Necessary and sufficient condition for the existence of a Fréchet mean on the circle

Benjamin Charlier
Institut de Mathématiques de Toulouse
Université de Toulouse et CNRS (UMR 5219)
Benjamin.Charlier@math.univ-toulouse.fr

March 7, 2012

Abstract

Let $(\mathbb{S}^1, d_{\mathbb{S}^1})$ be the unit circle in $\mathbb{R}^2$ endowed with the arclength distance. We give a sufficient and necessary condition for a general probability measure μ to admit a well defined Fréchet mean on $(\mathbb{S}^1, d_{\mathbb{S}^1})$. We derive a new sufficient condition of existence $P(\alpha, \varphi)$ with no restriction on the support of the measure. Then, we study the convergence of the empirical Fréchet mean to the Fréchet mean and we give an algorithm to compute it.

Keywords : circular data, Fréchet mean, uniqueness. *AMS classification*: 62H11.

1 Introduction

1.1 Statistics for non-Euclidean data.

In many fields of interest, results of an experiment are objects taking values in non-Euclidean spaces. A rather general framework to model such data is Riemannian geometry and more particularly quotient manifolds. As an illustration, in biology or geology, directional data are often used, see e.g. [14] or [6] and references therein. In this case, observations take their values in the circle or a sphere, that is, an Euclidean space quotiented by the action of scaling.

The usual definitions of basic statistical concepts were developed in an Euclidean framework. Therefore, these definitions must be adapted for random variables with values in non-Euclidean spaces such as manifolds. To describe the localization of a probability distribution, one needs to define a central value such as a mean or a median. There has been multiple attempts to give a definition of a mean in non Euclidean space, see among many others [2, 4, 12, 13, 8, 5, 16] or [7].

In this paper, we consider the so-called Fréchet mean, see [7, 8, 12] or [2] and references therein. We are particularly interested in the study of its uniqueness. The Fréchet mean is defined on general metric spaces by extending the fact that Euclidean mean minimizes the sum of square the distance to the data, see equation (1.3) below. To study the well definiteness of the Fréchet mean on a manifold, two facts must be taken into account: non uniqueness of geodesics from one point to another (existence of a cut locus) and the effect of curvature, see e.g. [4] for further discussion. Due to the cut-locus, the distance function is no longer convex and finding conditions to ensure the uniqueness of the Fréchet mean is not obvious. Two main directions have been explored in the literature: bounding the support of the measure in [3] for the n-spheres and in [8, 10, 12, 1] for manifolds, or consider special cases of absolutely continuous radial distributions, see [9] for the unit circle and or [10, 11] for projective spaces. In a sense, these two conditions control the concentration of the probability measure. The philosophy behind these works is to ensure a convexity property of the Fréchet functional given by equation (1.2) below, see e.g. the introduction of [1] for a review of the above cited papers.

1.2 Fréchet mean on the circle

A standard way to extend the definition of the Euclidean mean in non-Euclidean metric space is to use the minimization property of the Euclidean mean. This definition is usually credited to M. Fréchet in [7] although some authors credit it to E. Cartan, see e.g. [10]. Let $\mathbb{S}^1$ be the unit circle of the plane,

$$\mathbb{S}^1 = \{x_1^2 + x_2^2 = 1,\ (x_1, x_2) \in \mathbb{R}^2\}.$$

endowed with the arclength distance given for all $x = (x_1, x_2), p = (p_1, p_2) \in \mathbb{S}^1$ by

$$d_{\mathbb{S}^1}(x, p) = 2 \arcsin\left(\frac{\|x - p\|}{2}\right), \tag{1.1}$$

where $\|x - p\| = \sqrt{(x_1 - p_1)^2 + (x_2 - p_2)^2}$ is the Euclidean norm in $\mathbb{R}^2$. In the whole paper, μ is a probability measure on $\mathbb{S}^1$ and the Fréchet functional is defined for all $p \in \mathbb{S}^1$ by

$$F_\mu(p) = \frac{1}{2} \int_{\mathbb{S}^1} d^2(x, p) d\mu(x). \tag{1.2}$$

The Fréchet functional is Lipschitz since by the triangle inequality we have $|F_\mu(p_1) - F_\mu(p_2)| \leq 2\pi d_{\mathbb{S}^1}(p_1, p_2)$ for any $p_1, p_2 \in \mathbb{S}^1$. Thus, F_μ attains its minimum in at least one point and the only issue at hand is uniqueness.

Definition 1.1. *We say that the Fréchet mean of a probability measure μ in $(\mathbb{S}^1, d_{\mathbb{S}^1})$ is well defined if F_μ admits a* unique *argmin. That is, there exists a unique $p^* \in \mathbb{S}^1$ satisfying $F_\mu(p^*) = \min_{p \in M} F_\mu(p)$, and we note*

$$p^* = \underset{p \in \mathbb{S}^1}{\mathit{argmin}}\, F_\mu(p). \tag{1.3}$$

The argmins of F_μ are also called Riemannian center of mass [16] or intrinsic mean [2] as the $(\mathbb{S}^1, d_{\mathbb{S}^1})$ is a simple one dimensional compact Riemannian manifold. The advantage of dealing with a simple object such as the circle is that curvature problems disappear and we only face the cut-locus problem. In this sense, it allows us to completely understand its effect on the non-convexity of the distance function $d_{\mathbb{S}^1}$, and to give a complete answer about the problem of uniqueness. In what follows, we fully characterize probability measures that admit a well defined Fréchet mean on the circle $(\mathbb{S}^1, d_{\mathbb{S}^1})$. In particularly a necessary and sufficient condition is given in Theorem 4.1, which links the existence of a Fréchet mean for a measure μ to the comparison between the distribution μ and the uniform measure λ on $\mathbb{S}^1$. The surprising fact is that λ appears as a benchmark to discriminate measures having a well defined Fréchet mean. The uniform measure λ is the 'worst' possible case as all points of the circle is a Fréchet mean, indeed the Fréchet functional (1.2) is constant and equals to $\frac{\pi^3}{3}$.

In opposition to what have been done before we do not try to ensure convexity property on the Fréchet functional. Indeed, the definition of the Fréchet mean relies on the *global* optimization problem (1.3) which is, in general, non convex. The advantage of our approach is that we do not need to restrict the support or suppose restrictive conditions of symmetry on the density. As the geometry of flat manifold is simple, we can derive explicit form on the Fréchet functional and its derivative which can be hard to compute in non-flat manifolds such as n-dimensional spheres.

1.3 Organization of the paper

In Section 2, we introduce notations that will be used throughout the paper. In Section 3, we give explicit expressions for the Fréchet functional and its derivative and we discuss some properties of critical points of the Fréchet functional. Section 4 contains the main result with the necessary and sufficient condition of Theorem 4.1 for the existence of the Fréchet mean for a general measure. We also propose a new sufficient criterion $P(\alpha, \varphi)$ that ensures the well definiteness of the Fréchet mean. In Section 5, we study the convergence of the empirical Fréchet mean to the Fréchet mean, and describe an algorithm to compute the empirical Fréchet mean.

2 Notations

In what follows, $\mathbb{1}_A$ denotes the indicator function of the set $A \subset \mathbb{R}$ and the notation $\int_a^b f(t)d\mu_{p_0}(t)$ stands for the Lebesgue integral $\int_{[a,b[} f(t)d\mu_{p_0}(t)$ if $a \leq b$ and $\int_{[b,a[} f(t)d\mu_{p_0}(t)$ if $b > a$.

2.1 Normal coordinates

Given a base point $p \in \mathbb{S}^1$, there is a canonical chart called the exponential map defined from $T_p\mathbb{S}^1 \simeq \mathbb{R}$, the tangent space of $\mathbb{S}^1$ at p, to $\mathbb{S}^1$ and denoted by

$$\begin{array}{rcl} e_p : \mathbb{R} & \longrightarrow & \mathbb{S}^1 \\ \theta & \longmapsto & e_p(\theta) = R_\theta p, \end{array} \qquad \text{where } R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

This map is onto but not one to one as it is 2π-periodic. To guaranty the injectivity, we choose to restrict the domain of definition $\mathbb{R}$ of e_p to $[-\pi, \pi[$. Thus, for all $p_1, p_2 \in \mathbb{S}^1$ there is a now unique $\theta_{p_2}^{p_1} \in [-\pi, \pi[$ satisfying $e_{p_1}(\theta_{p_2}^{p_1}) = p_2$ and

$$e_p : [-\pi, \pi[\longrightarrow \mathbb{S}^1 \qquad \text{and} \qquad e_p^{-1} : \mathbb{S}^1 \longrightarrow [-\pi, \pi[, \qquad \text{for all } p \in \mathbb{S}^1.$$

Such parametrizations are called normal coordinates systems centered at p and $\theta_{p_2}^{p_1}$ is nothing else but the coordinate of p_2 read in a normal coordinate system centered in p_1. To simplify the notations, we will omit the exponent p_1 if no confusion is possible and we will write $\theta_{p_2}^{p_1} = \theta_{p_2}$.

The cut locus of a point $p_0 \in \mathbb{S}^1$ is denoted by $\tilde{p}_0$ and is equal to the opposite point (in $\mathbb{R}^2$) of p_0, that is $\tilde{p}_0 = -p_0$. In a normal coordinates system centered at $p \in \mathbb{S}^1$, the coordinate of $\tilde{p}_0$ is $\theta_{\tilde{p}_0}^p = \theta_{p_0}^p - \pi$ if $0 \leq \theta_{p_0}^p < \pi$ or $\theta_{\tilde{p}_0}^p = \theta_{p_0}^p + \pi$ if $\pi \leq \theta_{p_0}^p < 0$.

2.2 Distance function and probability measures and the Fréchet functional

The arclength distance between two points $p_1, p_2 \in \mathbb{S}^1$ was defined in (1.1). Given normal coordinates $\theta_{p_1}, \theta_{p_2} \in [-\pi, \pi[$ of these points in the chart centered at an arbitrary point ,

$$d_{\mathbb{S}_1}(p_1, p_2) = d_{\mathbb{R}/2\pi\mathbb{Z}}(\theta_{p_1}, \theta_{p_2}) := \min\{|\theta_{p_1} - \theta_{p_2} + 2\pi k|\,, k \in \mathbb{Z}\}.$$

This means that the circle $\mathbb{S}^1$ is locally isometric to the real line $\mathbb{R}$.

Unless specified, μ will denote a general probability measure on $(\mathbb{S}^1, \mathcal{B}(\mathbb{S}^1))$ where $\mathcal{B}(\mathbb{S}^1)$ is the Borel set of $\mathbb{S}^1 \subset \mathbb{R}^2$. Given a point $p_0 \in \mathbb{S}^1$, μ_{p_0} is the image measure of μ through $e_{p_0}^{-1} : \mathbb{S}^1 \longrightarrow [-\pi, \pi[$. This is a measure on $\mathbb{R}$ with a support in $[-\pi, \pi[$ which is defined by

$$\mu_{p_0}(A) = \mu \circ e_{p_0}(A \cap [-\pi, \pi[), \qquad \text{for all } A \in \mathcal{B}(\mathbb{R}), \tag{2.1}$$

where $\mathcal{B}(\mathbb{R})$ is the Borel set in $\mathbb{R}$. The usual Euclidean mean/expectation and variance of μ_{p_0} are denoted

$$m(\mu_{p_0}) = \int_{\mathbb{R}} t d\mu_{p_0}(t).$$

Finally, let us introduce for any $p_0 \in \mathbb{S}^1$, the map $F_{\mu_{p_0}} : [-\pi, \pi[\longrightarrow \mathbb{R}$ given by

$$\begin{aligned} F_{\mu_{p_0}}(\theta) := F_\mu(e_{p_0}(\theta)) &= \frac{1}{2}\int_{-\pi}^{\pi} d^2_{\mathbb{R}/2\pi\mathbb{Z}}(t, \theta) d\mu_{p_0}(t) \\ &= \frac{1}{2}\begin{cases} \int_{-\pi}^{\theta-\pi}(t + 2\pi - \theta)^2 d\mu_{p_0}(t) + \int_{\theta-\pi}^{\pi}(t-\theta)^2 d\mu_{p_0}(t), & \text{if } 0 \leq \theta < \pi, \\ \int_{-\pi}^{\theta+\pi}(t-\theta)^2 d\mu_{p_0}(t) + \int_{\theta+\pi}^{\pi}(t - 2\pi - \theta)^2 d\mu_{p_0}(t), & \text{if } -\pi \leq \theta < 0. \end{cases} \end{aligned}$$

3 The Fréchet functional on the Circle

3.1 The derivative of the Fréchet functional

A function $f : [-\pi, \pi[\longrightarrow \mathbb{R}$ is said left continuous on $[-\pi, \pi[$ if it is left continuous everywhere on $]-\pi, \pi[$ and with $\lim_{\varepsilon\to 0^-} f(\pi+\varepsilon) = f(-\pi)$. Similarly, f is said to be continuous on $[-\pi, \pi[$ if it is left and right continuous on $[-\pi, \pi[$. We provide here an explicit expression of the derivative of F_μ,

Proposition 3.1. *Let μ be a probability measure on $(\mathbb{S}^1, d_{\mathbb{S}^1})$ and fix an arbitrary $p_0 \in \mathbb{S}^1$. Then, $F_\mu : \mathbb{S}^1 \longrightarrow \mathbb{R}$ is differentiable in following sense :*

1. *Let $p \in \mathbb{S}^1$ be a point with a cut locus of μ-measure 0, i.e $\mu(\{-p\}) = 0$. Then $F_{\mu_{p_0}}$ is continuously differentiable at $\theta_p^{p_0}$ and we have*

$$\frac{d}{d\theta}F_{\mu_{p_0}}(\theta_p^{p_0}) = \begin{cases} \theta_p^{p_0} - 2\pi\mu_{p_0}([-\pi, -\pi+\theta_p^{p_0}[) - m(\mu_{p_0}), & \text{if } 0 \leq \theta_p^{p_0} < \pi, \\ \theta_p^{p_0} + 2\pi\mu_{p_0}([\pi+\theta_p^{p_0}, \pi[) - m(\mu_{p_0}), & \text{if } -\pi \leq \theta_p^{p_0} < 0. \end{cases} \tag{3.1}$$

2. *The function $\frac{d}{d\theta}F_{\mu_{p_0}}$ is left continuous on $[-\pi, \pi[$. Then we extend the definition of the derivative of F_μ by setting* for all $\theta \in]-\pi, \pi[$

$$\frac{d}{d\theta}F_{\mu_{p_0}}(\theta) := \lim_{\varepsilon\to 0^-} \frac{d}{d\theta}F_{\mu_{p_0}}(\theta+\varepsilon), \tag{3.2}$$

and $\frac{d}{d\theta}F_{\mu_{p_0}}(-\pi) = \lim_{\varepsilon\to 0^-} \frac{d}{d\theta}F_{\mu_{p_0}}(\pi+\varepsilon)$.

3. *Let $p \in \mathbb{S}^1$ be a point with a cut locus of positive measure, i.e $\mu(\{-p\}) > 0$. Then, p is a cusp point of F_μ in the sense that* $\lim_{\varepsilon\to 0^-} \frac{d}{d\theta}F_{\mu_{p_0}}(\theta_p^{p_0}+\varepsilon) - \lim_{\varepsilon\to 0^+} \frac{d}{d\theta}F_{\mu_{p_0}}(\theta_p^{p_0}+\varepsilon) = -\mu(\{-p\})$.

Note that the left-continuity comes from our convention on the exponential map which is defined on $[-\pi, \pi[$. If a measure μ is such that $\mu(\{p\}) = 0$ for all $p \in \mathbb{S}^1$ then F_μ is of class $\mathcal{C}^1$ on $[-\pi, \pi[$. Differentiability issues appear when the measure μ has atoms, see Figure 1.

Proof. For convenience we omit in this proof the superscript p_0 by writing $\theta_p = \theta_p^{p_0}$ for all $p \in \mathbb{S}^1$. In the coordinate system centered at p_0 we have for all $\theta_p \in [-\pi, \pi[$

$$F_{\mu_{p_0}}(\theta_p) = \frac{1}{2}\int_{\mathbb{R}} t^2 d\mu_{p_0}(t) - \theta_p m(\mu_{p_0}) + \frac{1}{2}\theta_p^2 + 2\pi\Big(g^+_{\mu_{p_0}}(\theta_p)\mathbb{1}_{[0,\pi[}(\theta_p) + g^-_{\mu_{p_0}}(\theta_p)\mathbb{1}_{[-\pi,0[,}(\theta_p)\Big) \tag{3.3}$$

where $g^+_{\mu_{p_0}}(\theta) = \int_{-\pi}^{-\pi+\theta}(\pi+t-\theta)d\mu_{p_0}(t)$ and $g^-_{\mu_{p_0}}(\theta) = \int_{\theta+\pi}^{\pi}(\pi-t+\theta)d\mu_{p_0}(t)$. Hence, to prove Proposition 3.1, we just have to study the derivative of $g^+_{\mu_{p_0}}$ and $g^-_{\mu_{p_0}}$.

For all $\theta_p \in]0, \pi[$ and $\varepsilon \in \mathbb{R}$ such that $\theta_p + \varepsilon \in]0, \pi[$ we have,

$$\frac{1}{\varepsilon}\Big(g^+_{\mu_{p_0}}(\theta_p+\varepsilon) - g^+_{\mu_{p_0}}(\theta_p)\Big) = \frac{1}{\varepsilon}\int_{-\pi+\theta_p}^{-\pi+\theta_p+\varepsilon}(\pi+t-\theta_p)d\mu_{p_0}(t) - \int_{-\pi}^{-\pi+\theta_p+\varepsilon} d\mu_{p_0}(t) \tag{3.4}$$

The limit from the left in equation (3.4) is $\lim_{\varepsilon\to 0^-} \frac{1}{\varepsilon}(g^+_{\mu_{p_0}}(\theta_p+\varepsilon) - g^+_{\mu_{p_0}}(\theta_p)) = -\mu_{p_0}([-\pi, -\pi+\theta_p[)$ when $0 < \theta_p < \pi$. The (left) derivative at $\theta_p = 0$ is given by $\lim_{\varepsilon\to 0^-} \frac{1}{\varepsilon}\Big(g^-_{\mu_{p_0}}(\theta_p+\varepsilon) - g^+_{\mu_{p_0}}(\theta_p)\Big) = 0$. Similarly, if $-\pi \leq \theta_p < 0$, we have $\lim_{\varepsilon\to 0^-} \frac{1}{\varepsilon}(g^-_{\mu_{p_0}}(\theta_p+\varepsilon) - g^-_{\mu_{p_0}}(\theta_p)) = \mu_{p_0}([\pi+\theta_p, \pi[)$ and statement 2 is proved.

To prove statement 1, suppose that the cut locus of p is of μ-measure 0. In this case, the limit from the left and from the right in equation (3.4) are equal as $\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\int_{-\pi+\theta_p}^{-\pi+\theta_p+\varepsilon}(\pi+t-\theta_p)d\mu_{p_0}(t) = 0$ since $\mu_{p_0}(\{\theta_p - \pi\}) = 0$. Thence, formula (3.3) yields $\frac{d}{d\theta}g^+_{\mu_{p_0}}(\theta_p) = -\mu_{p_0}([-\pi, -\pi+\theta_p[)$, if $\theta_p \in [0, \pi[$ and $\frac{d}{d\theta}g^-_{\mu_{p_0}}(\theta_p) = \mu_{p_0}([\pi+\theta_p, \pi[)$, if $\theta_p \in [-\pi, 0[$.

Finally, suppose that $p \in \mathbb{S}^1$ has a cut locus of positive measure. If $0 \leq \theta_p < \pi$, it means that $\mu_{p_0}(\{\theta_p - \pi\}) > 0$ and then, $\frac{d}{d\theta}F_{\mu_{p_0}}(\theta_p) - \lim_{\varepsilon\to 0^+} \frac{d}{d\theta}F_{\mu_{p_0}}(\theta_p+\varepsilon) = -\lim_{\varepsilon\to 0^+} \mu_{p_0}([-\pi+\theta_p, -\pi+\theta_p+\varepsilon[) = -\mu_{p_0}(\{-\pi+\theta_p\})$. The case $-\pi \leq \theta_p < 0$ is similar and the proof of statement 3 is completed. $\square$

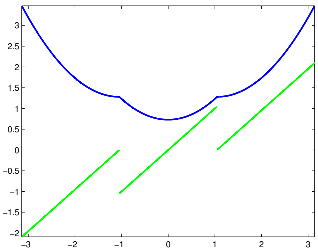


Figure 1: Let $\mu = \frac{1}{6}\delta_{p_1} + \frac{2}{3}\delta_{p^*} + \frac{1}{6}\delta_{p_2}$ with $p_1 = R_{\frac{2\pi}{3}} p^*$ and $p_2 = R_{-\frac{2\pi}{3}} p^*$. In blue: $F_{\mu_{p^*}}$. In green: $\frac{d}{d\theta} F_{\mu_{p^*}}$.

3.2 Local minimum of the Fréchet functional

The critical points of the Fréchet functional are the points at which the derivative of F_μ, in the sens of Proposition 3.1, is 0. As a immediate consequence of equation (3.1) we have

$$\frac{d}{d\theta} F_{\mu_{p_0}}(\theta_p^{p_0}) = -m(\mu_p) \tag{3.5}$$

Thus, the critical points are precisely the exponential barycenters (i.e. points $p \in \mathbb{S}^1$ at which $m(\mu_p) = 0$). This fact was already shown in [5] or [12] for general manifolds. Note that a critical point of the Fréchet functional is not in general an extremum (see an illustration Figure 1) but it is worth noticing that the minima of the Fréchet functional are regular points in the sens of the following result,

Corollary 3.1. *Let μ be a probability measure on $\mathbb{S}^1$. The cut locus of a (local or global) minimum of F_μ is of μ-measure 0.*

Proof. Let $p \in \mathbb{S}^1$ be a point satisfying $\mu(\{-p\}) > 0$. For any $p_0 \in \mathbb{S}^1$, Statement 3 of Proposition 3.1 ensure that the derivative $\frac{d}{d\theta} F_{\mu_{p_0}}$ of the Frćhet functional has a negative jump at $\theta_p^{p_0}$. Hence, the signs of $\left(\lim_{\varepsilon \to 0^-} \frac{d}{d\theta} F_{\mu_{p_0}}(\theta_{pm}^{p_0} + \varepsilon), \lim_{\varepsilon \to 0^+} \frac{d}{d\theta} F_{\mu_{p_0}}(\theta_{pm}^{p_0} + \varepsilon)\right)$ is either $(+,+)$, $(+,-)$ or $(-,-)$. This means that $\theta_p^{p_0}$ cannot be a minimum of $F_{\mu_{p_0}}$ since it would correspond to the case $(-,+)$. $\square$

Remark 3.1. *Note that assumptions of Corollary 1 in [17] and Theorem 1 in [2] contain a condition of the form $\mu(\{\tilde{p}\})$ to ensure (classical) differentiability of the Fréchet functional at its minimum. In the case of the circle, Corollary 3.1 shows us that the Fréchet functional is* automatically *differentiable at its minima.*

4 Necessary and sufficient condition for the existence of the Fréchet mean

4.1 Main result

Theorem 4.1. *Let μ be a probability measure and $p^* \in \mathbb{S}^1$ be a critical point of F_μ. Then, the following propositions are equivalent,*

1. *p^* is a well defined Fréchet mean of $(\mathbb{S}^1, \mu)$.*
2. *For all $0 < \theta < \pi$*

$$\int_0^\theta \lambda([-\pi, -\pi + t[) - \mu_{p^*}([-\pi, -\pi + t[)dt > 0,$$

and for all $-\pi \le \theta < 0$,

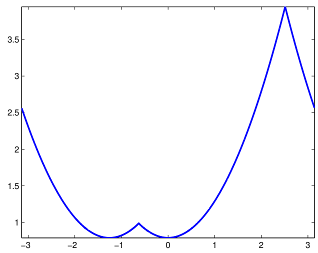
$$\int_{\theta}^{0} \lambda(]\pi + t, \pi[) - \mu_{p^*}(]\pi + t, \pi[)dt > 0,$$

where λ *is the uniform measure on* $[-\pi, \pi[$ *and* μ_{p^*} *is defined in* (2.1).

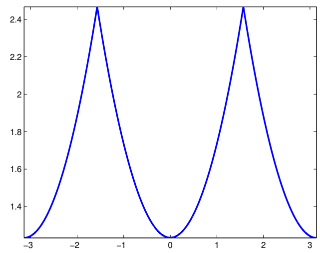
Theorem 4.1 gives a necessary and sufficient condition for the existence of the Fréchet mean of a general measure μ on the circle $\mathbb{S}^1$. This condition is given in terms of comparison between the μ-measure and the uniform measure λ of balls centered at the cut locus of a global minimum. The important point is that the μ-measure of a (small) neighborhood of the cut locus of p^* cannot be larger than the uniform measure of this neighborhood.

As μ is a probability measure, the functions $t \longmapsto \lambda([-\pi, -\pi + t[) - \mu_{p^*}([-\pi, -\pi + t[)$ and $t \longmapsto \lambda(]\pi - t, \pi[) - \mu_{p^*}(]\pi - t, \pi[)$ do not need to be always positive for $t \in [-\pi, 0[$ and $t \in [0, \pi[$ respectively. An example where this quantity is always positive is when μ admits a density which is a decreasing function of the distance to a point p^*, see [9]. In this case, the density is radially distributed around its mode p^* which is, by Theorem 4.1, the Fréchet mean of μ. Many classical probability distributions used in circular data analysis follow this framework: von Mises distribution, wrapped normal distribution, geodesic normal distribution [17] , etc...

Another well-known example of distributions that admit a well defined Fréchet mean is distributions with support included in an hemisphere, see [3]. More precisely, suppose that there exists a point $\hat{p} \in \mathbb{S}^1$ with $\mu(\{p \in \mathbb{S}^1, d_{\mathbb{S}^1}(p, \hat{p}) \le \frac{\pi}{2}\}) = 1$ and $\mu(\{p \in \mathbb{S}^1, d_{\mathbb{S}^1}(p, \hat{p}) < \frac{\pi}{2}\}) > 0$. In this case, Statement 2 of Theorem 4.1 holds since one can show that the minimum p^* of F_μ is in $\{p \in \mathbb{S}^1, d_{\mathbb{S}^1}(p, \hat{p}) < \frac{\pi}{2}\}$ and that $F_{\mu_{p^*}}(\theta) - F_{\mu_{p^*}}(0) > \frac{1}{2}(\theta + \pi - 2\theta_{\hat{p}})^2$ for $\theta \in [-\pi, \theta_{\hat{p}} - \frac{\pi}{2}[$, $F_{\mu_{p^*}}(\theta) - F_{\mu_{p^*}}(0) = \frac{\theta^2}{2}$ for $\theta \in [\theta_{\hat{p}} - \frac{\pi}{2}, \theta_{\hat{p}} + \frac{\pi}{2}[$ and $F_{\mu_{p^*}}(\theta) - F_{\mu_{p^*}}(0) > \frac{1}{2}(\theta - \pi - 2\theta_{\hat{p}})^2$ for all $\theta \in [\theta_{\hat{p}} + \frac{\pi}{2}, \pi[$. The case of equality corresponds to distributions with support in the boundary of the hemisphere, that is $\mu_{p^*} = (1 - \varepsilon)\delta_{\theta_{\hat{p}}^{p^*} - \frac{\pi}{2}} + \varepsilon\delta_{\theta_{\hat{p}}^{p^*} + \frac{\pi}{2}}$ with $\varepsilon = \frac{1}{\pi}\theta_{\hat{p}}^{p^*} + \frac{1}{2}$ and in this case, there are two global argmins at 0 and $2\theta_{\hat{p}} \pm \pi$, see Figure 2.



(a)



(b)

Figure 2: Let $\mu_\theta = (1 - \varepsilon)\delta_{\theta - \frac{\pi}{2}} + \varepsilon\delta_{\theta + \frac{\pi}{2}}$ with $\varepsilon = \frac{\theta}{\pi} + \frac{1}{2}$. (a) F_{μ_θ} with $\theta = 0$. (b) F_{μ_θ} with $\theta = \frac{3\pi}{10}$

4.2 Proof of Theorem 4.1

As already noticed, there exists at least one global argmin $p^* \in \mathbb{S}^1$ of F_μ since F_μ is a continuous function defined on the compact set $\mathbb{S}^1$. Moreover, Proposition 3.1 and Corollary 3.1 ensure that p^* is a regular critical point of F_μ, i.e. a zero of the derivative with a cut locus of μ-measure 0.

In the normal coordinate system centered at p^* the functional $G_{\mu_{p^*}}(\theta) = F_{\mu_{p^*}}(\theta) - F_{\mu_{p^*}}(0)$ has a particularly simple expression thanks to equation (3.5). Indeed, we have

$$G_{\mu_{p^*}}(\theta) = \frac{1}{2}\theta^2 + 2\pi\mathbb{1}_{[0,\pi[}(\theta)\int_{-\pi}^{\theta-\pi}(\pi + t - \theta)d\mu_{p^*}(t) + 2\pi\mathbb{1}_{[-\pi,0[}(\theta)\int_{\theta+\pi}^{\pi}(\pi - t + \theta)d\mu_{p^*}(t).$$

It is clear that Statement 1 is equivalent to the fact that the unique zero of $G_{\mu_{p^*}}$ is 0. Thence, Theorem 5.1 will be an easy consequence of Lemma 4.1 below since we have $\lambda([-\pi,-\pi+t[) = \frac{t}{2\pi}$ for all $0 \le t < \pi$. $\square$

Lemma 4.1. *Let μ be a probability measure on $\mathbb{S}^1$ and $p^* \in \mathbb{S}^1$ be an argmin of F_μ. Then for any $\theta \in [-\pi,\pi[$*

$$G_{\mu_{p^*}}(\theta) = 2\pi \begin{cases} \displaystyle\int_0^\theta \frac{t}{2\pi} - \mu_{p^*}([-\pi,-\pi+t[)dt, & \text{if } 0 \le \theta < \pi, \\ \displaystyle\int_\theta^0 \frac{-t}{2\pi} - \mu_{p^*}([\pi+t,\pi[)dt, & \text{if } -\pi \le \theta < 0 \end{cases}$$

Proof. The probability measure μ can be decomposed as follow,

$$\mu = a\mu^d + (1-a)\mu^\delta, \qquad 0 \le a \le 1, \tag{4.1}$$

where μ^d is a probability measure such that $\mu_d(\{p\}) = 0$ for all $p \in \mathbb{S}^1$ and $\mu_\delta = \sum_{j=1}^{+\infty} \omega_j \delta_{p_j}$ where $\sum_{j=0}^{+\infty} \omega_j = 1$ and the p_j's are in $\mathbb{S}^1$. Hence, we consider the two cases separately : first, when the measure is non atomic, and then, when it is purely atomic. The general case follows immediately in view of equation (4.1).

First, assume that μ is an atomless measure of $\mathbb{S}^1$. Proposition 3.1 ensures that F_μ is continuously differentiable everywhere and the real function $F_{\mu_{p^*}}$ is of class $\mathcal{C}^1$ on $[-\pi,\pi[$. Formula (3.1) and the fundamental theorem of calculus gives for all $\theta \in [-\pi,\pi[$

$$G_{\mu_{p^*}}(\theta) = \int_0^\theta t dt - 2\pi \begin{cases} \int_0^\theta \mu_{p^*}([-\pi,-\pi+t[)dt, & \text{if } \quad 0 \le \theta < \pi, \\ \int_\theta^0 \mu_{p^*}([\pi+t,\pi[)dt, & \text{if } \quad -\pi \le \theta < 0. \end{cases} \tag{4.2}$$

Consider now the case where μ is a purely atomic measure. First, we treat the case where the number of mass of Dirac in the sum is finite, i.e $\mu = \sum_{j=1}^n \omega_j \delta_{p_j}$, $n \in \mathbb{N}$. Recall that $F_{\mu_{p^*}}$ is a Lipschitz function on $[-\pi,\pi[$. Proposition 3.1 ensures that the derivative is piecewise continuous and formula (3.1) holds for all $\theta \in [-\pi,\pi[\backslash\{\theta_{\tilde{p}_j}\}_{j=1}^n$, i.e points that have a cut locus of μ-measure 0. Hence for all $\theta \in [-\pi,\pi[$, equation (4.2) holds too.

To treat the case where $\mu = \sum_{j=1}^{+\infty} \omega_j \delta_{x_j}$ we proceed by approximation. Let $\phi(n) = \{j \in \mathbb{N} \mid \omega_j \ge \frac{1}{2^n}\}$ and remark that $\mathrm{Card}(\phi(n)) < +\infty$ for all $n \in \mathbb{N}$ since $\sum_{j=1}^{+\infty} \omega_j = 1$. Then if $\nu_{p^*}^n = \frac{1}{c(n)} \sum_{j\in\phi(n)} \omega_j \delta_{x_j}$, where $c(n) = \sum_{j\in\phi(n)} \omega_j$ is a normalizing constant, we have for all $\theta \in [-\pi,\pi[$,

$$G_{\nu_{p^*}^n}(\theta) = \int_0^\theta t dt - 2\pi \begin{cases} \int_0^\theta \nu_{p^*}^n([-\pi,-\pi+t[)dt, & \text{if } \quad 0 \le \theta < \pi, \\ \int_\theta^0 \nu_{p^*}^n([\pi+t,\pi[)dt, & \text{if } \quad -\pi \le \theta < 0. \end{cases}$$

The sequence $(\nu_{p^*}^n)_{n\ge1}$ converges to μ in total variation. By the dominated convergence Theorem for all $\theta_p \in [-\pi,\pi[$, $G_{\nu_{p^*}^n}(\theta)$ converge as $n \to \infty$ to (4.2). $\square$

4.3 The criterion $P(\alpha,\varphi)$

Although the necessary and sufficient condition of Theorem 4.1 is a key step to understand the problem of non uniqueness of the Fréchet mean on $\mathbb{S}^1$, it is of little practice interest: we have to know *a priori* a critical point p^* of the Fréchet functional. In this section, we derive sufficient conditions of existence with no restriction on the support of the probability measure and that are easily usable. Let us introduce the following definition :

Definition 4.1. *Let $f : \mathbb{S}^1 \longrightarrow \mathbb{R}^+$ be a probability density, $p \in \mathbb{S}^1$, $\alpha \in]0,1]$ and $\varphi \in]0,\pi[$. We say that f satisfies the property $P(p,\alpha,\varphi)$ if for all $|\theta| \ge \varphi$*

$$f_p(\theta) \le \frac{1-\alpha}{2\pi}, \tag{4.3}$$

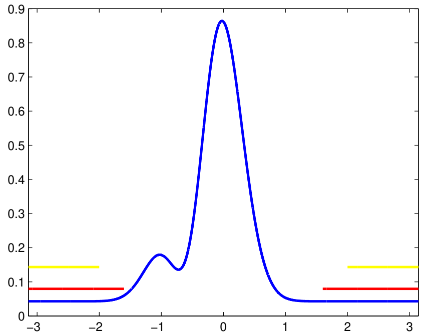


Figure 3: An example of distribution satisfying $P(\alpha, \varphi)$. Plot of the density f_p in blue with the bounds $P(p, 0.1, 2)$ in yellow and $P(p, 0.5, 1.6)$ in red.

where $f_p = f \circ e_p : [-\pi, \pi[\longrightarrow \mathbb{R}^+$. *Moreover, we say that* $f \in P(\alpha, \varphi)$ *if there is a* $p \in \mathbb{S}^1$ *such that* f *satisfies* $P(p, \alpha, \varphi)$*.*

The parameters α and φ control the concentration of μ around p. The idea is to control the mass lying in the complementary of the ball $B(p, \varphi)$, see Figure 3. We have the following properties :

Lemma 4.2. *Let* $f : \mathbb{S}^1 \longrightarrow \mathbb{R}$ *be a probability density on the circle. Then*

1. $P(p, \alpha_1, \varphi_1) \Longrightarrow P(p, \alpha_2, \varphi_2)$ *if* $\alpha_1 \geq \alpha_2$ *and* $\varphi_1 \leq \varphi_2$*.*
2. *Let* $p_1, p_2 \in \mathbb{S}^1$ *and* $\varphi < \frac{\pi}{2}$*. If* $d_{\mathbb{S}^1}(p_1, p_2) < \pi - \varphi$ *then* $P(p_1, \alpha, \varphi) \Longrightarrow P(p_2, \alpha, \varphi + d_{\mathbb{S}^1}(p_1, p_2))$*.*
3. *If* f *satisfies* $P(p, \alpha, \varphi)$ *then* $|m(\mu_p)| \leq \varphi + \frac{1-\alpha}{4\pi}(\pi - \varphi)^2$*.*

Proof. The first proposition is obvious in view of the Definition 4.1.

To prove the second claim suppose that $0 < \theta_{p_2}^{p_1} = -\theta_{p_1}^{p_2} \leq \pi - \varphi$ (the other case is similar) and write

$$f_{p_2}(\theta) = f_{p_1}(\theta + \theta_{p_2}^{p_1})\mathbb{1}_{[-\pi, \pi - \theta_{p_2}^{p_1}[}(\theta) + f_{p_1}(\theta + \theta_{p_2}^{p_1} - 2\pi)\mathbb{1}_{[\pi - \theta_{p_2}^{p_1}, \pi[}(\theta).$$

In particular, it implies that $\pi > \pi - \theta_{p_2}^{p_1} \geq \varphi$ and since $P(p_1, \alpha, \varphi)$ holds, we have $f_{p_2}(\theta) \leq \frac{1-\alpha}{2\pi}$, if $\theta \geq \varphi - \theta_{p_2}^{p_1}$ or $\theta \leq -\varphi - \theta_{p_2}^{p_1}$. This is equivalent to the fact that $P(p_2, \alpha, \min\{| - \varphi - \theta_{p_2}^{p_1}|, |\varphi - \theta_{p_2}^{p_1}|\}) = P(p_2, \alpha, \varphi + \theta_{p_2}^{p_1})$ holds. The case $\varphi - \pi \leq \theta_{p_2}^{p_1} \leq 0$ is similar and we have $P(p_2, \alpha, \varphi - \theta_{p_2}^{p_1})$. Finally recall that $|\theta_{p_2}^{p_1}| = d_{\mathbb{S}^1}(p_1, p_2)$ and the property is proved.

To show the last claim, we only need to consider the case where μ_p has its support on $[0, \pi[$. Indeed, $\mu_p = \omega\mu_p^- + (1 - \omega)\mu_p^+$ where $\mu_p^-([-\pi, 0[) = \mu_p^+([0, \pi[) = 1$ and $0 \leq \omega \leq 1$. It yields that $m(\mu_p) = \int_{\mathbb{R}} td(\omega\mu_p^- + (1-\omega)\mu_p^+) = \int_{\mathbb{R}^+} td(-\omega\mu_p^- + (1-\omega)\mu_p^+) \leq \int_{\mathbb{R}^+} td\mu_p^+$. Then, if the density f_p of μ_p has its support in $[0, \pi[$ we have

$$m(\mu_p) \leq \varphi\left(1 - \int_{\varphi}^{\pi} f_p(t)dt\right) + \int_{\varphi}^{\pi} tf_p(t)dt \leq \varphi + \frac{1-\alpha}{2\pi}\int_{\varphi}^{\pi}(t - \varphi)dt,$$

which gives the result. □

If the density f is sufficiently concentrated around a critical point p^* of F_μ then, this point is the Fréchet mean of μ. More precisely we have the following result,

Proposition 4.1. *Let* μ *be a probability measure with density* $f : \mathbb{S}^1 \longrightarrow \mathbb{R}^+$ *and* $p^* \in \mathbb{S}^1$ *be a critical point of* F_μ*. If* f *satisfies* $P(p^*, \alpha, \varphi)$ *with* $\alpha \in]0, 1]$ *and* $0 < \varphi < \varphi_\alpha = \pi\frac{\sqrt{\alpha}}{1+\sqrt{\alpha}}$ *then,* μ *admits a well defined Fréchet mean at* p^**.*

For all $\alpha \in]0,1]$, we have $\varphi_0 = 0 \leq \varphi_\alpha < \frac{\pi}{2} = \varphi_1$. Note that if $\alpha = 1$ then μ has its support included in the ball $B(p^*, \frac{\pi}{2}) = \{p \in \mathbb{S}, d_{\mathbb{S}^1}(p^*, p) \leq \frac{\pi}{2}\}$ and when $\alpha < 1$, the support of μ can be the entire circle $\mathbb{S}^1$.

Proof. Let $G_{\mu_{p^*}}(\theta) = F_{\mu_{p^*}}(\theta) - F_{\mu_{p^*}}(0)$ for all $\theta \in [-\pi, \pi[$. As the measure μ admits a density f, $G_{\mu_{p^*}}$ is twice differentiable and equation (3.1) implies that $\frac{d^2}{d\theta^2} G_{\mu_{p^*}}(\theta) = 1 - 2\pi f(-\pi + \theta)$, if $0 \leq \theta < \pi$, and $\frac{d^2}{d\theta^2} G_{\mu_{p^*}}(\theta) = 1 - 2\pi f(\pi + \theta)$, if $-\pi \leq \theta < 0$. Since $f \in P(p^*, \alpha, \varphi)$, the function $G_{\mu_{p^*}}$ is convex on $[-\pi + \varphi, \pi - \varphi]$ and has a unique minimum at 0. Let us show that 0 is the only argmin of $F_{\mu_{p^*}}$ on $[-\pi, \pi[$. If $\theta \in [\pi - \varphi, \pi[$, we have thanks to Lemma 4.1

$$G_{\mu_{p^*}}(\theta) = G_{\mu_{p^*}}(\pi - \varphi) + \int_{\pi-\varphi}^{\theta} t - 2\pi \mu_{p^*}([-\pi, -\pi + t[)dt.$$

Since $f \in P(p^*, \alpha, \varphi)$ we have $G_{\mu_{p^*}}(\pi - \varphi) \geq \frac{\alpha}{2}(\pi - \varphi)^2$ and the second term is bounded from below by $\int_{\pi-\varphi}^{\pi} t - 2\pi\nu([-\pi, -\pi + t[)dt$ where $\nu = \frac{1}{2}(\delta_{-\varphi} + \delta_\varphi)$. It yields for $\theta \in [\pi - \varphi, \pi[$,

$$G_{\mu_{p^*}}(\theta) \geq \frac{1}{2}(\alpha(\pi - \varphi)^2 - \varphi^2). \tag{4.4}$$

The right hand side of (4.4) is strictly positive if $\varphi < \varphi_\alpha = \pi\frac{\sqrt{\alpha}}{1+\sqrt{\alpha}}$. Similarly, the same condition implies $G_{\mu_{p^*}}(\theta) > 0$ for $\theta \in [-\pi, -\pi + \varphi[$. □

We are now able to define a functional class of densities that admit a well defined Fréchet mean *without* restriction on the support of the measure.

Theorem 4.2. *Let $0 < \delta < \frac{1}{2}$ be a parameter of concentration and μ a probability measure with density $f \in P(\alpha, \varphi)$ (see Definition 4.1) with*

$$\alpha_\delta \leq \alpha \leq 1 \qquad \text{and} \qquad \varphi \leq \delta\varphi_\alpha$$

where α_δ be the square of the root of $(5 - 6\delta + \delta^2)X^3 + (1 - \delta^2)X^2 - (2\delta + 1)X - 1$ that lies in $]0,1]$ and $\varphi_\alpha = \pi\frac{\sqrt{\alpha}}{1+\sqrt{\alpha}}$. Then μ admits a well defined Fréchet means.

Firstly, remark that there is no need to know a critical point *a priori*. Secondly, the parameter δ controls the concentration of f via the inequality $\alpha_\delta < \alpha$ and $\varphi \leq \delta\varphi_\alpha$. There is a tradeoff between α and the possible value of φ: the smaller α is (i.e the less f is concentrated) the smaller φ must be (i.e we need to control the value of the density on a bigger interval). In Tabular 1 we give examples of numerical values. Note that the column corresponding to $\delta = 0$ is given as a reference only as the set $P(\alpha_\delta, \delta\varphi_\alpha)$ is empty for this values of δ.

$\delta =$	0	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{1}{3}$	$\frac{1}{2}$
$\alpha_\delta \leq$	0.39	0.46	0.54	0.69	1
$\delta\varphi_{\alpha_\delta} \geq$	0	0.12	0.26	0.47	$\frac{\pi}{4}$

Table 1: Some values of α_δ and $\delta\varphi_{\alpha_\delta}$ depending on $\delta \in]0, \frac{1}{2}[$.

Proof. We show that under the hypothesis of the Theorem 4.2, there is a critical point p^* of F_μ satisfying $d_{\mathbb{S}^1}(p, p^*) \leq (1 - \delta)\varphi_\alpha$ where $p \in \mathbb{S}^1$ is a point satisfiying $f \in P(p, \alpha, \varphi)$. Thence, by Lemma 4.2, f belongs to $P(p^*, \alpha, \delta\varphi_\alpha + (1 - \delta)\varphi_\alpha) = P(p^*, \alpha, \varphi_\alpha)$ and Proposition 4.1 will ensure that p^* is the Fréchet mean of μ.

In the rest of the proof, we show that there is a $p^* \in \mathbb{S}^1$ such that $\frac{d}{d\theta}F_{\mu_p}(\theta^p_{p^*}) = 0$ with $d_{\mathbb{S}^1}(p, p^*) \leq (1-\delta)\varphi_\alpha$. To this end, suppose that $m(\mu_p) \geq 0$ (the case $m(\mu_p) < 0$ is similar). Equation (3.5) implies that $\frac{d}{d\theta}F_{\mu_p}(0) = -m(\mu_p) \leq 0$, and let us check that, under the hypothesis of Theorem 4.2, we have $\frac{d}{d\theta}F_{\mu_p}((1-\delta)\varphi_\alpha) \leq 0$. Since $\frac{d}{d\theta}F_{\mu_p}$ is a continuous function, the intermediate value Theorem will ensure the existence of a critical point p^* such that $|\theta^p_{p^*}| \leq (1-\delta)\varphi_\alpha$. Equation (3.1) gives

$$\frac{d}{d\theta}F_{\mu_p}((1-\delta)\varphi_\alpha) = (1-\delta)\varphi_\alpha - 2\pi\mu_p([-\pi, -\pi + (1-\delta)\varphi_\alpha[) - m(\mu_p).$$

We have $-2\pi\mu_p([-\pi, -\pi+(1-\delta)\varphi_\alpha[) \geq (\alpha-1)(1-\delta)\varphi_\alpha$ since $f \in P(p, \alpha, \delta\varphi_\alpha)$ and $|-\pi + (1-\delta)\varphi_\alpha| \geq \delta\varphi_\alpha$. Moreover, $-m(\mu) \geq \delta\varphi_\alpha - \frac{1-\alpha}{4\pi}(\pi - \delta\varphi_\alpha)^2$ by Lemma 4.2 Statement 3. It gives,

$$\frac{d}{d\theta}F_{\mu_p}((1-\delta)\varphi_\alpha) \geq \pi\frac{(5-6\delta+\delta^2)\alpha\sqrt{\alpha} + (1-\delta^2)\alpha - (2\delta+1)\sqrt{\alpha} - 1}{1+\sqrt{\alpha}}$$

This quantity is positive as soon as $1 \geq \alpha > \alpha_\delta$, where $\sqrt{\alpha_\delta}$ is the root of the polynomial $X \longmapsto (5-6\delta+\delta^2)X^3 + (1-\delta^2)X^2 - (2\delta+1)X - 1$ that lies in $]0, 1]$. It is easy to see that $\alpha_{\frac{1}{2}} = 1$ and numerical experiments show that the (increasing) function $\delta \longmapsto \alpha_\delta$ takes its value in $]0.39, 1[$ for $\delta \in]0, \frac{1}{2}[$. $\square$

5 Fréchet mean of an empirical measure

5.1 Existence

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with value in $(\mathbb{S}^1, d_{\mathbb{S}^1})$ and of probability distribution μ. The empirical measure is defined as usual by

$$\mu^n = \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}$$

and we note p_n^* the empirical Fréchet mean defined as the unique argmin of $F_{\mu^n}(p) = \frac{1}{2n}\sum_{i=1}^n d^2_{\mathbb{S}^1}(p, X_i)$, $p \in \mathbb{S}^1$. In [18] a strong law of large number is given for the empirical Fréchet mean in a semi metric space which is the case of $(\mathbb{S}^1, d_{\mathbb{S}^1})$. In particular, if p_n^* exists for each $n \in \mathbb{N}$, the empirical Fréchet mean is a consistent estimator of the Fréchet mean. Indeed, the empirical Fréchet mean is well defined almost surely for a wide class of probability measures as the following fact from [2] Remark 2.6 shows,

Lemma 5.1. *Let μ be a non atomic probability measure on the circle, i.e satisfying $\mu(\{p\}) = 0$ for all $p \in \mathbb{S}^1$. Then for all $n \in \mathbb{N}$ the empirical Fréchet mean exists almost surely.*

Hence, the empirical Fréchet mean p_n^* of a probability measure μ can be computed even if μ does not possess a well defined Fréchet mean.

5.2 Consistency

If the Fréchet mean p^* of μ is well defined, we study the rate of convergence of the empirical Fréchet mean p_n^* to p^*.

Proposition 5.1. *Let μ be a measure with density $f : \mathbb{S}^1 \longrightarrow \mathbb{R}$ that admits a well defined Fréchet mean p^*. Then, there exists a strictly increasing function $\rho :]0, \pi[\longrightarrow]0, +\infty[$ such that for all $p \in \mathbb{S}^1$, $p \neq p^*$*

$$F_\mu(p) - F_\mu(p^*) \geq \rho(d_{\mathbb{S}^1}(p, p^*)).$$

If p_n^* denotes the empirical Fréchet mean, we have for all $x > 0$

$$\mathbb{P}\left(\rho(d_{\mathbb{S}^1}(p_n^*, p^*)) \geq C(s)\sqrt{\frac{x}{n}}\right) \leq 2e^{-x}. \tag{5.1}$$

where $s = \max\{|x-y|\,,\ x, y \in \text{support}(\mu)\}$ and $C(s) = (4\pi^2 + 4\pi^2 s + 2s) \leq 4\pi(2\pi^2 + \pi + 1)$.

The function ρ in the statement of the Proposition 5.1 determines the rate of convergence of p_n^* to p^*.

Proof. The first claim about the lower bound ρ is a direct consequence of the following Lemma:

Lemma 5.2. *Let $f : [-\pi, \pi[\longrightarrow \mathbb{R}^+$ be a continuous function on $[-\pi, \pi[$ (see Section 3). If f vanishes at a unique point $\theta_0 \in [-\pi, \pi[$, then there exists a strictly increasing function $\rho :]0, \pi[\longrightarrow]0, +\infty[$ such that for all $\theta \in [-\pi, \pi[\backslash\{0\}$,*

$$f(\theta) \geq \rho(d_{\mathbb{S}^1}(\theta_0, \theta)).$$

Proof. As f is the restriction on $[-\pi, \pi[$ of a continuous periodic function we can assume, without loss of generality, that $\theta_0 = 0$. Then, define for all $\theta \in]0, \pi[$

$$\rho(\theta) = \frac{1}{|\theta|}\int_0^\theta g(t)dt, \qquad \text{where } g(t) = \min\left\{f(-t), f(t), \min_{|\tau|>t} f(\tau)\right\}, \text{ for all } t \in [-\pi, \pi[. \qquad \square$$

We now focus on the proof of the concentration inequality (5.1) which is divided in two steps: first we show the uniform convergence in probability of $F_{\mu_p^n}$ to F_{μ_p} and then, we deduce the convergence of their argmins by using the lower bound given by the function ρ. Equation (3.1) implies that

$$\begin{aligned}
2\sup_{\theta\in[-\pi,\pi[}\left|F_{\mu_p^n}(\theta) - F_{\mu_p}(\theta)\right| &\leq 2\pi \sup_{\theta\in[-\pi,\pi[}\left|\frac{d}{d\theta}F_{\mu_p^n}(\theta) - \frac{d}{d\theta}F_{\mu_p}(\theta)\right| + 2\left|F_{\mu_p^n}(0) - F_{\mu_p}(0)\right| \\
&\leq 4\pi^2\left|m(\mu_p) - m(\mu_p^n)\right| + 2\left|m_2(\mu_p) - m_2(\mu_p^n)\right| \qquad (5.2)\\
&\quad + 4\pi^2 \sup_{\theta\in[-\pi,\pi[}\left|\mu_p([-\pi,\theta[) - \mu_p^n([-\pi,\theta[)\right| \qquad (5.3)
\end{aligned}$$

where $m_2(\nu) = \int_{\mathbb{R}} t^2 d\nu(t)$, for a measure ν on $\mathbb{R}$. The term (5.3) can be controlled in probability using the Dvoretzky-Kiefer-Wolfowitz inequality (see e.g [15]), and we have for all $x > 0$, $\mathbb{P}\Big(4\pi^2 \sup_{\theta\in[-\pi,\pi[}\left|\mu_p([-\pi,\theta[) - \mu_p^n([-\pi,\theta[)\right| \geq 4\pi^2\sqrt{\frac{x}{n}}\Big) \leq 2e^{-x}$. For the terms of (5.2) which involve the first and second moment of μ_p and μ_p^n, we use an Hoeffding type inequality which gives for all $x > 0$, $\mathbb{P}\Big(4\pi^2\left|m(\mu_p) - m(\mu_p^n)\right| + 2\left|m_2(\mu_p) - m_2(\mu_p^n)\right| \geq s(4\pi^2 + 2s)\sqrt{\frac{x}{n}}\Big) \leq 2e^{-x}$, where $s = \max\{|x-y|,\ x, y \in \text{support } \mu\}$ is the diameter of the support of μ. Combining these two concentration inequalities we have for all $x > 0$,

$$\mathbb{P}\left(2\sup_{\theta\in\mathbb{R}}\left|F_{\mu_p}(\theta) - F_{\mu_p^n}(\theta)\right| \geq (4\pi^2 + 4\pi^2 s + 2s)\sqrt{\frac{x}{n}}\right) \leq 2e^{-x}. \qquad (5.4)$$

We now use a classical inequality in M-estimation, $\left|F_{\mu_p}(\theta_{p_n^*}) - F_{\mu_p}(\theta_{p^*})\right| \leq 2\sup_{\theta\in[-\pi,\pi[}\left|F_{\mu_p^n}(\theta) - F_{\mu_p}(\theta)\right|$. By Lemma 5.2, there exists an increasing function $\rho : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ such that $F_{\mu_p}(\theta_{p_n^*}) - F_{\mu_p}(\theta_{p^*}) \geq \rho(d_{\mathbb{S}^1}(\theta_{p_n^*}, \theta_{p^*}))$. Plugging this in equation (5.4) we have,

$$\mathbb{P}\left(\rho(d_{\mathbb{S}^1}(\theta_{p_n^*}, \theta_{p^*})) \geq (4\pi^2 + 4\pi^2 s + 2s)\sqrt{\frac{x}{n}}\right) \leq 2e^{-x},$$

and the proof of Proposition 5.1 is completed. $\square$

The function ρ that appears in the statement of Proposition 5.1 can be explicitly computed if the density $f \in P(\alpha, \varphi)$. The parameter $\alpha \in]0, 1]$ can be interpreted as a measure of the convexity of F_{μ_p} on the interval $[-\varphi, \varphi]$. For example, if $\alpha = 1$ and $\varphi = \varphi_\alpha = \frac{\pi}{2}$, then μ has its support contained in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and F_{μ_p} is quadratic on $[-\frac{\pi}{2}, \frac{\pi}{2}]$.

Proposition 5.2. *Let μ be a probability measure with density $f : \mathbb{S}^1 \longrightarrow \mathbb{R}^+$ satisfying the hypothesis of Theorem 4.2. Note p^* the Fréchet mean of μ and for all $x > 0$ we have*

$$\mathbb{P}\left(d_{\mathbb{S}^1}(p_n^*, p^*) \geq \sqrt{B(\alpha, \varphi)}\left(\frac{x}{n}\right)^{\frac{1}{4}}\right) \leq 2e^{-x},$$

where $B(\alpha, \varphi) = C\max\left\{\frac{\pi^2}{\gamma(\alpha,\varphi)}, \frac{2}{\alpha}\right\}$ with $\gamma(\alpha, \varphi) = \frac{1}{2}(\alpha(\pi - \varphi)^2 - \varphi^2)$ and $C = 4\pi(2\pi^2 + \pi + 1)$.

Proof. This result is a direct consequence of Proposition 5.1 and we only have to find a strictly increasing function $\rho : [0, \pi] \longrightarrow \mathbb{R}^+$ satisfying for all $\theta \in [-\pi, \pi[$, $F_\mu(p) - F_\mu(p^*) \geq \rho(d_{\mathbb{S}^1}(p, p^*))$. As μ_{p^*} admits a density f_{p^*}, the Fréchet functional $F_{\mu_{p^*}}$ is twice differentiable and $f \in P(p^*, \alpha, \varphi_\alpha)$, see proof of Theorem 4.2. Thus, for all $\theta \in [-\pi + \varphi_\alpha, \varphi_\alpha - \pi]$, a second order Taylor expansion of $F_{\mu_{p^*}}$ at 0 ensures that for some $\tilde{\theta} \in [-\pi + \varphi_\alpha, \pi - \varphi_\alpha]$,

$$F_{\mu_{p^*}}(\theta) - F_{\mu_{p^*}}(0) = \frac{1}{2}\theta^2 \frac{d^2}{d\theta^2} F_{\mu_{p^*}}(\tilde{\theta}) \geq \frac{\alpha}{2}\theta^2.$$

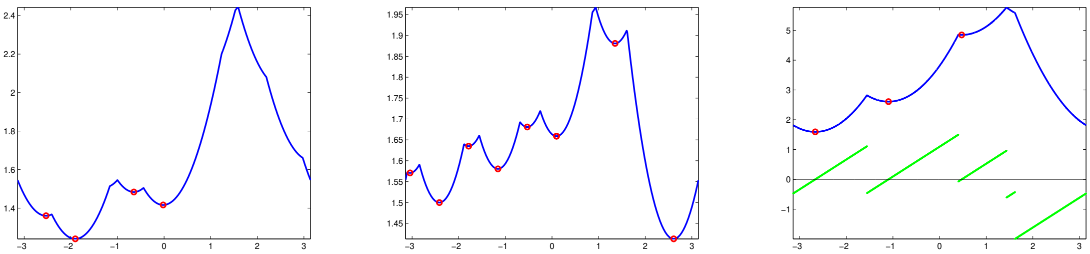
For all $\theta \in [-\pi, -\varphi_\alpha[\cup]\varphi_\alpha, \pi[$ we have by inequality (4.4),

$$F_{\mu_{p^*}}(\theta) - F_{\mu_{p^*}}(0) \geq \frac{1}{2}(\alpha(\pi - \varphi_\alpha)^2 - \varphi_\alpha^2) = \gamma(\alpha, \varphi_\alpha) > 0.$$

Then, let $\rho(t) = t^2 \min\{\frac{\gamma(\alpha,\varphi_\alpha)}{\pi^2}, \frac{\alpha}{2}\} \geq t^2 \min\{\frac{\gamma(\alpha,\varphi)}{\pi^2}, \frac{\alpha}{2}\}$ for any $\varphi \leq \varphi_\alpha$. $\square$

5.3 Computation

Computation of the Fréchet mean of a general probability measure may not be an easy task as it is a *global* optimization problem. In practice the Fréchet functional is not a convex function and a gradient descent algorithm will only give a *local* minimum which depends on the initialization point.



(a) (b) (c)

Figure 4: (a) and (b) Plots of F_{μ^n} where $n = 10$ and $\mu^n = \sum_{i=1}^n \delta_{X_i}$ where the X_i i.i.d of uniform law λ. The red points are the local minima computed with the described algorithm. (c) In blue: plot of F_{μ^n} where $n = 4$. In green: the derivative of F_{μ^n}. The critical points are given by the intersection between the green curve and the x-axis in black.

The Fréchet functional of an empirical measure is a continuous piecewise quadratic function as it can be written $F_{\mu_p^n}(\theta) = \frac{1}{2n}\sum_{i=1}^n \left(\theta - \theta^p_{X_i} + 2\pi(\mathbb{1}_{[\pi+\theta,\pi[}(\theta^p_{X_i})\mathbb{1}_{[-\pi,0[}(\theta) - \mathbb{1}_{[-\pi,\theta-\pi[}(\theta^p_{X_i})\mathbb{1}_{[0,\pi[}(\theta))\right)^2$. This formula together with Corollary 3.1 implies that the regular critical points (i.e point with no mass at their cut locus) of F_μ are precisely the local minima of F_μ. Moreover, the cumulative distribution function of μ^n is, here, piecewise constant with jumps of size $\frac{1}{n}$ and we have

$$\mu_p^n([-\pi, t[) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{[-\pi,t[}(\theta^p_{X_i}) = \frac{1}{n}\mathrm{Card}\{\theta^p_{X_i} < t\}.$$

Thence, the derivative of the empirical Fréchet functional is piecewise linear and to find the critical points amounts to solve n affine relations given by equation (3.1), see also Figure 4(c) for an illustration. Note, that in practice, there are less than n solutions, see Figure 4(a) and 4(b).

The following algorithm takes as input the values $\{X_i\}_{i=1}^n$ and returns the Fréchet mean of $\mu^n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$.

Initialization Step: Choose an arbitrarily point $p \in \mathbb{S}^1$.
Compute the coordinates $\{\theta^p_{X_i}\}_{i=1}^n$ and reorder them is increasing order. We denote $\tau_0^- = -\pi \leq \tau_1^- \leq \tau_2^- \leq \ldots \leq \tau_{n_1}^- < 0 = \tau_{n_1+1}^-$ the n_1 negative sorted terms and $\tau_0^+ = \pi > \tau_1^+ \geq \ldots \geq \tau_{n_2}^+ \geq 0 = \tau_{n_2+1}^+$ the $n_2 = n - n_1$ positive sorted terms.
Compute the mean $m(\mu_p^n) = \frac{1}{n}(\tau_1^- + \ldots + \tau_{n_1}^- + \tau_1^+ + \ldots + \tau_{n_2}^+)$ and initialize $\theta^p_{p^*}$ to 0, says.

`# The first step compares all the local minima in` $[0, \pi[$
Step 1: For i from 0 to n_1 do
`#` $\theta^p_{p^*,new}$ `is the candidate to be a critical point between` τ_i^- `and` τ_{i+1}^-
    Let $\theta^p_{p^*,new} = 2\pi\frac{i}{n} + m(\mu_p^n)$
`# verify if` $\theta^p_{p^*,new}$ `is a critical point and then test its value.  If better, keep it.`
    if $\tau_i^- + \pi \leq \theta^p_{p^*,new} \leq \tau_{i+1}^- + \pi$ and $F_{\mu_p^n}(\theta^p_{p^*,new}) \leq F_{\mu_p^n}(\theta^p_{p^*})$ then $\theta^p_{p^*} := \theta^p_{p^*,new}$ end if
end for.
`# The Step 2 is the same as Step 1 but for local minima in` $[-\pi, 0[$
Step 2: For i = 0 to n_2 do
    Let $\theta^p_{p^*,new} = -2\pi\frac{i}{n} + m(\mu_p^n)$
    if $\tau_{i+1}^+ - \pi \leq \theta^p_{p^*,new} \leq \tau_i^+ - \pi$ and $F_{\mu_p^n}(\theta^p_{p^*,new}) \leq F_{\mu_p^n}(\theta^p_{p^*})$ then $\theta^p_{p^*} := \theta^p_{p^*,new}$ end if
end for.
`# The value of` $\theta^p_{p^*}$ `is the best argmin`
Output: Return $p^* = e_p(\theta^p_{p^*})$.

This algorithm can be extended to more general measures that the empirical one. The approach will be the same: find the critical points of the Fréchet functional with formula 3.1 and compare the values of the local minima. Unfortunately, there may be some computational issues as general cumulative distribution function will be not piecewise constant anymore.

6 Conclusion

It is not straightforward to extend criterion such as the one given in Theorem 4.1 to more general spaces, e.g. for the n dimensional sphere $\mathbb{S}^n$. Recall that the circle $\mathbb{S}^1$ is a flat space in the sense that it is locally isometric to the Euclidean space $\mathbb{R}$. Then, the only phenomenon that induces uniqueness issues of the Fréchet mean is the presence of a cut locus. The criterion presented in this note relies on an explicit formula for the gradient of the Fréchet mean. Curvature has an extra effect on the metric and makes difficult to derive exact computation on the Fréchet functional and its gradient. Moreover, it is not clear if the role played by the uniform measure as a benchmark in the well definiteness of the Fréchet mean in $\mathbb{S}^1$ can be extended to n-spheres or non flat manifolds.

Acknowledgements. The author would like to thank Dominique Bakry for help and encouragement during the writing of this paper and Jérémie Bigot for his careful reading of the manuscript. The author acknowledges the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-JCJC-SIMI1 DEMOS.

References

[1] Bijan Afsari. Riemannian L^p center of mass: existence, uniqueness, and convexity. *Proc. Amer. Math. Soc.*, 139(2):655–673, 2011.

[2] Rabi Bhattacharya and Vic Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds. i. *The Annals of Statistics*, 31(1):pp. 1–29, 2003.

[3] Samuel R. Buss and Jay P. Fillmore. Spherical averages and applications to spherical splines and interpolation. *ACM Trans. Graph.*, 20:95–126, April 2001.

[4] José Manuel Corcuera and Wilfrid S. Kendall. Riemannian barycentres and geodesic convexity. *Math. Proc. Cambridge Philos. Soc.*, 127(2):253–269, 1999.

[5] Michel Émery and G. Mokobodzki. Sur le barycentre d'une probabilité dans une variété. In *Séminaire de Probabilités, XXV*, volume 1485 of *Lecture Notes in Math.*, pages 220–233. Springer, Berlin, 1991.

[6] N. I. Fisher. *Statistical analysis of circular data*. Cambridge University Press, Cambridge, 1993.

[7] Maurice Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré*, 10:215–310, 1948.

[8] H. Karcher. Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.*, 30(5):509–541, 1977.

[9] David Kaziska and Anuj Srivastava. The karcher mean of a class of symmetric distributions on the circle. *Statistics & Probability Letters*, 78(11):1314 – 1316, 2008.

[10] D. G. Kendall, D. Barden, T. K. Carne, and H. Le. *Shape and shape theory*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1999.

[11] Huiling Le. On the consistency of procrustean mean shapes. *Adv. in Appl. Probab.*, 30(1):53–63, 1998.

[12] Huiling Le. Locating Fréchet means with application to shape spaces. *Adv. in Appl. Probab.*, 33(2):324–338, 2001.

[13] Huiling Le. Estimation of Riemannian barycentres. *LMS J. Comput. Math.*, 7:193–200 (electronic), 2004.

[14] Kanti V. Mardia and Peter E. Jupp. *Directional statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 2000. Revised reprint of ıt Statistics of directional data by Mardia [MR0336854 (49 #1627)].

[15] P. Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.*, 18(3):1269–1283, 1990.

[16] J. M. Oller and J. M. Corcuera. Intrinsic analysis of statistical estimation. *Ann. Statist.*, 23(5):1562–1581, 1995.

[17] Xavier Pennec. Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. *J. Math. Imaging Vision*, 25(1):127–154, 2006.

[18] Herbert Ziezold. On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians (Tech. Univ. Prague, Prague, 1974), Vol. A*, pages 591–602. Reidel, Dordrecht, 1977.