



# The Fisher-Rao geometry of beta distributions applied to the study of canonical moments

Alice Le Brigant, Stéphane Puechmorel

## ► To cite this version:

Alice Le Brigant, Stéphane Puechmorel. The Fisher-Rao geometry of beta distributions applied to the study of canonical moments. 2019. hal-02100897

HAL Id: hal-02100897

<https://hal.archives-ouvertes.fr/hal-02100897>

Submitted on 16 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The Fisher-Rao geometry of beta distributions applied to the study of canonical moments

Alice Le Brigant and Stéphane Puechmorel

## Abstract

This paper studies the Fisher-Rao geometry on the parameter space of beta distributions. We derive the geodesic equations and the sectional curvature, and prove that it is negative. This leads to uniqueness for the Riemannian centroid in that space. We use this Riemannian structure to study canonical moments, an intrinsic representation of the moments of a probability distribution. Drawing on the fact that a uniform distribution in the regular moment space corresponds to a product of beta distributions in the canonical moment space, we propose a mapping from the space of canonical moments to the product beta manifold, allowing us to use the Fisher-Rao geometry of beta distributions to compare and analyze canonical moments.

## 1 Introduction

The differential geometric approach to probability theory and statistics has met increasing interest in the past years, from the theoretical point of view as well as in applications. In this approach, probability distributions are seen as elements of a differentiable manifold, on which a metric structure is defined through the choice of a Riemannian metric. Two very important ones are the Wasserstein metric, central in optimal transport, and the Fisher-Rao metric (also called Fisher information metric), essential in information geometry. Unlike optimal transport, information geometry is foremost concerned with parametric families of probability distributions, and defines a Riemannian structure on the parameter space using the Fisher information matrix [5]. In parameter estimation, the Fisher information can be interpreted as the quantity of information on the unknown parameter contained in the model. As the Hessian of the well-known Kullback-Leibler divergence, it measures through the notion of curvature the capacity to distinguish between two different values of the parameter. Rao [9] showed that it could be used to locally define a scalar product on the space of parameters, interpretable as a Riemannian metric. An important feature of this metric is that it is invariant under any diffeomorphic change of parameterization. In fact, considering the infinite-dimensional space of probability densities on a given manifold  $M$ , there is a unique metric, which also goes by the name Fisher-Rao, that is invariant with respect to the action of the diffeomorphism group of  $M$  [3, 2]. This metric induces the regular Fisher information metric on the finite dimensional submanifolds corresponding to the parameterized statistical models of interest in information geometry. Arguably the most famous example of Fisher-Rao geometry of a statistical model is that of the Gaussian model, which is hyperbolic. The multivariate Gaussian case, among other models, has also received a lot of attention [1, 11].

In this work, we are interested in beta distributions, a family of probability measures on  $[0, 1]$  used to model random variables defined on a compact interval in a wide variety of applications. Up to our knowledge, the information geometry of beta distributions has not yet received much attention. In this paper, we give new results and properties for this geometry, and its curvature in

particular. Interestingly, this geometric framework yields new by-product tools to study the set of all moments of compactly supported probability measures on the real line. This is achieved through the so-called canonical moments representation [4], an alternative to the usual moment representation of a probability distribution that satisfies interesting symmetries and invariance properties.

The paper is organized as follows. Section 2 deals with the study of the Fisher-Rao geometry of beta distributions. We derive the geodesic equations, prove that sectional curvature is negative, give some bounds and observe a geometrical manifestation of the central limit theorem. Section 3 deals with the application to canonical moments. After a brief presentation of these objects, we propose a representation in the product beta manifold, allowing us to use the Fisher-Rao geometry of beta distributions to compare and analyze canonical moments.

## 2 Geometry of the beta manifold

### 2.1 The beta manifold

Information geometry is concerned with parametric families of probability distributions, i.e. sets of distributions with densities with respect to a common dominant measure  $\mu$  parameterized by a parameter  $\theta$  member of a given set  $\Theta$ . That is, a collection of measures of the kind

$$\mathcal{P}_\Theta = \{p_\theta \mu, \theta \in \Theta\}.$$

We assume that  $\Theta$  is a non empty open subset of  $\mathbb{R}^d$ . Associated to any such family is the Fisher information matrix, defined for all  $\theta$  as

$$I(\theta) = \left[ E \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln p(X; \theta) \right) \right]_{1 \leq i, j \leq d}.$$

As an open subset of  $\mathbb{R}^d$ ,  $\Theta$  is a differentiable manifold and can be equipped with a Riemannian metric using this quantity. This gives the Fisher information metric on the parameter space  $\Theta$

$$G_\theta^F(u, v) = u^t I(\theta) v, \quad \theta \in \Theta, \quad u, v \in T_\theta \Theta \simeq \mathbb{R}^d,$$

where  $u^t$  denotes the transpose of the vector  $u$ . By extension, we talk of the Fisher geometry of the parameterized family  $\mathcal{P}_\Theta$ , and of the Riemannian manifold  $(\mathcal{P}_\Theta, g^F)$ .

In this paper, we are interested in beta distributions, a family of probability distributions on  $[0, 1]$  with density with respect to the Lebesgue measure parameterized by two positive scalars  $\alpha, \beta > 0$

$$p_{\alpha, \beta}(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad x \in [0, 1].$$

We consider the Riemannian manifold composed of the parameter space  $\Theta = \mathbb{R}_+^* \times \mathbb{R}_+^*$  and the Fisher metric  $g^F$ , and by extension denote by beta manifold the pair  $(\mathcal{B}, g^F)$ , where  $\mathcal{B}$  is the family of beta distributions

$$\mathcal{B} = \{B(\alpha, \beta) = p_{\alpha, \beta}(\cdot) dx, \alpha > 0, \beta > 0\}.$$

Here  $dx$  denotes the Lebesgue measure on  $[0, 1]$ . The distance between two beta distributions is then defined as the geodesic distance associated to the Fisher metric in the parameter space

$$d^F(B(\alpha, \beta), B(\alpha', \beta')) = \inf_\gamma \int_0^1 \sqrt{g^F(\dot{\gamma}(t), \dot{\gamma}(t))} dt,$$

where the infimum is taken over all paths  $\gamma : [0, 1] \rightarrow \Theta$  such that  $\gamma(0) = (\alpha, \beta)$  and  $\gamma(1) = (\alpha', \beta')$ .

## 2.2 The Fisher-Rao metric

The beta distributions are part of an exponential family and so the general term of the Fisher-Rao metric depends on second order derivatives of the underlying potential function. Denoting by  $g^F(\alpha, \beta)$  the matrix form of  $G_{\alpha, \beta}^F$ ,

$$g^F(\alpha, \beta) = -\text{Hess } \varphi(\alpha, \beta), \quad (1)$$

where  $\varphi$  is the potential function

$$\varphi(\alpha, \beta) = \ln \Gamma(\alpha) + \ln \Gamma(\beta) - \ln \Gamma(\alpha + \beta).$$

Proposition 1 describes the metric tensor and Proposition 2 the geodesic equations.

**Proposition 1.** *The matrix representation of the Fisher-Rao metric on the space of beta distributions is given by*

$$g^F(\alpha, \beta) = \begin{bmatrix} \psi'(\alpha) - \psi'(\alpha + \beta) & -\psi'(\alpha + \beta) \\ -\psi'(\alpha + \beta) & \psi'(\beta) - \psi'(\alpha + \beta) \end{bmatrix}$$

where  $\psi$  denotes the digamma function, i.e.  $\psi(x) = \frac{d}{dx} \ln \Gamma(x)$ .

*Proof.* This follows from straightforward computations.  $\square$

**Proposition 2.** *The geodesic equations are given by*

$$\begin{aligned} \ddot{\alpha} + a(\alpha, \beta)\dot{\alpha}^2 + b(\alpha, \beta)\dot{\alpha}\dot{\beta} + c(\alpha, \beta)\dot{\beta}^2 &= 0, \\ \ddot{\beta} + a(\beta, \alpha)\dot{\beta}^2 + b(\beta, \alpha)\dot{\alpha}\dot{\beta} + c(\beta, \alpha)\dot{\alpha}^2 &= 0, \end{aligned}$$

where

$$\begin{aligned} a(x, y) &= \frac{1}{2d(x, y)}(\psi''(x)\psi'(y) - \psi''(x)\psi'(x+y) - \psi'(y)\psi''(x+y)), \\ b(x, y) &= -\frac{1}{d(x, y)}\psi'(y)\psi''(x+y), \\ c(x, y) &= \frac{1}{2d(x, y)}(\psi''(y)\psi'(x+y) - \psi'(y)\psi''(x+y)), \\ d(x, y) &= \psi'(x)\psi'(y) - \psi'(x+y)(\psi'(x) + \psi'(y)). \end{aligned}$$

*Proof.* The geodesic equations are given by

$$\begin{aligned} \ddot{\alpha} + \Gamma_{\alpha\alpha}^{\alpha}\dot{\alpha}^2 + 2\Gamma_{\alpha\beta}^{\alpha}\dot{\alpha}\dot{\beta} + \Gamma_{\beta\beta}^{\alpha}\dot{\beta}^2 &= 0 \\ \ddot{\beta} + \Gamma_{\alpha\alpha}^{\beta}\dot{\alpha}^2 + 2\Gamma_{\alpha\beta}^{\beta}\dot{\alpha}\dot{\beta} + \Gamma_{\beta\beta}^{\beta}\dot{\beta}^2 &= 0 \end{aligned} \quad (2)$$

where the  $\Gamma_{ij}^k$ 's denote the Christoffel symbols of the second kind. These can be obtained from the Christoffel symbols of the first kind  $\Gamma_{ij}^k$  and the coefficients  $g^{ij}$  of the inverse of the metric matrix

$$\Gamma_{ij}^k = \Gamma_{ijl}g^{kl}.$$

Here we have used the Einstein summation convention. Since the Fisher metric is a Hessian metric, the Christoffel symbols of the first kind can be obtained as

$$\Gamma_{ijk} = \frac{1}{2}\varphi_{ijk},$$

where  $\varphi$  is the potential function (1). Straightforward computation yields the desired equations.  $\square$

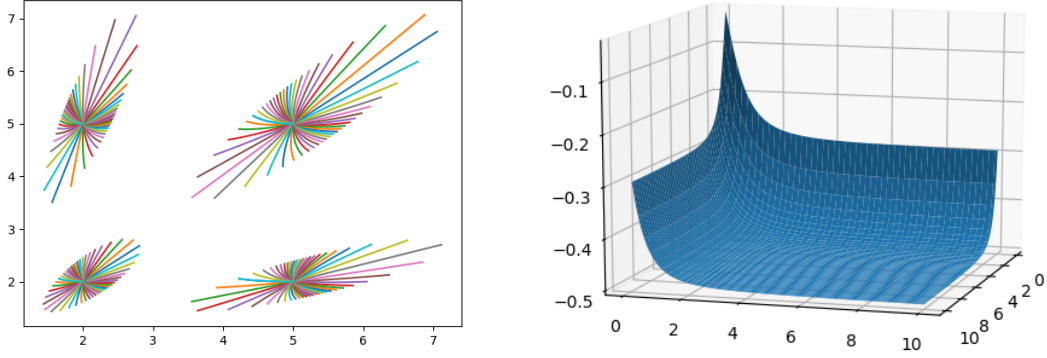


Figure 1: Geodesic balls (left) and sectional curvature (right) of the beta manifold.

Notice that when  $\alpha = \beta = \gamma$ , both geodesic equations (2) yield a unique ordinary differential equation

$$\ddot{\gamma} + (a(\gamma, \gamma) + b(\gamma, \gamma) + c(\gamma, \gamma))\dot{\gamma}^2 = 0.$$

The line of equation  $\alpha = \beta$  is therefore a geodesic for the Fisher metric. More precisely, we have the following corollary obtained directly from Proposition 2.

**Corollary 1.** *The line of equation  $\alpha(t) = \beta(t) = \gamma(t)$ , where*

$$\ddot{\gamma} + \frac{\psi'(\gamma)\psi''(\gamma) - 4\psi'(\gamma)\psi''(2\gamma)}{2(\psi'(\gamma)^2 - 2\psi'(\gamma)\psi'(2\gamma))}\dot{\gamma}^2 = 0,$$

*is a geodesic for the Fisher metric.*

### 2.3 Some properties of the polygamma functions

In order to further study the geometry of the beta manifold, we will need a few technical results on the polygamma functions. The polygamma functions are the successive derivatives of the logarithm of the Euler Gamma function  $\Gamma(x)$ , i.e.

$$\psi^{(m-1)}(x) := \frac{d^m}{dx^m} \ln \Gamma(x), \quad m \geq 1.$$

Their series representation is given by:

$$\psi^{(m)} = (-1)^{m+1} m! \sum_{k \geq 0} \frac{1}{(k+x)^{m+1}}.$$

In the sequel, we are mostly interested by the first three, i.e.

$$\psi'(x) = \sum_{k \geq 0} \frac{1}{(k+x)^2}, \quad \psi''(x) = -2 \sum_{k \geq 0} \frac{1}{(k+x)^3}, \quad \psi'''(x) = 6 \sum_{k \geq 0} \frac{1}{(k+x)^4},$$

and we will use the following equivalents in the neighborhood of zero, given by the first term of their series

$$\psi'(x) \underset{x \rightarrow 0}{\sim} \frac{1}{x^2}, \quad \psi''(x) \underset{x \rightarrow 0}{\sim} -\frac{2}{x^3}, \quad \psi'''(x) \underset{x \rightarrow 0}{\sim} \frac{6}{x^4}. \quad (3)$$

In the neighborhood of infinity, we will need the following expansions

$$\begin{aligned} \psi(x) &\underset{x \rightarrow +\infty}{=} \ln(x) - \frac{1}{2x} + o\left(\frac{1}{x^2}\right), \\ \psi'(x) &\underset{x \rightarrow +\infty}{=} \frac{1}{x} + \frac{1}{2x^2} + o\left(\frac{1}{x^2}\right), \\ \psi''(x) &\underset{x \rightarrow +\infty}{=} -\frac{1}{x^2} - \frac{1}{x^3} + o\left(\frac{1}{x^3}\right). \end{aligned} \quad (4)$$

## 2.4 Curvature of the Fisher-Rao metric

In this section, we prove our main result, that is that the sectional curvature of the beta manifold is negative.

**Proposition 3.** *The sectional curvature of the Fisher metric is given by:*

$$K(\alpha, \beta) = \frac{\psi''(\alpha)\psi''(\beta)\psi''(\alpha + \beta)}{4d(\alpha, \beta)^2} \left( \frac{\psi'(\alpha)}{\psi''(\alpha)} + \frac{\psi'(\beta)}{\psi''(\beta)} - \frac{\psi'(\alpha + \beta)}{\psi''(\alpha + \beta)} \right),$$

*Proof.* The sectional curvature of a Hessian metric is given by

$$K = \frac{1}{4(\det g)^2} R_{1212}$$

where

$$R_{1212} = -\varphi_{\beta\beta}(\varphi_{\alpha\alpha\alpha}\varphi_{\alpha\beta\beta} - \varphi_{\alpha\alpha\beta}^2) + \varphi_{\alpha\beta}(\varphi_{\alpha\alpha\alpha}\varphi_{\beta\beta\beta} - \varphi_{\alpha\alpha\beta}\varphi_{\alpha\beta\beta}) + \varphi_{\alpha\alpha}(\varphi_{\alpha\alpha\beta}\varphi_{\beta\beta\beta} - \varphi_{\alpha\beta\beta}^2).$$

Computing the partial derivatives of the potential function  $\varphi$  gives

$$\begin{aligned} \varphi_{\alpha\alpha\alpha} &= \psi''(\alpha) - \psi''(\alpha + \beta), \\ \varphi_{\beta\beta\beta} &= \psi''(\beta) - \psi''(\alpha + \beta), \\ \varphi_{\alpha\alpha\beta} &= \varphi_{\alpha\beta\beta} = -\psi''(\alpha + \beta), \end{aligned}$$

and the determinant of the metric is given by

$$\det g(\alpha, \beta) = \psi'(\alpha)\psi'(\beta) - \psi'(\alpha + \beta)(\psi'(\alpha) + \psi'(\beta)).$$

This gives

$$K = \frac{\psi''(\alpha + \beta)(\psi'(\alpha)\psi''(\beta) + \psi''(\alpha)\psi'(\beta)) - \psi'(\alpha + \beta)\psi''(\alpha)\psi''(\beta)}{4(d(\alpha, \beta))^2}.$$

Factorizing the numerator by  $\psi''(\alpha)\psi''(\beta)\psi''(\alpha + \beta)$  yields the desired result.  $\square$

**Proposition 4.** *The asymptotic behavior of the sectional curvature is given by*

$$\begin{aligned} \lim_{\beta \rightarrow 0} K(\alpha, \beta) &= \lim_{\beta \rightarrow 0} K(\beta, \alpha) = \frac{3}{4} - \frac{\psi'(\alpha)\psi'''(\alpha)}{2\psi''(\alpha)^2}, \\ \lim_{\beta \rightarrow \infty} K(\alpha, \beta) &= \lim_{\beta \rightarrow \infty} K(\beta, \alpha) = \frac{\alpha\psi''(\alpha) + \psi'(\alpha)}{4(\alpha\psi'(\alpha) - 1)^2}. \end{aligned}$$

Moreover, we have the following limits

$$\begin{aligned}\lim_{\alpha, \beta \rightarrow 0} K(\alpha, \beta) &= 0, & \lim_{\alpha, \beta \rightarrow \infty} K(\alpha, \beta) &= -\frac{1}{2}, \\ \lim_{\alpha \rightarrow 0, \beta \rightarrow \infty} K(\alpha, \beta) &= \lim_{\alpha \rightarrow \infty, \beta \rightarrow 0} K(\alpha, \beta) &= -\frac{1}{4}.\end{aligned}$$

*Proof.* Let us fix  $\alpha \in \mathbb{R}_+^*$ , and denote  $x = \beta$  the varying parameter of the beta distribution. The asymptotic behavior of the sectional curvature can be obtained by separately examining its numerator and the metric determinant appearing at the denominator

$$K(\alpha, x) = \frac{N(\alpha, x)}{4d(\alpha, x)^2}.$$

Using a first order Taylor development of  $\psi'$  in  $\alpha$  and the equivalent (3), we deduce the following expansion for the determinant around zero

$$d(\alpha, x) = \psi'_x(\psi'_\alpha - \psi'_{\alpha+x}) - \psi'_{\alpha+x}\psi'_\alpha \underset{x \rightarrow 0}{=} -\frac{\psi''_\alpha}{x} + o\left(\frac{1}{x}\right).$$

Similarly, writing the numerator of the sectional curvature as

$$\begin{aligned}N(\alpha, x) &:= \psi''_{\alpha+x}(\psi''_\alpha\psi'_x + \psi'_\alpha\psi''_x) - \psi'_{\alpha+x}\psi''_\alpha\psi''_x \\ &= (\psi''_{\alpha+x} - \psi''_\alpha)(\psi''_\alpha\psi'_x + \psi'_\alpha\psi''_x) + (\psi'_\alpha - \psi'_{\alpha+x})\psi''_\alpha\psi''_x + \psi''_\alpha(\psi''_\alpha\psi'_x + \psi'_\alpha\psi''_x) - \psi'_\alpha\psi''_\alpha\psi''_x,\end{aligned}$$

we get the following behavior around zero

$$\begin{aligned}N(\alpha, x) &\underset{x \rightarrow 0}{=} x\psi''_{\alpha+x} \left( \frac{\psi''_\alpha}{x^2} - \frac{2\psi'_\alpha(\alpha)}{x^3} \right) + x(\psi''_\alpha)^2 \frac{2}{x^3} + \psi''_\alpha \left( \frac{\psi''_\alpha}{x^2} - \frac{2\psi'_\alpha}{x^3} \right) + \frac{2\psi'_\alpha\psi''_\alpha}{x^3} + o\left(\frac{1}{x^2}\right) \\ &\underset{x \rightarrow 0}{=} \frac{3(\psi''_\alpha)^2 - 2\psi'_\alpha\psi''_\alpha}{x^2} + o\left(\frac{1}{x^2}\right).\end{aligned}$$

This yields the desired expression for the limit of  $K(\alpha, x)$  as  $x \rightarrow 0$ . Now, in the neighborhood of infinity, the expansions (4) yield the following behavior for the determinant

$$\begin{aligned}d(\alpha, x) &= \psi'_\alpha\psi'_\beta - \psi'_{\alpha+\beta}(\psi'_\alpha + \psi'_\beta) \\ &\underset{x \rightarrow +\infty}{=} \psi'_\alpha \left( \frac{1}{x} + \frac{1}{2x^2} \right) - \left( \frac{1}{\alpha+x} + \frac{1}{2(\alpha+x)^2} \right) \left( \psi'_\alpha + \frac{1}{x} \right) + o\left(\frac{1}{x^2}\right) \\ &\underset{x \rightarrow +\infty}{=} \frac{\alpha\psi'_\alpha - 1}{x^2} + o\left(\frac{1}{x^2}\right),\end{aligned}$$

while an expansion of the numerator gives

$$\begin{aligned}N(\alpha, x) &\underset{x \rightarrow +\infty}{=} -\left( \frac{1}{(\alpha+x)^2} + \frac{1}{(\alpha+x)^3} \right) \left( \frac{\psi''_\alpha}{x} + \frac{\psi''_\alpha}{2x^2} - \frac{\psi'_\alpha}{x^2} \right) \\ &\quad + \left( \frac{1}{\alpha+x} + \frac{1}{2(\alpha+x)^2} \right) \psi''_\alpha \left( \frac{1}{x^2} + \frac{1}{x^4} \right) + o\left(\frac{1}{x^3}\right) \\ &\underset{x \rightarrow +\infty}{=} \frac{\alpha\psi''_\alpha + \psi'_\alpha}{x^4} + o\left(\frac{1}{x^4}\right),\end{aligned}$$

yielding again the desired limit for  $K$ . Finally, approximating  $\psi_x'''$  by  $6/x^4$  when  $x \rightarrow 0$ , we get

$$\begin{aligned}\lim_{\alpha \rightarrow 0, \beta \rightarrow +\infty} K(\alpha, \beta) &= \lim_{\alpha \rightarrow 0} \frac{\alpha \psi''_{\alpha} + \psi'_{\alpha}}{4(\alpha \psi'_{\alpha} - 1)^2} = \lim_{\alpha \rightarrow 0} \frac{-1/\alpha^2}{4/\alpha^2} = -\frac{1}{4}, \\ \lim_{\alpha, \beta \rightarrow +\infty} K(\alpha, \beta) &= \lim_{\alpha, \beta \rightarrow +\infty} \frac{-1/(2\alpha^2)}{1/\alpha^2} = -\frac{1}{2}, \\ \lim_{\alpha, \beta \rightarrow 0} K(\alpha, \beta) &= \frac{3}{4} - \frac{1}{2} \lim_{\alpha, \beta \rightarrow 0} \frac{6/\alpha^6}{(-2/\alpha^3)^2} = 0,\end{aligned}$$

which completes the proof.  $\square$

We can now show the following property.

**Proposition 5.** *The sectional curvature is negative and bounded from below.*

*Proof.* Recall that in its most factorized form, the sectional curvature is given by

$$K(\alpha, \beta) = \frac{\psi''(\alpha)\psi''(\beta)\psi''(\alpha + \beta)}{4d(\alpha, \beta)^2} \left( \frac{\psi'(\alpha)}{\psi''(\alpha)} + \frac{\psi'(\beta)}{\psi''(\beta)} - \frac{\psi'(\alpha + \beta)}{\psi''(\alpha + \beta)} \right),$$

Since  $\psi''$  is negative, the first factor is negative and so there remains to prove that the function  $x \mapsto \frac{\psi'(x)}{\psi''(x)}$  is sub-additive, i.e.

$$\frac{\psi'(\alpha)}{\psi''(\alpha)} + \frac{\psi'(\beta)}{\psi''(\beta)} - \frac{\psi'(\alpha + \beta)}{\psi''(\alpha + \beta)} \geq 0.$$

This has been shown recently in [12] (Corollary 4). Now, to show that it is bounded from below, set

$$\begin{aligned}k_1(\alpha) &:= \lim_{\beta \rightarrow 0} K(\alpha, \beta) = \frac{3}{4} - \frac{\psi'(\alpha)\psi'''(\alpha)}{2\psi''(\alpha)^2}, \\ k_2(\alpha) &:= \lim_{\beta \rightarrow +\infty} K(\alpha, \beta) = \frac{\alpha\psi''(\alpha) + \psi'(\alpha)}{4(\alpha\psi'(\alpha) - 1)^2}.\end{aligned}$$

$k_1$  and  $k_2$  are continuous functions on  $\mathbb{R}_+^*$ , and according to Proposition 4 they have finite limits at the boundaries

$$\begin{aligned}\lim_{\alpha \rightarrow 0} k_1(\alpha) &= 0, & \lim_{\alpha \rightarrow +\infty} k_1(\alpha) &= -\frac{1}{4}, \\ \lim_{\alpha \rightarrow 0} k_2(\alpha) &= -\frac{1}{4}, & \lim_{\alpha \rightarrow +\infty} k_2(\alpha) &= -\frac{1}{2}.\end{aligned}$$

Therefore, they are bounded, i.e., there exist negative finite constants  $M_1$  and  $M_2$  such that for all  $\alpha \in \mathbb{R}_+^*$ ,

$$\lim_{\beta \rightarrow 0} K(\alpha, \beta) > M_1, \quad \lim_{\beta \rightarrow +\infty} K(\alpha, \beta) > M_2.$$

Setting  $f(\beta) := \inf_{\alpha \in \mathbb{R}_+^*} K(\alpha, \beta)$ , notice that  $f$  is a continuous function on  $\mathbb{R}_+^*$  due to the continuity of  $K$  in both its variables and the invertibility of the limit and infimum. For this last reason, we also obtain

$$\begin{aligned}\lim_{\beta \rightarrow 0} \inf_{\alpha \in \mathbb{R}_+^*} K(\alpha, \beta) &= \inf_{\alpha \in \mathbb{R}_+^*} \lim_{\beta \rightarrow 0} K(\alpha, \beta) > M_1, \\ \lim_{\beta \rightarrow +\infty} \inf_{\alpha \in \mathbb{R}_+^*} K(\alpha, \beta) &= \inf_{\alpha \in \mathbb{R}_+^*} \lim_{\beta \rightarrow +\infty} K(\alpha, \beta) > M_2,\end{aligned}$$



i.e.,  $f$  has finite limits at the boundaries and is therefore bounded, in particular from below

$$\inf_{\beta \in \mathbb{R}_+^*} \inf_{\alpha \in \mathbb{R}_+^*} K(\alpha, \beta) > -\infty.$$

□

The fact that the beta manifold has negative curvature is particularly interesting for the computation of Riemannian centroids such as Fréchet or Karcher means [6, 7]. In general, the Fréchet mean on a Riemannian manifold is not unique. However, when the curvature is negative, there is no cut locus and uniqueness holds. In this context, it is defined for any given sequence of probability measure  $B_1, \dots, B_n$  as

$$\bar{B} = \operatorname{argmin}_{B \in \mathcal{B}} \sum_{i=1}^n d^F(B, B_i)^2.$$

This quantity can be computed using a gradient descent algorithm the Karcher flow algorithm.

## 2.5 A lower bound on the determinant of the metric

The determinant of the metric is the key ingredient to volume computations. In this section, a lower bound of this determinant is computed, which is also its asymptotic value.

**Proposition 6.** *The determinant of the information metric matrix admits the following integral representation:*

$$|g(\alpha, \beta)| = \int_{\mathbb{R}_+} \int_0^1 \frac{x(1-x)}{(1-e^{-tx})(1-e^{-t(1-x)})} \left( (e^{\beta tx} - 1) (e^{-\alpha t(1-x)} - 1) \right) e^{-(\alpha+\beta)t} dx dt \quad (5)$$

*Proof.* The polygamma function of order  $n$  can be expressed as an integral [8] :

$$\psi^{(n)}(x) = (-1)^{n+1} \int_{\mathbb{R}_+} \frac{t^n}{1-e^{-t}} e^{-xt} dt, \quad x > 0 \quad (6)$$

The determinant  $|g(\alpha, \beta)|$  expands as:

$$|g(\alpha, \beta)| = (\psi'(\alpha) - \psi'(\alpha + \beta)) (\psi'(\beta) - \psi'(\alpha + \beta)) - \psi'(\alpha + \beta)^2 \quad (7)$$

Using the integral 6, it comes:

$$\psi'(\alpha) - \psi'(\alpha + \beta) = \int_{\mathbb{R}_+} \frac{t}{1-e^{-t}} e^{-\alpha t} - e^{-(\alpha+\beta)t} dt = \int_{\mathbb{R}_+} \frac{t}{1-e^{-t}} (e^{\alpha t} - 1) e^{-(\alpha+\beta)t} dt \quad (8)$$

The difference  $\psi'(\alpha) - \psi'(\alpha + \beta)$  is thus equal to the laplace transform at  $\alpha + \beta$  of the function:

$$\frac{t}{1-e^{-t}} (e^{\alpha t} - 1) \quad (9)$$

Using the convolution theorem [10], it comes:

$$\begin{aligned} & (\psi'(\alpha) - \psi'(\alpha + \beta)) (\psi'(\beta) - \psi'(\alpha + \beta)) = \\ & \int_{\mathbb{R}_+} \left( \int_0^t \frac{x(t-x)}{(1-e^{-x})(1-e^{-(1-x)})} (e^{\beta x} - 1) (e^{\alpha(t-x)} - 1) dx \right) e^{-(\alpha+\beta)t} dt = \\ & \int_{\mathbb{R}_+} t^3 \left( \int_0^1 \frac{x(1-x)}{(1-e^{-tx})(1-e^{-t(1-x)})} (e^{\beta tx} - 1) (e^{\alpha t(1-x)} - 1) dx \right) e^{-(\alpha+\beta)t} dt \end{aligned} \quad (10)$$

The same procedure can be applied to the integral expression of  $\psi'(\alpha + \beta)$  to obtain:

$$\begin{aligned} \psi'^2(\alpha + \beta) &= \\ \int_{\mathbb{R}^+} t^3 \left( \int_0^1 \frac{x(1-x)}{(1-e^{-tx})(1-e^{-t(1-x)})} dx \right) e^{-(\alpha+\beta)t} dt \end{aligned} \quad (11)$$

Combining 10 and 11 gives:

$$\begin{aligned} |g(\alpha, \beta)| &= \\ \int_{\mathbb{R}^+} t^3 \left( \int_0^1 \frac{x(1-x)}{(1-e^{-tx})(1-e^{-t(1-x)})} \left[ (e^{\beta tx} - 1)(e^{\alpha t(1-x)} - 1) - 1 \right] dx \right) e^{-(\alpha+\beta)t} dt \end{aligned} \quad (12)$$

□

Building on the integral representation of Proposition 6, it is possible to derive a lower bound for the determinant, which is also its asymptotic value.

**Proposition 7.** *The following lower bound holds:*

$$|g(\alpha, \beta)| > \frac{1 + \alpha + \beta}{2\alpha\beta(\alpha + \beta)^2} \quad (13)$$

*Proof.* The hyperbolic cotangent  $\coth$  satisfies:

$$\coth \frac{x}{2} = \frac{1 + e^{-x}}{1 - e^{-x}} \quad (14)$$

and so:

$$\frac{1}{1 - e^{-x}} = \frac{1}{2} + \frac{1}{2} \coth \frac{x}{2} \quad (15)$$

Letting:

$$h(x, t) = \frac{x}{2} \left( 1 + \coth \frac{x}{2} \right) \quad (16)$$

The integral expression 12 is rewritten as:

$$\begin{aligned} |g(\alpha, \beta)| &= \\ \int_{\mathbb{R}^+} t^3 \left( \int_0^1 h(x, t) h((1-x)t) \left[ (e^{\beta tx} - 1)(e^{\alpha t(1-x)} - 1) - 1 \right] dx \right) e^{-(\alpha+\beta)t} dt \end{aligned} \quad (17)$$

Since:

$$\coth x > \frac{1}{x}, \quad x > 0$$

it comes:

$$\begin{aligned} h(x, t) h((1-x)t) &> \frac{x(1-x)}{4} \left( 1 + \frac{2}{tx} \right) \left( 1 + \frac{2}{t(1-x)} \right) \\ &> \frac{1}{4} \left( x + \frac{2}{t} \right) \left( 1 - x + \frac{2}{t} \right) \end{aligned} \quad (18)$$

a lower bound for 17 is thus given by:

$$|g(\alpha, \beta)| > \int_{\mathbb{R}^+} t^3 \left( \int_0^1 \frac{1}{4} \left( x + \frac{2}{t} \right) \left( 1 - x + \frac{2}{t} \right) \left[ (e^{\beta tx} - 1)(e^{\alpha t(1-x)} - 1) - 1 \right] dx \right) e^{-(\alpha+\beta)t} dt \quad (19)$$

The inner term:

$$I(t) = t^3 \int_0^1 \frac{1}{4} \left( x + \frac{2}{t} \right) \left( 1 - x + \frac{2}{t} \right) \left[ (e^{\beta tx} - 1)(e^{\alpha t(1-x)} - 1) - 1 \right] dx$$

admits a closed form expression:

$$I(t) = \frac{1}{4a^3b^3(a-b)^3} I_1(t) + I_2(t) - I_3(t) - I_4(t) + I_5(t) - I_6(t) \quad (20)$$

with:

$$I_1(t) = a^6 \left( - (2b^2(t+2) (e^{bt} - 1) + bt (e^{bt} + 1) - 2e^{bt} + 2) \right) \quad (21)$$

$$I_2(t) = a^5 b \left( 4b^2(t+2) (e^{bt} - 1) + 3bt (e^{bt} + 1) - 6e^{bt} + 6 \right) \quad (22)$$

$$I_3(t) = 2a^4 b^2 \left( b^2(t+2) (- (e^{at} - e^{bt})) + bt (e^{bt} + 2) - 3e^{bt} + 3 \right) \quad (23)$$

$$I_4(t) = 2a^3 b^4 \left( 2b(t+2) (e^{at} - 1) - t (e^{at} + 2) \right) \quad (24)$$

$$I_5(t) = a^2 b^4 \left( 2b^2(t+2) (e^{at} - 1) - 3bt (e^{at} + 1) - 6e^{at} + 6 \right) \quad (25)$$

$$I_6(t) = 2b^6 (e^{at} - 1) + ab^5 (e^{at}(bt+6) + bt - 6) \quad (26)$$

Performing the outer integration yields finally:

$$|g(\alpha, \beta)| > \frac{1 + \alpha + \beta}{2\alpha\beta(\alpha + \beta)^2} \quad (27)$$

thus completing the proof.  $\square$

## 2.6 A geometric view point of the central limit theorem

The central limit theorem tells us that once re-centered, a beta distribution  $B(n\alpha, n\beta)$  converges at rate  $\sqrt{n}$  to a centered normal distribution

$$\sqrt{n} \left( B(n\alpha, n\beta) - \frac{\alpha}{\alpha + \beta} \right) \xrightarrow{n \rightarrow \infty} \mathcal{N} \left( 0, \frac{ab}{(a+b)^3} \right).$$

For a fixed  $\lambda > 0$ , the line  $\beta = \lambda\alpha$  corresponds to all the beta distributions of mean  $1/(1+\lambda)$ . Asymptotically, we retrieve a hyperbolic distance between two distributions on this line.

**Proposition 8.** *When  $\beta = \lambda\alpha$  for a fixed  $\lambda > 0$ , the metric is asymptotically*

$$ds^2 = \frac{d\alpha^2}{2\alpha^2} + o\left(\frac{1}{\alpha^2}\right).$$

*This means*

$$d^F(B(n\alpha, n\lambda\alpha), B(n\alpha', n\lambda\alpha')) \xrightarrow{n \rightarrow \infty} d^F(\mathcal{N}(0, \alpha), \mathcal{N}(0, \alpha'))$$

*Proof.* The infinitesimal element of length is given by

$$ds^2 = (\psi'(\alpha) - \psi'(\alpha + \beta))d\alpha^2 + (\psi'(\beta) - \psi'(\alpha + \beta))d\beta^2 - 2\psi'(\alpha + \beta)d\alpha d\beta,$$

and so when  $\beta = \lambda\alpha$  for a fixed  $\lambda > 0$ ,

$$ds^2 = G(\alpha)d\alpha^2$$

where

$$G(\alpha) = \psi'(\alpha) + \lambda^2\psi'(\lambda\alpha) - (1 + \lambda)^2\psi'((1 + \lambda)\alpha).$$

When  $\alpha \rightarrow \infty$ , we have asymptotically using (4)

$$G(\alpha) = \frac{1}{\alpha} + \frac{1}{2\alpha^2} + \frac{\lambda}{\alpha} + \frac{1}{2\alpha^2} - \frac{1 + \lambda}{\alpha} - \frac{1}{2\alpha^2} + o\left(\frac{1}{\alpha^2}\right) = \frac{1}{2\alpha^2} + o\left(\frac{1}{\alpha^2}\right),$$

yielding the desired result. □

### 3 Geometric structure of canonical moments

In this section we use the previously described geometry to equip the space of canonical moments with a natural geometric structure. Indeed, beta distributions naturally arise in the study of these quantities. The canonical moments of a probability distribution offer an alternative to the usual moment representation, that can be considered more intrinsic. Qualitative properties such as symmetry or the fact that two probability distributions are identical up to linear transformation are easier seen on canonical moments than on regular moments. We first start by recalling the definition and some useful properties of the canonical moments. For an overview on the subject, we refer the reader to the monograph [4].

#### 3.1 Canonical moments: definition and useful properties

Here we are interested in probability distributions defined on finite intervals  $[a, b]$  with  $a < b$ . Since any such distribution can be uniquely mapped to one on  $[0, 1]$  through translation and rescaling  $x \mapsto (x - a)/(b - a)$ , we restrict to the case  $[a, b] = [0, 1]$ . Let  $\mathcal{P}$  denote the set of all probability measures on  $[0, 1]$ , and  $\mu \in \mathcal{P}$  be such a measure. Its  $n$ -th order moment is given by

$$c_n(\mu) = \int_a^b x^n d\mu(x).$$

Following [4], we respectively denote its infinite sequence of moments and the truncated sequence of its  $n$  first moments by

$$\begin{aligned} \mathbf{c}(\mu) &= (c_1(\mu), c_2(\mu), \dots) \\ \mathbf{c}_n(\mu) &= (c_1(\mu), \dots, c_n(\mu)) \end{aligned}$$

The spaces of all moment sequences and sequences of size  $n$  corresponding to the  $n$  first moments of a probability measure are respectively denoted by

$$M = \{\mathbf{c}(\mu), \mu \in \mathcal{P}\}, \quad M_n = \{\mathbf{c}_n = (c_1, \dots, c_n) \in [0, 1]^n, \exists \mu \in \mathcal{P} \mathbf{c}_n = \mathbf{c}_n(\mu)\}.$$

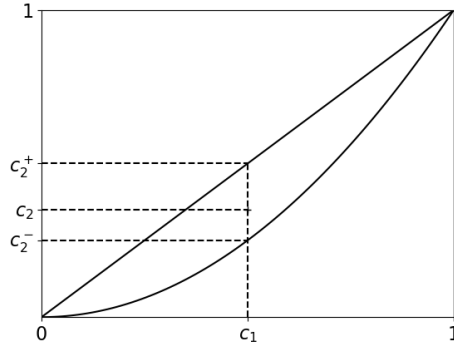


Figure 2: The space  $M_2$  of all first two moments on  $[0, 1]$ .

Obvious elements of  $M_n$  are the  $n$ -tuples  $\mathbf{c}_n(x) = (x, \dots, x^n)$  corresponding to the first  $n$  moments of Dirac distributions at points  $x \in [0, 1]$ . In fact, it can be shown that  $M_n$  is the convex hull of the curve  $\mathbf{c}_n(x) = (x, \dots, x^n)$  of  $[0, 1]^n$  [4, Theorem 1.2.1.]. For  $n = 2$ , this yields the area contained in-between the diagonal of the square  $[0, 1]^2$  and the parabola  $x \mapsto x^2$ , as shown in Figure 2. The points of the diagonal with coordinates  $(x, x)$  uniquely correspond to the first two moments of Bernoulli distributions of parameter  $x \in [0, 1]$ , while the points of the parabola uniquely represent Dirac distributions. More generally, any point of the boundary of  $M_n$  corresponds to the first moments of a unique probability measure, which is discrete (i.e., a linear combination of Dirac distributions), whereas any point of the interior corresponds to an infinity of probability distributions. In other words, for each sequence  $\mathbf{c}_n = (c_1, \dots, c_n) \in M_n$ , the set of measures

$$\mathcal{P}(\mathbf{c}_n) = \{\mu \in \mathcal{P}, \mathbf{c}_n(\mu) = \mathbf{c}_n\},$$

is infinite if  $\mathbf{c}_n \in \text{Int}M$  and it is a singleton if  $\mathbf{c}_n \in \partial M_n$ . For a fixed  $\mathbf{c}_{n-1} \in M_{n-1}$ , the intersection of the vertical line above  $\mathbf{c}_{n-1}$  and the compact and convex set  $M_n$  in  $[0, 1]^n$  yields an interval of all possible values for the following  $n^{\text{th}}$  moment, with minimum and maximum values

$$c_n^- = \min_{\mu \in \mathcal{P}(\mathbf{c}_{n-1})} c_n(\mu), \quad c_n^+ = \max_{\mu \in \mathcal{P}(\mathbf{c}_{n-1})} c_n(\mu).$$

Note that  $c_n^+$  and  $c_n^-$  are equal if  $\mathbf{c}_{n-1}$  is a boundary point. For  $\mathbf{c} \in M$ , let  $N(\mathbf{c}) = \min\{n \in \mathbb{N}, \mathbf{c}_n \in \partial M_n\}$ . Then for all  $n \leq N(\mathbf{c})$ , the  $n^{\text{th}}$  canonical moment describes the relative position of the  $n^{\text{th}}$  regular moment  $c_n$  with respect to these lower and upper bounds

$$p_n = \frac{c_n - c_n^-}{c_n^+ - c_n^-}.$$

The canonical moments provide an alternative representation of the underlying distribution that can be seen as more intrinsic in some sense. Indeed, the canonical moments representation benefits from the following interesting properties [4, Theorem 1.3.2 and Corollary 1.3.4].

**Theorem 1.** *Let  $\mu \in \mathcal{P}$  and  $a < b$ . We denote by  $\mu_{ab}$  the image measure of  $\mu$  by the linear transformation  $x \mapsto (b - a)x + a$ , by  $\mu^{(r)}$  the reflection of  $\mu$  with respect to the point  $1/2$ . Then we have the following properties.*

(i) The canonical moments of  $\mu$  remain invariant under linear transformation, i.e.

$$p_k(\mu) = p_k(\mu_{ab}), \quad k = 1, \dots, N(\mathbf{c}).$$

(ii) The measure  $\mu$  is symmetric, i.e.  $\mu = \mu^{(\tau)}$ , if and only if

$$p_k = 1/2, \quad 1 \leq k \leq (N(\mathbf{c}) + 1)/2.$$

The canonical moments can be expressed using the so-called Hankel determinants, defined for any  $n \geq 0$  by

$$\underline{H}_{2k} = \begin{vmatrix} c_0 & \dots & c_k \\ \vdots & \ddots & \vdots \\ c_k & \dots & c_{2k} \end{vmatrix}, \quad \overline{H}_{2k} = \begin{vmatrix} c_1 - c_2 & \dots & c_k - c_{k+1} \\ \vdots & \ddots & \vdots \\ c_k - c_{k+1} & \dots & c_{2k-1} - c_{2k} \end{vmatrix}$$

$$\underline{H}_{2k+1} = \begin{vmatrix} c_0 & \dots & c_{k+1} \\ \vdots & \ddots & \vdots \\ c_{k+1} & \dots & c_{2k+1} \end{vmatrix}, \quad \overline{H}_{2k+1} = \begin{vmatrix} c_0 - c_1 & \dots & c_k - c_{k+1} \\ \vdots & \ddots & \vdots \\ c_k - c_{k+1} & \dots & c_{2k} - c_{2k+1} \end{vmatrix}$$

The Hankel determinants verify the following relation [4, Theorem 1.4.5].

**Theorem 2.** For all  $n \geq 0$ ,

$$\underline{H}_n \overline{H}_n = \underline{H}_{n-1} \overline{H}_{n+1} + \overline{H}_{n-1} \underline{H}_{n+1}$$

The relative position of the  $n^{\text{th}}$  regular moment with respect to the lower and upper bounds  $c_n^-$  and  $c_n^+$  can be expressed in terms of quotients of Hankel determinants [4, Theorem 1.4.4]

**Theorem 3.** If  $\mathbf{c}_{n-1} \in \text{Int } M_{n-1}$  then

$$c_n - c_n^- = \underline{H}_n / \underline{H}_{n-2}, \quad c_n^+ - c_n = \overline{H}_n / \overline{H}_{n-2}.$$

Using these two results, it is possible to show that the Lebesgue measure in the interior of the moment space corresponds to a product measure of unnormalized beta distributions. More precisely, choosing an element  $\mathbf{c}_n = (c_1, \dots, c_n)$  at random in the interior of  $M_n$  is equivalent to choosing each canonical moment  $p_k$  independently in  $]0, 1[$  according to a beta distribution  $B(n-k+1, n-k+1)$  for  $k = 1, \dots, n$ . This is expressed by the following result, which corresponds to [4, Example 1.4.12], and for which we give the proof here for the sake of completeness.

**Theorem 4.**

$$dc_1 \dots dc_n = \prod_{k=1}^n p_k^{n-k} (1-p_k)^{n-k} dp_k.$$

*Proof.* Using the two previous results, the  $n^{\text{th}}$  canonical moment  $p_n$  and  $q_n = 1-p_n$  can be expressed as

$$p_n = \frac{\underline{H}_n \overline{H}_{n-2}}{\underline{H}_{n-1} \overline{H}_{n-1}}, \quad q_n = 1 - p_n = \frac{\overline{H}_n \underline{H}_{n-2}}{\underline{H}_{n-1} \overline{H}_{n-1}},$$

yielding the following recurrence relation

$$c_n - c_n^- = \frac{\underline{H}_n}{\underline{H}_{n-2}} = \frac{\underline{H}_n \underline{H}_{n-3}}{\underline{H}_{n-2} \underline{H}_{n-1}} \frac{\underline{H}_{n-1}}{\underline{H}_{n-3}} = q_{n-1} p_n (c_{n-1} - c_{n-1}^-).$$

This gives

$$c_n - c_{n-} = \prod_{k=1}^n q_{k-1} p_k.$$

Since  $p_n$  is a relative position,  $c_n^-$  is independent of  $p_n$  and since  $p_0 = c_0 = 1$ ,

$$\frac{\partial c_n}{\partial p_n} = \prod_{k=1}^{n-1} p_k q_k$$

Since for all  $k$ ,  $c_k$  only depends on  $p_\ell$  for  $\ell \leq k$ , the jacobian of the mapping  $\phi : \text{Int } M_n \rightarrow ]0, 1[^n$  that associates to any sequence  $(c_1, \dots, c_n)$  in the interior of  $M_n$  the corresponding canonical moments  $(p_1, \dots, p_n)$  is lower triangular, and its determinant is given by

$$\left| \frac{\partial(c_1, \dots, c_n)}{\partial(p_1, \dots, p_n)} \right| = \prod_{k=1}^n \frac{\partial c_k}{\partial p_k} = \prod_{k=1}^{n-1} (p_k q_k)^{n-k}.$$

□

Therefore the beta distribution naturally arises in the structure of the canonical moments space. Based on the Hankel determinants and similar quantities, the canonical moments can be computed from the regular moments using the so-called  $Q$ - $D$  algorithm. A description and proof and this algorithm can be found in [4].

### 3.2 Geometry of canonical moments

Now we define a natural geometric structure for the space of canonical moments. The goal is to be able to manipulate probability distributions, e.g. compute distances or centroids, through this geometric structure in the representation space given by the canonical moments. We propose a construction that associates to each sequence of canonical moments  $(p_1, \dots, p_n)$  a point in the product beta manifold  $(\mathcal{B}^n, g^n)$ , where  $g^n$  is the product of Fisher metrics on  $\mathcal{B}$ . More precisely, we associate to each  $p_k$  a beta distribution with the corresponding mean value, i.e., a point of the parameter space  $\Theta = \{(\alpha, \beta), \alpha > 0, \beta > 0\}$  belonging to  $\Delta_{p_k}$ , where  $\Delta_p$  is the straight line of equation

$$\Delta_p : \beta = \left( \frac{1}{p} - 1 \right) \alpha, \quad 0 < p < 1.$$

Let  $0 < p < 1$ , and  $B \in \mathcal{B}$ . We denote by  $\phi : \mathcal{B} \times (0, 1) \rightarrow \mathcal{B}$  the map that associates to  $B$  and  $p$  the closest neighbor of  $B$  on  $\Delta_p$

$$\phi(B; p) = \operatorname{argmin} \{d(B, B'), B' \in \Delta_p\}.$$

On the basis of Theorem 4, we define the following mapping

$$\Phi : \begin{cases} (0, 1)^n \rightarrow \mathcal{B}^n \\ (p_1, \dots, p_n) \mapsto (\phi(B(n, n); p_1), \dots, \phi(B(1, 1); p_n)), \end{cases}$$

Notice that  $\Phi$  associates the center  $(1/2, \dots, 1/2)$  of the cube  $[0, 1]^n$  to the sequence  $(B(n, n), \dots, B(1, 1))$ , and associates any sequence  $(p_1, \dots, p_n)$  in  $(0, 1)^n$  to the sequence of beta distributions for which each component  $B_k$  is the closest beta distribution to  $B(n - k + 1, n - k + 1)$  with mean value  $p_k$ .

Now let  $\mathbf{c}_n, \mathbf{c}'_n \in \text{Int } M_n$ . Then their canonical moments representations are in  $(0, 1)^n$ , and we can define the following measure of dissimilarity

$$\rho_n(\mathbf{c}_n, \mathbf{c}'_n) = d_n^F(\Phi(p_1, \dots, p_n), \Phi(p'_1, \dots, p'_n)),$$

where  $p_k = p_k(\mathbf{c}_n)$  and  $p'_k = p_k(\mathbf{c}'_n)$ , and

$$d_n^F((B_1, \dots, B_n), (B'_1, \dots, B'_n)) = (d^F(B_1, B'_1)^2 + \dots + d^F(B_n, B'_n)^2)^{1/2}$$

is the product distance on  $\mathcal{B}^n$ . Since  $\mathbf{c}_n$  uniquely determines  $(p_1, \dots, p_n)$  and  $\Phi$  is injective,  $\rho_n$  is a distance. It is now possible to define e.g. the centroid of several moment sequences  $\mathbf{c}_n^{(k)}$ ,  $k = 1, \dots, m$ , as the Fréchet mean

$$\bar{\mathbf{c}}_n = \underset{\mathbf{c}_n}{\text{argmin}} \sum_{k=1}^m \rho_n^2(\mathbf{c}_n, \mathbf{c}_n^{(k)}),$$

and to compute it using a Karcher flow in the product manifold  $\mathcal{B}^n$ . Due to the negative curvature of that Riemannian manifold, we know that it will be unique.

## 4 Conclusion and future work

The beta distribution is a natural exponential family, thus admitting a simple expression of the Fisher information metric as the Hessian of the log partition function. Nevertheless, explicit computations are difficult to conduct due to the presence of polygamma functions in the expression of the metric. In the present work, negative sectional curvature of the beta manifold was proved using recent results on ratios of polygamma functions and a lower bound on the determinant of the metric was obtained, that can further be used to derive rates of expansion for geodesic balls. Finally, the relationship with canonical moments allows us to better understand them from a geometrical standpoint.

## References

- [1] Colin Atkinson and Ann FS Mitchell. Rao's distance measure. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 345–365, 1981.
- [2] Martin Bauer, Martins Bruveris, and Peter W Michor. Uniqueness of the Fisher–Rao metric on the space of smooth densities. *Bulletin of the London Mathematical Society*, 48(3):499–506, 2016.
- [3] Nikolai Nikolaevich Cencov. *Statistical decision rules and optimal inference*. Number 53. American Mathematical Soc., 2000.
- [4] Holger Dette and William J Studden. *The theory of canonical moments with applications in statistics, probability, and analysis*, volume 338. John Wiley & Sons, 1997.
- [5] Ronald A Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594-604):309–368, 1922.
- [6] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré*, 10(3):215–310, 1948.



- [7] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on pure and applied mathematics*, 30(5):509–541, 1977.
- [8] F.W.J. Olver, National Institute of Standards, Technology (U.S.), D.W. Lozier, R.F. Boisvert, and C.W. Clark. *NIST Handbook of Mathematical Functions Paperback and CD-ROM*. Cambridge University Press, 2010.
- [9] C Radhakrishna Rao. Information and the accuracy attainable in the estimation of statistical parameters. In *Breakthroughs in statistics*, pages 235–247. Springer, 1992.
- [10] J.L. Schiff. *The Laplace Transform: Theory and Applications*. Undergraduate Texts in Mathematics. Springer New York, 2013.
- [11] Lene Theil Skovgaard. A Riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics*, pages 211–223, 1984.
- [12] Zhen-Hang Yang. Some properties of the divided difference of psi and polygamma functions. *Journal of Mathematical Analysis and Applications*, 455(1):761 – 777, 2017.