



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



ARTICLE INFO

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

ABSTRACT

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.

© 2017 Elsevier B.V. All rights reserved.

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).

\mathcal{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 \mathcal{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 \mathcal{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 \mathcal{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 $d : \chi \times \chi \rightarrow \mathbb{R}_{\geq 0}$ 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 $P = \{p_1, p_2, \dots, p_n\}$ ($p_i \geq 0 \forall i, \sum_{i=1}^n p_i = 1$). 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 $f : \mathbb{R}_+ \mapsto \mathbb{R}$ which are finite on \mathbb{R}_0 and continuous on \mathbb{R}_+ , where $\mathbb{R} = (-\infty, \infty)$, $\mathbb{R}_+ = [0, \infty)$ and $\mathbb{R}_0 = (0, \infty)$. The Csiszár’s f –divergence between the probability distributions $P = \{p_1, p_2, \dots, p_n\}$ and $Q = \{q_1, q_2, \dots, q_n\}$ 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:

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

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*: $f(1) \leq D_f(P, Q) \leq f(0) + f^*(0)$.

Bearing in mind the properties of a metric described in Section 2.1 it can be seen that properties 1 and 2 are essential in order to seek for a metric based on a Csiszár divergence. The following theorem addresses precisely the additional conditions that a Csiszár divergence needs to satisfy in order for $[D_f(P, Q)]^\alpha$ to be a metric (cf. Section 2.1).

Theorem 1. Let $D_f(P, Q)$ denote a Csiszár divergence (cf. Section 2.2 and Section 2.2.1). If $f(1) = 0$ and f is strictly convex at 1, $f^*(u) = f(u)$, and there exists $\alpha \in \mathbb{R}_0$ such that the function

$$h_\alpha(u) = \frac{(1 - u^\alpha)^{1/\alpha}}{f(u)} \quad (3)$$

is nonincreasing for $u \in [0, 1)$, then $d_\alpha(P, Q) \doteq [D_f(P, Q)]^\alpha$, in addition to properties 1 and 2, also satisfies the triangle inequality $d_\alpha(P, Q) \leq d_\alpha(P, R) + d_\alpha(R, Q)$. Therefore, $d_\alpha(P, Q)$ is a metric [23–25]. ■

2.2.2. KL divergence

The Kullback–Leibler divergence (KLD) between two probability distributions $P = \{p_1, p_2, \dots, p_n\}$ and $Q = \{q_1, q_2, \dots, q_n\}$ is defined as [3,4]:

$$D_{KL}(P, Q) = \sum_{i=1}^n p_i \log_2 \left(\frac{p_i}{q_i} \right). \quad (4)$$

It is straightforward to show that $D_{KL}(P, Q)$ is a Csiszár divergence (cf. Eq. (1)) taking the function $f(u)$ as:

$$f_{KL}(u) = u \log_2(u) \quad (5)$$

for $u \in \{\mathbb{R}_+ \cup \{0\}\}$.

The KLD constitutes a paradigmatic case of a widely used measure of dissimilarity between probability distributions which is not a metric. For example, $D_{KL}(P, Q)$ does not fulfill the basic property of symmetry (2) essential for a metric. In addition, it should be noted that this divergence also possesses another undesired features. For example, if for any j there exists a $q_j = 0$ for which $p_j \neq 0$, then $D_{KL}(P, Q)$ is undefined. Therefore, the probability distribution P must be absolutely continuous with respect to the probability distribution Q in order for $D_{KL}(P, Q)$ to be well defined [4]. However, it is worth to mention that this latter undesired feature of KLD can be avoided if the convention $0(-\infty) = 0$ is adopted.

2.2.3. Classical Jensen–Shannon divergence

The Jensen–Shannon divergence (JSD) between two probability distributions is defined as follows [1,2]

$$D_{JS}(P, Q) = \frac{1}{2} \left[D_{KL} \left(P, \frac{P+Q}{2} \right) + D_{KL} \left(Q, \frac{P+Q}{2} \right) \right]. \quad (6)$$

After some algebra, a more explicit version of JSD can be written in terms of Shannon entropy $H = -\sum_i p_i \log_2 p_i$ as follows

$$D_{JS}(P, Q) = H \left(\frac{P+Q}{2} \right) - \frac{1}{2} H(P) - \frac{1}{2} H(Q). \quad (7)$$

In last equation, the dissimilarity between both probability distributions P and Q is evaluated assuming that both distributions have the same weight (1/2). If, instead, we consider arbitrary weights $\pi_1 \geq 0$ and $\pi_2 \geq 0$, such that $\pi_1 + \pi_2 = 1$, then Eq. (7) can be generalized as follows [1,2]:

$$D_{JS}^{(\pi_1, \pi_2)}(P, Q) = H(\pi_1 P + \pi_2 Q) - \pi_1 H(P) - \pi_2 H(Q). \quad (8)$$

It is easy to prove that the expression (7) can be written in the form of a Csiszár divergence (cf. Eq. (1)) with the function $f(u)$ defined as:

$$f_{JS}(u) = \frac{1}{2} \left[(1+u) + u \log_2(u) - (1+u) \log_2(1+u) \right] \quad (9)$$

for $u \in \mathbb{R}_+$.

From Eq. (6) it can be seen that $D_{JS}(P, Q)$ is a symmetric version of $D_{KL}(P, Q)$. Originated in the field of Information Theory [1,2], JSD is always well defined and bounded. Unlike KLD, as JSD satisfies the properties 1, 2 and 3 (cf. Section 2.1), it is in fact a distance. Additionally, JSD has several interesting interpretations. For example, in statistical inference theory it gives both the lower and upper bounds to Bayes' probability error, whereas in the framework of information theory JSD can be related to mutual information [26].

3. Monoparametric family of metrics from classical Jensen–Shannon divergence

It is well known that the square root of the classical JSD, i.e., $[D_{JS}(P, Q)]^{1/2}$ (cf. Eq. (7)) is a (true) metric, i.e., it satisfies the properties 1, 2, 3 and 4 (cf. Section 2.1) [25,27]. In what follows we will extend this result to show that $d_\alpha(P, Q) \doteq [D_{JS}(P, Q)]^\alpha$

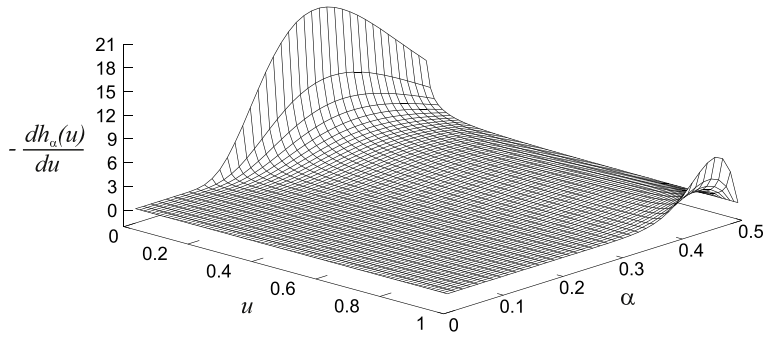


Fig. 1. Plot of $-\frac{dh_\alpha(u)}{du}$ (cf. Eq. (11)), as a function of $u \in [0, 1]$ and $\alpha \in (0, 1/2]$, in the case of Jensen–Shannon divergence. It can be seen that the quantity $h_\alpha(u)$ is nonincreasing for $\alpha \in (0, 1/2]$. According to [Theorem 1](#) this result indicates that the quantity $[D_{JS}(P, Q)]^\alpha$ for $\alpha \in (0, 1/2]$ is a metric.

for $\alpha \in (0, 1/2]$ is a (true) metric, i.e., we will prove that is possible to construct a monoparametric family of metrics from the classical expression of JSD. A key aspect of our approach for the proof is to consider the JSD as a particular example of a Csiszár divergence. This allows us to use the [Theorem 1](#) (cf. Section 2.2.1) to derive the conditions on the values of the parameter α for the function $d_\alpha(P, Q)$ to be a (true) metric.

[Fig. 1](#) shows a plot of the quantity $-dh_\alpha/du$, as a function of $u \in [0, 1]$ and $\alpha \in (0, 1/2]$ in the case of the JSD (cf. Eqs. (3) and (11)). It can be seen that the quantity $h_\alpha(u)$ is nonincreasing for $\alpha \in (0, 1/2]$. Thus, according to [Theorem 1](#), the quantity $[D_{JS}(P, Q)]^\alpha$ should be a metric for $\alpha \in (0, 1/2]$. In what comes next, we state this hypothesis as [Proposition 1](#) and next we formally prove it.

Proposition 1. Let $D_{JS}(P, Q)$ denote the Jensen–Shannon divergence between two probability distributions $P = \{p_i \in \mathbb{R} \mid p_i \geq 0; \sum_{i=1}^n p_i = 1\}$ and $Q = \{q_j \in \mathbb{R} \mid q_j \geq 0; \sum_{j=1}^n q_j = 1\}$. Then, the quantity

$$d_\alpha(P, Q) \doteq [D_{JS}(P, Q)]^\alpha \tag{10}$$

is a (true) metric for all $\alpha \in (0, 1/2]$. ■

Proof. The proof is based on [Theorem 1](#) (cf. Section 2.2.1). Therefore, we need to prove that the function $h_\alpha(u)$ (cf. Eq. (3)) corresponding to the JSD is non-increasing for all $u \in [0, 1] \subset \mathbb{R}$ and $\alpha \in (0, 1/2]$. Thus, we shall analyze the sign of the derivative of the function $h_\alpha(u)$. After some algebra, we obtain:

$$\frac{dh_\alpha(u)}{du} = -\frac{(1 - u^\alpha)^{\frac{1}{\alpha}-1} \cdot \{u \log_2 u + (u + u^\alpha)[1 - \log_2(1 + u)]\}}{2u[f_{JS}(u)]^2}. \tag{11}$$

Clearly, the sign of $\frac{dh_\alpha}{du}$ depends upon the sign of

$$u \log_2 u + (u + u^\alpha)[1 - \log_2(1 + u)] \tag{12}$$

because $(1 - u^\alpha)^{\frac{1}{\alpha}-1}$ and $u [f_{JS}(u)]^2$ are positive for $u \in [0, 1]$ and $\alpha > 0$. Thus, in order to determine the conditions for $\frac{dh_\alpha}{du} \leq 0$ we shall analyze the behavior of Eq. (12) as a function of α and u .

Starting from Eq. (12) we have the following sequence of inequalities:

$$u \log_2 u + (u + u^\alpha)[1 - \log_2(1 + u)] \geq 0 \Leftrightarrow \tag{13}$$

$$u^\alpha [1 - \log_2(1 + u)] \geq u \log_2(1 + u) - u - u \log_2(u) \Leftrightarrow \tag{14}$$

$$u^\alpha \log_2 \left(\frac{2}{1 + u} \right) \geq u \log_2 \left(\frac{1 + u}{2u} \right) \Leftrightarrow \tag{15}$$

$$u^\alpha \geq u \frac{\log_2 \left(\frac{1+u}{2u} \right)}{\log_2 \left(\frac{2}{1+u} \right)} = u \frac{\ln \left(\frac{1+u}{2u} \right)}{\ln \left(\frac{2}{1+u} \right)}. \tag{16}$$

In the case $\alpha = 1/2$ it has been proven that $[D_{JS}(P, Q)]^\alpha$ is a metric [\[25,27\]](#). Thus, inequality (16) is satisfied for the particular case $\alpha = 1/2$. Next, making use of the fact that for $u \in (0, 1)$, if $\alpha \geq \beta$ then $u^\alpha \leq u^\beta$, we obtain the following inequality

$$u^\beta \geq u^{1/2} \geq u \frac{\log_2 \left(\frac{1+u}{2u} \right)}{\log_2 \left(\frac{2}{1+u} \right)}, \tag{17}$$

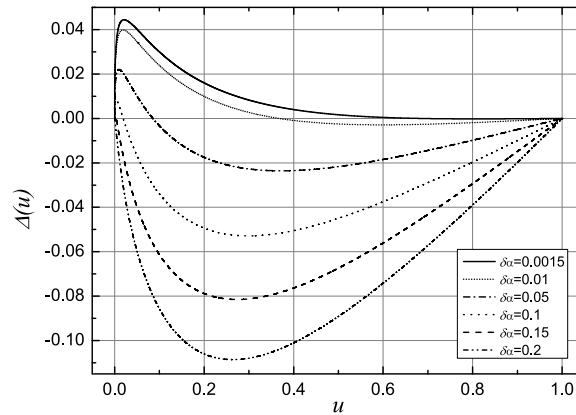


Fig. 2. Plot of $\Delta(u)$ (cf. Eq. (19)) as a function of u for some values of $\alpha = 1/2 + \delta\alpha$ with $\delta\alpha \in (0, 1/2)$. These results support the conjecture that $[D_{JS}(P, Q)]^\alpha$ is not a metric for $\alpha \in (1/2, 1)$.

for $\beta \in (0, 1/2]$. As a consequence, $h_\alpha(u)$ turns out to be non-increasing in $[0, 1]$ for all $\alpha \in (0, 1/2]$. Therefore,

$$d_\alpha(P, Q) = [D_{JS}(P, Q)]^\alpha$$

is a metric for all $\alpha \in (0, 1/2]$. □

Proposition 2. For $\alpha \geq 1$, $[D_{JS}(P, Q)]^\alpha$ is not a metric. ■

Proof. For $\alpha \geq 1$ we need to prove that inequality (16) is violated. Making use of the fact that for $u \in (0, 1)$, if $\alpha \geq 1$ then $u^\alpha \leq u$, $\ln(x)$ is a monotonically increasing function of x , and $(1 + u)/2u > 2/(1 + u) > 1$ for $u \in (0, 1)$, it follows that

$$u^\alpha < u \frac{\ln\left(\frac{1+u}{2u}\right)}{\ln\left(\frac{2}{1+u}\right)}. \tag{18}$$

Thus, inequality (16) is not satisfied in this case. Therefore, $[D_{JS}(P, Q)]^\alpha$ is not a metric for $\alpha \geq 1$. □

This result is in agreement with the obtained by Khosravifard and co-workers in [28]. They conclude that Variational Distance (also known as Kolmogorov Distance) is the unique Csiszár divergence which additionally is a metric [28].

Conjecture 1. For $1/2 < \alpha < 1$, $[D_{JS}(P, Q)]^\alpha$ is not a metric. ■

In the case $\alpha \in (1/2, 1)$ a number of counter-examples can be found showing that inequality (16) is violated for some sub-intervals of values of $u \in (0, 1)$, depending on the particular value of $\alpha \in (1/2, 1)$. These results support the conjecture that $[D_{JS}(P, Q)]^\alpha$ is not a metric for $\alpha \in (1/2, 1)$. However, so far we have not been able to prove this conjecture in a closed way. In order to present some examples of Conjecture 1, we define the following function:

$$\Delta(u) \doteq u^\alpha - u \frac{\ln\left(\frac{1+u}{2u}\right)}{\ln\left(\frac{2}{1+u}\right)}. \tag{19}$$

Fig. 2 shows some plots of the function $\Delta(u)$ for different values of the exponent $\alpha \doteq 1/2 + \delta\alpha$ with $\delta\alpha \in (0, 1/2)$. It can be seen that $\Delta(u) < 0$ for some subintervals of $u \in [0, 1]$ showing that inequality (16) cannot be fulfilled for all values of $u \in [0, 1]$. The size of the subinterval of values of u for which $\Delta(u) < 0$ is enlarged as the value of $\delta\alpha$ is increased from 0^+ to $1/2^-$. In fact, for values of $\delta\alpha \geq 0.2$ inequality (16) begins to be unsatisfied for almost all values of $u \in [0, 1]$. Thus, $[D_{JS}(P, Q)]^\alpha$ is not a metric for at least some values of the exponent α belonging to the interval $(1/2, 1)$.

4. Applications

In this section we briefly explore two possible applications of our main result, i.e., $[D_{JS}(P, Q)]^\alpha$ is a metric for $\alpha \in (0, 1/2]$. In Section 4.1 we explore the segmentation of symbolic sequences whereas in Section 4.2 we study the possibility to obtain a monparametric family of metrics in the quantum realm.

4.1. Segmentation of symbolic sequences

A symbolic sequence is said statistically stationary if the frequency (probability) of occurrence of different symbols is the same along the entire sequence. In symbolic sequences resulting from real processes stationarity is not the most common situation. Thus, it is of great practical interest to provide a method to detect non-stationarity. In Ref. [26] I. Grosse and co-workers developed a method based on the Jensen–Shannon divergence that allows the detection of changes in the statistical properties of symbolic sequences. This method and some modifications have been extensively used in the analysis of real and simulated time series [8]. Typical applications range from the detection of epileptic crisis to the study of the alignment of the axis of an electric motor [29].

The monoparametric family of metrics introduced in the previous section (cf. Proposition 1) motivated us to investigate the powers of $D_{JS}^{(\pi_1, \pi_2)}(P, Q)$ (cf. Eq. (8)), in the range of the exponent $0 < \alpha \leq \frac{1}{2}$, as suitable quantities for studying the presence of non-stationarity in symbolic sequences. To this end, we implemented Monte Carlo simulations described as follows. We generated 500 binary sequences of 1000 symbols each one. The first 500 symbols have a probability of occurrence $p_0 = 0.8$ and $p_1 = 0.2$ for symbols 0 and 1, respectively. The remaining 500 symbols have a probability of occurrence $q_0 = 0.2$ and $q_1 = 0.8$. We introduced a mobile cursor along each sequence, denoting the position of this cursor with the letter ℓ , $1 \leq \ell \leq 1000$. For each position given by ℓ we define two weights $\pi_1 = \frac{\ell}{1000}$ and $\pi_2 = \frac{1000-\ell}{1000}$. Then, we introduced the quantity

$$d'_\alpha(\ell) \equiv \left[- \sum_{i=1}^2 (\pi_1 f_i + \pi_2 g_i) \log(\pi_1 f_i + \pi_2 g_i) + \pi_1 \sum_{i=1}^2 f_i \log f_i + \pi_2 \sum_{j=1}^2 g_j \log g_j \right]^\alpha \tag{20}$$

where f_j is the calculated frequency of occurrence of the symbol j to the left of the cursor and g_j is the calculated frequency of the symbol j to the right of the cursor. Finally, we evaluated the average $d'_\alpha(\ell)$ over all the realizations of the sequences. In Fig. 3 we plot $d'_\alpha(\ell)$ as a function of ℓ for different values of α . As the quantity $d'_{\alpha=1}(\ell)$ has been used in other works (see for example Ref. [26]), for the sake of comparison, in Fig. 3 (a) and (b) we also show the results corresponding to $\alpha = 1$. In Fig. 3 (a) we clearly observe that the maximum occurs at the position $\ell = 500$, i.e., just the place where the probability distribution of the generated sequences changes from $\{p_0, p_1\}$ to $\{q_0, q_1\}$.

Now, when we face the problem of analyzing an unknown sequence, we do not have in general any *a priori* knowledge about its possible stationary character, i.e., the location of probable segmentation points. Thus, when a likely segmentation point is found, the procedure needs to decide if such a point is *statistically significant*, i.e., if it is significantly greater than expected by chance. Following Ref. [26], for an observed value $d'_\alpha = x$, the *statistical significance* $s_\alpha(x)$ is defined as

$$s_\alpha(x) \doteq \text{Prob}\{d'_\alpha \leq x\}. \tag{21}$$

Looking at Eq. (21), $s_\alpha(x)$ is interpreted as the probability of obtaining x or a lower value under the hypothesis that all subsequences are generated from the same probability distribution. As a result, following Ref. [26], the significance $s_\alpha(x)$ for large values of $N (\geq 10^2)$ can be estimated as [26]

$$s_\alpha(x) \sim \frac{\gamma(\nu/2, L(\ln 2)x^{1/\alpha})}{\Gamma(\nu/2)} \tag{22}$$

being $\nu = (N - 1)(m - 1)$, L the length of the sequence, m the number of subsequences and N the number of symbols in the used alphabet, whereas $\gamma(a; x)$ and $\Gamma(a)$ represent the incomplete and complete gamma function, respectively [30].

Once we established the criterion to decide the statistical significance of potential segmentation points, the procedure for the analysis of stationary features in an unknown sequence proceeds as follows. We calculate the value of $d'_\alpha(\ell)$ for each position ℓ of the cursor along the entire sequence, as described earlier in this section. Next, we select the particular point at which d'_α reaches its maximum value d'^{\max}_α and we compute its statistical significance s_α^{\max} . If the value s_α^{\max} exceeds a given threshold s_0 the sequence is partitioned at this point and the procedure continues recursively for each one of the two resulting subsequences. Otherwise, the sequence remains unpartitioned. The process is finished when none of the subsequent potential cutting points has a statistical significance exceeding s_0 . In this case, we say that the sequence has been segmented at a significance threshold s_0 .

Fig. 3(b) shows the dependence of the maximum value of d'_α found during the entire segmentation process of the simulated sequences described earlier in this section, as a function of α . It can be seen that this maximum increases as the exponent α decreases. In addition, for a given threshold value s_0 , it is straightforward to see from Eq. (22) that the statistical significance of d'_α monotonically increases as the value of α is decreased. This behavior is in complete agreement with the results showed in Fig. 3(b). Thus, the method based on the statistical significance of the maximum values of d'_α reached along the entire procedure turns out to be sensitive to the choice of the parameter α . The results also suggest that a suitable election of the value of α might improve the proposed scheme. However, it is worth to mention that another methods can be found in literature for this kind of tasks. For example, a procedure based upon the calculation of the derivatives around the maximum value of a segmentation process is presented in Refs. [31,32]. Further studies on this matter will be the subject of future research by the authors.

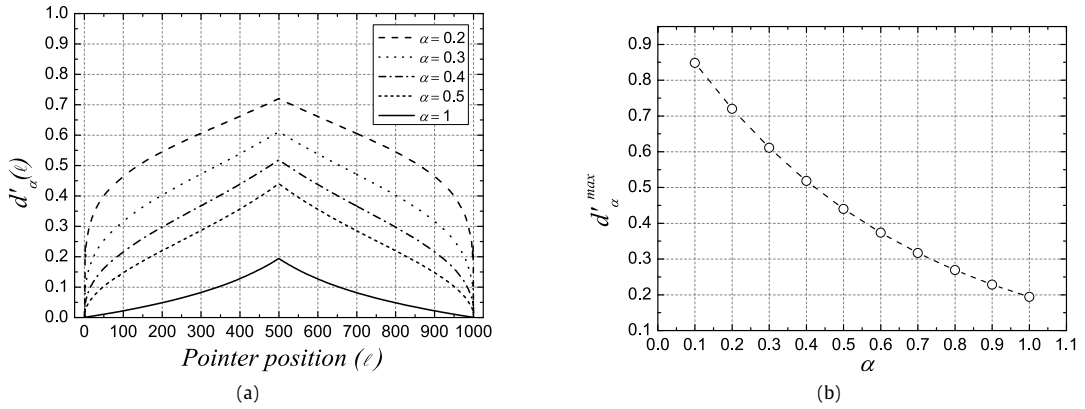


Fig. 3. (a) Average results for the segmentation of 500 binary sequences by means of d'_α (cf. Eq. (20)). Each sequence of 1000 symbols was generated by means of the Monte Carlo method. The maximum segmentation value occurs precisely for the cursor position $\ell = 500$, i.e., the position where the probability distribution for the generated sequences changes from $\{p_0, p_1\}$ to $\{q_0, q_1\}$. (b) Maximum segmentation value of d'_α as a function of the exponent α .

4.2. Mono-parametric family of metrics in the quantum realm

It is well known that two quantum states can be discriminated unambiguously if, and only if, they are orthogonal. Thus, in the realm of Quantum Mechanics, distance measures need to be devised to allow us to determine how close two quantum states are from each other. At present, there is no general agreement about the use of a unique or ideal measure of distinguishability between quantum states. Moreover, different distance measures seem to be useful depending on the particular application, whether a theoretical one, like a bound of what can be physically feasible for a given process, or the output of a quantum protocol experimentally implemented. Therefore, from a conceptual point of view it is useful to develop new measures of distance between quantum states and then analyze their properties and possible applications.

On one hand, the probability distributions which determine the possible results to be obtained performing measurements on a quantum system with a state represented by a density matrix ρ depend not only on ρ but also on the set of measurements which can be performed on the system. On the other hand, the most general way of representing measurements in the quantum realm is by means of the *Positive Operator-Valued Measurement* (POVM) formalism [19,33,34]. Therefore, it is possible to extend the use of classical JSD into the quantum realm by means of two probability distributions defined as follows:

$$P(\mathbb{E}, \rho) = \{p_i | p_i = \text{Tr}(E_i \rho)\} \tag{23}$$

$$Q(\mathbb{E}, \sigma) = \{q_i | q_i = \text{Tr}(E_i \sigma)\} \tag{24}$$

where $\mathbb{E} = \{E_i\}_{i=1}^K$ represent some POVM measurement ($\sum_{i=1}^K E_i = \mathbb{I}$) [19,34].

Thus, a distinguishability measure between the quantum states ρ and σ can be obtained from the classical JSD by assigning probabilities according to Eqs. (23) and (24), and then optimizing over all possible POVMs. Since this procedure has the freedom of choosing the particular POVM distinguishing between the two probability distributions with more certainty, the following quantity can be introduced [35,36]:

$$D_{JS1}(\rho, \sigma) = \max_{\{\mathbb{E}_i\}} D_{JS}(p_i, q_i), \tag{25}$$

where the maximum is taken over the entire set of POVM's. Physically D_{JS1} yields the best discrimination between the states ρ and σ that can be achieved by means of measurements. It is clear from the definition of $D_{JS1}(\rho, \sigma)$ that the results of our Proposition 1 (cf. Section 3) can be used to obtain a monoparametric family of metrics in the quantum realm in the form $[D_{JS1}(\rho, \sigma)]^\alpha$ with $\alpha \in (0, 1/2]$.

Among the potential applications of the metric property of $[D_{JS1}(\rho, \sigma)]^\alpha$ we can mention its use as a tool for testing the convergence of iterative algorithms in quantum computation tasks [36–38] and to evaluate the performance of complex tasks of quantum information processing which can be decomposed into sequences of operations of lesser complexity [39].

5. Concluding remarks

In this work we extended the previous and well-known result that the square root of the classical Jensen–Shannon divergence, i.e., $[D_{JS}(P, Q)]^{1/2}$, between two probability distributions P and Q is a metric by explicitly proving that it is possible to derive an entire monoparametric family of metrics from the classical JSD. Indeed, we demonstrated that the

quantity $[D_{JS}(P, Q)]^\alpha$ is a metric for all $\alpha \in (0, 1/2]$ in the sense that, in addition to the basic properties of a distance (Positive Definiteness, Symmetry and Identity of Indiscernibles, cf. Section 2.1), it also satisfies the Triangle Inequality. Furthermore, we explicitly demonstrated that the quantity $[D_{JS}(P, Q)]^\alpha$ is not a metric for all $\alpha \geq 1$. The key aspects of our proofs were to consider the Jensen–Shannon divergence as a particular case of a Csiszár divergence. We also conjectured that this quantity is not a metric for $\alpha \in (1/2, 1)$ by providing some general examples supporting this hypothesis. However, so far we have not been able to obtain an analytical proof of this conjecture. This issue will be the subject of further studies. Also, motivated by our findings we briefly explored an application into the field of segmentation of symbolic sequences by introducing the α -power of a quantity based on the generalized Jensen–Shannon divergence. Thus, the quantity d'_α (cf. Section 4.1) was studied as a tool for the detection of possible stationary features in symbolic sequences. The method used for segmentation is based on the statistical significance of the maximum values of d'_α reached along the entire procedure. On one hand, our results indicate that this method is sensitive to the particular choice of the parameter α . On the other hand, they also suggest that a suitable election of the parameter α might improve this segmentation scheme. Additionally, we analyzed the possibility of extending the monoparametric family of metrics we derived from classical Jensen–Shannon divergence into the quantum realm. As a result, we also found a monoparametric family of metrics in the quantum realm, based upon the classical JSD, which can be derived from a quantum distance introduced in Refs. [35,36] (cf. Eq. (25), Section 4.2). This last finding also deserves further study and the results will be presented elsewhere.

Acknowledgments

The authors are grateful to F. Österreicher for gently providing a hardcopy of Ref. [23]. T.M.O and P.W.L are members of the National Research Council of Argentina (CONICET). D.G.B. has a fellowship from CONICET. The authors are grateful to Secretaría de Ciencia y Técnica de la Universidad Nacional de Córdoba (SECyT-UNC, Argentina, grant code 30720150100113CB) and CONICET (grant code PIP 11220120100237) for financial support.

References

- [1] C. Rao, A general class of coefficients of divergence of one distribution from another, *IMS-Lectures Notes* 10 (1987) 217–240.
- [2] J. Lin, Divergence measures based on the Shannon entropy, *IEEE Trans. Inform. Theory* 37 (1991) 145–151. <http://dx.doi.org/10.1109/18.61115>.
- [3] S. Kullback, R.A. Leibler, On information and sufficiency, *Ann. Math. Statist.* 22 (1) (1951) 79–86. <http://dx.doi.org/10.1214/aoms/1177729694>.
- [4] S. Kullback, *Information Theory and Statistics*, Dover Publications New York, 1968.
- [5] E.H. Feng, G.E. Crooks, Length of time's arrow, *Phys. Rev. Lett.* 101 (2008) 090602. <http://dx.doi.org/10.1103/physrevlett.101.090602>.
- [6] O.A. Rosso, H.A. Larrondo, M.T. Martin, A. Plastino, M.A. Fuentes, Distinguishing noise from chaos, *Phys. Rev. Lett.* 99 (2007) 154102. <http://dx.doi.org/10.1103/physrevlett.99.154102>.
- [7] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*, Springer, 2009.
- [8] P.W. Lamberti, A.P. Majtey, Non-logarithmic jensen–shannon divergence, *Physica A* 329 (1) (2003) 81–90. [http://dx.doi.org/10.1016/S0378-4371\(03\)00566-1](http://dx.doi.org/10.1016/S0378-4371(03)00566-1).
- [9] A.P. Majtey, P.W. Lamberti, A. Plastino, A monoparametric family of metrics for statistical mechanics, *Physica A* 344 (3) (2004) 547–553. <http://dx.doi.org/10.1016/j.physa.2004.06.030>.
- [10] S.M. Ali, S.D. Silvey, A general class of coefficients of divergence of one distribution from another, *J. Roy. Statist. Soc. Ser. B* 28 (1966) 131–142.
- [11] I. Csiszár, Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizität von markoffschen ketten, *Magyar. Tud. Akad. Mat. Kutató Int. Közl* 8 (1963) 85–108.
- [12] I. Csiszár, On topological properties of f -divergences, *Studia Math. Hungar.* 2 (1967) 329–339.
- [13] S.-I. Amari, H. Nagaoka, *Methods of Information Geometry*, Vol. 191, Translations of Mathematical Monographs, American Mathematical Society, Oxford University Press, 1993.
- [14] S.-I. Amari, *Information Geometry and Its Applications*, Vol. 194, Applied Mathematical Sciences, Springer Japan, 2016.
- [15] N.N. Čencov, *Statistical Decision Rules and Optimal Inference*, Vol. 53, Translations of Mathematical Monographs, American Mathematical Society, Providence, 1982.
- [16] V. Bryant, *Metric Spaces: Iteration and Application*, Cambridge University Press, 1985.
- [17] M. Hayashi, S. Ishizaka, A. Kawachi, G. Kimura, T. Ogawa, *Introduction to Quantum Information Science*, Springer-Verlag, 2015.
- [18] M.M. Deza, E. Deza, *Encyclopedia of Distances*, fourth ed., Springer-Verlag Berlin Heidelberg, 2016.
- [19] I. Bengtsson, K. Życzkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement*, Cambridge University Press, 2006. <http://dx.doi.org/10.1017/CBO9780511535048>.
- [20] I. Vajda, On f -divergence and singularity of probability measures, *Period. Math. Hungar.* 2 (1972) 223–234.
- [21] I. Csiszár, Information measures: A critical survey, in: *Trans. 7th Prague Conf. on Information Theory*, Vol. A, 1974, 73–86.
- [22] F. Liese, I. Vajda, *Convex Statistical Distances*, in: *Band 95, Teubner Texte zur Mathematik*, Leipzig, 1987.
- [23] P. Kafka, F. Österreicher, I. Vincze, On powers of f -divergences defining a distance, *Studia Sci. Math. Hungar.* 26 (1991) 415–422.
- [24] F. Österreicher, I. Vajda, A new class of metric divergences on probability spaces and its applicability in statistics, *Ann. Inst. Statist. Math.* 55 (2003) 639–653. <http://dx.doi.org/10.1007/BF02517812>.
- [25] F. Österreicher, On a class of perimeter-type distances of probability distributions, *Kybernetika* 32 (1996) 389–393.
- [26] I. Grosse, P. Bernaola-Galván, P. Carpena, R. Román-Roldán, J. Oliver, H.E. Stanley, Analysis of symbolic sequences using the Jensen–Shannon divergence, *Phys. Rev. E* 65 (2002) 041905. <http://dx.doi.org/10.1103/physreve.65.041905>.
- [27] J. Endres, D.M. Schindelin, A new metric for probability distributions, *IEEE Trans. Inform. Theory* 49 (2003) 1858–1860. <http://dx.doi.org/10.1109/tit.2003.813506>.
- [28] M. Khosravifard, D. Fooladivanda, T.A. Gulliver, Conflict of the convexity and metric properties in f -divergences, *IEICE Trans. Fundam.* E90–A (2007) 1848–1853. <http://dx.doi.org/10.1093/ietfec/e90a.9.1848>.
- [29] D.M. Mateos, L.E. Riveaud, P.W. Lamberti, Detecting dynamical changes in time series by using the jensen shannon divergence, *Chaos* 27 (2017) 083118. <http://dx.doi.org/10.1063/1.4999613>.
- [30] M. Abramowitz, *Handbook of Mathematical Functions*, Dover, 1972.

- [31] A.D.C. Nascimento, M.M. Horta, A.C. Frery, R.J. Cintra, Comparing edge detection methods based on stochastic entropies and distances for polar imagery, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 7 (2014) 648–663. <http://dx.doi.org/10.1109/JSTARS.2013.2266319>.
- [32] J. Naranjo-Torres, J. Gambini, A.C. Frery, The geodesic distance between G^0 models and its application to region discrimination, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 10 (2017) 987–997. <http://dx.doi.org/10.1109/jstars.2017.2647846>.
- [33] C. Helstrom, *Quantum Detection and Estimation Theory*, in: *Mathematics in Science and Engineering : a series of monographs and textbooks*, Academic Press, 1976.
- [34] M.A. Nielsen, I.L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, Cambridge University Press, 2010. <http://dx.doi.org/10.1017/CBO9780511976667>.
- [35] A.P. Majtey, P.W. Lamberti, D.P. Prato, Jensen-shannon divergence as a measure of distinguishability between mixed quantum states, *Phys. Rev. A* 72 (2005) 052310. <http://dx.doi.org/10.1103/physreva.72.052310>.
- [36] P.W. Lamberti, A.P. Majtey, A. Borras, M. Casas, A. Plastino, Metric character of the quantum jensen-shannon divergence, *Phys. Rev. A* 77 (2008) 052311. <http://dx.doi.org/10.1103/PhysRevA.77.052311>.
- [37] A. Galindo, M.A. Martín-Delgado, Information and computation: Classical and quantum aspects, *Rev. Modern Phys.* 74 (2002) 347–423. <http://dx.doi.org/10.1103/RevModPhys.74.347>.
- [38] T.M. Osán, P.W. Lamberti, Purification-based metric to measure the distance between quantum states and processes, *Phys. Rev. A* 87 (2013) 062319. <http://dx.doi.org/10.1103/PhysRevA.87.062319>.
- [39] A. Gilchrist, N.K. Langford, M.A. Nielsen, Distance measures to compare real and ideal quantum processes, *Phys. Rev. A* 71 (2005) 062310. <http://dx.doi.org/10.1103/PhysRevA.71.062310>.