Physica A 495 (2018) 336–344

Contents lists available at ScienceDirect

Physica A

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

Monoparametric family of metrics derived from classical Jensen–Shannon divergence

Tristán M. Osán [a,b,*], Diego G. Bussandri [c,b], Pedro W. Lamberti [c,b]

[a] Instituto de Física Enrique Gaviola (IFEG), Consejo Nacional de Investigaciones Científicas y Técnicas de la República Argentina (CONICET), Av. Medina Allende s/n, X5000HUA, Córdoba, Argentina
[b] Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba, Av. Medina Allende s/n, Ciudad Universitaria, X5000HUA, Córdoba, Argentina
[c] Consejo Nacional de Investigaciones Científicas y Técnicas de la República Argentina (CONICET), Av. Rivadavia 1917, C1033AAJ, CABA, Argentina

A R T I C L E I N F O

Article history:
Received 25 September 2017
Received in revised form 28 November 2017
Available online 20 December 2017

Keywords:
Jensen–Shannon divergence
Metrics
Information theory
Quantum distances

A B S T R A C T

Jensen–Shannon divergence is a well known multi-purpose measure of dissimilarity between probability distributions. It has been proven that the square root of this quantity is a true metric in the sense that, in addition to the basic properties of a distance, it also satisfies the triangle inequality. In this work we extend this last result to prove that in fact it is possible to derive a monoparametric family of metrics from the classical Jensen–Shannon divergence. Motivated by our results, an application into the field of symbolic sequences segmentation is explored. Additionally, we analyze the possibility to extend this result into the quantum realm.

1. Introduction

Measures of dissimilarity between probability distributions constitute an important topic of research in Probability Theory, Statistics and Information Geometry. Among some fields of application of this kind of measures we can mention, evaluation of risks in statistical decision problems, signal detection, data compression, coding, pattern classification, cluster analysis, etc. Furthermore, many problems of statistical physics can be established in terms of a measure of distance or distinguishability between two probability distributions.

In the realms of Statistics and Information Theory, an extensively applied measure of dissimilarity between probability distributions, is the Jensen–Shannon divergence (JSD) [1,2]. This measure turns out to be a symmetrized, smoothed, well-behaved and bounded version of the Kullback–Leibler divergence [3,4]. JSD has been successfully applied in a wide variety of research fields, such as, analysis and characterization of symbolic sequences and segmentation of digital images. Particularly, it has been exhaustively used in the study of segmentation of DNA sequences. Remarkably, in statistical physics JSD has been used as a measure of the length of the time's arrow [5] and also in the definition of a measure of complexity [6]. In addition, the generalization of JSD within the framework of the non-extensive Tsallis statistics [7] has been studied in [8,9].

In this work we show that it is possible to derive a monoparametric family of metrics from classical Jensen–Shannon divergence. A key aspect of our approach for the demonstration of this assertion is to consider the JSD as a particular case of a Csiszár divergence [10–12].

In information geometry there exists a natural Riemannian structure associated with a local metric known as Fisher's metric [13,14]. Čencov showed that Fisher's metric is the only Riemannian metric on the probability distributions space

* Corresponding author.
E-mail addresses: tosan@famaf.unc.edu.ar, tristan.osan@unc.edu.ar (T.M. Osán).

https://doi.org/10.1016/j.physa.2017.12.073
0378-4371/

$\mathscr{P}$ for which certain natural statistical embeddings are isometries [15]. Besides this formal fact, Fisher's metric is directly related with the practical parameters' estimation problem via the Cramer–Rao bound [13,14]. Furthermore, the existence of Fisher's metric allows the space $\mathscr{P}$ to possess the character of metric space. Indeed, by evaluating the length of a geodesic associated with Fisher's metric, joining two points on the probability distributions space, we can provide a way of measuring the distance between two arbitrary points belonging to the space $\mathscr{P}$. It should be emphasized that we make a distinction between the local metric (which measures how separated are two near points from each other) and a measure of the distance between two arbitrary points. The distance defined through this procedure verifies the properties of a metric (cf. Section 2.1 where we summarize these properties). It is worth to mention that the inverse procedure, i.e., to derive a Riemannian metric from a metric, is not always possible. Additionally, it has been shown that having a metric defined on $\mathscr{P}$ (and over any arbitrary space) is of crucial importance to establish convergence criteria in iterative processes [16].

This paper is organized as follows. In Section 2, we introduce the basic theoretical background related to our work. Next, in Section 3 we prove the main result of this work, i.e., that it is possible to construct a monoparametric family of metrics from the classical expression of JSD. Then, in Section 4 we briefly explore the possibilities of applying our results in two different contexts. On one hand, in Section 4.1 we explore the segmentation of symbolic sequences. On the other hand, in Section 4.2 we study the extension of the monoparametric family of metrics into the quantum realm. Finally, in Section 5 we summarize our results.

2. Theoretical framework

2.1. Divergences, distances and metrics

From a mathematically rigorous viewpoint, a *metric* (or sometimes, a *true* metric) d on a set χ is a function $d : \chi \times \chi \to \mathbb{R}_{\geq 0}$ such that for any $x, y, z \in \chi$ the following properties are satisfied

1. *Non-negativity*: $d(x, y) \geq 0$
2. *Identity of indiscernibles*: $d(x, y) = 0$ if and only if $x = y$
3. *Symmetry*: $d(x, y) = d(y, x)$
4. *Triangle inequality*: $d(x, y) \leq d(x, z) + d(z, y)$.

In the context of classical information, usually χ represents the set of probability distributions and x or y represent an entire probability distribution such as $P = \{p_1, p_2, \ldots, p_n\}$ ($p_i \geq 0\ \forall i$, $\sum_{i=1}^{n} p_i = 1$). Often, if a distance measure d only satisfies the property 1, is called a *divergence*. If, additionally, d satisfies the properties 2 and 3 then d is called a *distance* [16–18]. It is worth to mention that throughout literature the term distance is used many times as equivalent to metric. Due to this use can be misleading in some contexts, for the sake of clarity, throughout this work we will use the terms divergence, distance or metric, according to the specific meaning needed.

2.2. Csiszár's divergences

Csiszár's divergences, also known as $f-$divergences, constitute an important class of measures of distinguishability between probability distributions [10,11]. Let $\mathcal{F}$ be the set of convex functions $f : \mathbb{R}_+ \mapsto \bar{\mathbb{R}}$ which are finite on $\mathbb{R}_0$ and continuous on $\mathbb{R}_+$, where $\bar{\mathbb{R}} = (-\infty, \infty]$, $\mathbb{R}_+ = [0, \infty)$ and $\mathbb{R}_0 = (0, \infty)$. The Csiszár's $f-$divergence between the probability distributions $P = \{p_1, p_2, \ldots, p_n\}$ and $Q = \{q_1, q_2, \ldots, q_n\}$ is defined as [10–12]

$$D_f(P, Q) = \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right). \tag{1}$$

From its definition, it can be seen that $D_f(P, Q)$ is a useful functional form which encompasses most of the commonly used divergence measures between probability distributions, such as, Kullback–Leibler divergence, Variational Distance, Hellinger distance, χ^2-divergence, Jensen–Shannon divergence, among others [2–4,18,19].

2.2.1. Basic properties of Csiszár's divergences

In what follows we will summarize some basic properties of Csiszár's divergences directly related with the main result of this work. Further analysis of the properties of $f-$divergences can be found in references [20–24].

Let $f^* \in \mathcal{F}$, the *–conjugate (convex) function of f, be defined as:

$$f^*(u) = u f\left(\frac{1}{u}\right) \quad \text{for} \quad u \in \mathbb{R}_0. \tag{2}$$

If $f(1) = 0$, f is strictly convex at 1, and $f^*(u) = f(u)$, then $D_f(P, Q)$ satisfies the following basic properties [23–25]:

1. *Non-negativity and Identity of indiscernibles*: $D_f(P, Q) \geq 0$ with $D_f(P, Q) = 0 \iff P = Q$
2. *Symmetry*: $D_f(P, Q) = D_f(Q, P)$
3. *Uniqueness*: $D_{f_1}(P, Q) = D_f(P, Q)$, $\iff \exists c \in \mathbb{R} \,/\, f_1(u) = f(u) + c(u - 1)$
4. *Range of values*: $f(1) \leq D_f(P, Q) \leq f(0) + f^*(0)$.