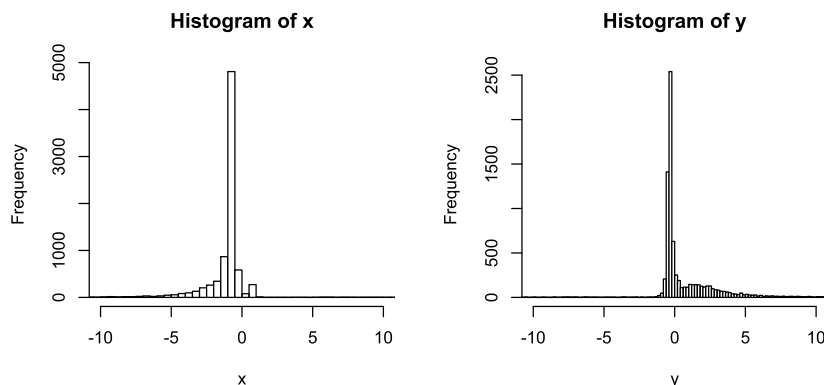


FIG 18. *Stereographic projection of the marginals of the distributions of the 8190 Conformation Angles.*

since the model is good only in the region of right handed α -helices, but is not sensitive enough in the area of left handed α -helices. However, this dataset contains 1208 PDB structures and so it would be probably too much requesting a better performance for a model with so few parameters.

This example shows a possible procedure to adopt in the cases where more common models, like VM or WN do not provide a good fit and/or it is inconvenient to increase too much the number of parameters (using mixtures, for example).

Remark 6.2. Using the novel hybrid QTAIM Ramachandran plots (see [27]), we think that the BISND could potentially be useful in modelling the glycine amino acid monomer, which largely occupies the “forbidden” regions of the Ramachandran plot.

Remark 6.3. VMs are successfully used in Projective Shape Analysis (see for example [9] and [24]). It is definitely important to see if the ISND family is a good model in that context as well. Because of the Comparison Theorems presented in Section 3, we expect similar performances in the high-concentration limit.

7. Discussion

In this section, we critically discuss the advantages and disadvantages in modelling distribution as *Inverse Stereographic Distributions*.

The cut-point

The construction of an inverse stereographic projected distribution might not seem completely satisfactory for the possible ambiguity left in the choice of the cut-point. Consider for simplicity the case of $n = 2$, namely $\Theta \in \mathbb{S}^1$, and let θ_0 be the North Pole of the corresponding *Stereographic Projection*. For generic $\alpha, \beta \in \mathbb{S}^1$, we have that $P_\Theta(\Theta \in [\alpha, \beta])$ depends on θ_0 . This does not seem a

desirable property. However, the choice of the North Pole corresponds to the choice of the point at infinity in the real line. Also that is arbitrary and “breaks the symmetry” of the real line. Very similarly, the choice of the North Pole “breaks the symmetry” of the circle. Analogously can be argued for the choice of the origin of a coordinate system of the real line. Note that the problem of the cut-point is strictly connected to the fact that the circle \mathbb{S}^1 is homeomorphic to the real projective line $P\mathbb{R}^1$, but not to the real line \mathbb{R} . This seems to us the intrinsic reason for which this problem cannot be fully resolved.

A richer family of distributions can be considered by introducing a translation parameter θ_0 , in which case the fit can be optimized over θ_0 . In this perspective, the corresponding family of normal distributions might be considered more naturally on $P\mathbb{R}^1$, with θ_0 corresponding to the point at infinity. The extra flexibility has been proven to be useful in the applications considered in Section 6.

Population and Sample ISMs

The definitions of ISMs turn out to be somehow independent of the choice of the *North Pole*. To fix the ideas, consider the case $n = 2$. The choice of the *North Pole* N is as arbitrary as the choice of the origin or the point at infinity in the real line. A line does not know anything about the system of coordinates that we put on. In the same way, the circle has no intrinsically well defined *North Pole*. However, we want that characteristic quantities like the expected value and the variance are as independent as possible with respect to the choice of the *North Pole*.

If we choose a North Pole $N' \neq N$, you can go from one *Stereographic Projection* to another by a simple change in the angles $\theta_i \mapsto \theta + \theta_i^0$ for $i = 1, \dots, n$. This choice of N' produces a new *Inverse Stereographic Projection* and so a new pdf $f_{\Theta}(\theta + \theta_0) \neq f_{\Theta}(\theta)$. The measure also changes consistently $\left(\frac{d\theta}{1+\cos\theta} \mapsto \frac{d\theta}{1+\cos(\theta+\theta_0)}\right)$, and so it compensates the change in the pdf.

This procedure of choosing a geometric definition of moments leaves every characteristic quantity to be invariant. There is still the dependence on the choice of the origin (or of the point at infinity) in the Euclidean Space and so the definition is not perfectly independent of any coordinate system. However, this sort of dependence on the coordinate system is the same as the Euclidean one, where it is well accepted and unavoidable.

If we pose our attention to the sample moments, instead of the population moments, we need to consider the possibility of data points crossing the cut-point. Note that the ISND approaches zero heavily in the neighbourhood of the cut-point and makes the crossing less likely to happen. Nevertheless, confidence bands around the cut-point can get extremely large, as shown in Figure 2. Distributions with positive probability in the neighbourhood of the cut-point do not have finite population moments, as shown in Proposition 4.6.

ISMs vs CMs

The canonical definition of CMs run into some problems due to the fact that they do not take into consideration the different geometry of the circle and torus.

Consider the sample mean of two angles. If the angles are $\theta_1 = -0.01$ and $\theta_2 = +0.01$, we have no problem to say that the mean is 0. Suppose, instead, the angles are $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. Then, it is not clear if it is more reasonable to say that the mean is 0 or π .

Knowing the mean is much more meaningful in the Euclidean setting than in a periodic setting, where $-\infty$ and $+\infty$ join together. The ISMs address this issue by “breaking the symmetry” of the circle and keeping $-\infty$ and $+\infty$ distinct. In this way, the problem mentioned above no longer exists. This procedure also suggests that it might be more intuitive to give an interpretation of parameters of circular/toroidal distributions after stereographic projection of the distribution, as we have done above.

Parameter Estimation vs cut-point

Another important issue which naturally arises from the above comments is how the parameter estimation is influenced by the choice of the cut-point.

Suppose we have a set of observations on \mathbb{S}^1 highly concentrated around a point θ and we believe the underlying probability distribution come from an inverse stereographic family. If we believe that unimodality needs to be highlighted in our model, it is not recommended to choose the cut-point in the middle of the cluster, because that would in fact risk to split the single mode into two modes highly concentrated around the cut-point. It is indeed more reasonable to choose the cut-point opposite to the centre of the cluster.

For real valued random variables, the parameter estimation is also influenced by the choice of the coordinate system, but in a more linear way and so, the influence is more intuitive, especially in the case of a normal distribution. The stereographic projection is nonlinear and so does not preserve linearity. As a consequence, different choices of cut-points lead to different parameter estimations.

8. Conclusion

In this paper, we have discussed several aspects of the ISND. We have derived some theoretical properties, like the unimodality conditions, and proved the Comparison Theorems with the VM and the WN. Then, we discussed the ISMs, the ISMGF and an intrinsic version of the CLT. We concluded with the illustrations, and some applications to molecular biology.

ISMs bring a novel perspective to the parametric statistics on manifolds and need to be investigated further. The performance of inverse stereographic estimators with respect to classical ones deserves future attention.

To verify the potential of ISND especially in higher dimensional models, further applications need to be tested. The possibility for the BISND to have more than two peaks might make the BISND flexible enough to describe the secondary structure of some of the proteins whose dihedral angles have multimodal distributions without the need of mixture models.

Appendix A: Proofs

In Appendix A, we give the proofs of the theorems presented in the paper.

Proof of Theorem 2.3

We give the proof just in the case $n = 2$, which is more interesting for the applications that we have in mind, but the proof for the general case follows in a similar way.

Suppose a random variable $\Theta := \{\Theta_1, \dots, \Theta_n\} \sim \mathcal{MISN}(\mu, \Sigma)$. Its pdf is given by

$$f_{\Theta}(\theta) := (2\pi)^{-n/2} |\Sigma|^{-1/2} \prod_{i=1}^n \frac{1}{1 + \cos \theta_i} \times e^{-\frac{1}{2} \left(\frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1, \dots, \frac{\sin \theta_n}{1 + \cos \theta_n} - \mu_n \right)^T \Sigma^{-1} \left(\frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1, \dots, \frac{\sin \theta_n}{1 + \cos \theta_n} - \mu_n \right)},$$

for some $\mu := (\mu_1, \dots, \mu_n) \in \mathbb{R}^n$, some $\Sigma \in \text{Sym}_n^+$ and for any $\theta = (\theta_1, \dots, \theta_n) \in [-\pi, +\pi]^n$. In the case $n = 2$, we have:

$$f_{\Theta_1, \Theta_2}(\theta_1, \theta_2) := \frac{1}{2\pi |\Sigma|^{1/2}} \frac{1}{(1 + \cos \theta_1)} \frac{1}{(1 + \cos \theta_2)} \times e^{-\frac{1}{2} \left[\frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1, \frac{\sin \theta_2}{1 + \cos \theta_2} - \mu_2 \right]^T \Sigma^{-1} \left[\frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1, \frac{\sin \theta_2}{1 + \cos \theta_2} - \mu_2 \right]}.$$

Therefore,

$$f_{\Theta_1}(\theta_1) = \int_{-\pi}^{+\pi} d\theta_2 f_{\Theta_1, \Theta_2}(\theta_1, \theta_2) = \int_{-\pi}^{+\pi} d\theta_2 \frac{1}{2\pi |\Sigma|^{1/2}} \frac{1}{(1 + \cos \theta_1)} \frac{1}{(1 + \cos \theta_2)} \times e^{-\frac{1}{2} \left[\frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1, \frac{\sin \theta_2}{1 + \cos \theta_2} - \mu_2 \right]^T \Sigma^{-1} \left[\frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1, \frac{\sin \theta_2}{1 + \cos \theta_2} - \mu_2 \right]}.$$

Now, define $B := \Sigma^{-1}$ and so $B_{ij} = (\Sigma^{-1})_{ij}$ for $j = 1, 2$ and $2b = b_{12} + b_{21}$. This implies that

$$f_{\Theta_1}(\theta_1) = \frac{1}{2\pi |\Sigma|^{1/2}} \frac{e^{-\frac{b_{11}}{2} \left\{ \frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1 \right\}^2}}{(1 + \cos \theta_1)} \times \int_{-\pi}^{+\pi} \frac{d\theta_2}{(1 + \cos \theta_2)} e^{-\frac{b_{22}}{2} \left\{ \frac{\sin \theta_2}{1 + \cos \theta_2} - \mu_2 \right\}^2} e^{-\frac{2b}{2} \left\{ \frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1 \right\} \left\{ \frac{\sin \theta_2}{1 + \cos \theta_2} - \mu_2 \right\}}.$$

Now, we do the following change of variables:

$$z := \frac{\sin \theta_2}{1 + \cos \theta_2} - \mu_2, \quad dz = \frac{d\theta_2}{1 + \cos \theta_2}.$$

Then, we obtain:

$$f_{\Theta_1}(\theta_1) = \frac{1}{2\pi |\Sigma|^{1/2}} \frac{e^{-\frac{b_{11}}{2} \left\{ \frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1 \right\}^2}}{(1 + \cos \theta_1)} \int_{-\infty}^{+\infty} dz e^{-\frac{1}{2} [2b\alpha(\theta_1)z + b_{22}z^2]},$$

with $\alpha(\theta_1) := \frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1$. Upon completing the square, we get

$$f_{\Theta_1}(\theta_1) = \frac{1}{2\pi|\Sigma|^{1/2}} \frac{e^{-\frac{b_{11}}{2} \left\{ \frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1 \right\}^2} e^{+\frac{b^2}{2b_{22}} \left\{ \frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1 \right\}^2}}{(1 + \cos \theta_1)} \times \int_{-\infty}^{+\infty} dz e^{-\frac{1}{2} \left[b_{22}z + \frac{b\alpha(\theta_1)}{\sqrt{b_{22}}} \right]^2}.$$

By performing another change of variables

$$x := b_{22}z + \frac{b\alpha(\theta_1)}{\sqrt{b_{22}}}, \quad dx = b_{22}dz,$$

we get

$$\begin{aligned} f_{\Theta_1}(\theta_1) &= \frac{1}{2\pi|\Sigma|^{1/2}\sqrt{b_{22}}} \frac{e^{-\frac{b_{11}}{2} \left\{ \frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1 \right\}^2} e^{+\frac{b^2}{2b_{22}} \left\{ \frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1 \right\}^2}}{(1 + \cos \theta_1)} \int_{-\infty}^{+\infty} dx e^{-\frac{1}{2}x^2} \\ &= \frac{\sqrt{b_{22}b_{11} - b^2}}{\sqrt{2\pi}\sqrt{b_{22}}} \frac{e^{-\frac{1}{2} \frac{b_{11}b_{22} - b^2}{b_{22}} \left\{ \frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1 \right\}^2}}{(1 + \cos \theta_1)} = \frac{1}{\sqrt{2\pi\Sigma_{11}}} \frac{e^{-\frac{1}{2\Sigma_{11}} \left\{ \frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1 \right\}^2}}{(1 + \cos \theta_1)}. \end{aligned}$$

Since

$$f_{\Theta_1}(\theta_1) = \frac{1}{\sqrt{2\pi\Sigma_{11}}} \frac{e^{-\frac{1}{2\Sigma_{11}} \left\{ \frac{\sin \theta_1}{1 + \cos \theta_1} - \mu_1 \right\}^2}}{(1 + \cos \theta_1)},$$

then $\Theta_1 \sim ISN(\mu_1, \Sigma_{11})$. Thus, the family of *MISND* is closed under marginalization, as required.

Proof of Theorem 2.6

We need to solve the equation

$$\sigma^2 \sin \theta - \left(\frac{\sin \theta}{1 + \cos \theta} - \mu \right) = 0$$

for $\mu \in \mathbb{R}$, $\mu \neq 0$ and $\sigma^2 > 0$ and deduce the conditions on μ, σ^2 for which it has only a single solution for $\theta \in (-\pi, +\pi)$. We have already discussed the case $\theta = -\pi$, and so we can assume $\theta \in (-\pi, +\pi)$. In this domain the transformation $t = \tan(\theta/2)$ is a diffeomorphism, so we can use the parametric equations

$$\sin \theta = \frac{2t}{1 + t^2}, \quad \cos \theta = \frac{1 - t^2}{1 + t^2},$$

(there is a one-to-one correspondence between the solutions in θ and the solutions in t). We get:

$$\sigma^2 \frac{2t}{1 + t^2} - \left(\frac{\frac{2t}{1 + t^2}}{1 + \frac{1 - t^2}{1 + t^2}} - \mu \right) = 0,$$

which simplifying becomes

$$\sigma^2 \frac{2t}{1+t^2} - t + \mu = 0$$

and so

$$\sigma^2 2t - t(1+t^2) + \mu(1+t^2) = 0$$

or better

$$t^3 - \mu t^2 + (1 - 2\sigma^2)t - \mu = 0.$$

We want to find conditions on μ and σ^2 so that this polynomial admits only one real solution in t . For this, we need to examine the *Discriminant of the Polynomial*. We recall the following lemma.

Lemma A.1. *Consider the polynomial*

$$P[t; a, b, c, d] := at^3 + bt^2 + ct + d = 0$$

with $a \neq 0$ and $a, b, c, d \in \mathbb{R}$. Equation $P[t; a, b, c, d] = 0$ has at least one real solution t_1 . Moreover, let

$$\Delta := 18abcd - 4b^3d + b^2c^2 - 4ac^3 - 27a^2d^2$$

be the Discriminant of $P[t; a, b, c, d]$. Then:

- If $\Delta > 0$, $P[t; a, b, c, d]$ has three distinct real roots;
- If $\Delta = 0$, $P[t; a, b, c, d]$ has multiple roots and all the roots are real;
- If $\Delta < 0$, $P[t; a, b, c, d]$ has one real root and two complex conjugate roots.

We use this lemma for the polynomial

$$P[t; a, b, c, d] = t^3 - \mu t^2 + (1 - 2\sigma^2)t - \mu.$$

In the notation of Lemma A.1, we have

$$a = 1, \quad b = -\mu, \quad c = 1 - 2\sigma^2, \quad d = -\mu.$$

Therefore, the discriminant $\Delta = \Delta(\mu, \sigma^2)$ in this case becomes

$$\Delta(\mu, \sigma^2) = 18\mu^2(1 - 2\sigma^2) - \mu^4 + \mu^2(1 - 2\sigma^2)^2 - 4(1 - 2\sigma^2)^3 - 27\mu^2.$$

Note that for $\mu = 0$, we recover the case already treated above and in [1]. Lemma A.1 tells us that, when $\Delta(\mu, \sigma^2) < 0$, we have one single root in t , and so we have unimodality in t and so in θ , and that, when $\Delta(\mu, \sigma^2) > 0$, we have three distinct roots in t and hence we have multimodality in t and so in θ . The proof of the theorem is not complete yet, because it is still possible that for $\Delta(\mu, \sigma^2) = 0$, there are three identical roots. Note that since our polynomial is monic, then to have three identical roots $\beta \in \mathbb{R}$ it needs to be of the form:

$$t^3 - 3\beta t^2 + 3\beta^2 t - \beta^3 = 0.$$

This implies the following conditions on μ and σ^2 :

$$-\mu = -3\beta, \quad 1 - 2\sigma^2 = 3\beta^2, \quad -\mu = -\beta^3.$$

Note that the first and third condition together imply either $\beta = 0$ and so $\mu = 0$ and $\sigma^2 = \frac{1}{2}$ which has been treated before, or $\beta^2 = 3$. But $\beta^2 = 3$ together with the second condition implies $1 - 2\sigma^2 = 9$ which is a contradiction. This completes the proof of the theorem.

Remark A.2. The solutions of this system can be calculated explicitly and pretty easily using some symbolic software like Maple or Wolfram Alpha. However, they have a complicated form and do not add any extra insight. Therefore, we have decided to not report them in the paper.

Proofs of Comparisons Theorems

Proof of Theorem 3.1 We have for $K := [-\kappa^{-1/3}, \kappa^{-1/3}]$

$$\begin{aligned} & \|f_{VM}(\theta | 0, \kappa) - f_{ISN}(\theta | 0, \sigma^2)\|_{L^\infty([- \pi, +\pi])} = \\ & \max \left\{ \|f_{VM}(\theta | 0, \kappa) - f_{ISN}(\theta | 0, \sigma^2)\|_{L^\infty(K)}, \right. \\ & \left. \|f_{VM}(\theta | 0, \kappa) - f_{ISN}(\theta | 0, \sigma^2)\|_{L^\infty([- \pi, +\pi] \setminus K)} \right\}. \end{aligned}$$

Consider the first argument of the max. We get (for $\kappa \gg 1$):

$$\begin{aligned} & \left\| \frac{e^{\kappa \cos(\theta)}}{2\pi I_0(\kappa)} - \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{1 + \cos \theta} e^{-\frac{1}{2\sigma^2} \left(\frac{\sin \theta}{1 + \cos \theta}\right)^2} \right\|_{L^\infty(K)} \leq \\ & \sup_{|\theta| \leq \kappa^{-1/3}} \left| \sqrt{\frac{\kappa}{2\pi}} \left(e^{\kappa(\cos \theta - 1)} - e^{-2\kappa \left(\frac{\sin \theta}{1 + \cos \theta}\right)^2} \right) \right| = \\ & \sup_{|\theta| \leq \kappa^{-1/3}} \sqrt{\frac{\kappa}{2\pi}} e^{-2\kappa \left(\frac{\sin \theta}{1 + \cos \theta}\right)^2} \left| \left(e^{\kappa(\cos \theta - 1 - 2\left(\frac{\sin \theta}{1 + \cos \theta}\right)^2)} - 1 \right) \right| \leq \\ & \sup_{|\theta| \leq \kappa^{-1/3}} \sqrt{\frac{\kappa}{2\pi}} e^{-\kappa \frac{\theta^2}{2}} * \sup_{|\theta| \leq \kappa^{-1/3}} \left| \left(e^{\kappa(\cos \theta - 1 - \theta^2/2)} - e^{\kappa \frac{\theta^2}{2} - 2\kappa \left(\frac{\sin \theta}{1 + \cos \theta}\right)^2} \right) \right| \\ & \leq C\kappa^{-1/3} \rightarrow 0, \text{ as } \kappa \rightarrow +\infty. \end{aligned}$$

for some $C > 0$ by Lagrange remainder's theorem.

Consider the second argument of the max. We get (for $\kappa \gg 1$):

$$\begin{aligned} & \left\| \frac{e^{\kappa \cos(\theta)}}{2\pi I_0(\kappa)} - \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{1 + \cos \theta} e^{-\frac{1}{2\sigma^2} \left(\frac{\sin \theta}{1 + \cos \theta}\right)^2} \right\|_{L^\infty([- \pi, +\pi] \setminus [-\kappa^{-1/3}, \kappa^{-1/3}])} = \\ & \left\| \frac{e^{\kappa \cos(\theta)}}{2\pi I_0(\kappa)} - \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{1 + \cos \theta} e^{-\frac{1}{2\sigma^2} \left(\frac{\sin \theta}{1 + \cos \theta}\right)^2} \right\|_{L^\infty([\kappa^{-1/3}, +\pi])} \leq \end{aligned}$$

$$\begin{aligned} & \left\| \frac{e^{\kappa \cos(\theta)}}{2\pi I_0(\kappa)} \right\|_{L^\infty([\kappa^{-1/3}, +\pi))} + \left\| \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{1 + \cos \theta} e^{-\frac{1}{2\sigma^2} \left(\frac{\sin \theta}{1 + \cos \theta}\right)^2} \right\|_{L^\infty([\kappa^{-1/3}, +\pi))} \\ & \simeq \frac{1}{\sqrt{2\pi/\kappa}} e^{\kappa(1-\kappa^{-2/3}-1)} + \frac{1}{\sqrt{2\pi/\kappa}} e^{-\kappa^{1/3}/2} \rightarrow 0, \text{ as } \kappa \rightarrow 0, \end{aligned}$$

where in the last inequality we used the fact that the function

$$\frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{1 + \cos \theta} e^{-\frac{1}{2\sigma^2} \left(\frac{\sin \theta}{1 + \cos \theta}\right)^2}$$

is monotone decreasing in θ . This completes the proof of the theorem.

Remark A.3. Due to the extra parameters of the BVM, there are several possible regimes in which the BVM distribution and BISND are close to each other.

With the choice of the parameters in Theorem 3.2, the BVM and BISND match at second order near the origin. If we want the two families to match at cubic order near the origin, we need $a = 0$ (which is the coefficient of the term $-\theta_1\theta_2^2 - \theta_2\theta_1^3$). It is not possible for the two families to match at quartic order since that would require $\kappa_1 = -12b_{11}$ (coefficients of θ_1^4) and $\kappa_2 = -12b_{22}$ (coefficients of θ_2^4) which would be in contradiction with the matching conditions at second order $\kappa_1 = \frac{b_{11}}{4}$ and $\kappa_2 = \frac{b_{22}}{4}$.

Proof of Theorem 4.12

By construction of the ISMGF, we have

$$M_{\Theta_i}^{\mathbb{S}^1}(t) = M_{P(\Theta_i)}(t),$$

for every $i \in \mathbb{N}$. This implies that

$$E^{\mathbb{S}^1}[\Theta_i] = \mu, \quad Var^{\mathbb{S}^1}[\Theta_i] = \sigma^2.$$

By definition of ISMGF, we obtain

$$M_{\Phi_n}^{\mathbb{S}^1}(t) = M_{P^{-1}\left(\frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n P(\Theta_i) - \mu\right)}{\sigma}\right)}^{\mathbb{S}^1}(t) = M_{\frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n P(\Theta_i) - \mu\right)}{\sigma}}(t).$$

The random variable $\frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n P(\Theta_i) - \mu\right)}{\sigma}$ satisfies all the hypotheses of the classical CLT. Moreover, independence is preserved by the *Stereographic Projection*, as it can be seen by the following chain of equalities:

$$\begin{aligned} f_{\Theta_1, \Theta_2}(\theta_1, \theta_2) &= f_{X_1, X_2} \left(\frac{\sin \theta_1}{1 + \cos \theta_1}, \frac{\sin \theta_2}{1 + \cos \theta_2} \right) \frac{d\theta_1}{1 + \cos \theta_1} \frac{d\theta_2}{1 + \cos \theta_2} \\ &= f_{X_1} \left(\frac{\sin \theta_1}{1 + \cos \theta_1} \right) f_{X_2} \left(\frac{\sin \theta_2}{1 + \cos \theta_2} \right) \frac{d\theta_1}{1 + \cos \theta_1} \frac{d\theta_2}{1 + \cos \theta_2} = f_{\Theta_1}(\theta_1) f_{\Theta_2}(\theta_2). \end{aligned}$$

Therefore,

$$\frac{\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n P(\Theta_i) - \mu \right)}{\sigma} \rightarrow X \sim \mathcal{N}(0, 1)$$

in distribution. By the continuous mapping theorem,

$$g \left(\frac{\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n P(\Theta_i) - \mu \right)}{\sigma} \right) \rightarrow g(X)$$

for every continuous g . We have that

$$M_{\frac{\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n P(\Theta_i) - \mu \right)}{\sigma}} \rightarrow M_{\mathcal{N}(0,1)} = M_{P^{-1}(\mathcal{N}(0,1))}(t) = M_{\mathcal{ISN}(0,1)}(t)$$

and so that

$$M_{\Phi_n^{\mathbb{S}^1}}(t) \rightarrow M_{\mathcal{ISN}(0,1)}(t) \text{ as } n \rightarrow +\infty.$$

Also

$$\Phi_n \rightarrow \mathcal{ISN}(0, 1) \text{ as } n \rightarrow +\infty,$$

by the continuous mapping theorem with $g = P^{-1}$. This concludes the proof of the theorem.

Appendix B: Supplementary material

We provide several of the codes used in the manuscript. The missing ones are simple adaptations of the ones presented here.

Here is the code for the Unimodality Plot

```
#mu^2 vs 1-2*sigma^2
import numpy as np
import matplotlib.pyplot as plt
xlist = np.linspace(0, 10, 500)
ylist = np.linspace(-9, 1, 500)
X, Y = np.meshgrid(xlist, ylist)
Z = np.sqrt(18*X*Y-X**2+X*Y**2-4*Y**3-27*X)
plt.figure()
contour = plt.contour(X, Y, Z, levels=[0])
contour_filled = plt.contourf(X, Y, Z, levels, colors=['g'])
plt.title('Unimodality vs Multimodality')
plt.xlabel('mu^2')
plt.ylabel('1-2*sigma^2')
plt.show()
#mu vs sigma^2
import numpy as np
import matplotlib.pyplot as plt
xlist = np.linspace(-10, 10, 500)
ylist = np.linspace(0, 20, 500)
```

```

X, Y = np.meshgrid(xlist, ylist)
Z = np.sqrt(18*X**2*(1-2*Y)-X**4+X**2*(1-2*Y)**2
-4*(1-2*Y)**3-27*X**2)
plt.figure()
contour = plt.contour(X, Y, Z, levels=[0])
contour_filled = plt.contourf(X, Y, Z, levels, colors=['g'])
plt.title('Unimodality vs Multimodality')
plt.xlabel('mu')
plt.ylabel('sigma^2')
plt.show()

```

The following is the code to plot ISND at different level of concentration. This is for the case $\mu = 0$, but a similar one can be produced when $\mu = 1$.

```
##Plots ISND##
```

```

par(mfrow=c(2,2))
##Unimodal
sigma2=0.1
mu=0
eq = function(x){1/(2*pi*sigma2)^0.5*1/(1+cos(x))
*exp(-1/(2*sigma2)*(sin(x)/(1+cos(x))-mu)^2)}
plot(eq, from = -pi, to = pi, type='l',xlab="angle theta",
ylab="ISND(0,0.1)-unimodal-")

##Unimodal (umbilic top-threshold)
sigma2=0.5
mu=0
eq = function(x){1/(2*pi*sigma2)^0.5*1/(1+cos(x))
*exp(-1/(2*sigma2)*(sin(x)/(1+cos(x))-mu)^2)}
plot(eq, from = -pi, to = pi, type='l',xlab="angle theta",
ylab="ISND(0,0.5)-unimodal-")

##Multimodal (Standard ISND)
sigma2=1
mu=0
eq = function(x){1/(2*pi*sigma2)^0.5*1/(1+cos(x))
*exp(-1/(2*sigma2)*(sin(x)/(1+cos(x))-mu)^2)}
plot(eq, from = -pi, to = pi, type='l',xlab="angle theta",
ylab="(standard) ISND(0,1)-multimodal-")

#Multimodal (low concentration)
sigma2=10
mu=0
eq = function(x){1/(2*pi*sigma2)^0.5*1/(1+cos(x))
*exp(-1/(2*sigma2)*(sin(x)/(1+cos(x))-mu)^2)}
plot(eq, from = -pi, to = pi, type='l',xlab="angle theta",

```

```
ylab="ISND(0,10)-multimodal-")
```

This is the code for the Correlated Bivariate ISND:

```

%%%%
%%Bivariate ISND
%%%%

%% Correlated, different values Sigma and mu. %%

%Figure 1
x = linspace(-pi,pi);
y = linspace(-pi,pi);
[X,Y] = meshgrid(x,y);
sigma12=0.1;
sigma22=0.1;
t=0;
b11=sigma22/(sigma12*sigma22-t^2);
b=-t/(sigma12*sigma22-t^2);
b22=sigma12/(sigma12*sigma22-t^2);
mu1=1;
mu2=0.3;
b=0;
Z = exp( -(0.5*b11)*(tan(X/2)-mu1).^2).
*exp(-(0.5*b22)*(tan(Y/2)-mu2).^2)
.*exp( -(b)*(tan(X/2)-mu1).*(tan(Y/2)-mu2))
.*((sigma12*sigma22-t^2)*(1+cos(X)).*(1+cos(Y))).^(-1);
figure
subplot(3,2,1);
contour(X,Y,Z)
title('Sigma=[0.1,0;0,0.1] and mu=(1,0.3)')
hold on
%Figure 2
x = linspace(-pi,pi);
y = linspace(-pi,pi);
[X,Y] = meshgrid(x,y);
sigma12=0.1;
sigma22=1;
t=0;
b11=sigma22/(sigma12*sigma22-t^2);
b=-t/(sigma12*sigma22-t^2);
b22=sigma12/(sigma12*sigma22-t^2);
mu1=1;
mu2=0.3;
b=0;
Z = exp( -(0.5*b11)*(tan(X/2)-mu1).^2).
*exp(-(0.5*b22)*(tan(Y/2)-mu2).^2)

```

```

.*exp( -(b)*(tan(X/2)-mu1).*(tan(Y/2)-mu2))
.*((sigma12*sigma22-t^2)*(1+cos(X)).*(1+cos(Y))).^(-1);
subplot(3,2,2);
contour(X,Y,Z)
title('Sigma=[0.1,0;0,1] and mu=(1,0.3)')
hold on
%Figure 3
x = linspace(-pi,pi);
y = linspace(-pi,pi);
[X,Y] = meshgrid(x,y);
sigma12=1;
sigma22=1;
t=0;
b11=sigma22/(sigma12*sigma22-t^2);
b=-t/(sigma12*sigma22-t^2);
b22=sigma12/(sigma12*sigma22-t^2);
mu1=0.1;
mu2=0.1;
b=0;
Z = exp( -(0.5*b11)*(tan(X/2)-mu1).^2).
*exp(-(0.5*b22)*(tan(Y/2)-mu2).^2)
.*exp( -(b)*(tan(X/2)-mu1).*(tan(Y/2)-mu2))
.*((sigma12*sigma22-t^2)*(1+cos(X)).*(1+cos(Y))).^(-1);
subplot(3,2,6);
contour(X,Y,Z)
title('Sigma=[1,0;0,1] and mu=(0.1,0.1)')
hold on
%Figure 4
x = linspace(-pi,pi);
y = linspace(-pi,pi);
[X,Y] = meshgrid(x,y);
sigma12=1;
sigma22=1;
t=0;
b11=sigma22/(sigma12*sigma22-t^2);
b=-t/(sigma12*sigma22-t^2);
b22=sigma12/(sigma12*sigma22-t^2);
mu1=0.1;
mu2=0.1;
b=0.5;
Z = exp( -(0.5*b11)*(tan(X/2)-mu1).^2).
*exp(-(0.5*b22)*(tan(Y/2)-mu2).^2)
.*exp( -(b)*(tan(X/2)-mu1).*(tan(Y/2)-mu2))
.*((sigma12*sigma22-t^2)*(1+cos(X)).*(1+cos(Y))).^(-1);
subplot(3,2,5);
contour(X,Y,Z)

```

```

title('Sigma=[1,0.5;0.5,1] and mu=(0.1,0.1)')
hold on
%Figure 5
x = linspace(-pi,pi);
y = linspace(-pi,pi);
[X,Y] = meshgrid(x,y);
sigma12=1;
sigma22=1.5;
t=0;
b11=sigma22/(sigma12*sigma22-t^2);
b=-t/(sigma12*sigma22-t^2);
b22=sigma12/(sigma12*sigma22-t^2);
mu1=1;
mu2=0.3;
b=0.05;
Z = exp( -(0.5*b11)*(tan(X/2)-mu1).^2).
*exp(-(0.5*b22)*(tan(Y/2)-mu2).^2)
.*exp( -(b)*(tan(X/2)-mu1).*(tan(Y/2)-mu2))
.*((sigma12*sigma22-t^2)*(1+cos(X)).*(1+cos(Y))).^(-1);
subplot(3,2,4);
contour(X,Y,Z)
title('Sigma=[1,0.05;0.05,1.5] and mu=(1,0.3)')
hold on
%Figure 6
x = linspace(-pi,pi);
y = linspace(-pi,pi);
[X,Y] = meshgrid(x,y);
sigma12=1;
sigma22=1;
t=0;
b11=sigma22/(sigma12*sigma22-t^2);
b=-t/(sigma12*sigma22-t^2);
b22=sigma12/(sigma12*sigma22-t^2);
mu1=1;
mu2=0.3;
b=0.5;
Z = exp( -(0.5*b11)*(tan(X/2)-mu1).^2).
*exp(-(0.5*b22)*(tan(Y/2)-mu2).^2)
.*exp( -(b)*(tan(X/2)-mu1).*(tan(Y/2)-mu2))
.*((sigma12*sigma22-t^2)*(1+cos(X)).*(1+cos(Y))).^(-1);
subplot(3,2,3);
contour(X,Y,Z)
title('Sigma=[1,0.5;0.5,1] and mu=(1,0.3)')
hold on

```


Here, we give the codes for the comparison between the bimodal ISND and a mixture of VM:

```
#mixture Separated
x<-seq(-pi,+pi, length=1000);
head(x)
kappa1<-15;
kappa2<-15;
e<-0.7;
mu1<-(pi-e);
mu2<-(pi+e);
p1<-0.5;
y<-p1*exp(kappa1*(cos(x-mu1)))/(2*pi*besselI(kappa1,0))+(1-p1)
*exp(kappa2*(cos(x-mu2)))/(2*pi*besselI(kappa2,0))
plot(x,y)
sigma2<-kappa1*p1+kappa2*(1-p1);
mu<-0;
y<-((2*pi*sigma2)^0.5*(1+cos(x)))^(-1)
*exp(-0.5*((sigma2)^(-1))*(tan(x/2)-mu)^2);
plot(x,y)
```

```
#mixture Closeby
x<-seq(-pi,+pi, length=1000);
head(x)
kappa1<-5;
kappa2<-5;
e<-1;
mu1<-(-e);
mu2<-(+e);
p1<-0.5;
y<-p1*exp(kappa1*(cos(x-mu1)))/(2*pi*besselI(kappa1,0))+(1-p1)
*exp(kappa2*(cos(x-mu2)))/(2*pi*besselI(kappa2,0))
plot(x,y)
sigma2=1;
#sigma2<-kappa1*p1+kappa2*(1-p1);
mu<-0;
y<-((2*pi*sigma2)^0.5*(1+cos(x)))^(-1)
*exp(-0.5*((sigma2)^(-1))*(tan(x/2)-mu)^2);
plot(x,y)
```

This is the code for the comparison of the VM and the ISDN.

```
#Comparison von Mises ISND
curve(sqrt(101)/(sqrt(2*pi)*(1+cos(x+pi/2)))
x = -pi:pi/100:pi;
kappa=10;
y = (2*kappa/pi)^0.5*(1+cos(x)).^(-1).
```

```

*exp(-2*kappa.*(tan(x/2)).^2);
plot(x,y)
hold on
y = exp(kappa*cos(x))./(2*pi*besseli(0,kappa));
plot(x,y)
legend('ISND(0,1/(4kappa))','VM(0, kappa)')

```

This is the code for confidence bands.

```

e<-1;
plot(x,dnorm(x, mean=0, sd=10),
      xlim=c(-pi, pi), ylim=c(-2, 2), col="red")
lines(x,dnorm(x, mean=0, sd=10)
      +e, xlim=c(-pi, pi), ylim=c(-2, 2), col="blue")
lines(x,dnorm(x, mean=0, sd=10)
      -e, xlim=c(-pi, pi), ylim=c(-2, 2))
plot(x,(1+cos(x))^-1*dnorm(tan(x/2), mean=0, sd=10),
      xlim=c(-pi, pi), ylim=c(-2, 2), col="red")
lines(x,(1+cos(x))^-1*(dnorm(tan(x/2), mean=0, sd=10)+e),
      xlim=c(-pi, pi), ylim=c(-2, 2), col="blue")
lines(x,(1+cos(x))^-1*(dnorm(tan(x/2), mean=0, sd=10)-e),
      xlim=c(-pi, pi), ylim=c(-2, 2))

```

This is the code for the analysis of the Gomesin's dihedral angle ψ :

```

###Gomesin
##2 parameters

phi<-(pi/180)*na.omit(tor$phi)
psi<-(pi/180)*na.omit(tor$psi[2:19])
plot(phi, psi)
#Variance estimate
sigma2=0;
err<-10;
m<-max(density(psi)$y);
mu<-0;
for (k in 1:1000) {
  sigma2<-k/1000;
  for (j in 1:1000) {
    mu<- -pi+2*pi*j/1000;
    z<-((1+cos(psi))*(sqrt(2*pi*sigma2)))^-1
    *exp(-(0.5/sigma2)*(tan(-psi/2)-mu)^2);
    if (max(abs(m-max(z)))<err) {
      t<-k;
      s<-j;
      err<-max(abs(m-max(z)));
    }
  }
}

```

```

    }
}

print(t)

#Psi Marginals S1
plot(density(psi),
      xlim=c(min(psi), max(psi)), ylim=c(0,3), col="red")
x<-seq(-pi, +pi, length=1000);
mu=-pi+2*pi*s/1000;
sigma2=t/1000;
w<-((1+cos(x))*(sqrt(2*pi*sigma2)))^(-1)
*exp(-(0.5/sigma2)*(tan(x/2)-mu)^2)
lines(x+pi/2,w,
      xlim=c(min(psi)-density(psi)$x[which.max(density(psi)$y)],
            max(psi)-density(psi)$x[which.max(density(psi)$y)]), col="black")
err

This code loads and analyzes the data relative to the example in molecular
biology from CDB.

#Data
require(gdata);
DihedralAnglesData = read.xls ("/DehidralAnglesData_JustAngles
_CorrectDataGoodFormat.xls",
  sheet = 1, header = TRUE);
DAD<-DihedralAnglesData;
head(DAD);
plot(DAD[,1],DAD[,2]);
plot(DAD[,1],DAD[,2],
      main="Ramachandran Plot", xlab="phi-angle(degrees)",
      ylab="psi-angle(degrees)");
AngoliMedi<-c(mean(DAD[,1]), mean(DAD[,2]));
var(DAD);
smoothScatter(DAD[,1],DAD[,2],
              main="Ramachandran Plot-SmoothScatter",
              xlab="phi-angle(degrees)", ylab="psi-angle(degrees)")
x=sin(2*pi*DAD[,1]/360)/(1+cos(2*pi*DAD[,1]/360));
y=sin(2*pi*DAD[,2]/360)/(1+cos(2*pi*DAD[,2]/360));
Projected<- matrix(c(x,y), nrow = 8190, ncol=2);
dim(Projected);
plot(Projected[,1], Projected[,2],
      xlim=c(-20,20), ylim=c(-20,20))

#Estimation:
mean(x)
[1] -2.092306

```

```

mean(y)
[1] 2.257566
InvProjMeanX<- 2*atan(mean(x))
InvProjMeanY<- 2*atan(mean(y))
InvProjMeanX
[1] -2.249901
InvProjMeanY
[1] 2.307633

```

```

#Test
library(mvnormtest)
require(mvnormtest)
mshapiro.test(t(Projected[1:4000,]))

```

```

Shapiro-Wilk normality test
data: Z
W = 0.1063, p-value < 2.2e-16

```

#Reject the test of Normality without any doubt,
accordingly to the Shapiro Test.

This code produces the marginals for the dataset in the example from CDB.

```

par(mfrow=c(2,2)); hist(x,breaks=10000, xlim=c(-10,10));
  hist(y,breaks=10000, xlim=c(-10,10));

```

We test our proposed distributions with with *Kolmogorov-Smirnov*.

```
ks.test(y,mean(y))
```

Two-sample Kolmogorov-Smirnov test

```

data: y and mean(y)
D = 0.8172, p-value = 0.5165
alternative hypothesis: two-sided

```

```
ks.test(-x,mean(-x))
```

Two-sample Kolmogorov-Smirnov test

```

data: -x and mean(-x)
D = 0.8641, p-value = 0.4442
alternative hypothesis: two-sided

```

#The Projected look like exponential (the y) and reversed
exponential (the x). This may suggest a stereographic double
exponential as a good fit. Can't reject, it seems a good fit.

Acknowledgements

I warmly thank my family and Kathleen Foster for their constant support. I thank my supervisor Prof. Narayanaswamy Balakrishnan for his help and inspiring guidance. I thank Prof. Viveros-Aguilera, Prof. Min-Oo and Prof. Balakrishnan for useful comments in the preparation of this manuscript. Many thanks also to the referees whose suggestions improved significantly the content of the paper and for bringing several relevant references to my attention.

References

- [1] T. ABE, K. SHIMIZU AND A. PEWSEY, Symmetric unimodal models for directional data motivated by inverse stereographic projection, *Journal of the Japan Statistical Society*, **40** (1) (2010) 045-061. [MR2956783](#)
- [2] J. BESAG, Statistical analysis of non-lattice data, *Journal of the Royal Statistical Society. Series D (The Statistician)*, Vol. **24** No. 3 (1975) 179-195.
- [3] D. J. BEST AND N. I. FISHER, Efficient simulation methods of the Von Mises Distribution, *Journal of the Royal Statistical Society. Series C (Appl. Statist.)*, **28** No. 2 (1979) 152-157.
- [4] G. CASELLA AND R. L. BERGER, *Statistical Inference*, The Wadsworth & Brooks/Cole Statistics/Probability Series, Pacific Grove (1990). [MR1051420](#)
- [5] <http://cluster.physics.iisc.ernet.in/cgi-bin/cadb/callapplet1.pl?identity=25&experiment=N&mode=F&Entire=1&Fully=0&Additional=0&Generous=0&Disallowed=0&Alpha=0&Beta=0&Threeten=0&residue0=ALA>
- [6] D. B. DIX, Mathematical Models of Protein Folding, <http://people.math.sc.edu/dix/fold.pdf>
- [7] J. -L. DORTET-BERNADET AND N. WICKER, A Note on Inverse Stereographic Projection of Elliptical Distributions, *Sankhya A*, Volume **80** Issue 1 (2018) 138-151. [MR3772095](#)
- [8] W. FELLER, An Introduction to Probability Theory and its Applications, *John Wiley and Sons*, New York (1970). [MR0270403](#)
- [9] C. R. GOODALL AND K. V. MARDIA, Projective Shape Analysis, *Journal of Computational and Graphical Statistics*, **8:2** (1999) 143-168. [MR1706381](#)
- [10] M. M. HINGORANI AND M. O'DONNELL, Toroidal proteins: Running rings around DNA, *Current Biology*, Volume **8** Issue 3 (January 1998) R83-R86.
- [11] M. M. HINGORANI AND M. O'DONNELL, A tale of toroids in DNA metabolism *Nature Reviews Molecular Cell Biology*, **1** (October 2000) 22-30.
- [12] S. HUCKEMANN, K. R. KIM, A. MUNK, F. REHFELDT, M. SOMMERFELD, J. WEICKEERT AND C. WOLLNIK, The circular sizer, inferred persistence of shape parameters and application to early stem cell differentiation, *Bernoulli*, **22** (4) 2113-2142. [MR3498025](#)

- [13] J. T. KENT, K. V. MARDIA AND C. C. TAYLOR, Modelling strategies for bivariate circular data. In: Barber, S., Baxter, P.D., Gusnanto, A., Mardia, K.V. (eds.) *The Art and Science of Statistical Bioinformatics*, Leeds University Press, Leeds (2008) 70-74.
- [14] P. E. JUPP AND K. V. MARDIA, A general correlation coefficient for directional data and related regression problems, *Biometrika*, **67** 1 (1980) 163-173. [MR0570518](#)
- [15] P. E. JUPP AND K. V. MARDIA, *Directional Statistics*, John Wiley and Sons, New York (2008). [MR1828667](#)
- [16] P. M. LEE, *Bayesian Statistics: An Introduction*, 3rd edition, John Wiley and Sons, New York (2004). [MR1475928](#)
- [17] K. V. MARDIA, Statistics of directional data (with discussion), *J. Roy. Statist. Soc. B*, **37** (1975) 349-393. [MR0402998](#)
- [18] K. V. MARDIA, Characterization of directional distributions. In: Patil, G.P., Kotz, S., Ord, J.K. (eds.) *Statistical Distributions in Scientific Work*, D. Reidel Publishing Company, Dordrecht (1975) 365-386.
- [19] K. V. MARDIA, Bayesian Analysis for Bivariate Von Mises Distributions, *J. Appl. Stat.*, **37** (2010) 515-528. [MR2756649](#)
- [20] K. V. MARDIA AND S. A. M. EL-ATOUM, Bayesian inference for the von Mises-Fisher distribution, *Biometrika*, **63** (1976) 203-206. [MR0415862](#)
- [21] K. V. MARDIA AND J. FRELLSEN, Statistics of Bivariate von Mises Distributions, In: *Bayesian Methods in Structural Bioinformatics. Statistics for Biology and Health*, Springer, Heidelberg (2012) 159-178. [MR3060435](#)
- [22] K. V. MARDIA, J. T. KENT, G. HUGHES AND C. C. TAYLOR, Maximum likelihood estimation using composite likelihoods for closed exponential families, *Biometrika*, **96** (2009) 975-982. [MR2767282](#)
- [23] K. V. MARDIA, J. T. KENT, G. HUGHES, C. C. TAYLOR AND H. SINGH, A multivariate von Mises distribution with applications to bioinformatics, *Can. J. Stat.*, **36** (2008) 99-109. [MR2432195](#)
- [24] K. V. MARDIA AND V. PATRANGENARU, Directions and Projective Shapes, *Annals of Statistics*, **33** No. 4 (2005) 1666-1699. [MR2166559](#)
- [25] K. V. MARDIA AND T. W. SUTTON, On the modes of a mixture of two von Mises distributions, *Biometrika*, **62** 3 (1975) 699-701. [MR0388494](#)
- [26] K. V. MARDIA, C. C. TAYLOR AND G. K. SUBRAMANIAM, Protein Bioinformatics and Mixtures of Bivariate von Mises Distributions for Angular Data, *Biometrics*, **63** (2007) 505-512. [MR2370809](#)
- [27] R. MOMEN, A. AZIZI, L. WANG, Y. PING, T. XU, S. R. KIRK, W. LI, S. MANZHOS AND S. JENKINS, Exploration of the forbidden regions of the Ramachandran plot ($\phi - \psi$) with QTAIM, *Phys. Chem. Chem. Phys.*, **19** (2017) 26423.
- [28] R. J. MUIRHEAD, *Aspects of multivariate statistical theory*, John Wiley and Sons, New York (2005). [MR0652932](#)
- [29] V. PATRANGENARU AND L. ELLINGTON, *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*, CRC/Chapmann Hall (2015). [MR3444169](#)
- [30] Y. PHANI, S. GIRIJA AND A. DATTATREYA RAO, Circular model induced

- by inverse stereographic projection on extreme-value distribution, *Engineering Science and Technology*, **2** (5) (2012) 881-888.
- [31] Y. PHANI, Y. SRIHARI, S. GIRIJA AND A. DATTATREYA RAO, Stereographic circular normal moment distribution, *Applied Mathematics and Sciences: An International Journal*, **1** (2014) 65-72.
- [32] *Protein Database*, <http://pdb101.rcsb.org/>
- [33] A. RADHIKA, S. GIRIJA AND A. DATTATREYA RAO, On rising Sun von mises and rising Sun wrapped cauchy circular models, *Journal of Applied Mathematics, Statistics and Informatics*, **9** (2) (2013) 41-55. [MR3252482](#)
- [34] G. N. RAMACHANDRAN, V. RAMAKRISHNAN AND C. SASISEKHARAN, Stereochemistry of polypeptide chain configurations *J. Mol. Biol.*, **7** (1963) 95-99.
- [35] L. RIVEST, A distribution for dependent unit vectors, *Comm. Stat. Theor. Meth.*, **17** (1988) 461-483. [MR0939661](#)
- [36] P. J. ROUSSEUW AND K. VAN DRIESSEN, A Fast Algorithm for the Minimum Covariance Determinant Estimator, *Technometrics*, **41** no. 3 (1999) 212-223.
- [37] D. J. SELKOE, Folding proteins in fatal ways, *Nature*, **426** (6968) (December 2003) 9004.
- [38] S. S. SHAPIRO AND M. B. WILK, Analysis of variance test for normality (complete samples), *Biometrika*, **52** (1965) 591-611. [MR0205384](#)
- [39] H. SINGH, V. HNIZDO AND E. DEMCHUK, Probabilistic Model for two dependent circular variables, *Biometrika*, **89** no. 3 (2002) 719-723. [MR1929175](#)
- [40] C. C. TAYLOR, Directional Data on the torus, with applications to protein structure, *Proceedings of the SIS 2009 Statistical Conference on Statistical Methods for the Analysis of Large Data-Sets*, (2013) 105-108.
- [41] G. S. WATSON, *Statistics on Spheres*, Wiley, New York (1983). [MR0709262](#)