



# On closed-form expressions for the Fisher–Rao distance

Henrique K. Miyamoto<sup>1</sup> · Fábio C. C. Meneghetti<sup>2</sup> · Julianna Pinele<sup>3</sup> · Sueli I. R. Costa<sup>2</sup>

Received: 29 February 2024 / Revised: 31 July 2024 / Accepted: 1 August 2024  
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd. 2024

## Abstract

The Fisher–Rao distance is the geodesic distance between probability distributions in a statistical manifold equipped with the Fisher metric, which is a natural choice of Riemannian metric on such manifolds. It has recently been applied to supervised and unsupervised problems in machine learning, in various contexts. Finding closed-form expressions for the Fisher–Rao distance is generally a non-trivial task, and those are only available for a few families of probability distributions. In this survey, we collect examples of closed-form expressions for the Fisher–Rao distance of both discrete and continuous distributions, aiming to present them in a unified and accessible language. In doing so, we also: illustrate the relation between negative multinomial distributions and the hyperbolic model, include a few new examples, and write a few more in the standard form of elliptical distributions.

**Keywords** Fisher metric · Fisher–Rao distance · Geodesic distance · Parametric distributions · Statistical manifolds

---

Communicated by Tomonari Sei.

✉ Henrique K. Miyamoto  
henrique.miyamoto@centralesupelec.fr

Julianna Pinele  
julianna.pinele@ufrb.edu.br

Sueli I. R. Costa  
sueli@unicamp.br

- <sup>1</sup> Laboratoire des Signaux et Systèmes (L2S), Université Paris-Saclay, CentraleSupélec, CNRS, Gif-sur-Yvette, France
- <sup>2</sup> Instituto de Matemática, Estatística e Computação Científica (IMECC), Universidade Estadual de Campinas (Unicamp), Campinas, SP, Brazil
- <sup>3</sup> Centro de Ciências Exatas e Tecnológicas, Universidade Federal do Recôncavo da Bahia (UFRB), Cruz das Almas, BA, Brazil

## Contents

1	Introduction	.....
2	Information geometry preliminaries	.....
3	Hyperbolic geometry results	.....
4	Discrete distributions	.....
4.1	Binomial	.....
4.2	Poisson	.....
4.3	Geometric	.....
4.4	Negative binomial	.....
4.5	Categorical	.....
4.6	Multinomial	.....
4.7	Negative multinomial	.....
5	Continuous distributions	.....
5.1	Exponential	.....
5.2	Rayleigh	.....
5.3	Erlang	.....
5.4	Univariate elliptical distributions	.....
5.4.1	Gaussian	.....
5.4.2	Laplace	.....
5.4.3	Generalised Gaussian	.....
5.4.4	Logistic	.....
5.4.5	Cauchy	.....
5.4.6	Student's $t$	.....
5.5	Log-Gaussian	.....
5.6	Inverse Gaussian	.....
5.7	Extreme-value distributions	.....
5.7.1	Gumbel	.....
5.7.2	Fréchet	.....
5.7.3	Weibull	.....
5.8	Pareto	.....
5.9	Power function	.....
5.10	Wishart	.....
5.11	Inverse Wishart	.....
6	Product distributions	.....
7	Final remarks	.....
	References	.....

## 1 Introduction

Information geometry [1–3] uses the tools of differential geometry to study spaces of probability distributions by regarding them as differential manifolds, called *statistical manifolds*. When these distributions are parametric, a structure of interest is the *Fisher metric*, a Riemannian metric induced by the Fisher information matrix. This is essentially the unique Riemannian metric on statistical manifolds that is invariant by sufficient statistics [4, 5], making it a natural choice to study the geometry of these manifolds. Moreover, this structure allows one to define the *Fisher–Rao distance* between two probability distributions on the same statistical manifold as the geodesic distance between them, i.e., the length of the minimising path, according to the Fisher metric.

The idea of considering the geodesic distance in Riemannian manifolds equipped with the Fisher metric was first suggested by Hotelling in 1930 [6] (reprinted in [7]),

and later in 1945 in a landmark paper by Rao [8] (reprinted in [9]). This has influenced many authors to study the Fisher–Rao distance in different families of probability distributions in the following years [10–23]. It is important to note that, contrary to commonly used divergence measures, such as the Kullback–Leibler divergence, the Fisher–Rao distance is a proper distance, i.e., it is symmetric and the triangle inequality holds—properties that could be required when comparing two distributions, depending on the application.

More recently, the Fisher–Rao distance has gained attention especially in applications to machine learning problems. In the context of unsupervised learning, it has been used for clustering different types of data: shape clustering applied to morphometry [24], clustering of financial returns [25], image segmentation [20] and identification of diseases from medical data [26, 27]. When it comes to supervised learning, it has been used to analyse the geometry in the latent space of generative models [28], to enhance robustness against adversarial attacks [29, 30], in detection of out-of-distribution samples [31], as a loss function for learning under label noise [32], and to classify EEG signals of brain-computer interfaces [33].

However, finding closed-form expressions for the Fisher–Rao distance of arbitrary distributions is not a trivial task—as a matter of fact, more generally, finding geodesics in an arbitrary manifold is a difficult problem in differential geometry. Having that in mind, in this work, we collect examples of statistical models for which closed-form expressions for the Fisher–Rao distance are available. Most of them have been published over the last decades, in different places, and we aim to present these results in a unified and accessible language, hoping to bring them to a broader audience. In curating this collection, we also add a few contributions, such as: illustrating the relation between the manifold of negative multinomial distributions and the hyperbolic model; including a few new examples (Rayleigh, Erlang, Laplace, generalised Gaussian, power function, inverse Wishart), to the best of the authors’ knowledge; and writing more examples in the standard form of univariate elliptical distributions (Laplace, generalised Gaussian, logistic). Finally, we note that numerical methods have been proposed to compute the Fisher–Rao distance when no closed-form expression is available, as in [26, 27, 34–38], but those techniques are beyond the scope of the present work.

We review preliminaries of information geometry in Sect. 2 and of hyperbolic geometry in Sect. 3. We then collect closed-form expressions for discrete distributions in Sect. 4, and for continuous distributions in Sect. 5. Product distributions are discussed in Sect. 6, and Sect. 7 concludes the paper.

### **Notation.**

We denote the sets  $\mathbb{N} = \{0, 1, 2, \dots\}$ ,  $\mathbb{N}^* = \{1, 2, 3, \dots\}$ ,  $\mathbb{R}_+ = [0, \infty[$ , and  $\mathbb{R}_+^* = ]0, \infty[$ .  $\mathbb{1}_A(x)$  is the indicator function, that takes value 1 if  $x \in A$ , and 0 otherwise.  $\delta_{ij} := \mathbb{1}_{\{j\}}(i)$  denotes the Kronecker delta. We denote  $\dot{x}(t) := \frac{d}{dt}x(t)$ .  $\mathcal{P}_n(\mathbb{R})$  denotes the cone of  $n \times n$  real symmetric definite-positive matrices.

## 2 Information geometry preliminaries

Let  $(\Omega, \mathcal{G}, P)$  be a probability space and  $X: \Omega \rightarrow \mathcal{X}$  a random variable in the  $\sigma$ -finite measure space  $(\mathcal{X}, \mathcal{F}, \mu)$ . The push-forward measure of  $P$  by  $X$  is given by  $X_*P(E) := P(X^{-1}(E))$ , for  $E \in \mathcal{F}$ , and we assume that  $X_*P$  is absolutely continuous with respect to  $\mu$ . The Radon–Nikodym derivative  $p := \frac{dX_*P}{d\mu}: \mathcal{X} \rightarrow \mathbb{R}$  can be seen as the probability mass or density function (p.m.f. or p.d.f.), respectively, in the cases that  $\mathcal{X}$  is discrete or continuous. When  $\mathcal{X}$  is discrete, we take  $\mu$  as the counting measure, and the integral with respect to  $\mu$  becomes a summation; when  $\mathcal{X} = \mathbb{R}^n$ , we take  $\mu$  as the Lebesgue measure.

A statistical model [1]

$$\mathcal{S} := \left\{ p_\xi = p(x; \xi) \mid \xi = (\xi^1, \dots, \xi^n) \in \Xi \subseteq \mathbb{R}^n \right\} \tag{1}$$

is a family of probability distributions  $p_\xi$  parametrised by  $n$ -dimensional vectors  $\xi = (\xi^1, \dots, \xi^n)$  such that the mapping  $\xi \mapsto p_\xi$  is injective and  $\Xi$  is an open set of  $\mathbb{R}^n$ . We consider statistical models in which the support of  $p_\xi$  does not depend on  $\xi$  and we take  $\mathcal{X} = \text{supp } p_\xi$ , unless otherwise stated. Note that  $\mathcal{S}$  is contained in the infinite-dimensional space  $\mathcal{P}(\mathcal{X}) := \{p \in L^1(\mu) \mid p > 0, \int_{\mathcal{X}} p \, d\mu = 1\}$  of positive,  $\mu$ -integrable functions of unit total measure.

To introduce a differentiable structure in  $\mathcal{S}$ , we consider the following assumptions: 1) the parametrisation  $\varphi: \Xi \rightarrow \mathcal{P}(\mathcal{X})$ ,  $\varphi(\xi) = p_\xi$  is a homeomorphism on its image; 2) denoting  $\partial_i := \frac{\partial}{\partial \xi^i}$ , the functions  $\{\partial_1 p_\xi, \dots, \partial_n p_\xi\}$  are linearly independent; 3) the mapping  $\xi \mapsto p_\xi(x)$  is smooth, for all  $x \in \mathcal{X}$ ; 4) the partial derivatives  $\partial_i p_\xi(x)$  commute with the integrals. Moreover, by considering diffeomorphic parametrisations as equivalent,  $\mathcal{S}$  becomes a differentiable manifold, that we may call a *statistical manifold*.<sup>1</sup> Note that the parametrisation  $\xi \mapsto p_\xi$  is a global coordinate system for this manifold (see Fig. 1).

We can further equip the statistical manifold  $\mathcal{S}$  with a Riemannian metric. Denoting  $\ell(\xi) := \log p_\xi$  the log-likelihood function, the elements of the *Fisher information matrix* (or simply *Fisher matrix*)  $G(\xi) = [g_{ij}(\xi)]_{i,j}$  are defined as

$$g_{ij} := g_{ij}(\xi) := \mathbb{E} [\partial_i \ell(\xi) \partial_j \ell(\xi)], \tag{2}$$

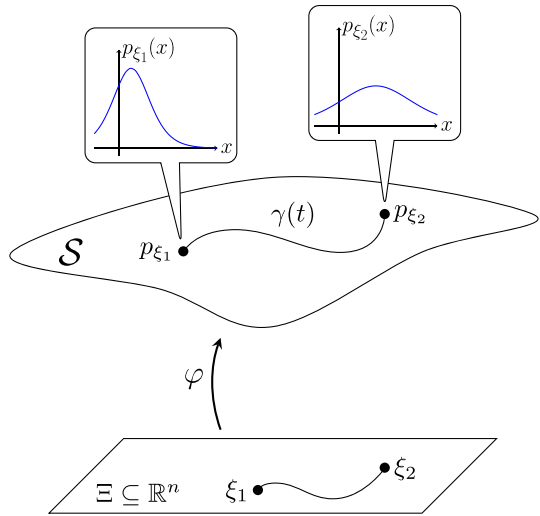
for  $1 \leq i, j \leq n$ , where the expectation is taken with respect to  $p_\xi$ . Explicitly,

$$g_{ij}(\xi) = \int_{\mathcal{X}} p_\xi \left( \frac{\partial}{\partial \xi^i} \log p_\xi \right) \left( \frac{\partial}{\partial \xi^j} \log p_\xi \right) d\mu.$$

Alternatively, the Fisher matrix can be written as the negative expectation of the Hessian of the log-likelihood function, a result that can make the computation of the Fisher matrix easier in some cases:

<sup>1</sup> We follow the nomenclature from [1], but remark that, more generally, *statistical manifold* refers to a manifold equipped with a Riemannian metric and a 3-symmetric tensor, from a purely geometric point of view [39] (see also [40, § 4.5]). We will restrict ourselves to the case of parametric statistical models presented above.

**Fig. 1** Schematic representing the parametrisation  $\varphi$  from the parameter space  $\Xi$  to the statistical manifold  $\mathcal{S}$ . The curve  $\gamma(t)$  joins two points in the manifold, which are probability density (or mass) functions



**Proposition 1** ([2, Prop. 1.6.3]) *The elements of the Fisher matrix can be expressed as*

$$g_{ij}(\xi) = -\mathbb{E}[\partial_j \partial_i \ell(\xi)]. \tag{3}$$

**Proof** As  $\int_{\mathcal{X}} p_{\xi} \, d\mu = 1$ , taking the derivative yields  $\int_{\mathcal{X}} p_{\xi} \partial_i \log p_{\xi} \, d\mu = \int_{\mathcal{X}} \partial_i p_{\xi} \, d\mu = 0$ . By taking the derivative again, we have

$$\begin{aligned} 0 &= \partial_j \int_{\mathcal{X}} p_{\xi} \partial_i \log p_{\xi} \, d\mu = \int_{\mathcal{X}} \partial_j p_{\xi} \partial_i \log p_{\xi} \, d\mu + \int_{\mathcal{X}} p_{\xi} \partial_j \partial_i \log p_{\xi} \, d\mu \\ &= \int_{\mathcal{X}} p_{\xi} (\partial_j \log p_{\xi}) (\partial_i \log p_{\xi}) \, d\mu + \int_{\mathcal{X}} p_{\xi} \partial_j \partial_i \log p_{\xi} \, d\mu \\ &= \mathbb{E}[(\partial_j \log p_{\xi})(\partial_i \log p_{\xi})] + \mathbb{E}[\partial_j \partial_i \log p_{\xi}]. \end{aligned}$$

□

Since the Fisher matrix is symmetric and positive-definite, it defines a Riemannian metric  $g_{p_{\xi}}$  (also denoted simply  $g_{\xi}$ ), called the *Fisher metric*; that is, a family of inner products  $g_{\xi} : T_{p_{\xi}} \mathcal{S} \times T_{p_{\xi}} \mathcal{S} \rightarrow \mathbb{R}$  that vary smoothly on the statistical manifold. Applying the Fisher metric to two vectors  $v_1 = d\varphi_{p_{\xi}}(\xi_1)$  and  $v_2 = d\varphi_{p_{\xi}}(\xi_2)$  in the tangent space  $T_{p_{\xi}} \mathcal{S}$  is equivalent to computing an inner product mediated by the Fisher matrix  $G(\xi)$  between the respective local-coordinate vectors  $\xi_1, \xi_2 \in \mathbb{R}^n$ :

$$\langle v_1, v_2 \rangle_{G(\xi)} := g_{\xi}(v_1, v_2) := g_{p_{\xi}}(v_1, v_2) = \xi_1^T G(\xi) \xi_2. \tag{4}$$

The following results will help some derivations in the rest of the text. First, we note that, as any Riemannian metric, the Fisher metric is covariant under reparametrisation of the parameter space:

**Proposition 2** ([2, Thm. 1.6.5]) *The Fisher matrix is covariant under reparametrisation of the parameters space, that is, given two coordinate systems  $\xi = (\xi^1, \dots, \xi^n)$  and  $\theta = (\theta^1, \dots, \theta^n)$ , related by the bijection  $\xi = \xi(\theta)$ , the Fisher matrix transforms its coordinates as*

$$\tilde{G}(\theta) = \left[ \frac{d\xi}{d\theta} \right]^T G(\xi(\theta)) \left[ \frac{d\xi}{d\theta} \right], \tag{5}$$

where  $\left[ \frac{d\xi}{d\theta} \right]$  denotes the Jacobian matrix of the transformation  $\theta \mapsto \xi$ .

**Proof** Denote  $\tilde{p}_\theta := p_{\xi(\theta)} = \varphi(\xi(\theta))$  and note that, by the chain rule,  $\frac{\partial \tilde{p}_\theta}{\partial \theta^i} = \sum_{k=1}^n \frac{\partial \xi^k}{\partial \theta^i} \frac{\partial p_\xi}{\partial \xi^k}$  and  $\frac{\partial \tilde{p}_\theta}{\partial \theta^j} = \sum_{r=1}^n \frac{\partial \xi^r}{\partial \theta^j} \frac{\partial p_\xi}{\partial \xi^r}$ . Thus

$$\begin{aligned} \tilde{g}_{ij}(\theta) &= \int_{\mathcal{X}} \tilde{p}_\theta \left( \frac{\partial}{\partial \theta^i} \log \tilde{p}_\theta \right) \left( \frac{\partial}{\partial \theta^j} \log \tilde{p}_\theta \right) d\mu = \int_{\mathcal{X}} \frac{1}{\tilde{p}_\theta} \frac{\partial \tilde{p}_\theta}{\partial \theta^i} \frac{\partial \tilde{p}_\theta}{\partial \theta^j} d\mu \\ &= \sum_{k=1}^n \sum_{r=1}^n \left( \int_{\mathcal{X}} \frac{1}{p_{\xi(\theta)}} \frac{\partial p_\xi}{\partial \xi^k} \frac{\partial p_\xi}{\partial \xi^r} d\mu \right) \frac{\partial \xi^k}{\partial \theta^i} \frac{\partial \xi^r}{\partial \theta^j} = \sum_{k=1}^n \sum_{r=1}^n g_{kr}(\xi(\theta)) \frac{\partial \xi^k}{\partial \theta^i} \frac{\partial \xi^r}{\partial \theta^j}. \end{aligned}$$

□

Second, we state an invariance property specific to the Fisher metric:

**Proposition 3** ([2, Thm. 1.6.4]) *Let  $X: \Omega \rightarrow \mathcal{X} \subseteq \mathbb{R}^n$  be a random variable distributed according to  $p_\xi$ . The Fisher metric is invariant under reparametrisations of the sample space  $\mathcal{X}$ .*

**Proof** Consider the reparametrisation by the bijection  $f: \mathcal{X} \rightarrow \mathcal{Y} \subseteq \mathbb{R}^n$  and denote  $\tilde{p}_\xi$  the distribution associated to the random variable  $Y := f(X)$ . The Jacobian determinant  $\left| \frac{df}{dx} \right|$  of the transformation  $f$  relates the relation between the densities:

$$p_\xi(x) = \tilde{p}_\xi(y) \left| \frac{df}{dx} \right|. \tag{6}$$

The log-likelihood functions are  $\tilde{\ell}(\xi) = \log \tilde{p}_\xi(y) = \log \tilde{p}_\xi(f(x))$  and  $\ell(\xi) = \log p_\xi(x) = \log \tilde{p}_\xi(y) + \log \left| \frac{df}{dx} \right|$ . As  $f$  does not depend on the parameter  $\xi$ , we have  $\partial_i \ell(\xi) = \partial_i \tilde{\ell}(\xi)$ , whence

$$\begin{aligned} g_{ij}(\xi) &= \int_{\mathcal{X}} p_\xi \partial_i \ell(\xi) \partial_j \ell(\xi) d\mu = \int_{\mathcal{X}} (\tilde{p}_\xi \circ f) \left| \frac{df}{dx} \right| \partial_i \tilde{\ell}(\xi) \partial_j \tilde{\ell}(\xi) d\mu \\ &= \int_{\mathcal{Y}} \tilde{p}_\xi \partial_i \tilde{\ell}(\xi) \partial_j \tilde{\ell}(\xi) d(f_*\mu) = \tilde{g}_{ij}(\xi). \end{aligned}$$

□

Furthermore, it can be shown that the Fisher metric is the unique Riemannian metric (up to a multiplicative constant) in statistical manifolds that is invariant under sufficient statistics [4, Thm. 1.2] (see also [5]). This invariance characterisation justifies the choice of this metric to study the geometry of statistical models.

The Fisher metric induces a notion of distance between two distributions in the same statistical manifold, called *Fisher–Rao distance*, and given by the geodesic distance between these points. Specifically, consider a curve  $\xi : [0, 1] \rightarrow \Xi$  in the parameter space and its image  $\gamma : [0, 1] \rightarrow \mathcal{S}$  by the parametrisation  $\varphi$ , i.e.,  $\gamma(t) = (\varphi \circ \xi)(t)$ , for  $t \in [0, 1]$ . Since  $\dot{\gamma}(t) = d\varphi_{\xi(t)}(\dot{\xi}(t))$ , in the Fisher geometry, the length of  $\gamma$  can be computed as

$$l(\gamma) := \int_0^1 \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{G(\xi(t))}} dt = \int_0^1 \sqrt{\dot{\xi}(t)^\top G(\xi(t)) \dot{\xi}(t)} dt. \tag{7}$$

Then, given two distributions  $p_{\xi_1}$  and  $p_{\xi_2}$  in  $\mathcal{S}$ , the Fisher–Rao distance between them<sup>2</sup> is the infimum of the length of piecewise differentiable curves  $\gamma$  joining these two points:

$$d_{FR}(\xi_1, \xi_2) := d_{FR}(p_{\xi_1}, p_{\xi_2}) := \inf_{\gamma} \{l(\gamma) \mid \gamma(0) = p_{\xi_1}, \gamma(1) = p_{\xi_2}\}. \tag{8}$$

A curve  $\gamma(t) = (\varphi \circ \xi)(t)$  is a *geodesic* if, in local coordinates  $\xi = (\xi^1, \dots, \xi^n)$ , the curve  $\xi(t)$  is a solution to the geodesic differential equations

$$\ddot{\xi}^k(t) + \sum_{i=1}^n \sum_{j=1}^n \Gamma_{ij}^k \dot{\xi}^i(t) \dot{\xi}^j(t) = 0, \quad k \in \{1, \dots, n\}, \tag{9}$$

where  $\Gamma_{ij}^k$  are the Christoffel symbols of the second kind, which can be obtained from the equations

$$\sum_{k=1}^n g_{lk} \Gamma_{ij}^k = \frac{1}{2} \left( \frac{\partial}{\partial \xi^i} g_{j\ell} + \frac{\partial}{\partial \xi^j} g_{\ell i} - \frac{\partial}{\partial \xi^\ell} g_{ij} \right), \quad i, j, \ell \in \{1, \dots, n\}$$

The Hopf–Rinow theorem (e.g., [41, Thm. 6.4.6]) provides a sufficient condition for the minimum length, as in (8), to be realised by a geodesic: if  $(\mathcal{S}, d_{FR})$  is connected and complete as a metric space, then any two points  $p, q$  in the manifold  $\mathcal{S}$  can be joined by a minimising curve which is a geodesic, that is, a curve whose length is equal to the Fisher–Rao distance  $d_{FR}(p, q)$ . This condition is satisfied in all statistical manifolds considered in this paper.

**Remark 1** The Fisher–Rao distance is related to the Kullback–Leibler divergence [2, Thm. 4.4.5]:

$$D_{KL}(p_{\xi_1} \parallel p_{\xi_2}) = \frac{1}{2} d_{FR}^2(p_{\xi_1}, p_{\xi_2}) + o\left(d_{FR}^2(p_{\xi_1}, p_{\xi_2})\right),$$

<sup>2</sup> We shall also refer to the Fisher–Rao distance between two parametric distributions as the distance between their respective parameters.

where  $o(x)$  represents a quantity such that  $\lim_{x \rightarrow 0} \frac{o(x)}{x} = 0$ . This means that the Kullback–Leibler divergence locally behaves as a ‘squared distance’. Differently from general statistical divergences, the Fisher–Rao distance is a proper distance, i.e., it is symmetric and the triangle inequality holds.

Unfortunately, finding the Fisher–Rao distance between two distributions in a statistical manifold usually is a non-trivial task, since it involves finding the minimising geodesics, potentially by solving the geodesics differential equations (9), and then evaluating the integral in (7). In the case of one-dimensional manifolds, i.e., those which are parametrised by a single real number, computing the Fisher–Rao distance is easier, since the geodesics are immediately given. In such cases, the Fisher matrix  $G(\xi) = [g_{11}(\xi)]$  contains a single element (also called *Fisher information*). Given two parameters  $\xi_1$  and  $\xi_2$ , there is only one path joining them, whose length does not depend on the chosen parametrisation. In particular, we can consider the arc length parametrisation  $\xi(t) = t$ , with  $t \in [\xi_1, \xi_2]$ , so that  $|\dot{\xi}(t)| = 1$ . Thus the expression for the length of the curve  $\gamma(t) = (\varphi \circ \xi)(t)$  in (7) becomes

$$l(\gamma) = \int_{\xi_1}^{\xi_2} \sqrt{g_{11}(\xi(t))} dt = \int_{\xi_1}^{\xi_2} \sqrt{g_{11}(\xi)} d\xi,$$

and the Fisher–Rao distance between distributions parametrised by  $\xi_1$  and  $\xi_2$  is

$$d_{\text{FR}}(\xi_1, \xi_2) = \left| \int_{\xi_1}^{\xi_2} \sqrt{g_{11}(\xi)} d\xi \right|. \quad (10)$$

For higher-dimensional manifolds, the techniques to find the geodesics and the Fisher–Rao distance consist in directly solving the geodesic differential equations, or in doing an analogy with some well-known geometry (e.g., spherical, hyperbolic), as we shall see in the next sections.

**Remark 2** The expression (10) allows one to find the Fisher–Rao distance in one-dimensional submanifolds of higher dimensional statistical manifolds, i.e., when only one parameter is allowed to vary and the others are fixed. This has been done for Gamma, Weibull and power function distributions in [10, 12].

### 3 Hyperbolic geometry results

We recall in this section some classical results from hyperbolic geometry [42–44], since many of the statistical manifolds studied in this work are related to that geometry. These will be extensively used particularly in Sects. 4.7 and 5.4–5.9. If desired, the reader may skip this section for now, and return to it when reading those subsections to get the details of the derivations.

We start with the hyperbolic geometry in dimension two, to be used in the approach of the statistical manifolds in Sects. 5.4–5.9, analogously to what is done in [45]. In this case, we consider the *Poincaré half-plane*  $\mathcal{H}^2 := \{(x, y) \in \mathbb{R}^2 \mid y > 0\}$  as a



model for hyperbolic geometry, with the metric given in matrix form by

$$G_{\mathcal{H}^2}(x, y) = \begin{pmatrix} \frac{1}{y^2} & 0 \\ 0 & \frac{1}{y^2} \end{pmatrix}. \tag{11}$$

The geodesics in this manifold are vertical half-lines and half-circles centred at  $y = 0$ , and the geodesic distance between two points is given by the following equivalent expressions:

$$\begin{aligned} d_{\mathcal{H}^2}((x_1, y_1), (x_2, y_2)) &= \log \left( \frac{\sqrt{(x_1 - x_2)^2 + (y_1 + y_2)^2} + \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}}{\sqrt{(x_1 - x_2)^2 + (y_1 + y_2)^2} - \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}} \right) \\ &= \operatorname{arccosh} \left( 1 + \frac{(x_1 - x_2)^2 + (y_1 - y_2)^2}{2y_1y_2} \right) \\ &= 2 \operatorname{arctanh} \left( \sqrt{\frac{(x_1 - x_2)^2 + (y_1 - y_2)^2}{(x_1 - x_2)^2 + (y_1 + y_2)^2}} \right). \end{aligned}$$

The next lemma allows one to relate the geodesic distance in a class of two-dimensional Riemannian manifold to that in the Poincaré half-plane.

**Lemma 4** Consider the Poincaré half-plane model  $\mathcal{H}^2$  with the hyperbolic metric (11), and a two-dimensional Riemannian manifold  $(\mathcal{M}, g_{\mathcal{M}})$ , parametrised by a global coordinate system  $\varphi: \Pi \subseteq \mathbb{R} \times \mathbb{R}_+^* \rightarrow \mathcal{M}$ . If the metric  $g_{\mathcal{M}}$  is given in matrix form by

$$G_{\mathcal{M}}(x, y) = \begin{pmatrix} \frac{a}{y^2} & 0 \\ 0 & \frac{b}{y^2} \end{pmatrix},$$

with  $a, b$  positive constants, then the geodesic distance between two points in  $\mathcal{M}$  is

$$d_{\mathcal{M}}(\varphi(x_1, y_1), \varphi(x_2, y_2)) = \sqrt{b} d_{\mathcal{H}^2}((\sqrt{a}x_1, \sqrt{b}y_1), (\sqrt{a}x_2, \sqrt{b}y_2)). \tag{12}$$

**Proof** Consider a curve  $\pi(t) = (x(t), y(t))$  in the parameter space  $\Pi$  and its image  $\alpha(t) = (\varphi \circ \pi)(t)$  by the parametrisation. Applying the metric  $G_{\mathcal{M}}(x, y)$  to  $\dot{\alpha}(t) = d\varphi_{\pi(t)}(\dot{\pi}(t))$  in  $\mathcal{M}$  gives

$$\begin{aligned} \langle \dot{\alpha}(t), \dot{\alpha}(t) \rangle_{G(\pi(t))} &= \begin{pmatrix} \dot{x}(t) & \dot{y}(t) \end{pmatrix} \begin{pmatrix} \frac{a}{(y(t))^2} & 0 \\ 0 & \frac{b}{(y(t))^2} \end{pmatrix} \begin{pmatrix} \dot{x}(t) \\ \dot{y}(t) \end{pmatrix} \\ &= \frac{1}{(y(t))^2} \left( a(\dot{x}(t))^2 + b(\dot{y}(t))^2 \right). \end{aligned}$$

On the other hand, consider the diffeomorphism  $\psi: \mathcal{M} \rightarrow \mathcal{H}^2, \varphi(x, y) \mapsto (\sqrt{a}x, \sqrt{b}y)$ , and the image curve  $\beta(t) := (\psi \circ \alpha)(t) = (\sqrt{a}x(t), \sqrt{b}y(t))$ . Applying

$G_{\mathcal{H}^2}(\beta(t))$  to  $\dot{\beta}(t) = d\psi_{\alpha(t)}(\dot{\alpha}(t))$  in  $\mathcal{H}^2$  gives

$$\begin{aligned} \langle \dot{\beta}(t), \dot{\beta}(t) \rangle_{G_{\mathcal{H}^2}(\beta(t))} &= (\sqrt{a}\dot{x}(t) \quad \sqrt{b}\dot{y}(t)) \begin{pmatrix} \frac{1}{b(y(t))^2} & 0 \\ 0 & \frac{1}{b(y(t))^2} \end{pmatrix} \begin{pmatrix} \sqrt{a}\dot{x}(t) \\ \sqrt{b}\dot{y}(t) \end{pmatrix} \\ &= \frac{1}{b(y(t))^2} \left( a(\dot{x}(t))^2 + b(\dot{y}(t))^2 \right). \end{aligned}$$

Thus,  $\langle \dot{\alpha}(t), \dot{\alpha}(t) \rangle_{G_{\mathcal{M}}(\alpha(t))} = b \langle \dot{\beta}(t), \dot{\beta}(t) \rangle_{G_{\mathcal{H}^2}(\beta(t))}$ , that is,  $\|\dot{\alpha}(t)\|_{G_{\mathcal{M}}(\alpha(t))} = \sqrt{b} \|\dot{\beta}(t)\|_{G_{\mathcal{H}^2}(\beta(t))}$ . This implies that a curve  $\alpha(t)$  connecting two points in  $\mathcal{M}$  is a geodesic (minimises the length) if, and only if, its image  $\beta(t)$  is a geodesic in  $\mathcal{H}^2$ . Taking (7) and (8) into account concludes the proof.  $\square$

**Remark 3** It is possible to deduce what the geodesics in  $\mathcal{M}$  look like in the parameter space  $\Pi$  (i.e., their preimages by  $\varphi$ ), since they are the inverse image by  $\psi$  of geodesics in  $\mathcal{H}^2$ . Consider a geodesic in  $\mathcal{M}$  connecting the points  $\varphi(x_1, y_1)$  and  $\varphi(x_2, y_2)$ . Its preimage in the parameter space is given either by the vertical line joining  $y_1$  to  $y_2$ , if  $x_1 = x_2$ , or, otherwise, by the arc of the half-ellipse joining  $(x_1, y_1)$  to  $(x_2, y_2)$ , centred at  $(C, 0)$  and given by  $\left(\frac{R}{\sqrt{a}} \cos(t) + C, \frac{R}{\sqrt{b}} \sin(t)\right)$ , where  $C = \frac{a(x_1^2 - x_2^2) + b(y_1^2 - y_2^2)}{2a(x_1 - x_2)}$  and

$$R = \frac{1}{2} \sqrt{a(x_1 - x_2)^2 + \frac{b^2}{a} \left(\frac{y_1^2 - y_2^2}{x_1 - x_2}\right)^2} + 2b(y_1^2 + y_2^2).$$

More generally, the  $n$ -dimensional hyperbolic half-space model is the Riemannian manifold  $\mathcal{H}^n := \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid x_n > 0\}$ , equipped with the metric given in matrix form by

$$G_{\mathcal{H}^n}(x_1, \dots, x_n) = \begin{pmatrix} \frac{1}{x_n^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{x_n^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{x_n^2} \end{pmatrix}, \tag{13}$$

that is,  $\langle (u_1, \dots, u_n), (v_1, \dots, v_n) \rangle_{G_{\mathcal{H}^n}(x_1, \dots, x_n)} = \frac{1}{x_n^2} \sum_{i=1}^n u_i v_i$ . This manifold has constant negative curvature. The geodesics in this manifold are vertical half-lines and vertical half-circles centred at the hyperplane  $x_n = 0$ , and the geodesic distance between two points in  $\mathcal{H}^n$  is given by

$$d_{\mathcal{H}^n}((x_1, \dots, x_n), (y_1, \dots, y_n)) = \operatorname{arccosh} \left( 1 + \frac{\sum_{i=1}^n (x_i - y_i)^2}{2x_n y_n} \right). \tag{14}$$

Restricted to points in the  $(n - 1)$ -dimensional unit half-sphere in  $\mathcal{H}^n$ , that is,  $S_1^{n-1} := \{(x_1, \dots, x_n) \in \mathcal{H}^n \mid \sum_{i=1}^n x_i^2 = 1\}$ , the expression for the distance becomes

$$\tilde{d}_{S_1^{n-1}}((x_1, \dots, x_n), (y_1, \dots, y_n)) = \operatorname{arccosh} \left( \frac{1 - \sum_{i=1}^{n-1} x_i y_i}{x_n y_n} \right). \tag{15}$$

From (14), we immediately see that the distance  $d_{\mathcal{H}^n}$  is invariant by dilation or contraction, that is, denoting  $\mathbf{x} := (x_1, \dots, x_n) \in \mathcal{H}^n$  and  $\mathbf{y} := (y_1, \dots, y_n) \in \mathcal{H}^n$ , we have  $d_{\mathcal{H}^n}(\mathbf{x}, \mathbf{y}) = d_{\mathcal{H}^n}(\lambda\mathbf{x}, \lambda\mathbf{y})$ , for  $\lambda \neq 0$ . For points  $\mathbf{x}$  and  $\mathbf{y}$  in the half-sphere  $S_r^{n-1} := \{(x_1, \dots, x_n) \in \mathcal{H}^n \mid \sum_{i=1}^n x_i^2 = r^2\}$  of radius  $r$ , we can see that the distance between them, restricted to the sphere  $S_r^{n-1}$ , is the distance given by (15) between their projections onto the radius-1 sphere, that is,

$$\tilde{d}_{S_r^{n-1}}(\mathbf{x}, \mathbf{y}) = \tilde{d}_{S_1^{n-1}}\left(\frac{\mathbf{x}}{r}, \frac{\mathbf{y}}{r}\right). \tag{16}$$

The central projection  $\pi : S_1^n \subseteq \mathcal{H}^{n+1} \rightarrow \mathcal{H}^n$  given by  $(x_1, \dots, x_n, x_{n+1}) \mapsto \left(\frac{2x_2}{x_1+1}, \dots, \frac{2x_{n+1}}{x_1+1}\right)$  is an isometry between the unit half-sphere  $S_1^n \subseteq \mathcal{H}^{n+1}$ , with the restriction of the ambient hyperbolic metric, and the hyperbolic space  $\mathcal{H}^n$ . It provides a *hemisphere model* in dimension  $n + 1$  for the  $n$ -dimensional hyperbolic space [43]. The geodesics in the hemisphere model are the inverse image by  $f$  of the geodesics in  $\mathcal{H}^n$ , namely, semicircles orthogonal to the hyperplane  $x_{n+1} = 0$ .

### 4 Discrete distributions

In the following, we begin with examples of one-dimensional statistical manifolds (Sects. 4.1–4.4), and then consider high-dimensional manifolds (Sects. 4.5–4.7). The results are summarised in Table 1.

#### 4.1 Binomial

A binomial distribution [2, 10, 12] models the probability of having  $x$  successes in  $n$  independent and identically distributed (i.i.d.) Bernoulli experiments with parameter  $\theta$ . Its p.m.f. is given by  $p(x) = \binom{n}{x}\theta^x(1 - \theta)^{n-x}$ , defined for  $x \in \{0, 1, \dots, n\}$ , and parametrised by  $\theta \in ]0, 1[$ . In this case,  $\partial_\theta \ell(\theta) = \frac{x}{\theta} - \frac{n-x}{1-\theta}$ , and the Fisher information is

$$\begin{aligned} g_{11}(\theta) &= \mathbb{E} \left[ (\partial_\theta \ell(\theta))^2 \right] = \mathbb{E} \left[ \left( \frac{X}{\theta} - \frac{n-X}{1-\theta} \right)^2 \right] \\ &= \frac{\mathbb{E}[X^2]}{\theta^2} + \frac{2\mathbb{E}[X^2] - 2n\mathbb{E}[X]}{\theta(1-\theta)} + \frac{n - 2n\mathbb{E}[X] + \mathbb{E}[X^2]}{(1-\theta)^2} \\ &= \frac{n}{\theta(1-\theta)}, \end{aligned} \tag{17}$$

where we have used that  $\mathbb{E}[X] = n\theta$  and  $\mathbb{E}[X^2] = n\theta - n\theta^2 + n^2\theta^2$ . The Fisher–Rao distance is then

$$d_{FR}(\theta_1, \theta_2) = \left| \int_{\theta_1}^{\theta_2} \sqrt{\frac{n}{\theta(1-\theta)}} \, d\theta \right| = 2\sqrt{n} \left| \arcsin \sqrt{\theta_1} - \arcsin \sqrt{\theta_2} \right|. \tag{18}$$

**Table 1** Fisher–Rao distances for discrete distributions

	Distribution	Support	Parameters	Fisher matrix	Fisher–Rao distance	References
Binomial	$\binom{n}{x} \theta^x (1 - \theta)^{n-x}$	$x \in \{0, 1, \dots, n\}$	$\theta \in ]0, 1[$	$\left(\frac{n}{\theta(1-\theta)}\right)$	$2\sqrt{n}  \arcsin \sqrt{\theta_1} - \arcsin \sqrt{\theta_2} $	[2, 10, 12]
Poisson	$\frac{\lambda^x e^{-\lambda}}{x!}$	$x \in \mathbb{N}$	$\lambda \in \mathbb{R}_+^*$	$\left(\frac{1}{\lambda}\right)$	$2 \sqrt{\lambda_1} - \sqrt{\lambda_2} $	[2, 10, 12]
Geometric	$\theta(1 - \theta)^{x-1}$	$x \in \mathbb{N}^*$	$\theta \in ]0, 1[$	$\left(\frac{1}{\theta^2(1-\theta)}\right)$	$2 \operatorname{arctanh} \sqrt{1-\theta_1} - \operatorname{arctanh} \sqrt{1-\theta_2} $	[2, 19]
Negative binomial	$\frac{\Gamma(x+r)}{x! \Gamma(r)} \theta^r (1 - \theta)^x$	$x \in \mathbb{N}$	$\theta \in ]0, 1[$	$\left(\frac{r}{\theta^2(1-\theta)}\right)$	$2\sqrt{r}  \operatorname{arctanh} \sqrt{1-\theta_1} - \operatorname{arctanh} \sqrt{1-\theta_2} $	[12, 19]
Categorical	$\sum_{i=1}^n p_i \mathbb{1}_{\{i\}}(x)$	$x \in \{1, \dots, n\}$	$(p_1, \dots, p_{n-1}) \in ]0, 1[^{n-1}, \sum_i p_i < 1,$ $p_n := 1 - \sum_i p_i$	$\left(\frac{\delta_{ij}}{p_i} + \frac{1}{p_n}\right)_{i,j}$	$2 \arccos \left(\sum_{i=1}^n \sqrt{p_i q_i}\right)$	[2, 4, 10, 12, 46]
Multinomial	$m! \prod_{i=1}^n \frac{p_i^{x_i}}{x_i!}$	$(x_1, \dots, x_n) \in \mathbb{N}^n, \sum_i x_i = m$	$(p_1, \dots, p_{n-1}) \in ]0, 1[^{n-1}, \sum_i p_i < 1,$ $p_n := 1 - \sum_i p_i$	$\left(\frac{m \delta_{ij}}{p_i} + \frac{m}{p_n}\right)_{i,j}$	$2\sqrt{m} \arccos \left(\sum_{i=1}^n \sqrt{p_i q_i}\right)$	[2, 10, 12]
Negative multinomial	$p_n \frac{\Gamma(\sum_{i=1}^n x_i)}{\prod_{i=1}^{n-1} x_i!}$	$(x_1, \dots, x_{n-1}) \in \mathbb{N}^{n-1}$	$(p_1, \dots, p_{n-1}) \in ]0, 1[^{n-1}, \sum_i p_i < 1,$ $p_n := 1 - \sum_i p_i$	$\left(\frac{x_n \delta_{ij}}{p_i p_n} + \frac{x_n}{p_n}\right)_{i,j}$	$2\sqrt{x_n} \operatorname{arccosh} \left(\frac{1 - \sum_{i=1}^{n-1} \sqrt{p_i q_i}}{\sqrt{p_n q_n}}\right)$	[12, 19, 46]

### 4.2 Poisson

A Poisson distribution [2, 10, 12] has p.m.f.  $p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$ , defined for  $x \in \mathbb{N}$ , and parametrised by  $\lambda \in \mathbb{R}_+^*$ . In this case, we have  $\partial_\lambda \ell(\lambda) = \frac{x}{\lambda} - 1$ , and the Fisher information is given by

$$\begin{aligned} g_{11}(\lambda) &= \mathbb{E} \left[ (\partial_\lambda \ell(\lambda))^2 \right] = \mathbb{E} \left[ \left( \frac{X}{\lambda} - 1 \right)^2 \right] \\ &= \frac{1}{\lambda^2} \mathbb{E} \left[ X^2 \right] - \frac{2}{\lambda} \mathbb{E} [X] + 1 \\ &= \frac{1}{\lambda}, \end{aligned} \tag{19}$$

where we have used that  $\mathbb{E}[X] = \lambda$  and  $\mathbb{E}[X^2] = \lambda(\lambda + 1)$ . Thus the Fisher–Rao distance is

$$d_{FR}(\lambda_1, \lambda_2) = \left| \int_{\lambda_1}^{\lambda_2} \frac{1}{\sqrt{\lambda}} \, d\lambda \right| = 2 \left| \sqrt{\lambda_1} - \sqrt{\lambda_2} \right|. \tag{20}$$

### 4.3 Geometric

A geometric distribution [2, 19] models the number of i.i.d. Bernoulli trials with parameter  $\theta$  needed to obtain one success. Its p.m.f. is  $p(x) = \theta(1 - \theta)^{x-1}$ , defined for  $x \in \mathbb{N}^*$ , and parametrised by  $\theta \in ]0, 1[$ . We have  $\partial_\theta \ell(\theta) = \frac{1}{\theta} - \frac{x-1}{1-\theta}$ , and the Fisher information is

$$\begin{aligned} g_{11}(\theta) &= \mathbb{E} \left[ (\partial_\theta \ell(\theta))^2 \right] = \mathbb{E} \left[ \left( \frac{1}{\theta} - \frac{X-1}{1-\theta} \right)^2 \right] \\ &= \frac{1}{\theta^2} + \frac{2 - 2\mathbb{E}[X]}{\theta(1-\theta)} + \frac{\mathbb{E}[X^2] - 2\mathbb{E}[X] + 1}{(1-\theta)^2} \\ &= \frac{1}{\theta^2(1-\theta)}, \end{aligned} \tag{21}$$

where we have used that  $\mathbb{E}[X] = \frac{1}{\theta}$  and  $\mathbb{E}[X^2] = \frac{2-\theta}{\theta^2}$ . The Fisher–Rao distance is

$$d_{FR}(\theta_1, \theta_2) = \left| \int_{\theta_1}^{\theta_2} \frac{1}{\theta \sqrt{1-\theta}} \, d\theta \right| = 2 \left| \operatorname{arctanh} \sqrt{1-\theta_1} - \operatorname{arctanh} \sqrt{1-\theta_2} \right|. \tag{22}$$

### 4.4 Negative binomial

A negative binomial distribution [12, 19] models the excess of i.i.d. Bernoulli experiments with parameter  $\theta$  needed until a number of  $r$  successes occur. It has p.m.f.  $p(x) = \frac{\Gamma(x+r)}{x! \Gamma(r)} \theta^r (1 - \theta)^x$ , defined for  $x \in \mathbb{N}$ , and is parametrised by  $\theta \in ]0, 1[$ , for a fixed  $r$ , that can be extended to  $r \in \mathbb{R}_+^*$ . We have  $\partial_\theta \ell(\theta) = \frac{r}{\theta} - \frac{x}{1-\theta}$ , and the Fisher

information is

$$\begin{aligned}
 g_{11}(\theta) &= \mathbb{E} \left[ (\partial_\theta \ell(\theta))^2 \right] = \mathbb{E} \left[ \left( \frac{r}{\theta} - \frac{X}{1-\theta} \right)^2 \right] \\
 &= \frac{r^2}{\theta^2} - \frac{2r\mathbb{E}[X]}{\theta(1-\theta)} + \frac{\mathbb{E}[X^2]}{(1-\theta)^2} \\
 &= \frac{r}{\theta^2(1-\theta)},
 \end{aligned} \tag{23}$$

where we have used that  $\mathbb{E}[X] = \frac{r(1-\theta)}{\theta}$  and  $\mathbb{E}[X^2] = \frac{r(1-\theta)+r^2(1-\theta)^2}{(1-\theta)^2}$ . The Fisher-Rao distance is then

$$d_{\text{FR}}(\theta_1, \theta_2) = \left| \int_{\theta_1}^{\theta_2} \frac{\sqrt{r}}{\theta\sqrt{1-\theta}} \, d\theta \right| = 2\sqrt{r} \left| \operatorname{arctanh} \sqrt{1-\theta_1} - \operatorname{arctanh} \sqrt{1-\theta_2} \right|. \tag{24}$$

### 4.5 Categorical

A categorical distribution [2, 4, 10, 12, 46] models a random variable taking values in the sample space  $\mathcal{X} = \{1, 2, \dots, n\}$  with probabilities  $p_1, \dots, p_n$ , and has p.m.f.  $p(x) = \sum_{i=1}^n p_i \mathbb{1}_{\{i\}}(x)$ . The associated  $(n - 1)$ -dimensional statistical manifold

$$\mathcal{S} = \left\{ p = \sum_{i=1}^n p_i \mathbb{1}_{\{i\}} \mid p_i \in ]0, 1[, \sum_{i=1}^n p_i = 1 \right\}$$

is in correspondence with the interior of the probability simplex  $\hat{\Delta}^{n-1} := \{p = (p_1, \dots, p_n) \mid p_i \in ]0, 1[, \sum_{i=1}^n p_i = 1\}$  through the bijection  $\iota: \hat{\Delta}^{n-1} \rightarrow \mathcal{S}$ , given by  $(p_1, \dots, p_n) \mapsto \sum_{i=1}^n p_i \mathbb{1}_{\{i\}}$ . Both these manifolds can be parametrised by the set

$$\Xi = \left\{ \xi = (\xi^1, \dots, \xi^{n-1}) \in \mathbb{R}^{n-1} \mid \xi^i > 0, \sum_{i=1}^{n-1} \xi^i < 1 \right\}, \tag{25}$$

by taking  $p_i = \xi^i$ , for  $1 \leq i \leq n - 1$ , and  $p_n = 1 - \sum_{i=1}^{n-1} \xi^i$ .

To compute the Fisher matrix, it is useful to write  $p(x) = \sum_{i=1}^{n-1} \xi^i \mathbb{1}_{\{i\}}(x) + (1 - \sum_{i=1}^{n-1} \xi^i) \mathbb{1}_{\{n\}}(x)$ , so that

$$\partial_i \ell(\xi) = \frac{\mathbb{1}_{\{i\}}(x) - \mathbb{1}_{\{n\}}(x)}{\sum_{k=1}^n p_k \mathbb{1}_{\{k\}}(x)},$$

with  $p_k = p_k(\xi)$  as above. The elements of the Fisher matrix are, for  $1 \leq i, j \leq n - 1$ ,

$$\begin{aligned}
 g_{ij}(\xi) &= \mathbb{E} \left[ (\partial_i \ell(\xi)) (\partial_j \ell(\xi)) \right] = \mathbb{E} \left[ \frac{(\mathbb{1}_{\{i\}}(X) - \mathbb{1}_{\{n\}}(X))(\mathbb{1}_{\{j\}}(X) - \mathbb{1}_{\{n\}}(X))}{(\sum_k p_k \mathbb{1}_{\{k\}}(X))^2} \right] \\
 &= \mathbb{E} \left[ \frac{\mathbb{1}_{\{i\}}(X)\mathbb{1}_{\{j\}}(X)}{(\sum_k p_k \mathbb{1}_{\{k\}}(X))^2} - \frac{\mathbb{1}_{\{i\}}(X)\mathbb{1}_{\{n\}}(X)}{(\sum_k p_k \mathbb{1}_{\{k\}}(X))^2} - \frac{\mathbb{1}_{\{j\}}(X)\mathbb{1}_{\{n\}}(X)}{(\sum_k p_k \mathbb{1}_{\{k\}}(X))^2} + \frac{(\mathbb{1}_{\{n\}}(X))^2}{(\sum_k p_k \mathbb{1}_{\{k\}}(X))^2} \right]
 \end{aligned}$$

$$\begin{aligned}
 &= \frac{\delta_{ij}}{p_i} + \frac{1}{p_n} \\
 &= \frac{\delta_{ij}}{\xi^i} + \frac{1}{1 - \sum_{k=1}^{n-1} \xi^k},
 \end{aligned} \tag{26}$$

where we have used that  $\mathbb{E} \left[ \frac{\mathbb{1}_{\{i\}}(X)\mathbb{1}_{\{j\}}(X)}{(\sum_k p_k \mathbb{1}_{\{k\}}(X))^2} \right] = \frac{\delta_{ij}}{p_i}$ .

To obtain the geodesics and the Fisher–Rao distance, it is convenient to consider the mapping  $p_i \mapsto 2\sqrt{p_i}$  that takes the simplex  $\Delta^{n-1} \subseteq \mathbb{R}^n$  (in correspondence with the statistical manifold) to the positive part of the radius-two Euclidean sphere, denoted  $S_{2,+}^{n-1} \subseteq \mathbb{R}^n$ . In fact, this bijection is an isometry [40, 46]:

**Proposition 5** *The diffeomorphism*

$$\begin{aligned}
 f: \mathcal{S} \subseteq \mathcal{P}(\mathcal{X}) &\rightarrow S_{2,+}^{n-1} \subseteq \mathbb{R}^n \\
 p = \sum_{i=1}^n p_i \mathbb{1}_{\{i\}} &\mapsto (2\sqrt{p_1}, \dots, 2\sqrt{p_n})
 \end{aligned} \tag{27}$$

is an isometry between the statistical manifold  $\mathcal{S}$  equipped with the Fisher metric  $g_p$  and  $S_{2,+}^{n-1}$  with the restriction of the ambient Euclidean metric.

**Proof** We will show that  $g_p(u, v) = \langle df_p(u), df_p(v) \rangle$ , for all  $p \in \mathcal{S}$ ,  $u, v \in T_p\mathcal{S}$ . Let  $\varphi$  denote the parametrisation of the statistical manifold  $\mathcal{S}$ . Consider the curve  $\alpha_i(t) = \varphi(\xi^1, \dots, \xi^i + t, \dots, \xi^{n-1}) \in \mathcal{S}$  and take  $p = \varphi(\xi^1, \dots, \xi^{n-1})$ . We have that

$$\beta_i(t) := (f \circ \alpha_i)(t) = \left( 2\sqrt{\xi^1}, \dots, 2\sqrt{\xi^i + t}, \dots, 2\sqrt{\xi^{n-1}}, 2\sqrt{1 - \sum_{k=1}^{n-1} \xi^k - t} \right).$$

Now, we can compute the differential applied to the tangent vector  $\frac{\partial}{\partial \xi^i}(p)$ :

$$df_p \left( \frac{\partial}{\partial \xi^i}(p) \right) = \frac{d}{dt} \beta_i(t) \Big|_{t=0} = \left( 0, \dots, 0, \frac{1}{\sqrt{\xi^i}}, 0, \dots, 0, \frac{-1}{\sqrt{1 - \sum_{k=1}^{n-1} \xi^k}} \right).$$

Therefore,

$$\left\langle df_p \left( \frac{\partial}{\partial \xi^i}(p) \right), df_p \left( \frac{\partial}{\partial \xi^j}(p) \right) \right\rangle = \frac{\delta_{ij}}{\xi^i} + \frac{1}{1 - \sum_{k=1}^{n-1} \xi^k},$$

which is equal to  $g_{ij}(\xi) = g_p \left( \frac{\partial}{\partial \xi^i}(p), \frac{\partial}{\partial \xi^j}(p) \right)$  in (26). Since  $\left\{ \frac{\partial}{\partial \xi^1}(p), \dots, \frac{\partial}{\partial \xi^{n-1}}(p) \right\}$  is a basis of  $T_p\mathcal{S}$ , this is enough to show that  $f$  is indeed an isometry.  $\square$

Thus, the Fisher metric in  $\mathcal{S}$  coincides with the Euclidean metric restricted to the positive part of the sphere  $S_{2,+}^{n-1}$ , that is, the Fisher–Rao distance between distributions  $p = \varphi(p_1, \dots, p_{n-1})$  and  $q = \varphi(q_1, \dots, q_{n-1})$  in  $\mathcal{S}$  is equal to the length of geodesic joining  $f(p)$  and  $f(q)$  on the sphere, which is great circle arc. This length is double the angle  $\alpha$  between the vectors  $f(p)$  and  $f(q)$ , i.e.,  $2\alpha = 2 \arccos \left\langle \frac{f(p)}{2}, \frac{f(q)}{2} \right\rangle = 2 \arccos \left( \sum_{i=1}^n \sqrt{p_i q_i} \right)$ , with  $p_n = 1 - \sum_{i=1}^{n-1} p_i$  and  $q_n = 1 - \sum_{i=1}^{n-1} q_i$ . Therefore, the Fisher–Rao distance between these two distributions is

$$d_{\text{FR}}(p, q) = 2 \arccos \left( \sum_{i=1}^n \sqrt{p_i q_i} \right). \tag{28}$$

Note that the isometry  $f$  also allows extending the Fisher metric to the boundaries of the statistical manifold  $\mathcal{S}$ .

**Remark 4** The reparametrisation to the sphere provides a nice geometrical interpretation for the relation between the Fisher–Rao distance and the Hellinger distance [47, § 2.4]  $d_{\text{H}}(p, q) = \sqrt{\sum_{i=1}^n (\sqrt{p_i} - \sqrt{q_i})^2}$ . While the Fisher–Rao distance between distributions  $p$  and  $q$  is the length of the radius-two circumference arc between  $f(p)$  and  $f(q)$ , the Hellinger distance is half the Euclidean distance between  $f(p)$  and  $f(q)$ , i.e.,  $2d_{\text{H}}(p, q) = \|f(p) - f(q)\|_2$ . In other words, double the Hellinger distance is the arc-chord approximation for the Fisher–Rao distance.

### 4.6 Multinomial

Consider  $m$  i.i.d. experiments that follow a categorical distribution with  $n$  possible outcomes and probabilities  $p_1, \dots, p_n$ . A multinomial distribution [2, 10, 12] gives the probability of getting  $x_i$  times the  $i$ -th outcome, for  $1 \leq i \leq n$  and  $\sum_{i=1}^n x_i = m$ .

Its p.m.f. is  $p(\mathbf{x}) = p(x_1, \dots, x_n) = m! \prod_{i=1}^n \frac{p_i^{x_i}}{x_i!}$  and is defined on the sample space  $\mathcal{X} = \{\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{N}^n \mid \sum_{i=1}^n x_i = m\}$ . This distribution is parametrised by the same  $\xi = (\xi^1, \dots, \xi^{n-1}) \in \Xi$  as in the categorical distribution, cf. (25), with  $p_i = \xi^i$ , for  $1 \leq i \leq n - 1$  and  $p_n = 1 - \sum_{i=1}^{n-1} \xi^i$ . In this case, we have  $\partial_i \ell(\xi) = \frac{x_i}{p_i} - \frac{x_n}{p_n}$  and  $\partial_j \partial_i \ell(\xi) = -\frac{x_i}{p_i^2} \delta_{ij} - \frac{x_n}{p_n^2}$ . Thus, the elements of the Fisher matrix are given by

$$\begin{aligned} g_{ij}(\xi) &= -\mathbb{E} [\partial_j \partial_i \ell(\xi)] = \mathbb{E} \left[ \frac{X_i}{p_i^2} \delta_{ij} + \frac{X_n}{p_n^2} \right] \\ &= m \left( \frac{\delta_{ij}}{p_i} + \frac{1}{p_n} \right) \\ &= m \left( \frac{\delta_{ij}}{\xi^i} + \frac{1}{1 - \sum_{k=1}^{n-1} \xi^k} \right), \end{aligned} \tag{29}$$



where we have used that  $\mathbb{E}[X_i] = mp_i$ . Note that this is the same metric as for the categorical distribution (26), up to the factor  $m$ . Therefore, the Fisher–Rao distance between two distributions  $p = \varphi(p_1, \dots, p_{n-1})$  and  $q = \varphi(q_1, \dots, q_{n-1})$ , with  $p_n = 1 - \sum_{i=1}^{n-1} p_i$  and  $q_n = 1 - \sum_{i=1}^{n-1} q_i$  is

$$d_{FR}(p, q) = 2\sqrt{m} \arccos \left( \sum_{i=1}^n \sqrt{p_i q_i} \right). \tag{30}$$

### 4.7 Negative multinomial

A negative multinomial distribution [12, 19, 48] generalises the negative binomial distribution. Consider a sequence of i.i.d. categorical experiments with  $n$  possible outcomes. A negative multinomial distribution models the number of times  $x_1, \dots, x_{n-1}$  that the first  $n - 1$  outcomes occur before the  $n$ -th outcome occurs  $x_n$  times. It is characterised by the p.m.f.  $p(\mathbf{x}) = p(x_1, \dots, x_{n-1}) = p_n^{x_n} \frac{\Gamma(\sum_{i=1}^n x_i)}{\Gamma(x_n)} \prod_{i=1}^{n-1} \frac{p_i^{x_i}}{x_i!}$ , and defined for  $\mathbf{x} = (x_1, \dots, x_{n-1}) \in \mathbb{N}^{n-1}$ . It has the same parametrisation as the categorical distribution, cf. (25), with  $p_i = \xi^i$ , for  $1 \leq i \leq n - 1$  and  $p_n = 1 - \sum_{i=1}^{n-1} \xi^i$ , for a fixed  $x_n$  that can be extended to  $x_n \in \mathbb{R}_+^*$ . In this case, we have  $\partial_i \ell(\xi) = \frac{x_i}{p_i} - \frac{x_n}{p_n}$  and  $\partial_j \partial_i \ell(\xi) = -\frac{x_i}{p_i^2} \delta_{ij} - \frac{x_n}{p_n^2}$ . The elements of the Fisher matrix are then

$$\begin{aligned} g_{ij}(\xi) &= -\mathbb{E}[\partial_j \partial_i \ell(\xi)] = \mathbb{E} \left[ \frac{X_i}{p_i^2} \delta_{ij} + \frac{x_n}{p_n^2} \right] \\ &= \frac{x_n}{p_n} \left( \frac{\delta_{ij}}{p_i} + \frac{1}{p_n} \right) \\ &= \frac{x_n}{1 - \sum_{k=1}^{n-1} \xi^k} \left( \frac{\delta_{ij}}{\xi^i} + \frac{1}{1 - \sum_{k=1}^{n-1} \xi^k} \right), \end{aligned} \tag{31}$$

where we used that  $\mathbb{E}[X_i] = x_n p_i / p_n$ .

To find the Fisher–Rao distance, we relate the geometry of this manifold to the radius-two hemisphere model with hyperbolic metric (cf. Sect. 3) using a similar diffeomorphism as (27). Consider

$$\begin{aligned} f: \mathcal{S} \subseteq \mathcal{P}(\mathcal{X}) &\rightarrow S_{2,+}^{n-1} \subseteq \mathcal{H}^n \\ p = \sum_{i=1}^n p_i \mathbb{1}_{\{i\}} &\mapsto (2\sqrt{p_1}, \dots, 2\sqrt{p_n}). \end{aligned} \tag{32}$$

Denote  $\varphi$  the parametrisation of the statistical manifold, and consider the point  $p = \varphi(\xi^1, \dots, \xi^{n-1})$ . Taking the curves  $\alpha_i(t) = \varphi(\xi^1, \dots, \xi^i + t, \dots, \xi^{n-1})$  in  $\mathcal{S}$

their images in  $S_{2,+}^{n-1}$  are

$$\beta_i(t) := (f \circ \alpha_i)(t) = \left( 2\sqrt{\xi^1}, \dots, 2\sqrt{\xi^i + t}, \dots, 2\sqrt{\xi^{n-1}}, 2\sqrt{1 - \sum_{k=1}^{n-1} \xi^k - t} \right).$$

We have

$$df_p \left( \frac{\partial}{\partial \xi^i}(p) \right) = \frac{d}{dt} \beta_i(t) \Big|_{t=0} = \left( 0, \dots, 0, \frac{1}{\sqrt{\xi^i}}, 0, \dots, \frac{-1}{\sqrt{1 - \sum_{k=1}^{n-1} \xi^k}} \right),$$

therefore

$$\left\langle df_p \left( \frac{\partial}{\partial \xi^i}(p) \right), df_p \left( \frac{\partial}{\partial \xi^j}(p) \right) \right\rangle_{G_{\mathcal{H}^n}(p)} = \frac{1}{4(1 - \sum_{k=1}^{n-1} \xi^k)} \left( \frac{\delta_{ij}}{\xi^i} + \frac{1}{1 - \sum_{k=1}^{n-1} \xi^k} \right),$$

and we conclude that  $g_{ij}(\xi) = g_{\xi} \left( \frac{\partial}{\partial \xi^i}(p), \frac{\partial}{\partial \xi^j}(p) \right) = 4x_n \left\langle df_p \left( \frac{\partial}{\partial \xi^i}(p) \right), df_p \left( \frac{\partial}{\partial \xi^j}(p) \right) \right\rangle_{G_{\mathcal{H}^n}(p)}$ . This means that, up to the factor  $4x_n$ ,  $f$  in (32) is an isometry. Then, using (16), we find that the Fisher–Rao distance between two distributions  $p = \varphi(p_1, \dots, p_{n-1})$  and  $q = \varphi(q_1, \dots, q_{n-1})$ , with  $p_n = 1 - \sum_{i=1}^{n-1} p_i$  and  $q_n = 1 - \sum_{i=1}^{n-1} q_i$ , is

$$\begin{aligned} d_{FR}(p, q) &= 2\sqrt{x_n} \tilde{d}_{S_2^{n-1}} \left( (2\sqrt{p_1}, \dots, 2\sqrt{p_n}), (2\sqrt{q_1}, \dots, 2\sqrt{q_n}) \right) \\ &= 2\sqrt{x_n} \operatorname{arccosh} \left( \frac{1 - \sum_{i=1}^{n-1} \sqrt{p_i q_i}}{\sqrt{p_n q_n}} \right). \end{aligned} \tag{33}$$

In the above equality, we used the fact that the hyperbolic distance in  $S_{2,+}^{n-1} \subseteq \mathcal{H}^n$  is invariant to dilation and contraction, as remarked in Sect. 3. We also conclude that the geodesics in the statistical manifold  $\mathcal{S}$  are associated to orthogonal semicircles in  $S_{2,+}^{n-1}$ .

It is interesting to note that the similar maps (27) and (32), respectively, for categorical and negative multinomial distributions are used to embed the statistical manifold (in correspondence with the simplex) in the radius-two sphere, but with different metrics. Figure 2 illustrates the geodesics according to these two metrics.

### 5 Continuous distributions

In the following, we start presenting one-dimensional examples (Sects. 5.1–5.3), then we consider two-dimensional statistical manifolds (Sects. 5.4–5.9), and multivariate models (Sects. 5.10–5.11). The results are summarised in Table 2.

**Table 2** Fisher–Rao distances for continuous distributions

Distribution	Support	Parameters	Fisher matrix	Fisher–Rao distance	References
Exponential $\lambda e^{-\lambda x}$	$x \in \mathbb{R}_+$	$\lambda \in \mathbb{R}_+^*$	$\begin{pmatrix} \frac{1}{\lambda^2} \end{pmatrix}$	$ \log \lambda_1 - \log \lambda_2 $	[10]
Rayleigh $\frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right)$	$x \in \mathbb{R}_+$	$\sigma \in \mathbb{R}_+^*$	$\begin{pmatrix} \frac{4}{\sigma^2} \end{pmatrix}$	$2  \log \sigma_1 - \log \sigma_2 $	*
Erlang $\frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}$	$x \in \mathbb{R}_+$	$\lambda \in \mathbb{R}_+^*$	$\begin{pmatrix} \frac{k}{\lambda^2} \end{pmatrix}$	$\sqrt{k}  \log \lambda_1 - \log \lambda_2 $	*
Gaussian $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$x \in \mathbb{R}$	$(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$	$\begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}$	$2\sqrt{2} \operatorname{arctanh}\left(\sqrt{\frac{(\mu_1 - \mu_2)^2 + 2(\sigma_1 - \sigma_2)^2}{(\mu_1 - \mu_2)^2 + 2(\sigma_1 + \sigma_2)^2}}\right)$	[2, 10, 12, 17, 34, 45]
Laplace $\frac{1}{2\sigma} \exp\left(-\frac{ x-\mu }{\sigma}\right)$	$x \in \mathbb{R}$	$(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$	$\begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{\sigma^2} \end{pmatrix}$	$2 \operatorname{arctanh}\left(\sqrt{\frac{(\mu_1 - \mu_2)^2 + (\sigma_1 - \sigma_2)^2}{(\mu_1 - \mu_2)^2 + (\sigma_1 + \sigma_2)^2}}\right)$	Particular case in [22]
Generalised Gaussian $\frac{\beta}{2\sigma\Gamma(1/\beta)} \exp\left(-\frac{ x-\mu ^\beta}{\sigma}\right)$	$x \in \mathbb{R}$	$(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$	$\begin{pmatrix} \frac{\beta}{\sigma^2} \Gamma(2-\frac{1}{\beta}) & 0 \\ 0 & \frac{\beta}{\sigma^2} \end{pmatrix}$	$\sqrt{\beta+1} \operatorname{arctanh}\left(\sqrt{\frac{(\mu_1 - \mu_2)^2 \Gamma\left(2-\frac{1}{\beta}\right) + \frac{\beta+1}{\beta} (\sigma_1 - \sigma_2)^2 \Gamma\left(1+\frac{1}{\beta}\right)}{(\mu_1 - \mu_2)^2 \Gamma\left(2-\frac{1}{\beta}\right) + \frac{\beta+1}{\beta} (\sigma_1 + \sigma_2)^2 \Gamma\left(1+\frac{1}{\beta}\right)}}\right)$	Particular case in [22]

Table 2 continued

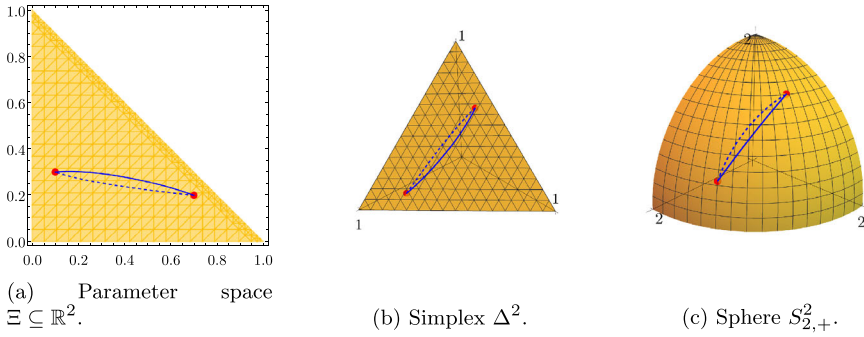
	Distribution	Support	Parameters	Fisher matrix	Fisher-Rao distance	References
Logistic	$\frac{\exp(-(x-\mu)/\sigma)}{\sigma(\exp(-(x-\mu)/\sigma)+1)^2}$	$x \in \mathbb{R}$	$(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$	$\begin{pmatrix} \frac{1}{3\sigma^2} & 0 \\ 0 & \frac{\pi^2+3}{9\sigma^2} \end{pmatrix}$	$\frac{2\sqrt{\pi^2+3}}{3} \operatorname{arctanh} \left( \sqrt{\frac{3(\mu_2-\mu_1)^2+(\pi^2+3)(\sigma_2-\sigma_1)^2}{3(\mu_2-\mu_1)^2+(\pi^2+3)(\sigma_2+\sigma_1)^2}} \right)$	[18]
Cauchy	$\frac{\sigma}{\pi [(x-\mu)^2+\sigma^2]}$	$x \in \mathbb{R}$	$(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$	$\begin{pmatrix} \frac{1}{2\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^2} \end{pmatrix}$	$\sqrt{2} \operatorname{arctanh} \left( \sqrt{\frac{(\mu_1-\mu_2)^2+(\sigma_1-\sigma_2)^2}{(\mu_1-\mu_2)^2+(\sigma_1+\sigma_2)^2}} \right)$	[17, 49]
Student's $t$	$\left(1 + \frac{1}{\nu} \left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}} \times \frac{\Gamma(\frac{\nu+1}{2})}{\sigma \sqrt{\pi \Gamma(\nu/2)}}$	$x \in \mathbb{R}$	$(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$	$\begin{pmatrix} \frac{\nu+1}{(\nu+3)\sigma^2} & 0 \\ 0 & \frac{2\nu}{(\nu+3)\sigma^2} \end{pmatrix}$	$2\sqrt{\frac{2\nu}{\nu+3}} \operatorname{arctanh} \left( \sqrt{\frac{(\nu+1)(\mu_2-\mu_1)^2+2\nu(\sigma_2-\sigma_1)^2}{(\nu+1)(\mu_2-\mu_1)^2+2\nu(\sigma_2+\sigma_1)^2}} \right)$	[17]
Log-Gaussian	$\frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right)$	$x \in \mathbb{R}_+^*$	$(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$	$\begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}$	$2\sqrt{2} \operatorname{arctanh} \left( \sqrt{\frac{(\mu_1-\mu_2)^2+2(\sigma_1-\sigma_2)^2}{(\mu_1-\mu_2)^2+2(\sigma_1+\sigma_2)^2}} \right)$	[2, 34]
Inverse Gaussian	$\sqrt{\frac{\lambda}{2\pi x^3}} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right)$	$x \in \mathbb{R}_+^*$	$(\lambda, \mu) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$	$\begin{pmatrix} \frac{1}{2\lambda^2} & 0 \\ 0 & \frac{\lambda}{\mu^2} \end{pmatrix}$	$2\sqrt{2} \operatorname{arctanh} \left( \sqrt{\frac{\mu_1\mu_2(\sqrt{\lambda_1}-\sqrt{\lambda_2})^2+2\lambda_1\lambda_2(\sqrt{\mu_1}-\sqrt{\mu_2})^2}{\mu_1\mu_2(\sqrt{\lambda_1}-\sqrt{\lambda_2})^2+2\lambda_1\lambda_2(\sqrt{\mu_1}+\sqrt{\mu_2})^2}} \right)$	[23, 48]
Gumbel	$\frac{1}{\sigma} e^{-\frac{x-\mu}{\sigma}} \exp\left(-e^{-\frac{x-\mu}{\sigma}}\right)$	$x \in \mathbb{R}$	$(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$	$\begin{pmatrix} \frac{1}{\sigma^2} & \frac{\gamma-1}{\sigma^2} \\ \frac{\gamma-1}{\sigma^2} & \frac{(\gamma-1)^2+\frac{\pi^2}{6}}{\sigma^2} \end{pmatrix}$	$\frac{2\gamma}{\sqrt{6}} \operatorname{arctanh} \left( \sqrt{\frac{[(\mu_1-\mu_2)-(1-\gamma)(\sigma_1-\sigma_2)]^2+\frac{\pi^2}{6}(\sigma_1-\sigma_2)^2}{[(\mu_1-\mu_2)-(1-\gamma)(\sigma_1-\sigma_2)]^2+\frac{\pi^2}{6}(\sigma_1+\sigma_2)^2}} \right)$	[18]

Table 2 continued

Distribution	Support	Parameters	Fisher matrix	Fisher–Rao distance	References
Fréchet	$x \in \mathbb{R}_+^*$	$(\beta, \lambda) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$	$\begin{pmatrix} \frac{\lambda^2}{\beta^2} & \frac{1-\gamma}{\beta} \\ \frac{1-\gamma}{\beta} & (\gamma-1)^2 + \frac{\pi^2}{6} \end{pmatrix} \frac{1-\gamma}{\lambda^2}$	$\frac{2\pi}{\sqrt{6}} \operatorname{arctanh} \left( \sqrt{\frac{\left[ \log \frac{\beta_1}{\beta_2} - (1-\gamma) \left( \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) \right]^2 + \frac{\pi^2}{6} \left( \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right)^2}{\left[ \log \frac{\beta_1}{\beta_2} - (1-\gamma) \left( \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) \right]^2 + \frac{\pi^2}{6} \left( \frac{1}{\lambda_1} + \frac{1}{\lambda_2} \right)^2}} \right)$	[18]
Weibull	$x \in \mathbb{R}_+^*$	$(\beta, \lambda) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$	$\begin{pmatrix} \frac{\lambda^2}{\beta^2} & \frac{\gamma-1}{\beta} \\ \frac{\gamma-1}{\beta} & (\gamma-1)^2 + \frac{\pi^2}{6} \end{pmatrix} \frac{\gamma-1}{\lambda^2}$	$\frac{2\pi}{\sqrt{6}} \operatorname{arctanh} \left( \sqrt{\frac{\left[ \log \frac{\beta_1}{\beta_2} - (1-\gamma) \left( \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) \right]^2 + \frac{\pi^2}{6} \left( \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right)^2}{\left[ \log \frac{\beta_1}{\beta_2} - (1-\gamma) \left( \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) \right]^2 + \frac{\pi^2}{6} \left( \frac{1}{\lambda_1} + \frac{1}{\lambda_2} \right)^2}} \right)$	[12, 18, 50]
Pareto	$x \in [\alpha, \infty[$	$(\theta, \alpha) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$	$\begin{pmatrix} \frac{1}{\theta^2} & 0 \\ 0 & \frac{\alpha^2}{\alpha^2} \end{pmatrix} \frac{1}{\alpha^2}$	$2 \operatorname{arctanh} \left( \sqrt{\frac{(\theta_1 \theta_2 \log(\alpha_1/\alpha_2))^2 + (\theta_1 - \theta_2)^2}{(\theta_1 \theta_2 \log(\alpha_1/\alpha_2))^2 + (\theta_1 + \theta_2)^2}} \right)$	[12, 14]
Power function	$x \in ]0, \alpha[$	$(\theta, \alpha) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$	$\begin{pmatrix} \frac{1}{\theta^2} & 0 \\ 0 & \frac{\alpha^2}{\alpha^2} \end{pmatrix} \frac{1}{\alpha^2}$	$2 \operatorname{arctanh} \left( \sqrt{\frac{(\theta_1 \theta_2 \log(\alpha_1/\alpha_2))^2 + (\theta_1 - \theta_2)^2}{(\theta_1 \theta_2 \log(\alpha_1/\alpha_2))^2 + (\theta_1 + \theta_2)^2}} \right)$	Particular case in [12]
Wishart	$X \in P_m(\mathbb{R})$	Same as for $\Sigma \in P_m(\mathbb{R})$	$\frac{n}{2} D_m^T (\Sigma^{-1} \otimes \Sigma^{-1}) D_m$	$\sqrt{\frac{n}{2} \sum_{k=1}^m (\log \lambda_k)^2}$	[11]
Inverse Wishart	$X \in P_m(\mathbb{R})$	Same as for $\Sigma \in P_m(\mathbb{R})$	$\frac{n}{2} D_m^T (\Sigma^{-1} \otimes \Sigma^{-1}) D_m$	$\sqrt{\frac{n}{2} \sum_{k=1}^m (\log \lambda_k)^2}$	*

Wishart distribution:  $\frac{(\det X)^{(n-m-1)/2} e^{-\operatorname{tr}(\Sigma^{-1} X)/2}}{2^{nm/2} (\det \Sigma)^{n/2} \Gamma_m(n/2)}$

Inverse Wishart distribution:  $\frac{(\det \Sigma)^{n/2} (\det X)^{-(n+m+1)/2} \exp\left(-\frac{1}{2} \operatorname{tr}(\Sigma X^{-1})\right)}{2^{nm/2} \Gamma_m(n/2)}$



**Fig. 2** Geodesics joining points  $p = (0.7, 0.2, 0.1)$  and  $q = (0.1, 0.3, 0.6)$  according to categorical (solid) and negative multinomial (dashed) metrics, seen in the parameter space, on the simplex, and on the sphere. The distance between the categorical distributions is  $d_{\text{cat}}(p, q) \approx 1.432$ , and between the negative multinomial distributions is  $d_{\text{neg-mult}}(p, q) \approx 2.637\sqrt{x_n}$

### 5.1 Exponential

An exponential distribution [10] has p.d.f.  $p(x) = \lambda e^{-\lambda x}$ , defined for  $x \in \mathbb{R}_+$ , and parametrised by  $\lambda \in \mathbb{R}_+^*$ . In this case, we have  $\partial_\lambda \ell(\lambda) = \frac{1}{\lambda} - x$ , and the Fisher information is

$$\begin{aligned}
 g_{11}(\lambda) &= \mathbb{E} \left[ (\partial_\lambda \ell(\lambda))^2 \right] \\
 &= \mathbb{E} \left[ \left( \frac{1}{\lambda} - X \right)^2 \right] \\
 &= \frac{1}{\lambda^2} - \frac{2}{\lambda} \mathbb{E}[X] + \mathbb{E}[X^2] \\
 &= \frac{1}{\lambda^2},
 \end{aligned}
 \tag{34}$$

where we have used that  $\mathbb{E}[X] = \frac{1}{\lambda}$  and  $\mathbb{E}[X^2] = \frac{2}{\lambda^2}$ . The Fisher–Rao distance is given by

$$d_{\text{FR}}(\lambda_1, \lambda_2) = \left| \int_{\lambda_1}^{\lambda_2} \frac{1}{\lambda} \, d\lambda \right| = |\log \lambda_1 - \log \lambda_2|.
 \tag{35}$$

### 5.2 Rayleigh

A Rayleigh distribution has p.d.f.  $p(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right)$ , defined for  $x \in \mathbb{R}_+$ , and parametrised by  $\sigma \in \mathbb{R}_+^*$ . We have  $\partial_\sigma \ell(\sigma) = \frac{x^2}{\sigma^3} - \frac{2}{\sigma}$ , and the Fisher information is

$$\begin{aligned}
 g_{11}(\sigma) &= \mathbb{E} \left[ (\partial_\sigma \ell(\sigma))^2 \right] = \mathbb{E} \left[ \left( \frac{X^2}{\sigma^3} - \frac{2}{\sigma} \right)^2 \right] \\
 &= \frac{\mathbb{E}[X^4]}{\sigma^6} - \frac{2\mathbb{E}[X^2]}{\sigma^4} + \frac{4}{\sigma^2}
 \end{aligned}$$

$$= \frac{4}{\sigma^2}, \tag{36}$$

where we have used that  $\mathbb{E}[X^2] = 4\sigma^2$  and  $\mathbb{E}[X^4] = 8\sigma^4$ . Thus, the Fisher–Rao distance is given by

$$d_{\text{FR}}(\sigma_1, \sigma_2) = \left| \int_{\sigma_1}^{\sigma_2} \frac{2}{\sigma} \, d\sigma \right| = 2 |\log \sigma_1 - \log \sigma_2|. \tag{37}$$

### 5.3 Erlang

An Erlang distribution has p.d.f.  $p(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}$ , defined for  $x \in \mathbb{R}_+$ , and parametrised by  $\lambda \in \mathbb{R}_+^*$ , for a fixed  $k \in \mathbb{N}^*$ . We have  $\partial_\lambda \ell(\lambda) = \frac{k}{\lambda} - x$ , so that the Fisher information is

$$\begin{aligned} g_{11}(\lambda) &= \mathbb{E} \left[ (\partial_\lambda \ell(\lambda))^2 \right] = \mathbb{E} \left[ \left( \frac{k}{\lambda} - X \right)^2 \right] \\ &= \mathbb{E}[X^2] - \frac{2k\mathbb{E}[X]}{\lambda} + \frac{k^2}{\lambda^2} \\ &= \frac{k}{\lambda^2}, \end{aligned} \tag{38}$$

where we have used that  $\mathbb{E}[X] = \frac{k}{\lambda}$  and  $\mathbb{E}[X^2] = \frac{k(k+1)}{\lambda^2}$ . The Fisher–Rao distance is then

$$d_{\text{FR}}(\lambda_1, \lambda_2) = \left| \int_{\lambda_1}^{\lambda_2} \frac{\sqrt{k}}{\lambda} \, d\lambda \right| = \sqrt{k} |\log \lambda_1 - \log \lambda_2|. \tag{39}$$

### 5.4 Univariate elliptical distributions

Elliptical distributions are a class of distributions that generalise Gaussian distributions [51]. Here we focus on univariate elliptical distributions, which are defined for  $x \in \mathbb{R}$ , parametrised by  $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$ , and have p.d.f. of the form

$$p(x) = \frac{1}{\sigma} h \left( \frac{(x - \mu)^2}{\sigma^2} \right), \tag{40}$$

for a fixed measurable function  $h: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  that satisfies  $\int_{-\infty}^{\infty} h(z^2) \, dz = 1$  and  $\lim_{z \rightarrow +\infty} zh(z^2) = 0$ . For a random variable  $X$  distributed according to (40), provided its mean and variance exist, they are given by  $\mathbb{E}[X] = \mu$  and  $\text{Var}(X) = \sigma^2 \int_{-\infty}^{\infty} z^2 h(z^2) \, dz$ . For each function  $h$ , the set of distributions of the form (40) forms a statistical manifold parametrised by  $(\mu, \sigma)$ . Some examples of these manifolds have been studied in [17], and we write some other examples in the same standard form.

For these distributions, we have  $\partial_\mu \ell := \partial_\mu \ell(\mu, \sigma) = -\frac{2(x-\mu)}{\sigma^2} \frac{h'(\sigma^{-2}(x-\mu)^2)}{h(\sigma^{-2}(x-\mu)^2)}$ , and  $\partial_\sigma \ell := \partial_\sigma \ell(\mu, \sigma) = -\frac{1}{\sigma} - \frac{2(x-\mu)^2}{\sigma^3} \frac{h'(\sigma^{-2}(x-\mu)^2)}{h(\sigma^{-2}(x-\mu)^2)}$ . Thus, the elements of the Fisher matrix are

$$g_{11} = \mathbb{E} \left[ (\partial_\mu \ell)^2 \right] \\ = \int_{-\infty}^{\infty} \left( -\frac{2(x-\mu)}{\sigma^2} \frac{h'(\sigma^{-2}(x-\mu)^2)}{h(\sigma^{-2}(x-\mu)^2)} \right)^2 \frac{1}{\sigma} h \left( \frac{(x-\mu)^2}{\sigma^2} \right) dx \\ = \frac{4}{\sigma^2} \int_{-\infty}^{\infty} z^2 \frac{[h'(z^2)]^2}{h(z^2)} dz,$$

$$g_{12} = g_{21} = \mathbb{E} \left[ (\partial_\mu \ell) (\partial_\sigma \ell) \right] \\ = \int_{-\infty}^{\infty} \left( -\frac{2(x-\mu)}{\sigma^2} \frac{h'(\sigma^{-2}(x-\mu)^2)}{h(\sigma^{-2}(x-\mu)^2)} \right) \left( -\frac{1}{\sigma} - \frac{2(x-\mu)^2}{\sigma^3} \frac{h'(\sigma^{-2}(x-\mu)^2)}{h(\sigma^{-2}(x-\mu)^2)} \right) \\ \times \frac{1}{\sigma} h \left( \frac{(x-\mu)^2}{\sigma^2} \right) dx \\ = \frac{2}{\sigma^2} \int_{-\infty}^{\infty} z \left( h'(z^2) + 2z^2 \frac{[h'(z^2)]^2}{h(z^2)} \right) dz \\ = 0,$$

$$g_{22} = \mathbb{E} \left[ (\partial_\sigma \ell)^2 \right] \\ = \int_{-\infty}^{\infty} \left( -\frac{1}{\sigma} - \frac{2(x-\mu)^2}{\sigma^3} \frac{h'(\sigma^{-2}(x-\mu)^2)}{h(\sigma^{-2}(x-\mu)^2)} \right)^2 \frac{1}{\sigma} h \left( \frac{(x-\mu)^2}{\sigma^2} \right) dx \\ = \int_{-\infty}^{\infty} \left( \frac{1}{\sigma^2} + \frac{4}{\sigma^2} z^4 \left( \frac{h'(z^2)}{h(z^2)} \right)^2 + \frac{4}{\sigma^2} z^2 \frac{h'(z^2)}{h(z^2)} \right) h(z^2) dz \\ = \frac{1}{\sigma^2} + \frac{4}{\sigma^2} \int_{-\infty}^{\infty} z^4 \frac{[h'(z^2)]^2}{h(z^2)} dz + \frac{4}{\sigma^2} \int_{-\infty}^{\infty} z^2 h'(z^2) dz \\ = \frac{4}{\sigma^2} \int_{-\infty}^{\infty} z^4 \frac{[h'(z^2)]^2}{h(z^2)} dz - \frac{1}{\sigma^2}.$$

The Fisher matrix for elliptical distributions has the form

$$G(\mu, \sigma) = \begin{pmatrix} \frac{a_h}{\sigma^2} & 0 \\ 0 & \frac{b_h}{\sigma^2} \end{pmatrix}, \tag{41}$$

where

$$a_h := 4 \int_{-\infty}^{\infty} z^2 \frac{[h'(z^2)]^2}{h(z^2)} dz,$$



and

$$b_h := 4 \int_{-\infty}^{\infty} z^4 \frac{[h'(z^2)]^2}{h(z^2)} dz - 1.$$

Applying Lemma 4 to a statistical manifold formed by univariate elliptical distributions (i.e., for a fixed  $h$ ) equipped with the metric (41), we get an expression for the Fisher–Rao distance in this manifold given by

$$\begin{aligned} d_{FR}((\mu_1, \sigma_1), (\mu_2, \sigma_2)) &= \sqrt{b_h} d_{\gamma^2} \left( (\sqrt{a_h} \mu_1, \sqrt{b_h} \sigma_1), (\sqrt{a_h} \mu_2, \sqrt{b_h} \sigma_2) \right) \\ &= 2\sqrt{b_h} \operatorname{arctanh} \left( \sqrt{\frac{a_h(\mu_1 - \mu_2)^2 + b_h(\sigma_1 - \sigma_2)^2}{a_h(\mu_1 - \mu_2)^2 + b_h(\sigma_1 + \sigma_2)^2}} \right). \end{aligned} \tag{42}$$

Note that, if  $\mu_1 = \mu_2 =: \mu$ , the expression simplifies to

$$d_{FR}((\mu, \sigma_1), (\mu, \sigma_2)) = \sqrt{b_h} |\log \sigma_1 - \log \sigma_2|. \tag{43}$$

### 5.4.1 Gaussian

A Gaussian distribution [2, 10, 12, 17, 34, 45] is characterised by the p.d.f.  $p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ , defined for  $x \in \mathbb{R}$ , and parametrised by the pair  $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$ . Note that Gaussian distributions are elliptical distribution with  $h(u) = \frac{1}{\sqrt{2\pi}} \exp(-u/2)$ ,  $a_h = 1$ , and  $b_h = 2$ . Thus, by (41), the corresponding Fisher matrix is

$$G(\mu, \sigma) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}. \tag{44}$$

And, by (42), the Fisher–Rao distance is obtained as

$$d_{FR}((\mu_1, \sigma_1), (\mu_2, \sigma_2)) = 2\sqrt{2} \operatorname{arctanh} \left( \sqrt{\frac{(\mu_1 - \mu_2)^2 + 2(\sigma_1 - \sigma_2)^2}{(\mu_1 - \mu_2)^2 + 2(\sigma_1 + \sigma_2)^2}} \right). \tag{45}$$

### 5.4.2 Laplace

A Laplace distribution has p.d.f.  $p(x) = \frac{1}{2\sigma} \exp\left(-\frac{|x-\mu|}{\sigma}\right)$ , defined for  $x \in \mathbb{R}$ , and parametrised by  $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$ . Laplace distributions are elliptical distribution with  $h(u) = \frac{1}{2} \exp(-\sqrt{|u|})$ ,  $a_h = 1$ , and  $b_h = 1$ . Using (41), we get the Fisher matrix as

$$G(\mu, \sigma) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{\sigma^2} \end{pmatrix}, \tag{46}$$

showing that this metric coincides with the hyperbolic metric (11) of the Poincaré half-plane. Using (42), we get that the Fisher–Rao distance in this manifold is

$$d_{FR}((\mu_1, \sigma_1), (\mu_2, \sigma_2)) = 2 \operatorname{arctanh} \left( \sqrt{\frac{(\mu_1 - \mu_2)^2 + (\sigma_1 - \sigma_2)^2}{(\mu_1 - \mu_2)^2 + (\sigma_1 + \sigma_2)^2}} \right). \tag{47}$$

The particular case of zero-mean Laplace distributions is included in [22].

### 5.4.3 Generalised Gaussian

A generalised Gaussian distribution<sup>3</sup> [22, 51, 53], also known as exponential power distribution, is characterised by the p.d.f.  $p(x) = \frac{\beta}{2\sigma\Gamma(1/\beta)} \exp\left(-\frac{|x-\mu|^\beta}{\sigma}\right)$ , defined for  $x \in \mathbb{R}$ , and parametrised by  $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$ , for a fixed  $\beta > 0$ . These can be seen as elliptical distributions with  $h(u) = \frac{\beta}{2\Gamma(1/\beta)} \exp(-u^{\beta/2})$ ,  $a_h = \beta \frac{\Gamma(2-1/\beta)}{\Gamma(1+1/\beta)}$ , and  $b_h = \beta$ . Using (41), we get the Fisher matrix as

$$G(\mu, \sigma) = \begin{pmatrix} \frac{\beta}{\sigma^2} \frac{\Gamma(2-1/\beta)}{\Gamma(1+1/\beta)} & 0 \\ 0 & \frac{\beta}{\sigma^2} \end{pmatrix}, \tag{48}$$

and, using (42), we get the Fisher–Rao distance as

$$\begin{aligned} d_{FR}((\mu_1, \sigma_1), (\mu_2, \sigma_2)) &= \sqrt{\beta + 1} \operatorname{arctanh} \left( \sqrt{\frac{\beta(\mu_1 - \mu_2)^2 \Gamma(2 - 1/\beta) + (\beta + 1)(\sigma_1 - \sigma_2)^2 \Gamma(1 + 1/\beta)}{\beta(\mu_1 - \mu_2)^2 \Gamma(2 - 1/\beta) + (\beta + 1)(\sigma_1 + \sigma_2)^2 \Gamma(1 + 1/\beta)}} \right). \end{aligned} \tag{49}$$

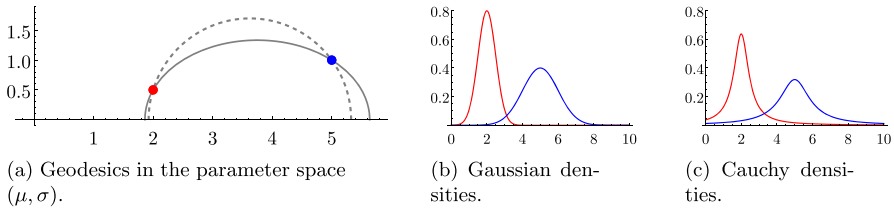
Note that choosing  $\beta = 2$  yields a Gaussian distribution with mean  $\mu$  and variance  $\sigma^2/2$ , while letting  $\beta = 1$  corresponds to a Laplace distribution with mean  $\mu$  and variance  $8\sigma^2$ . Multivariate, zero-mean generalised Gaussian distributions have been studied in [22].

### 5.4.4 Logistic

A logistic distribution [18] has p.d.f.  $p(x) = \frac{\exp(-(x-\mu)/\sigma)}{\sigma(\exp(-(x-\mu)/\sigma)+1)^2}$ , defined for  $x \in \mathbb{R}$  and parametrised by  $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$ . A logistic distribution is an elliptical distribution with  $h(u) = \frac{\exp(-\sqrt{u})}{(1+\exp(-\sqrt{u}))^2}$ ,  $a_h = \frac{1}{3}$ , and  $b_h = \frac{\pi^2+3}{9}$ . From (41), we have that the Fisher matrix is

$$G(\mu, \sigma) = \begin{pmatrix} \frac{1}{3\sigma^2} & 0 \\ 0 & \frac{\pi^2+3}{9\sigma^2} \end{pmatrix}, \tag{50}$$

<sup>3</sup> These generalised Gaussian distributions are in a different sense that those considered in [52].



**Fig. 3** Geodesics joining points  $(\mu_1, \sigma_1) = (2, 0.5)$  and  $(\mu_2, \sigma_2) = (5, 1)$  according to Gaussian metric (solid) and Cauchy metric (dashed), seen in the parameter space  $(\mu, \sigma)$ , and the corresponding densities. The distance between the two Gaussian distributions is  $d_{\text{Gaussian}}((\mu_1, \sigma_1), (\mu_2, \sigma_2)) \approx 3.443$ , and between the Cauchy distributions is  $d_{\text{Cauchy}}((\mu_1, \sigma_1), (\mu_2, \sigma_2)) \approx 1.721$

and, from (42), the Fisher–Rao distance is

$$d_{\text{FR}}((\mu_1, \sigma_1), (\mu_2, \sigma_2)) = \frac{2\sqrt{\pi^2 + 3}}{3} \operatorname{arctanh} \left( \sqrt{\frac{3(\mu_2 - \mu_1)^2 + (\pi^2 + 3)(\sigma_2 - \sigma_1)^2}{3(\mu_2 - \mu_1)^2 + (\pi^2 + 3)(\sigma_2 + \sigma_1)^2}} \right). \tag{51}$$

### 5.4.5 Cauchy

A Cauchy distribution [17, 49] has p.d.f.  $p(x) = \frac{\sigma}{\pi[(x-\mu)^2 + \sigma^2]}$ , defined for  $x \in \mathbb{R}$ , and parametrised by  $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$ . Cauchy distributions are elliptical distributions by choosing  $h(u) = \frac{1}{\pi(1+u^2)}$ ,  $a_h = 1/2$ ,  $b_h = 1/2$ . Recall that the mean and variance are not defined in this case. From (41), we get its the Fisher matrix

$$G(\mu, \sigma) = \begin{pmatrix} \frac{1}{2\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^2} \end{pmatrix}, \tag{52}$$

and (42) gives the Fisher–Rao distance as

$$d_{\text{FR}}((\mu_1, \sigma_1), (\mu_2, \sigma_2)) = \sqrt{2} \operatorname{arctanh} \left( \sqrt{\frac{(\mu_1 - \mu_2)^2 + (\sigma_1 - \sigma_2)^2}{(\mu_1 - \mu_2)^2 + (\sigma_1 + \sigma_2)^2}} \right). \tag{53}$$

Figure 3 illustrates geodesics between Gaussian and Cauchy distributions.

### 5.4.6 Student’s *t*

A location-scale Student’s *t* distribution [17] with  $\nu \in \mathbb{N}^*$  degrees of freedom generalises the Cauchy distribution. It has p.d.f.  $p(x) = \left(1 + \frac{1}{\nu} \left(\frac{x-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}} \frac{\Gamma((\nu+1)/2)}{\sigma\sqrt{\pi\nu}\Gamma(\nu/2)}$ , defined for  $x \in \mathbb{R}$ , and parametrised by  $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$ . This is an elliptical distribution, with  $h(u) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\Gamma(\nu/2)} \left(1 + \frac{u^2}{\nu}\right)^{-(\nu+1)/2}$ ,  $a_h = \frac{\nu+1}{\nu+3}$ , and  $b_h = \frac{2\nu}{\nu+3}$ . Then,

by (41) we obtain the Fisher matrix

$$G(\mu, \sigma) = \begin{pmatrix} \frac{\nu+1}{(\nu+3)\sigma^2} & 0 \\ 0 & \frac{2\nu}{(\nu+3)\sigma^2} \end{pmatrix}, \tag{54}$$

and by (42), the Fisher–Rao distance

$$d_{FR}((\mu_1, \sigma_1), (\mu_2, \sigma_2)) = 2\sqrt{\frac{2\nu}{\nu+3}} \operatorname{arctanh} \left( \sqrt{\frac{(\nu+1)(\mu_2 - \mu_1)^2 + 2\nu(\sigma_2 - \sigma_1)^2}{(\nu+1)(\mu_2 - \mu_1)^2 + 2\nu(\sigma_2 + \sigma_1)^2}} \right). \tag{55}$$

**Remark 5** We close this subsection with a remark on the general case of multivariate elliptical distributions. These are distributions of the form  $p(\mathbf{x}) = (\det \Sigma)^{-1/2} h((\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}))$ , for some function  $h: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ , defined for  $\mathbf{x} \in \mathbb{R}^n$ , and characterised by a vector  $\boldsymbol{\mu} \in \mathbb{R}^n$  and an  $n \times n$  positive-definite symmetric matrix  $\Sigma \in P_n(\mathbb{R})$ . Analogously to the univariate case, the set of elliptical distributions for a fixed  $h$  forms an  $(n + \frac{n(n+1)}{2})$ -dimensional statistical manifold. In the general case, however, no general closed-form expression for the Fisher–Rao distance is known; instead, only expressions for particular cases and bounds for this distance are available [17, 54–57]. Multivariate Gaussian distributions  $p(\mathbf{x}) = ((2\pi)^n \det \Sigma)^{-1/2} \exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}))$  have been particularly studied [10, 12, 13, 20, 58–60]. Special cases for which the Fisher–Rao distance can be written include: fixed mean, fixed covariance, diagonal covariance matrix, and mirrored distributions. The case of diagonal covariance matrix is equivalent to independent components and will be treated in Sect. 6. Special cases for the multivariate generalised Gaussian distributions have been studied in [22]. The Fisher–Rao distance between zero-mean complex elliptically symmetric distributions has been computed in [33, 61].

### 5.5 Log-Gaussian

A log-Gaussian distribution [2, 34] is the distribution of a random variable whose logarithm follows a Gaussian distribution. It has p.d.f.  $p(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right)$ , defined for  $x \in \mathbb{R}_+^*$ , and parametrised by  $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$ . In this case, we have  $\partial_\mu \ell := \partial_\mu \ell(\mu, \sigma) = \frac{\log x - \mu}{\sigma^2}$  and  $\partial_\sigma \ell := \partial_\sigma \ell(\mu, \sigma) = -\frac{1}{\sigma} + \frac{(\log x - \mu)^2}{\sigma^3}$ , so that the elements of the Fisher matrix are

$$\begin{aligned} g_{11} &= \mathbb{E} \left[ (\partial_\mu \ell)^2 \right] = \mathbb{E} \left[ \left( \frac{\log X - \mu}{\sigma^2} \right)^2 \right] \\ &= \frac{\mathbb{E}[(\log X - \mu)^2]}{\sigma^4} = \frac{1}{\sigma^2}, \\ g_{12} = g_{21} &= \mathbb{E} \left[ (\partial_\mu \ell) (\partial_\sigma \ell) \right] = \mathbb{E} \left[ \left( \frac{\log X - \mu}{\sigma^2} \right) \left( -\frac{1}{\sigma} + \frac{(\log X - \mu)^2}{\sigma^3} \right) \right] \end{aligned}$$

$$\begin{aligned}
 &= -\frac{\mathbb{E}[\log X - \mu]}{\sigma^3} + \frac{\mathbb{E}[(\log X - \mu)^3]}{\sigma^5} \\
 &= 0, \\
 g_{22} &= \mathbb{E}\left[(\partial_\sigma \ell)^2\right] = \mathbb{E}\left[\left(-\frac{1}{\sigma} + \frac{(\log X - \mu)^2}{\sigma^3}\right)^2\right] \\
 &= \frac{1}{\sigma^2} - \frac{2\mathbb{E}[(\log X - \mu)^2]}{\sigma^4} + \frac{\mathbb{E}[(\log X - \mu)^4]}{\sigma^6} = \frac{2}{\sigma^2},
 \end{aligned}$$

having used that  $\mathbb{E}[\log X] = \mu$ ,  $\mathbb{E}[\log X - \mu] = \mathbb{E}[(\log X - \mu)^3] = 0$ ,  $\mathbb{E}[(\log X - \mu)^2] = \sigma^2$ , and  $\mathbb{E}[(\log X - \mu)^4] = 3\sigma^4$ . Thus, the Fisher matrix is given by

$$G(\mu, \sigma) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}. \tag{56}$$

This is the same as for the Gaussian manifold (44), therefore the Fisher–Rao distance the same, namely,

$$d_{FR}((\mu_1, \sigma_1), (\mu_2, \sigma_2)) = 2\sqrt{2} \operatorname{arctanh} \left( \sqrt{\frac{(\mu_1 - \mu_2)^2 + 2(\sigma_1 - \sigma_2)^2}{(\mu_1 - \mu_2)^2 + 2(\sigma_1 + \sigma_2)^2}} \right). \tag{57}$$

### 5.6 Inverse Gaussian

An inverse Gaussian distribution [23, 48] has p.d.f.  $p(x) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right)$ , defined for  $x \in \mathbb{R}_+^*$ , and parametrised by  $(\lambda, \mu) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$ . In this case, we have  $\partial_\lambda \ell := \partial_\lambda \ell(\lambda, \mu) = \frac{1}{2\lambda} - \frac{(x-\mu)^2}{2\mu^2 x}$  and  $\partial_\mu \ell := \partial_\mu \ell(\lambda, \mu) = \frac{\lambda(x-\mu)}{\mu^3}$ . The elements of the Fisher matrix are then

$$\begin{aligned}
 g_{11} &= \mathbb{E}\left[(\partial_\lambda \ell)^2\right] = \mathbb{E}\left[\left(\frac{1}{2\lambda} - \frac{(x-\mu)^2}{2\mu^2 x}\right)^2\right] \\
 &= \frac{1}{4\lambda^2} + \frac{1}{\lambda\mu} + \frac{3}{2\mu^2} - \left(\frac{1}{2\lambda\mu^2} + \frac{1}{\mu^3}\right) \mathbb{E}[X] \\
 &\quad - \left(\frac{1}{2\lambda} + \frac{1}{\mu}\right) \mathbb{E}\left[\frac{1}{X}\right] + \frac{1}{4\mu^4} \mathbb{E}[X^2] + \frac{1}{4} \mathbb{E}\left[\frac{1}{X^2}\right] \\
 &= \frac{1}{2\lambda^2}, \\
 g_{12} &= g_{21} = \mathbb{E}\left[(\partial_\lambda \ell)(\partial_\mu \ell)\right] = \mathbb{E}\left[\left(\frac{1}{2\lambda} - \frac{(x-\mu)^2}{2\mu^2 x}\right)\left(\frac{\lambda(x-\mu)}{\mu^3}\right)\right] \\
 &= -\frac{1}{2\lambda\mu^2} - \frac{3}{2\mu^3} + \left(\frac{1}{2\lambda\mu^3} - \frac{3\lambda}{2\mu^4}\right) \mathbb{E}[X] + \frac{\lambda}{2\mu^2} \mathbb{E}\left[\frac{1}{X}\right] - \frac{\lambda}{2\mu^5} \mathbb{E}[X^2] \\
 &= 0,
 \end{aligned}$$

$$\begin{aligned}
 g_{22} &= \mathbb{E} \left[ (\partial_\mu \ell)^2 \right] = \mathbb{E} \left[ \left( \frac{\lambda(x - \mu)}{\mu^3} \right)^2 \right] \\
 &= \frac{\lambda^2}{\mu^4} - \frac{2\lambda^2}{\mu^5} \mathbb{E}[X] + \frac{\lambda^2}{\mu^6} \mathbb{E}[X^2] \\
 &= \frac{\lambda}{\mu^3},
 \end{aligned}$$

where we have used that  $\mathbb{E}[X] = \mu$ ,  $\mathbb{E}[X^2] = \frac{\mu^3}{\lambda} + \mu^2$ ,  $\mathbb{E} \left[ \frac{1}{X} \right] = \frac{1}{\lambda} + \frac{1}{\mu}$  and  $\mathbb{E} \left[ \frac{1}{X^2} \right] = \frac{3}{\lambda^2} + \frac{3}{\lambda\mu} + \frac{1}{\mu^2}$ . Thus, the Fisher matrix is given by

$$G(\lambda, \mu) = \begin{pmatrix} \frac{1}{2\lambda^2} & 0 \\ 0 & \frac{\lambda}{\mu^3} \end{pmatrix}. \tag{58}$$

To find the Fisher–Rao distance, we consider the change of coordinates  $u = 1/\sqrt{\lambda}$ ,  $v = \sqrt{2/\mu}$ . Applying Proposition 2 we find that the Fisher matrix in the new coordinates

$$\tilde{G}(u, v) = \begin{pmatrix} \frac{2}{u^2} & 0 \\ 0 & \frac{2}{v^2} \end{pmatrix} = 2 \begin{pmatrix} \frac{1}{u^2} & 0 \\ 0 & \frac{1}{v^2} \end{pmatrix}.$$

Applying Lemma 4 yields

$$\begin{aligned}
 d_{\text{FR}}((\lambda_1, \mu_1), (\lambda_2, \mu_2)) &= \sqrt{2} d_{\mathcal{H}^2} \left( (1/\sqrt{\lambda_1}, \sqrt{2/\mu_1}), (1/\sqrt{\lambda_2}, \sqrt{2/\mu_2}) \right) \\
 &= 2\sqrt{2} \operatorname{arctanh} \left( \sqrt{\frac{\mu_1\mu_2(\sqrt{\lambda_1} - \sqrt{\lambda_2})^2 + 2\lambda_1\lambda_2(\sqrt{\mu_1} - \sqrt{\mu_2})^2}{\mu_1\mu_2(\sqrt{\lambda_1} - \sqrt{\lambda_2})^2 + 2\lambda_1\lambda_2(\sqrt{\mu_1} + \sqrt{\mu_2})^2}} \right). \tag{59}
 \end{aligned}$$

### 5.7 Extreme-value distributions

Extreme-value distributions [62] are limit distribution for the maxima (or minima) of a sequence of i.i.d. random variables. They are usually considered to be of one of three families, all of which can be described in the form of the generalised extreme-value distributions:

$$p(x) = \frac{1}{\sigma} [t(x, \xi)]^{\xi+1} e^{-t(x, \xi)}, \tag{60}$$

where  $\sigma > 0$ ,  $\xi \in \mathbb{R}$ , and

$$t(x, \xi) = \begin{cases} \left( 1 + \xi \left( \frac{x-\mu}{\sigma} \right) \right)^{-\frac{1}{\xi}}, & \xi \neq 0, \\ \exp \left( -\frac{x-\mu}{\sigma} \right), & \xi = 0, \end{cases}$$

for  $\mu \in \mathbb{R}$ . Note that  $t(x, \xi)$  is continuous on  $\xi = 0$ , for every  $x \in \mathbb{R}$ . When  $\xi = 0$ , the support of (60) is  $x \in \mathbb{R}$ , and it is called type I or *Gumbel-type* distribution; when

$\xi > 0$ , the support is  $x \in [\mu - \frac{\sigma}{\xi}, +\infty[$ , and it is called type II or *Fréchet-type* distribution; when  $\xi < 0$ , the support is  $x \in ]-\infty, \mu - \frac{\sigma}{\xi}]$ , and it is called type III or *Weibull-type* distribution. Therefore, in general, these distributions are parametrised by the triple  $(\mu, \sigma, \xi) \in \mathbb{R} \times \mathbb{R}_+^* \times \mathbb{R}$ . Instead of treating the three-dimensional manifold of generalised extreme-value distributions in full generality [25, 63], for which no closed-form expression for the Fisher–Rao is available, we consider the usual two-dimensional versions of them, following [18].

### 5.7.1 Gumbel

A Gumbel distribution [18] has p.d.f.  $p(x) = \frac{1}{\sigma} \exp(-\frac{x-\mu}{\sigma}) \exp(-\exp(-\frac{x-\mu}{\sigma}))$ , defined for  $x \in \mathbb{R}$ , and parametrised by  $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}_+^*$ . This corresponds to (60) taking  $\xi = 0$ . In this case, we have  $\partial_\mu \ell := \partial_\mu \ell(\mu, \sigma) = \frac{1}{\sigma} - \frac{1}{\sigma} \exp(-\frac{x-\mu}{\sigma})$  and  $\partial_\sigma \ell := \partial_\sigma \ell(\mu, \sigma) = -\frac{x-\mu}{\sigma^2} \exp(-\frac{x-\mu}{\sigma}) + \frac{x-\mu}{\sigma^2} - \frac{1}{\sigma}$ . Denoting  $Z := \frac{X-\mu}{\sigma}$ , the elements of the Fisher matrix can be written as

$$\begin{aligned} g_{11} &= \mathbb{E}[(\partial_\mu \ell)^2] = \mathbb{E}\left[\left(\frac{1}{\sigma} - \frac{1}{\sigma} e^{-Z}\right)^2\right] \\ &= \frac{1}{\sigma^2} - \frac{2}{\sigma^2} \mathbb{E}[e^{-Z}] + \frac{1}{\sigma^2} \mathbb{E}[e^{-2Z}] \\ &= \frac{1}{\sigma^2}, \\ g_{12} &= g_{21} = \mathbb{E}[(\partial_\mu \ell)(\partial_\sigma \ell)] \\ &= \mathbb{E}\left[\left(\frac{1}{\sigma} - \frac{1}{\sigma} e^{-Z}\right) \left(\frac{Ze^{-Z}}{\sigma} + \frac{Z}{\sigma} + \frac{1}{\sigma^2}\right)\right] \\ &= -\frac{2}{\sigma^2} \mathbb{E}[Ze^{-Z}] + \frac{1}{\sigma^2} \mathbb{E}[Z] - \frac{1}{\sigma^2} + \frac{1}{\sigma^2} \mathbb{E}[Ze^{-2Z}] + \frac{1}{\sigma^2} \mathbb{E}[e^{-Z}] \\ &= \frac{\gamma - 1}{\sigma^2}, \\ g_{22} &= \mathbb{E}[(\partial_\sigma \ell)^2] = \mathbb{E}\left[\left(\frac{Ze^{-Z}}{\sigma} + \frac{Z}{\sigma} + \frac{1}{\sigma^2}\right)^2\right] \\ &= \frac{1}{\sigma^2} \mathbb{E}[Z^2 e^{-2Z}] - \frac{2}{\sigma^2} \mathbb{E}[Z^2 e^{-Z}] + \frac{2}{\sigma^2} \mathbb{E}[Ze^{-Z}] \\ &\quad + \frac{1}{\sigma^2} \mathbb{E}[Z^2] - \frac{2}{\sigma^2} \mathbb{E}[Z] + \frac{1}{\sigma^2} \\ &= \frac{1}{\sigma^2} \left( (\gamma - 1)^2 + \frac{\pi^2}{6} \right), \end{aligned}$$

where  $\gamma$  is the Euler constant and we have used that  $\mathbb{E}[Z] = \gamma$ ,  $\mathbb{E}[e^{-Z}] = 1$ ,  $\mathbb{E}[e^{-2Z}] = 2$ ,  $\mathbb{E}[Ze^{-Z}] = \gamma - 1$ ,  $\mathbb{E}[Ze^{-2Z}] = 2\gamma - 3$ ,  $\mathbb{E}[Z^2 e^{-2Z}] = 2\gamma^2 - 6\gamma + 2 + \frac{\pi^2}{3}$ ,  $\mathbb{E}[Z^2 e^{-Z}] = \gamma^2 - 2\gamma + \frac{\pi^2}{6}$  and  $\mathbb{E}[Z^2] = \gamma^2 + \frac{\pi^2}{6}$ . Thus, the Fisher matrix

is

$$G(\mu, \sigma) = \begin{pmatrix} \frac{1}{\sigma^2} & \frac{\gamma-1}{\sigma^2} \\ \frac{\gamma-1}{\sigma^2} & \frac{1}{\sigma^2} \left( (\gamma-1)^2 + \frac{\pi^2}{6} \right) \end{pmatrix}. \tag{61}$$

To find the Fisher–Rao distance, we consider the change of coordinates  $u = \mu - (1 - \gamma)\sigma$ ,  $v = \pi\sigma/\sqrt{6}$ . Applying Proposition 2 we find that the Fisher matrix in the new coordinates is

$$\tilde{G}(u, v) = \begin{pmatrix} \frac{\pi^2}{6v^2} & 0 \\ 0 & \frac{\pi^2}{6v^2} \end{pmatrix} = \frac{\pi^2}{6} \begin{pmatrix} \frac{1}{v^2} & 0 \\ 0 & \frac{1}{v^2} \end{pmatrix}$$

Then, applying Lemma 4, the Fisher–Rao distance is given by

$$\begin{aligned} d_{\text{FR}}((\mu_1, \sigma_1), (\mu_2, \sigma_2)) &= \frac{\pi}{\sqrt{6}} d_{\mathcal{H}^2} \left( \left( \mu_1 - (1 - \gamma)\sigma_1, \frac{\pi}{\sqrt{6}}\sigma_1 \right), \left( \mu_2 - (1 - \gamma)\sigma_2, \frac{\pi}{\sqrt{6}}\sigma_2 \right) \right) \\ &= \frac{2\pi}{\sqrt{6}} \operatorname{arctanh} \left( \sqrt{\frac{[(\mu_1 - \mu_2) - (1 - \gamma)(\sigma_1 - \sigma_2)]^2 + \frac{\pi^2}{6}(\sigma_1 - \sigma_2)^2}{[(\mu_1 - \mu_2) - (1 - \gamma)(\sigma_1 - \sigma_2)]^2 + \frac{\pi^2}{6}(\sigma_1 + \sigma_2)^2}} \right). \end{aligned} \tag{62}$$

### 5.7.2 Fréchet

A Fréchet distribution [18] has p.d.f.  $p(x) = \frac{\lambda}{\beta} \left(\frac{x}{\beta}\right)^{-\lambda-1} \exp\left(-\left(\frac{x}{\beta}\right)^{-\lambda}\right)$ , defined for  $x \in \mathbb{R}_+^*$ , and parametrised by  $(\beta, \lambda) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$ . Note that this corresponds to (60) taking  $\mu = \beta$ ,  $\sigma = \beta/\lambda$  and  $\xi = 1/\lambda$ . This distribution can be related to the Gumbel distribution by considering the reparametrisation of the sample space  $Y := \log X$ , which preserves the Fisher metric (Proposition 3). The p.d.f. of the new random variable is then

$$p(y) = \frac{1}{\left|\frac{dy}{dx}\right|} \frac{\lambda}{\beta} \left(\frac{e^y}{\beta}\right)^{-\lambda-1} \exp\left(-\left(\frac{e^y}{\beta}\right)^{-\lambda}\right) = \lambda \left(\frac{e^y}{\beta}\right)^{-\lambda} \exp\left(-\left(\frac{e^y}{\beta}\right)^{-\lambda}\right).$$

Now, considering the change of coordinates  $\alpha = \log \beta$ ,  $\theta = 1/\lambda$ , we find

$$p(y) = \frac{1}{\theta} \exp\left(-\frac{y - \alpha}{\theta}\right) \exp\left(-\left(-\frac{y - \alpha}{\theta}\right)\right),$$

which coincides with that of a Gumbel distribution with parameters  $(\alpha, \theta) \in \mathbb{R} \times \mathbb{R}_+^*$ . Comparing with the Fisher matrix (61) and applying Proposition 2 we find that the Fisher matrix in the  $(\beta, \lambda)$  coordinates is

$$G(\beta, \lambda) = \begin{pmatrix} \frac{\lambda^2}{\beta^2} & \frac{1-\gamma}{\beta} \\ \frac{1-\gamma}{\beta} & \frac{1}{\lambda^2} \left( (\gamma-1)^2 + \frac{\pi^2}{6} \right) \end{pmatrix}. \tag{63}$$



Note that, by considering the change of coordinates  $u = \log \beta - (1 - \gamma)/\lambda$ ,  $v = \pi/(\lambda\sqrt{6})$  and applying again Proposition 2, the Fisher matrix is found to be

$$\tilde{G}(u, v) = \begin{pmatrix} \frac{\pi^2}{6v^2} & 0 \\ 0 & \frac{\pi^2}{6v^2} \end{pmatrix} = \frac{\pi^2}{6} \begin{pmatrix} \frac{1}{v^2} & 0 \\ 0 & \frac{1}{v^2} \end{pmatrix},$$

Therefore, by Lemma 4 the Fisher–Rao distance for Fréchet distributions is

$$\begin{aligned} d_{\text{FR}}((\beta_1, \lambda_1), (\beta_2, \lambda_2)) &= \frac{\pi}{\sqrt{6}} d_{\mathcal{H}^2} \left( \left( \log \beta_1 - \frac{(1 - \gamma)}{\lambda_1}, \frac{\pi}{\lambda_1 \sqrt{6}} \right), \left( \log \beta_2 - \frac{(1 - \gamma)}{\lambda_2}, \frac{\pi}{\lambda_2 \sqrt{6}} \right) \right) \\ &= \frac{2\pi}{\sqrt{6}} \operatorname{arctanh} \left( \sqrt{\frac{\left[ \log \frac{\beta_1}{\beta_2} - (1 - \gamma) \left( \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) \right]^2 + \frac{\pi^2}{6} \left( \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right)^2}{\left[ \log \frac{\beta_1}{\beta_2} - (1 - \gamma) \left( \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) \right]^2 + \frac{\pi^2}{6} \left( \frac{1}{\lambda_1} + \frac{1}{\lambda_2} \right)^2}} \right)}. \end{aligned} \tag{64}$$

### 5.7.3 Weibull

A Weibull distribution [12, 18, 50] has p.d.f.  $p(x) = \frac{\lambda}{\beta} \left(\frac{x}{\beta}\right)^{\lambda-1} \exp\left(-\left(\frac{x}{\beta}\right)^\lambda\right)$ , defined for  $x \in \mathbb{R}_+^*$ , and parametrised by  $(\beta, \lambda) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$ . Note that this corresponds to the distribution of  $-X$  in (60) taking  $\mu = -\beta, \sigma = \beta/\lambda$  and  $\xi = -1/\lambda$ . It is possible to relate a Weibull distribution to Gumbel by considering the reparametrisation of the sample space  $Y := -\log X$ , which preserves the Fisher metric, (Proposition 3). The p.d.f. of the new random variable is

$$p(y) = \frac{1}{\left|\frac{dy}{dx}\right|} \frac{\lambda}{\beta} \left(\frac{e^{-y}}{\beta}\right)^{\lambda-1} \exp\left(-\left(\frac{e^{-y}}{\beta}\right)^\lambda\right) = \lambda \left(\frac{e^{-y}}{\beta}\right)^\lambda \exp\left(-\left(\frac{e^{-y}}{\beta}\right)^\lambda\right).$$

Moreover, with the change of coordinates  $\lambda = 1/\theta, \alpha = -\log \beta$ , we have

$$p(y) = \frac{1}{\theta} \exp\left(-\frac{y - \alpha}{\theta}\right) \exp\left(-\exp\left(-\frac{y - \alpha}{\theta}\right)\right),$$

which coincides with a Gumbel distribution with parameters  $(\alpha, \theta) \in \mathbb{R} \times \mathbb{R}_+^*$ . Again, comparing with the Fisher matrix (61) and applying Proposition 2 we find that the Fisher matrix in the  $(\beta, \lambda)$  coordinates is

$$G(\beta, \lambda) = \begin{pmatrix} \frac{\lambda^2}{\beta^2} & \frac{\gamma-1}{\beta} \\ \frac{\gamma-1}{\beta} & \frac{1}{\lambda^2} \left( (\gamma - 1)^2 + \frac{\pi^2}{6} \right) \end{pmatrix}. \tag{65}$$

The change of coordinates  $u = -\log \beta - (1 - \gamma)/\lambda$ ,  $v = \pi/(\lambda\sqrt{6})$  with Proposition 2 yields the following form for the Fisher matrix:

$$\tilde{G}(u, v) = \begin{pmatrix} \frac{\pi^2}{6v^2} & 0 \\ 0 & \frac{\pi^2}{6v^2} \end{pmatrix} = \frac{\pi^2}{6} \begin{pmatrix} \frac{1}{v^2} & 0 \\ 0 & \frac{1}{v^2} \end{pmatrix}.$$

In addition, by Lemma 4, the Fisher–Rao distance for Weibull distributions is

$$\begin{aligned} d_{FR}((\beta_1, \lambda_1), (\beta_2, \lambda_2)) &= \frac{\pi}{\sqrt{6}} d_{\mathcal{H}^2} \left( \left( -\log \beta_1 - \frac{(1 - \gamma)}{\lambda_1}, \frac{\pi}{\lambda_1\sqrt{6}} \right), \left( -\log \beta_2 - \frac{(1 - \gamma)}{\lambda_2}, \frac{\pi}{\lambda_2\sqrt{6}} \right) \right) \\ &= \frac{2\pi}{\sqrt{6}} \operatorname{arctanh} \left( \sqrt{\frac{\left[ \log \frac{\beta_2}{\beta_1} - (1 - \gamma) \left( \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) \right]^2 + \frac{\pi^2}{6} \left( \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right)^2}{\left[ \log \frac{\beta_2}{\beta_1} - (1 - \gamma) \left( \frac{1}{\lambda_1} - \frac{1}{\lambda_2} \right) \right]^2 + \frac{\pi^2}{6} \left( \frac{1}{\lambda_1} + \frac{1}{\lambda_2} \right)^2}} \right). \end{aligned} \tag{66}$$

The special case of fixed  $\lambda$  has been addressed in [12].

**Remark 6** If  $X$  is a random variable following a Weibull distribution, then  $-X$  follows a reversed Weibull distribution, which corresponds to the Weibull-type distribution from (60), and has the same geometry, and same Fisher–Rao distance as the Weibull distribution [18].

### 5.8 Pareto

A Pareto distribution [12, 14] has p.d.f.  $p(x) = \theta\alpha^\theta x^{-(\theta+1)}$ , defined for  $x \in [\alpha, \infty[$  and parametrised by  $(\theta, \alpha) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$ . In this case, the support depends on the parametrisation, thus violating one of the assumptions made in the definition of a statistical manifold (1). Nevertheless, it is still possible<sup>4</sup> to compute a Riemannian metric from the Fisher information matrix, as in (2). We thus have  $\partial_\theta \ell := \partial_\theta \ell(\theta, \alpha) = \frac{1}{\theta} + \log \alpha - \log x$  and  $\partial_\alpha \ell := \partial_\alpha \ell(\theta, \alpha) = \frac{\theta}{\alpha}$ . The elements of the Fisher matrix are

$$\begin{aligned} g_{11} &= \mathbb{E} \left[ (\partial_\theta \ell)^2 \right] = \mathbb{E} \left[ \left( \frac{1}{\theta} + \log \alpha - \log X \right)^2 \right] \\ &= \frac{1}{\theta^2} + (\log \alpha)^2 + \mathbb{E} \left[ (\log X)^2 \right] + \frac{2}{\theta} (\log \alpha - \mathbb{E}[\log X]) - 2 \log \alpha \mathbb{E}[\log X] \\ &= \frac{1}{\theta^2}, \\ g_{12} &= g_{21} = \mathbb{E} [(\partial_\theta \ell) (\partial_\alpha \ell)] = \mathbb{E} \left[ \left( \frac{1}{\theta} + \log \alpha - \log X \right) \left( \frac{\theta}{\alpha} \right) \right] \\ &= \frac{1}{\alpha} + \frac{\theta}{\alpha} \log \alpha - \frac{\theta}{\alpha} \mathbb{E}[\log X] = 0, \end{aligned}$$

<sup>4</sup> As noted in [14], what is not possible is to use the alternative expression (3), which would result in a ‘fake metric’.

$$g_{22} = \mathbb{E} \left[ (\partial_\alpha \ell)^2 \right] = \mathbb{E} \left[ \left( \frac{\theta}{\alpha} \right)^2 \right] = \frac{\theta^2}{\alpha^2},$$

where we have used that  $\mathbb{E}[\log X] = \frac{1}{\theta} + \log \alpha$  and  $\mathbb{E}[(\log X)^2] = \frac{2}{\theta^2} + \frac{2 \log \alpha}{\theta} + (\log \alpha)^2$ . Thus, the Fisher matrix is

$$G(\theta, \alpha) = \begin{pmatrix} \frac{1}{\theta^2} & 0 \\ 0 & \frac{\theta^2}{\alpha^2} \end{pmatrix}. \tag{67}$$

To find the Fisher–Rao distance, we consider the change of coordinates  $u = \log \alpha$ ,  $v = 1/\theta$ . Applying Proposition 2 we find that the Fisher matrix in the new coordinates

$$\tilde{G}(u, v) = \begin{pmatrix} \frac{1}{v^2} & 0 \\ 0 & \frac{1}{v^2} \end{pmatrix},$$

which coincides with the hyperbolic metric (11) restricted to the positive quadrant. Therefore the Fisher–Rao distance is given by

$$d_{FR}((\theta_1, \alpha_1), (\theta_2, \alpha_2)) = d_{\mathcal{H}^2}((\log \alpha_1, 1/\theta_1), (\log \alpha_2, 1/\theta_2)) = 2 \operatorname{arctanh} \left( \sqrt{\frac{(\theta_1 \theta_2 \log(\alpha_1/\alpha_2))^2 + (\theta_1 - \theta_2)^2}{(\theta_1 \theta_2 \log(\alpha_1/\alpha_2))^2 + (\theta_1 + \theta_2)^2}} \right). \tag{68}$$

The special case of fixed  $\alpha$  has been addressed in [12].

### 5.9 Power function

A power function distribution [12] has p.d.f.  $p(x) = \theta \beta^{-\theta} x^{\theta-1}$ , defined for  $x \in ]0, \beta]$ , and parametrised by  $(\theta, \beta) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$ . As in the previous example, the support depends on the parametrisation, but it is still possible to consider the Fisher metric as in (2). This distribution can be related to the Pareto distribution (Sect. 5.8) as follows. Consider the reparametrisation of the sample space given by  $Y := 1/X$  (cf. Proposition 3), and the change of coordinates  $\alpha = 1/\beta$ . Note that, since  $x \in ]0, \beta]$ , we have  $y \in [\alpha, \infty[$ . The p.d.f. of the new random variable, with the new coordinates, is

$$p(y) = \frac{1}{\left| \frac{dy}{dx} \right|} \theta \beta^{-\theta} y^{-(\theta-1)} = \theta \beta^{-\theta} y^{-(\theta+1)} = \theta \alpha^\theta y^{-(\theta+1)}.$$

which coincides with a Pareto distribution with parameters  $(\theta, \alpha)$ . Therefore, applying Proposition 2, we find

$$G(\theta, \beta) = \begin{pmatrix} \frac{1}{\theta^2} & 0 \\ 0 & \frac{\theta^2}{\beta^2} \end{pmatrix}, \tag{69}$$

and therefore

$$\begin{aligned} d_{FR}((\theta_1, \beta_1), (\theta_2, \beta_2)) &= d_{\gamma L^2}((\log \beta_1, 1/\theta_1), (\log \beta_2, 1/\theta_2)) \\ &= 2 \operatorname{arctanh} \left( \sqrt{\frac{(\theta_1 \theta_2 \log(\beta_1/\beta_2))^2 + (\theta_1 - \theta_2)^2}{(\theta_1 \theta_2 \log(\beta_1/\beta_2))^2 + (\theta_1 + \theta_2)^2}} \right). \end{aligned} \tag{70}$$

The special case of fixed  $\alpha$  has been addressed in [12].

**Remark 7** All the examples of two-dimensional statistical manifolds of continuous distributions presented so far in this section are related to the hyperbolic Poincaré half-plane, and have constant negative curvature. However, there are examples of two-dimensional statistical manifolds which are not of constant negative curvature, even if we do not have an explicit expression for the Fisher–Rao distances. We present some examples in the following.

- The statistical manifold of Gamma distributions  $p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ , defined for  $x \in \mathbb{R}_+^*$  and parametrised by  $(\alpha, \beta) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$ . The curvature of this manifold is [38, 39, 64]

$$\kappa = \frac{\psi^{(1)}(\alpha) + \alpha \psi^{(2)}(\alpha)}{4(\alpha \psi^{(1)}(\alpha) - 1)^2} < 0,$$

which is negative, but not constant, and where  $\psi^{(m)}(x) := \frac{d^{m+1}}{dx^{m+1}} \log \Gamma(x)$  denotes the polygamma function of order  $m$ . Bounds for the Fisher–Rao distance in this manifold have been studied in [64].

- The statistical manifold of Beta distributions  $p(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$ , defined for  $x \in [0, 1]$  and parametrised by  $(\alpha, \beta) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$ . The curvature of this manifold is [26]

$$\begin{aligned} \kappa &= \frac{\psi^{(2)}(\alpha)\psi^{(2)}(\beta)\psi^{(2)}(\alpha + \beta)}{4(\psi^{(1)}(\alpha)\psi^{(1)}(\beta) - \psi^{(1)}(\alpha + \beta)[\psi^{(1)}(\alpha) + \psi^{(1)}(\beta)])^2} \\ &\quad \times \left( \frac{\psi^{(1)}(\alpha)}{\psi^{(2)}(\alpha)} + \frac{\psi^{(1)}(\beta)}{\psi^{(2)}(\beta)} - \frac{\psi^{(1)}(\alpha + \beta)}{\psi^{(2)}(\alpha + \beta)} \right) \\ &< 0, \end{aligned}$$

which is negative, but not constant too. In fact, more generally, the sectional curvature of the statistical manifold of Dirichlet distributions (which are the multivariate generalisation of Beta distributions) is negative [65, Thm. 6].

- Finally, a construction of  $n$ -dimensional statistical manifolds, based on a Hilbert space representation of probability measures, was given in [16]. The geometry of these manifolds is spherical, that is, they have constant positive curvature.

### 5.10 Wishart

A Wishart distribution [11, 66] in dimension  $m$ , with  $n \geq m$  degrees of freedom,  $n \in \mathbb{N}$ , has p.d.f.

$$p(X) = \frac{(\det X)^{(n-m-1)/2} \exp\left(-\frac{1}{2} \operatorname{tr}(\Sigma^{-1} X)\right)}{2^{nm/2} (\det \Sigma)^{n/2} \Gamma_m(n/2)},$$

defined for  $X \in P_m(\mathbb{R})$ , and characterised by  $\Sigma \in P_m(\mathbb{R})$ , where  $\Gamma_m(z) := \pi^{\frac{m(m-1)}{4}} \prod_{j=1}^m \Gamma\left(z + \frac{1-j}{2}\right)$  denotes the multivariate Gamma function. Note that, for fixed  $m, n$ , the associated  $\left(\frac{m(m+1)}{2}\right)$ -dimensional statistical manifold  $\mathcal{S}$  is in correspondence with the cone  $P_m(\mathbb{R})$  of symmetric positive-definite matrices via the bijection  $\iota: P_m(\mathbb{R}) \rightarrow \mathcal{S}$ , given by  $\Sigma \mapsto \frac{(\det X)^{(n-m-1)/2} \exp\left(-\frac{1}{2} \operatorname{tr}(\Sigma^{-1} X)\right)}{2^{nm/2} (\det \Sigma)^{n/2} \Gamma_m(n/2)}$ . Denoting  $\sigma_{i,j}$  the  $(i, j)$ -th entry of the matrix  $\Sigma$ , we can write the parameter vector as  $\xi = (\xi^1, \dots, \xi^{m(m+1)/2}) = (\sigma_{1,1}, \dots, \sigma_{1,m}, \sigma_{2,2}, \dots, \sigma_{2,m}, \dots, \sigma_{m,m})$ . We then have  $\partial_i \ell := \partial_i \ell(\Sigma) = \frac{1}{2} \operatorname{tr}(\Sigma^{-1} X \Sigma^{-1} (\partial_i \Sigma)) - \frac{n}{2} \operatorname{tr}(\Sigma^{-1} (\partial_i \Sigma))$  and  $\partial_j \partial_i \ell = -\operatorname{tr}(\Sigma^{-1} (\partial_i \Sigma) \Sigma^{-1} (\partial_j \Sigma) \Sigma^{-1} X) + \frac{n}{2} \operatorname{tr}(\Sigma^{-1} (\partial_i \Sigma) \Sigma^{-1} (\partial_j \Sigma))$ , where the derivative in  $\partial_i \Sigma$  is taken entry-wise. The elements of the Fisher metric are then

$$\begin{aligned} g_{ij}(\xi) &= -\mathbb{E}[\partial_j \partial_i \ell] \\ &= -\mathbb{E}\left[-\operatorname{tr}\left(\Sigma^{-1} (\partial_i \Sigma) \Sigma^{-1} (\partial_j \Sigma) \Sigma^{-1} X\right) + \frac{n}{2} \operatorname{tr}\left(\Sigma^{-1} (\partial_i \Sigma) \Sigma^{-1} (\partial_j \Sigma)\right)\right] \\ &= \operatorname{tr}\left(\Sigma^{-1} (\partial_i \Sigma) \Sigma^{-1} (\partial_j \Sigma) \Sigma^{-1} \mathbb{E}[X]\right) - \frac{n}{2} \operatorname{tr}\left(\Sigma^{-1} (\partial_i \Sigma) \Sigma^{-1} (\partial_j \Sigma)\right) \\ &= \frac{n}{2} \operatorname{tr}\left(\Sigma^{-1} (\partial_i \Sigma) \Sigma^{-1} (\partial_j \Sigma)\right), \end{aligned} \tag{71}$$

where we have used that  $\mathbb{E}[X] = n\Sigma$ .

In view of the bijection  $\iota$ , the tangent space  $T_{p_\xi} \mathcal{S}$  can be identified with  $H_m$ , the set of  $m \times m$  symmetric matrices [67, Chapter 6]. Given two matrices  $U, \tilde{U} \in H_m$ , parametrized as  $\theta = (\theta^1, \dots, \theta^{m(m+1)/2}) = (u_{1,1}, \dots, u_{1,m}, u_{2,2}, \dots, u_{2,m}, \dots, u_{m,m})$  and  $\tilde{\theta} = (\tilde{\theta}^1, \dots, \tilde{\theta}^{m(m+1)/2}) = (\tilde{u}_{1,1}, \dots, \tilde{u}_{1,m}, \tilde{u}_{2,2}, \dots, \tilde{u}_{2,m}, \dots, \tilde{u}_{m,m})$ , we shall compute the inner product defined by the Fisher metric, cf. (4). In the following,  $\otimes$  denotes the Kronecker product, and, for an  $m \times m$  matrix  $A$ , whose columns are  $A_1, A_2, \dots, A_m$ , we denote  $\operatorname{vec}(A) := (A_1^\top A_2^\top \dots A_m^\top)^\top$  the  $m^2$ -dimensional vector formed by the concatenation of its columns. Moreover, if  $A$  is symmetric, denote  $\nu(A) := (a_{1,1}, \dots, a_{1,m}, a_{2,2}, \dots, a_{2,m}, \dots, a_{m,m})$ . We denote  $D_m$  the unique  $m^2 \times \frac{m(m+1)}{2}$

matrix that verifies  $D_m v(A) = \text{vec}(A)$ , for any symmetric  $A$  [68, § 7]. We thus have

$$\begin{aligned} \langle U, \tilde{U} \rangle_{G(\xi)} &= \theta^\top G(\xi) \tilde{\theta} \\ &= \sum_{i=1}^{\frac{m(m+1)}{2}} \sum_{j=1}^{\frac{m(m+1)}{2}} g_{ij}(\xi) \theta^i \tilde{\theta}^j \\ &= \sum_{i=1}^{\frac{m(m+1)}{2}} \sum_{j=1}^{\frac{m(m+1)}{2}} \frac{n}{2} \text{tr} \left( \Sigma^{-1} (\partial_i \Sigma) \Sigma^{-1} (\partial_j \Sigma) \right) \theta^i \tilde{\theta}^j \\ &= \frac{n}{2} \text{tr} \left( \Sigma^{-1} U \Sigma^{-1} \tilde{U} \right) \\ &= \frac{n}{2} \text{vec}(U) (\Sigma^{-1} \otimes \Sigma^{-1}) \text{vec}(\tilde{U}) \\ &= \frac{n}{2} \theta^\top D_m^\top (\Sigma^{-1} \otimes \Sigma^{-1}) D_m \tilde{\theta}, \end{aligned}$$

where we have used that  $\text{tr}(ABCD) = \text{vec}(D)^\top (A \otimes C^\top) \text{vec}(B^\top)$ , for  $A, B, C$  and  $D$  matrices such that the product  $ABCD$  is defined and square [68, Lemma 3]. We can thus conclude that the Fisher matrix is

$$G(\xi) = \frac{n}{2} D_m^\top (\Sigma^{-1} \otimes \Sigma^{-1}) D_m. \tag{72}$$

This metric turns out to coincide with the Fisher metric of the statistical manifold formed by multivariate Gaussian distributions with fixed mean [12, 20, 37, 60], up to the factor  $n$ . Therefore, the Fisher–Rao distance is proportional to the one of that manifold, that is,

$$d_{\text{FR}}(\Sigma_1, \Sigma_2) = \sqrt{\frac{n}{2}} \left\| \log \left( \Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2} \right) \right\|_F = \sqrt{\frac{n}{2} \sum_{k=1}^m (\log \lambda_k)^2}, \tag{73}$$

where  $\lambda_k$  are the eigenvalues of  $\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2}$ ,  $\log$  denotes the matrix logarithm, and  $\|A\|_F = \sqrt{\text{tr}(AA^\top)}$  is the Frobenius norm. This metric also coincides with the standard metric of  $P_m(\mathbb{R})$ , when endowed with the matrix inner product  $\langle A, B \rangle = \text{tr}(A^\top B)$  [67, Chapter 6], up to the factor  $\frac{n}{2}$ , and is in fact related to the metric of the Siegel upper space [69] (see also [37, Appendix D]). Note that when  $\Sigma$  is restricted to be diagonal this distance is, up to a factor  $\sqrt{n}$ , the product distance between univariate Gaussian distributions with fixed mean—see Example 2 in Sect. 6.

### 5.11 Inverse Wishart

An inverse Wishart distribution [70] in dimension  $m$ , with  $n \geq m$  degrees of freedom,  $n \in \mathbb{N}$ , has p.d.f.

$$p(X) = \frac{(\det \Sigma)^{n/2} (\det X)^{-(n+m+1)/2} \exp\left(-\frac{1}{2} \operatorname{tr}(\Sigma X^{-1})\right)}{2^{nm/2} \Gamma_m(n/2)},$$

defined for  $X \in P_m(\mathbb{R})$ , and characterised by  $\Sigma \in P_m(\mathbb{R})$ . We can relate an inverse Wishart distribution to a Wishart distribution by considering the reparametrisation of the sample space given by  $Y := X^{-1}$  (cf. Proposition 3), and the change of coordinates  $\Phi := \Sigma^{-1}$ . The p.d.f. of the new random variable, in the new coordinates, is

$$\begin{aligned} p(Y) &= \frac{1}{\left|\frac{dY}{dX}\right|} \frac{(\det(\Phi^{-1}))^{n/2} (\det(Y^{-1}))^{-(n+m+1)/2} \exp\left(-\frac{1}{2} \operatorname{tr}(\Phi^{-1} Y)\right)}{2^{nm/2} \Gamma_m(n/2)}, \\ &= \frac{1}{(\det(Y^{-1}))^{-(m+1)}} \frac{(\det(Y^{-1}))^{-(n+m+1)/2} \exp\left(-\frac{1}{2} \operatorname{tr}(\Phi^{-1} Y)\right)}{(\det \Phi)^{n/2} 2^{nm/2} \Gamma_m(n/2)} \\ &= \frac{(\det Y)^{(n-m-1)/2} \exp\left(-\frac{1}{2} \operatorname{tr}(\Phi^{-1} Y)\right)}{(\det \Phi)^{n/2} 2^{nm/2} \Gamma_m(n/2)}, \end{aligned}$$

where we have used that  $\left|\frac{dY}{dX}\right| = \det(X)^{-(m+1)}$  [68, § 12], which coincides with the p.d.f. of a Wishart distribution with parameter  $\Phi$ . Write  $\sigma_{i,j}$  and  $\phi_{i,j}$  the  $(i, j)$ -th entries of matrices  $\Sigma$  and  $\Phi$ , respectively. Write  $\xi = (\xi^1, \dots, \xi^{m(m+1)/2}) = (\sigma_{1,1}, \dots, \sigma_{1,m}, \sigma_{2,2}, \dots, \sigma_{2,m}, \dots, \sigma_{m,m})$ , and  $\theta = (\theta^1, \dots, \theta^{m(m+1)/2}) = (\phi_{1,1}, \dots, \phi_{1,m}, \phi_{2,2}, \dots, \phi_{2,m}, \dots, \phi_{m,m})$ . Denote  $G_W(\theta)$  the Fisher matrix of a Wishart distribution (72) in coordinates  $\theta$ . In the following,  $D_m$  is the matrix defined in the previous section, and  $A^+$  denotes the Moore–Penrose inverse of a matrix  $A$ . Applying Proposition 2 yields

$$\begin{aligned} G(\xi) &= \left[\frac{d\Sigma}{d\Phi}(\Phi)\right]^{-\top} G_W(\theta) \left[\frac{d\Sigma}{d\Phi}(\Phi)\right]^{-1} \\ &= \frac{n}{2} \left(-D_m^+(\Phi^{-1} \otimes \Phi^{-1}) D_m\right)^{-\top} D_m^\top (\Phi^{-1} \otimes \Phi^{-1}) \\ &\quad D_m \left(-D_m^+(\Phi^{-1} \otimes \Phi^{-1}) D_m\right)^{-1} \\ &= \frac{n}{2} (D_m^+(\Sigma \otimes \Sigma) D_m)^{-\top} D_m^\top (\Sigma \otimes \Sigma) D_m (D_m^+(\Sigma \otimes \Sigma) D_m)^{-1} \\ &= \frac{n}{2} \left(D_m^+(\Sigma^{-1} \otimes \Sigma^{-1}) D_m\right)^\top D_m^\top (\Sigma \otimes \Sigma) D_m D_m^+(\Sigma^{-1} \otimes \Sigma^{-1}) D_m \\ &= \frac{n}{2} D_m^\top (\Sigma^{-1} \otimes \Sigma^{-1}) (D_m^+)^\top D_m^\top (\Sigma \otimes \Sigma) D_m D_m^+(\Sigma^{-1} \otimes \Sigma^{-1}) D_m \\ &= \frac{n}{2} D_m^\top (\Sigma^{-1} \otimes \Sigma^{-1}) D_m, \end{aligned}$$

where we have used that  $\frac{d\Sigma}{d\Phi}(\Phi) = -D_m^+(\Phi^{-1} \otimes \Phi^{-1}) D_m$  [68, § 12],  $(D_m^+(\Sigma \otimes \Sigma) D_m)^{-1} = D_m^+(\Sigma^{-1} \otimes \Sigma^{-1}) D_m$ , and  $D_m D_m^+(\Sigma^{-1} \otimes \Sigma^{-1}) D_m = (\Sigma^{-1} \otimes \Sigma^{-1}) D_m$  [68, Lemma 11]. Therefore, the Fisher matrix is the same as for Wishart distributions in (72), and the Fisher–Rao distance is

$$d_{FR}(\Sigma_1, \Sigma_2) = \sqrt{\frac{n}{2}} \left\| \log \left( \Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2} \right) \right\|_F = \sqrt{\frac{n}{2} \sum_{k=1}^m (\log \lambda_k)^2}, \tag{74}$$

where  $\lambda_k$  are the eigenvalues of  $\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2}$ .

**Remark 8** Other matrix distributions have been recently studied in the literature. In [11], general Wishart elliptical distributions have been addressed, which also include  $t$ -Wishart and Kotz–Wishart, by noting that their metric coincides with that of zero-mean multivariate elliptical distributions [54].

### 6 Product distributions

In this short section, we address the Fisher–Rao distance for multivariate product distributions, i.e., distributions of random vectors whose components are independent. In this case, the distribution of the random vector is the product of the distributions of each component, and the associated statistical manifold is the product of the statistical manifold associated to each component.

Consider  $m$  Riemannian manifolds  $\{(M_i, g_i) : 1 \leq i \leq m\}$ , and the product manifold  $(M, g)$ , with  $M := M_1 \times \cdots \times M_m$ , and  $g := g_1 \oplus \cdots \oplus g_m$ . In matrix form, the product metric is given by the block-diagonal matrix

$$G = \begin{pmatrix} G_1 & 0 & \cdots & 0 \\ 0 & G_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & G_m \end{pmatrix},$$

where  $G_i$  is the matrix form of the metric  $g_i$ , for  $1 \leq i \leq m$ . Let  $d_i$  denote the geodesic distance in  $M_i$ . Then, the geodesic distance  $d$  in  $(M, g)$  is given by a Pythagorean formula [71, Prop. 1], [72]:

$$d((x_1, \dots, x_m), (y_1, \dots, y_m)) = \sqrt{\sum_{i=1}^m [d_i(x_i, y_i)]^2}, \tag{75}$$

where  $(x_1, \dots, x_m) \in M_1 \times \cdots \times M_m$ , and  $(y_1, \dots, y_m) \in M_1 \times \cdots \times M_m$ . This result can be used to write explicit forms for the Fisher–Rao distance in statistical manifolds of product distributions, as they are statistical product manifolds. This type of construction has been used, e.g., in [18, 23]. Some examples are given below.



**Example 1** The Fisher–Rao distance between the distributions of  $n$ -dimensional vectors formed independent Poisson distributions [12] (cf. Section 4.2) with parameters  $(\lambda_1, \dots, \lambda_n)$  and  $(\lambda'_1, \dots, \lambda'_n)$  is

$$d_{FR}((\lambda_1, \dots, \lambda_n), (\lambda'_1, \dots, \lambda'_n)) = 2 \sqrt{\sum_{i=1}^n (\sqrt{\lambda_i} - \sqrt{\lambda'_i})^2}. \tag{76}$$

**Example 2** Consider multivariate independent Gaussian distributions (cf. Sect. 5.4.1). In this case, the covariance matrix is diagonal, a case that has been addressed in [12, 20, 45]. The Fisher–Rao distance between such distributions parametrised by  $(\mu_1, \sigma_1, \dots, \mu_n, \sigma_n)$  and  $(\mu'_1, \sigma'_1, \dots, \mu'_n, \sigma'_n)$  is

$$\begin{aligned} & d_{FR}((\mu_1, \sigma_1, \dots, \mu_n, \sigma_n), (\mu'_1, \sigma'_1, \dots, \mu'_n, \sigma'_n)) \\ &= 2\sqrt{2} \sqrt{\sum_{i=1}^n \left[ \operatorname{arctanh} \left( \sqrt{\frac{(\mu_i - \mu'_i)^2 + 2(\sigma_i - \sigma'_i)^2}{(\mu_i - \mu'_i)^2 + 2(\sigma_i + \sigma'_i)^2}} \right) \right]^2}. \end{aligned} \tag{77}$$

**Example 3** More generally, consider a vector  $(X_1, \dots, X_n)$  of  $n$ -independent generalised Gaussian distributions, where  $X_i$  follows a generalised Gaussian with fixed  $\beta_i$ , parametrised by  $(\mu_i, \sigma_i)$ . For fixed values  $(\beta_1, \dots, \beta_n)$ , the distance between the distribution of two such vectors, parametrised by  $(\mu_1, \sigma_1, \dots, \mu_n, \sigma_n)$  and  $(\mu'_1, \sigma'_1, \dots, \mu'_n, \sigma'_n)$ , is

$$\begin{aligned} & d_{FR}((\mu_1, \sigma_1, \dots, \mu_n, \sigma_n), (\mu'_1, \sigma'_1, \dots, \mu'_n, \sigma'_n)) \\ &= \sqrt{\sum_{i=1}^n (\beta_i + 1) \left[ \operatorname{arctanh} \left( \sqrt{\frac{\beta_i (\mu_i - \mu'_i)^2 \Gamma(2 - \frac{1}{\beta_i}) + (\beta_i + 1)(\sigma_i - \sigma'_i)^2 \Gamma(1 + \frac{1}{\beta_i})}{\beta_i (\mu_i - \mu'_i)^2 \Gamma(2 - \frac{1}{\beta_i}) + (\beta_i + 1)(\sigma'_i + \sigma_i)^2 \Gamma(1 + \frac{1}{\beta_i})}} \right) \right]^2}. \end{aligned} \tag{78}$$

## 7 Final remarks

In this survey, we have collected closed-form expressions for the Fisher–Rao distance in different statistical manifolds of both discrete and continuous distributions. In curating this collection in a unified language, we also provided some punctual contributions. The results are summarised in Tables 1 and 2. We hope that providing these expressions readily available can be helpful not only to those interested in information geometry itself, but also to a broader audience, interested in using these distances in different applications.

**Acknowledgements** Part of this work was developed while the first author was with IMECC, Unicamp, Brazil. Discussions with Luiz Lara and Franciele Silva are gratefully acknowledged. We thank Gabriel Khan for bringing to our attention the example of inverse Gaussian distributions.

**Author Contributions** All authors contributed to this work. All authors wrote the first draft, and H.K.M., J.P. and S.I.R.C. reviewed the final manuscript.

**Funding** H.K.M. was supported by São Paulo Research Foundation (FAPESP) grant 2021/04516-8. F.C.C.M. was supported by Brazilian National Council for Scientific and Technological Development (CNPq) grant 141407/2020-4. S.I.R.C. was supported by FAPESP grant 2020/09838-0 (BIOS – Brazilian Institute of Data Science) and CNPq grants 314441/2021-2 and 308399/2023-4.

**Data availability** No datasets were generated or analysed during the current study.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

## References

1. Amari, S., Nagaoka, H.: *Methods of Information Geometry*. American Mathematical Society, Providence (2000)
2. Calin, O., Udriște, C.: *Geometric Modeling in Probability and Statistics*. Springer, Cham (2014)
3. Nielsen, F.: An elementary introduction to information geometry. *Entropy* **22**(10), 1100 (2020)
4. Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: Information geometry and sufficient statistics. *Probab. Theory Relat. Fields* **162**, 327–364 (2015)
5. Lê, H.V.: The uniqueness of the Fisher metric as information metric. *Ann. Inst. Stat. Math.* **69**, 879–896 (2017)
6. Hotelling, H.: Spaces of statistical parameters. *Bull. Am. Math. Soc.* **36**, 191 (1930)
7. Stigler, S.M.: The epic story of maximum likelihood. *Stat. Sci.* **22**(4), 598–620 (2007)
8. Rao, C.R.: Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **37**, 81–91 (1945)
9. Rao, C.R.: Information and the accuracy attainable in the estimation of statistical parameters. In: Kotz, S., Johnson, N.L. (eds.) *Breakthroughs in Statistics: Foundations and Basic Theory*, pp. 235–247. Springer, New York (1992)
10. Atkinson, C., Mitchell, A.F.S.: Rao’s distance measure. *Sankhyā Indian J. Stat. Ser. A* **43**(3), 345–365 (1981)
11. Ayadi, I., Bouchard, F., Pascal, F.: Elliptical Wishart distribution: maximum likelihood estimator from information geometry. In: *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5 (2023)
12. Burbea, J.: Informative geometry of probability spaces. *Expo. Math.* **4**, 347–378 (1986)
13. Calvo, M., Oller, J.M.: An explicit solution of information geodesic equations for the multivariate model. *Stat. Decis.* **9**(1–2), 119–138 (1991)
14. Li, M., Sun, H., Peng, L.: Fisher-Rao geometry and Jeffreys prior for Pareto distribution. *Commun. Stat. Theory Methods* **51**(6), 1895–1910 (2022)
15. Micchelli, C.A., Noakes, L.: Rao distances. *J. Multivar. Anal.* **92**(1), 97–115 (2005)
16. Minarro, A., Oller, J.M.: On a class of probability density functions and their information metric. *Sankhyā Indian J. Stat. Ser. A* **55**(2), 214–225 (1993)
17. Mitchell, A.F.S.: Statistical manifolds of univariate elliptic distributions. *Int. Stat. Rev.* **56**(1), 1–16 (1988)
18. Oller, J.M.: Information metric for extreme value and logistic probability distributions. *Sankhyā Indian J. Stat. Ser. A* **49**(1), 17–23 (1987)
19. Oller, J.M., Cuadras, C.M.: Rao’s distance for negative multinomial distributions. *Sankhyā Indian J. Stat. Ser. A* **47**(1), 75–83 (1985)
20. Pinele, J., Strapasson, J.E., Costa, S.I.R.: The Fisher–Rao distance between multivariate normal distributions: special cases, bounds and applications. *Entropy* **22**(4), 404 (2020)
21. Rao, C.R.: Differential metrics in probability spaces, chap. 1. In: Amari, S., Barndorff-Nielsen, O.E., Kass, R.E., Lauritzen, S.L., Rao, C.R. (eds.) *Differential Geometry in Statistical Inference*, vol. 10. Institute of Mathematical Statistics, Hayward (1987)
22. Verdoolaege, G., Scheunders, P.: On the geometry of multivariate generalized Gaussian models. *J. Math. Imaging Vis.* **43**, 180–193 (2012)

23. Villarroya, A., Oller, J.M.: Statistical tests for the inverse gaussian distribution based on Rao distance. *Sankhyā Indian J. Stat. Ser. A* **55**(1), 80–103 (1993)
24. Gattone, S.A., De Sanctis, A., Russo, T., Pulcini, D.: A shape distance based on the Fisher–Rao metric and its application for shapes clustering. *Physica A Stat. Mech. Appl.* **487**, 93–102 (2017)
25. Taylor, S.: Clustering financial return distributions using the Fisher information metric. *Entropy* **21**(2), 110 (2019)
26. Le Brigant, A., Guigui, N., Rebbah, S., Puechmorel, S.: Classifying histograms of medical data using information geometry of beta distributions. In: 24th International Symposium on Mathematical Theory of Networks and Systems MTNS 2020. IFAC-PapersOnLine, vol. 54, pp. 514–520 (2021)
27. Rebbah, S., Nicol, F., Puechmorel, S.: The geometry of the generalized Gamma manifold and an application to medical imaging. *Mathematics* **7**(8), 674 (2019)
28. Arvanitidis, G., González-Duque, M., Pouplin, A., Kalatzis, D., Hauberg, S.: Pulling back information geometry. In: Camps-Valls, G., Ruiz, F.J.R., Valera, I. (eds.) The 25th International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 151, pp. 4872–4894 (2022)
29. Picot, M., Messina, F., Boudiaf, M., Labeau, F., Ayed, I.B., Piantanida, P.: Adversarial robustness via Fisher–Rao regularization. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(3), 2698–2710 (2023)
30. Shi-Garrier, L., Bouaynaya, N.C., Delahaye, D.: Adversarial robustness with partial isometry. *Entropy* **26**(2), 103 (2024)
31. Gomes, E.D.C., Alberge, F., Duhamel, P., Piantanida, P.: Igeood: An information geometry approach to out-of-distribution detection. In: International Conference on Learning Representations (2022). [https://openreview.net/forum?id=mfwdY3U\\_9ea](https://openreview.net/forum?id=mfwdY3U_9ea)
32. Miyamoto, H.K., Meneghetti, F.C.C., Costa, S.I.R.: The Fisher–Rao loss for learning under label noise. *Inf. Geom.* **6**, 107–126 (2023)
33. Bouchard, F., Breloy, A., Collas, A., Renaux, A., Ginolhac, G.: The Fisher–Rao geometry of CES distributions. [arXiv:2310.01032](https://arxiv.org/abs/2310.01032) (2023)
34. Arwini, K.A., Dodson, C.T.J.: *Information Geometry: Near Randomness and Near Independence*. Springer, Heidelberg (2008)
35. Han, M., Park, F.C.: DTI segmentation and fiber tracking using metrics on multivariate normal distributions. *J. Math. Imaging Vis.* **49**, 317–334 (2014)
36. Le Brigant, A., Deschamps, J., Collas, A., Miolane, N.: Parametric information geometry with the package *Geomstats*. *ACM Trans. Math. Softw.* **49**(4), 1–26 (2023)
37. Nielsen, F.: A simple approximation method for the Fisher–Rao distance between multivariate normal distributions. *Entropy* **25**(4) (2023)
38. Reverter, F., Oller, J.M.: Computing the Rao distance for Gamma distributions. *J. Comput. Appl. Math.* **157**(1), 155–167 (2003)
39. Lauritzen, S.L.: Statistical manifolds, chap. 4. In: Amari, S., Barndorff-Nielsen, O.E., Kass, R.E., Lauritzen, S.L., Rao, C.R. (eds.) *Differential Geometry in Statistical Inference*, vol. 10. Institute of Mathematical Statistics, Hayward (1987)
40. Ay, N., Jost, J., Lê, H.V., Schwachhöfer, L.: *Information Geometry*. Springer, Cham (2017)
41. Klingenberg, W.: *A Course in Differential Geometry*. Springer, New York (1978)
42. Beardon, A.F.: *The Geometry of Discrete Groups*. Springer, New York (1983)
43. Cannon, J.W., Floyd, W.J., Kenyon, R., Parry, W.R.: Hyperbolic geometry. In: Levy, S. (ed.) *Flavors of Geometry*. MSRI Publications, vol. 31. Cambridge University Press, Cambridge (1997)
44. Ratcliffe, J.G.: *Foundations of Hyperbolic Manifolds*, 2nd edn. Springer, New York (2006)
45. Costa, S.I.R., Santos, S.A., Strapasson, J.E.: Fisher information distance: A geometrical reading. *Discrete Appl. Math.* **197**, 59–69 (2015)
46. Kass, R.E., Vos, P.W.: *Geometrical Foundations of Asymptotic Inference*. Wiley, New York (1997)
47. Tsybakov, A.B.: *Introduction to Nonparametric Estimation*. Springer, New York (2009)
48. Khan, G., Zhang, J.: A hall of statistical mirrors. *Asian J. Math.* **26**(6), 809–846 (2022)
49. Nielsen, F.: On Voronoi diagrams on the information-geometric Cauchy manifolds. *Entropy* **22**(7), 713 (2020)
50. Wauters, D., Vermeire, L.: Intensive numerical and symbolic computing in parametric test theory. In: Härdle, W., Simar, L. (eds.) *Computer Intensive Methods in Statistics*, pp. 62–72. Physica-Verlag, Heidelberg (1993)
51. Fang, K.-T., Kotz, S., Ng, K.W.: *Symmetric Multivariate and Related Distributions*. Chapman and Hall, London (1990)

52. Andai, A.: On the geometry of generalized Gaussian distributions. *J. Multivar. Anal.* **100**(4), 777–793 (2009)
53. Dytso, A., Bustin, R., Poor, H.V., Shamai, S.: Analytical properties of generalized Gaussian distributions. *J. Stat. Distrib. Appl.* **5**, 6 (2018)
54. Berkane, M., Oden, K., Bentler, P.M.: Geodesic estimation in elliptical distributions. *J. Multivar. Anal.* **63**(1), 35–46 (1997)
55. Calvo, M., Oller, J.M.: A distance between elliptical distributions based in an embedding into the Siegel group. *J. Comput. Appl. Math.* **145**(2), 319–334 (2002)
56. Chen, X., Zhou, J., Hu, S.: Upper bounds for Rao distance on the manifold of multivariate elliptical distributions. *Automatica* **129**, 109604 (2021)
57. Mitchell, A.F.S., Krzanowski, W.J.: The Mahalanobis distance and elliptic distributions. *Biometrika* **72**(2), 464–467 (1985)
58. Calvo, M., Oller, J.M.: A distance between multivariate normal distributions based in an embedding into the Siegel group. *J. Multivar. Anal.* **35**(2), 223–242 (1990)
59. Krzanowski, W.J.: Rao's distance between normal populations that have common principal components. *Biometrics* **4**, 1467–1471 (1996)
60. Skovgaard, L.T.: A Riemannian geometry of the multivariate normal model. *Scand. J. Stat.* **11**(4), 211–223 (1984)
61. Breloy, A., Ginolhac, G., Renaux, A., Bouchard, F.: Intrinsic Cramér–Rao bounds for scatter and shape matrices estimation in CES distributions. *IEEE Signal Process. Lett.* **26**(2), 262–266 (2019)
62. Kotz, S., Nadarajah, S.: *Extreme Value Distributions: Theory and Applications*. Imperial College Press, London (2000)
63. Prescott, P., Walden, A.T.: Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika* **67**(3), 723–724 (1980)
64. Burbea, J., Oller, J.M., Reverter, F.: Some remarks on the information geometry of the Gamma distribution. *Commun. Stat. Theory Methods* **31**(11), 1959–1975 (2002)
65. Le Brigant, A., Preston, S.C., Puechmorel, S.: Fisher-Rao geometry of Dirichlet distributions. *Differ. Geom. Appl.* **74**, 101702 (2021)
66. Gupta, A.K., Nagar, D.K.: *Matrix Variate Distributions*. Chapman and Hall/CRC, Boca Raton (2000)
67. Bhatia, R.: *Positive Definite Matrices*. Princeton University Press, Princeton (2007)
68. Magnus, J.R., Neudecker, H.: Symmetry, 0–1 matrices and Jacobians: a review. *Econom. Theory* **2**(2), 157–190 (1986)
69. Siegel, C.L.: Symplectic geometry. *Am. J. Math.* **65**(1), 1–86 (1943)
70. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B.: *Bayesian Data Analysis*. Chapman and Hall/CRC, Boca Raton (2013)
71. D'Andrea, F., Martinetti, P.: On Pythagoras theorem for products of spectral triples. *Lett. Math. Phys.* **103**, 469–492 (2012)
72. Oller, J.M.: Some geometrical aspects of data analysis and statistics. In: Dodge, Y. (ed.) *Statistical Data Analysis and Inference*, pp. 41–58. North-Holland, Amsterdam (1989)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.