Discrete Applied Mathematics 197 (2015) 59–69

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Fisher information distance: A geometrical reading[☆]

Sueli I.R. Costa [a], Sandra A. Santos [a,*], João E. Strapasson [b]

[a] Institute of Mathematics, University of Campinas, 13083-970, Campinas, São Paulo, Brazil
[b] School of Applied Sciences, University of Campinas, 13484-350, Limeira, São Paulo, Brazil

ARTICLE INFO

Article history:
Received 17 December 2013
Received in revised form 19 September 2014
Accepted 9 October 2014
Available online 4 November 2014

Keywords:
Fisher distance
Information geometry
Normal probability distribution functions
Kullback–Leibler divergence
Hyperbolic geometry

ABSTRACT

This paper presents a geometrical approach to the Fisher distance, which is a measure of dissimilarity between two probability distribution functions. The Fisher distance, as well as other divergence measures, is also used in many applications to establish a proper data average. The main purpose is to widen the range of possible interpretations and relations of the Fisher distance and its associated geometry for the prospective applications. It focuses on statistical models of the normal probability distribution functions and takes advantage of the connection with the classical hyperbolic geometry to derive closed forms for the Fisher distance in several cases. Connections with the well-known Kullback–Leibler divergence measure are also devised.

1. Introduction

Information geometry is a research field that has provided framework and enlarged the perspective of analysis for a wide variety of domains, such as statistical inference, information theory, mathematical programming, neurocomputing, to name a few. It is an outcome of the investigation of the differential geometric structure on manifolds of probability distributions, with the Riemannian metric defined by the Fisher information matrix [1]. R. A. Fisher introduced the theory of statistical inference and his concept of information appears in his remarkable article [14] (see also [30]). Rao's pioneering work [25] was subsequently followed by several authors (e.g. [3,20,29], among others). We quote [1] as a general reference for this matter.

Concerning specifically to information theory and signal processing, an important aspect of the Fisher matrix arises from its trace being related to the surface area of the typical set associated with a given probability distribution, whereas the volume of this set is related to the entropy. This was used to establish connections between inequalities in information theory and geometric inequalities [9,12].

The Fisher–Rao metric and the Kullback–Leibler divergence may be used to model experimental data in signal processing. As the underlying Fisher–Rao geometry of Gaussians is hyperbolic without a closed-form equation for the centroids, in [21, Chap. 16] the authors have adopted the hyperbolic model centroid approximation, showing its usage in a single-step clustering method. Another recent reference in image processing that also rests upon the hyperbolic geometric structure of the Gaussians is [2], where morphological operators were formulated for an image model where at each pixel is given a

☆ Partially supported by FAPESP (Grants 2011/01096-6, 2013/25977-7, 2013/05475-7 and 2013/07375-0) and CNPq (Grants 304032/2010-7 and 312926/2013-8).

* Corresponding author.
E-mail addresses: sueli@ime.unicamp.br (S.I.R. Costa), sandra@ime.unicamp.br (S.A. Santos), joao.strapasson@fca.unicamp.br (J.E. Strapasson).

http://dx.doi.org/10.1016/j.dam.2014.10.004
0166-218X/

univariate Gaussian distribution, properties of order invariance are explored and an application to morphological processing of univariate Gaussian distribution-valued images is illustrated.

Current applications of information geometry in statistics include the problem of dimensionality reduction through information geometric methods on statistical manifolds [7] as well as the preparation of samplers for sequential Monte Carlo techniques [28]. In the former, the fact that a manifold of probability density function is often intrinsically lower dimensional than the domain of the data realization provides the background for establishing two methods of dimensionality reduction; the proposed tools are illustrated for case studies on actual patient data sets in the clinical flow cytometric analysis. In the latter, the developed sampler with an information geometric kernel design has attained a higher level of statistical robustness in the inferred parameters of the analyzed dynamical systems than the standard adaptive random walk kernel.

In general, many applications demand a measure of dissimilarity between the probability distributions of the involved objects, or also require the replacement of a set of data by a proper average or a centroid [15]. In these cases, the Fisher distance in the model of the considered probability distributions may apply as well as other dissimilarity measures such as the Kullback–Leibler, Bregman and Burbea–Rao measures [19,22–24,26,27]. In the context of distance geometry, similar approaches may be useful in dealing with the Interval Distance Geometry Problem, cf. Sections 3.4 and 3.5 of [18], where probability distributions can be associated with the interval distances provided by practical problems, such as the calculation of protein structures using Nuclear Magnetic Resonance (NMR) data [17].

Our contribution in this paper is to present a geometrical view of the Fisher matrix, focusing on the parameters that describe the univariate and the multivariate normal distributions, with the aim of widen the range of possible interpretations for the prospective applications of information geometry in a variety of fields. Our geometrical reading of information geometry fundamentals, starting at Section 2.1, allows to employ results from the classical hyperbolic geometry and to derive closed expressions for the Fisher distance in special cases of the multivariate normal distributions. Connections with other dissimilarity measure are also deduced. To enhance the geometric approach, those results are deduced along the text, instead of being displayed in a "proposition–proof" format. A preliminary summary of some results presented here has appeared in [10].

This text is organized as follows: in Section 2 we explore the two dimensional statistical model of the Gaussian (normal) univariate probability distribution function (PDF). Closed forms for this distance are derived in the most common parameters (cf. (12)–(14) and Fig. 6) and a relationship with the Kullback–Leibler measure of divergence is presented (see (16)–(17) and Fig. 7). Section 3 is devoted to the Fisher information geometry of the multivariate normal PDF's. For the special cases of the round Gaussian distributions and normal distributions with diagonal covariance matrices, closed forms for the distances are derived (cf. (18) and (20), resp.). The Fisher information distance for the general bivariate case is discussed as well (Section 3.3).

2. Univariate normal distributions: a geometrical view

2.1. The hyperbolic model of the mean × standard deviation half-plane

The geometric model of the mean × standard deviation half-plane associates each point in the half upper plane of $\mathbb{R}^2$ with a univariate Gaussian PDF

$$f(x,\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-|x-\mu|^2}{2\sigma^2}\right).$$

Hence, a classic parametric space for this family of PDF's is

$$H=\{(\mu,\sigma)\in\mathbb{R}^2 \mid \sigma>0\}.$$

A distance between two points $P=(\mu_1,\sigma_1)$ and $Q=(\mu_2,\sigma_2)$ in the half-plane H should reflect the dissimilarity between the associated PDF's. We will not distinguish the notation of the point P in the parameter space and its associated PDF $f(x,P)$.

A comparison between univariate normal distributions is illustrated in Fig. 1. By fixing the means and increasing the standard deviation, we can see that the dissimilarity between the probabilities attached to the same interval concerning the PDF's associated with C and D is smaller than the one between the PDF's associated with A and B (left). This means that the distance between points in the upper half-plane (right) representing normal distributions cannot be Euclidean. Moreover, we can observe that such a metric must vary with the inverse of the standard deviation σ. The points C and D should be closer to each other than the points A and B, reflecting that the pair of distributions A and B is more dissimilar than the pair C and D.

A proper distance arises from the Fisher information matrix, which is a measure of the amount of information of the location parameter [11, ch. 12]. For univariate distributions parametrized by an n-dimensional space, the coefficients of this matrix, which define a metric, are calculated as the expectation of a product involving partial derivatives of the logarithm of the PDF's:

$$g_{ij}(\boldsymbol{\beta})=\int_{-\infty}^{\infty}f(x,\boldsymbol{\beta})\frac{\partial \ln f(x,\boldsymbol{\beta})}{\partial\beta_i}\frac{\partial \ln f(x,\boldsymbol{\beta})}{\partial\beta_j}dx. \tag{1}$$