



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Physica A 329 (2003) 81–90

PHYSICA A

www.elsevier.com/locate/physa

Non-logarithmic Jensen–Shannon divergence

Pedro W. Lamberti*, Ana P. Majtey

*Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba,
Ciudad Universitaria, 5000 Córdoba, Argentina*

Received 17 February 2003; received in revised form 4 April 2003

Abstract

The Jensen–Shannon divergence is a symmetrized and smoothed version of the Kullback–Leibler divergence. Recently it has been widely applied to the analysis and characterization of symbolic sequences. In this paper we investigate a generalization of the Jensen–Shannon divergence. This generalization is done in the framework of the non-extensive Tsallis statistics. We study its basic properties and we investigate its applicability as a tool for segmentating symbolic sequences.

© 2003 Elsevier B.V. All rights reserved.

PACS: 87.15.Cc

Keywords: Jensen–Shannon divergence; Nonextensivity; Segmentation

1. Introduction

Lin introduced an information-theoretical based divergence measure between two or more probability distributions [1]. This measure is known as Jensen–Shannon divergence (JSD). Its main property is that, unlike others divergence measures, it does not require the condition of absolute continuity for the probability distributions involved. Recently the Jensen–Shannon divergence has been exhaustively studied in the context of the analysis of symbolic sequences; in particular it was successfully applied to the study of segmentation of DNA sequences [2–6] and in edge detection of digital images [7]. Furthermore, it has been shown that the JSD defines a true metric in the space

* Corresponding author. Fax: +54-351-481-8360.

E-mail address: lamberti@fis.uncor.edu (P.W. Lamberti).

of probability distributions—in fact it is the square of a metric [8]. In this work we investigate the properties of a generalized version of the Jensen–Shannon divergence in the framework of Tsallis generalized statistics [9]. This kind of generalizations has been investigated for other divergence measures [10–12]. Inspired in the theory of multifractals, Tsallis proposed to generalize the standard Boltzmann–Gibbs statistical mechanics by introducing a non-extensive entropy. Instead of the Boltzmann–Gibbs–Shannon entropy, he introduced a non-logarithmic entropy, which reads

$$H_q(p_1, \dots, p_W) = -k \frac{1 - \sum_{j=1}^W p_j^q}{1 - q} = k \sum_{j=1}^W p_j^q \frac{1 - p_j^{1-q}}{1 - q}; \quad \left(\sum_{j=1}^W p_j = 1 \right), \quad (1.1)$$

where q is a real number and p_j are the usual probabilities of events j ; k is a positive constant (which we will take equal to one). The standard Boltzmann–Gibbs–Shannon entropy is recovered in the limit $q \rightarrow 1$. Similar expression corresponds to the case where we have a continuum of microscopic configurations. The index q characterizes the degree of *non-extensivity* through the rule

$$H_q(A + B) = H_q(A) + H_q(B) + (1 - q) \frac{H_q(A)H_q(B)}{k}, \quad (1.2)$$

where A and B are two independent systems in the sense that the probabilities of $A + B$ factorize into those of A and of B . Within this formalism, the standard expectation values are replaced with the so-called *unnormalized q -expectation value* of the form

$$\langle \omega \rangle_q = \sum_j p_j^q \omega_j, \quad (1.3)$$

where ω is some state variable of the system [13].

One of the most beautiful result arising in the Tsallis statistics is the connection established between the Lévy like anomalous diffusion and non-extensivity [14]. Anomalous diffusion has been observed in many physical systems as well as in biological ones. For example some sectors of genomic DNA sequences are known to exhibit statistical properties analogous to anomalous diffusion [15]. In this context our program consists in investigating the applicability of the non-logarithmic JSD to the segmentation procedure in sequences in which an “anomalous diffusion” behavior is present. In this paper we investigate the properties of a “non-logarithmic like extension” of the JSD and we applied it to the segmentation procedure in finite simulated sequences.

The paper is organized as follows: in Section 2 we review the main properties of the standard Jensen–Shannon divergence. In Section 3 we define the generalized version of JSD and we study its statistical properties. In Section 4 we present some results arising in Monte Carlo simulations and we deduce the branching property for the non-logarithmic JSD. Finally, some conclusions are addressed in Section 5.

2. The standard JSD

Let us consider a discrete random variable X and let P_1 and P_2 be two probability distributions for X . The Kullback information gain (*Kullback–Leibler divergence*) is given by

$$K_1(P_1, P_2) = \sum_j p_{1j} \ln \frac{p_{1j}}{p_{2j}}, \quad (2.1)$$

where p_{ij} is the probability of occurrence of the value $X=x_j$ for each of the probability distribution P_i , $i=1,2$. This definition requires that P_1 must be absolutely continuous with respect to P_2 , that is $p_{1j}=0$ whenever $p_{2j}=0$. To overcome this restriction it is possible to define another divergence measure between the probabilities P_1 and P_2 by

$$I_1(P_1, P_2) = \sum_j p_{1j} \ln \frac{p_{1j}}{\frac{1}{2}p_{1j} + \frac{1}{2}p_{2j}}, \quad (2.2)$$

which in term of K_1 can be written as

$$I_1(P_1, P_2) = K_1(P_1, \frac{1}{2}P_1 + \frac{1}{2}P_2). \quad (2.3)$$

K_1 and I_1 are obviously not symmetric. A symmetric divergence measure based on I_1 can be defined by

$$L_1(P_1, P_2) = I_1(P_1, P_2) + I_1(P_2, P_1). \quad (2.4)$$

This divergence measure L_1 can be expressed in term of the Boltzmann–Gibbs–Shannon entropy $H_1(P) = -\sum_j p_j \log p_j$ in the form

$$L_1(P_1, P_2) = 2H_1\left(\frac{P_1 + P_2}{2}\right) - H_1(P_1) - H_1(P_2). \quad (2.5)$$

In obtaining this measure we have assumed that the probability distribution P_1 and P_2 have the same *weight* ($\frac{1}{2}$). Let $\pi_1, \pi_2 \geq 0$, $\pi_1 + \pi_2 = 1$ be arbitrary weights for the probability distributions P_1 and P_2 , respectively. A direct generalization of the divergence L_1 is

$$JS_1^{(\pi_1, \pi_2)}(P_1, P_2) = H_1(\pi_1 P_1 + \pi_2 P_2) - \pi_1 H_1(P_1) - \pi_2 H_1(P_2). \quad (2.6)$$

This divergence measure is called the Jensen–Shannon divergence. For an arbitrary set of probability distributions P_1, P_2, \dots, P_n , with weights $\pi_1, \pi_2, \dots, \pi_n$ ($\pi_i > 0$, $\sum_{i=1}^n \pi_i = 1$), the Jensen–Shannon divergence is defined by

$$JS_1^{(\pi_1, \dots, \pi_n)}(P_1, P_2, \dots, P_n) = H_1\left(\sum_{i=1}^n \pi_i P_i\right) - \sum_{i=1}^n \pi_i H_1(P_i). \quad (2.7)$$

Now we will go through the most relevant properties of the JS divergence. We will write them for two probability distributions, P_1 and P_2 , but the extension to an arbitrary number of distributions is immediate. As it was mentioned previously the

probability distributions P_i in (2.7) need not to be absolutely continuous with respect to each other. This is a crucial difference between JS_1 and the Kullback divergence measure. Furthermore, the weights π_j allow to assign a different weight to each probability distribution according to their importance. The obvious extension of (2.6) for a continuous random variable X is

$$JS_1^{(\pi_1, \pi_2)}(P_1, P_2) = \int dx [\pi_1 p_1(x) + \pi_2 p_2(x)] \ln [\pi_1 p_1(x) + \pi_2 p_2(x)] \\ - \int dx \pi_1 p_1(x) \ln p_1(x) - \int dx \pi_2 p_2(x) \ln p_2(x),$$

where P_1 and P_2 are two probability distributions for the variable X .

Since H_1 is a concave function, JS_1 is a positive definite quantity

$$JS_1(P_1, P_2) \geq 0 \quad (2.8)$$

and $JS_1(P_1, P_2) = 0$ if and only if $P_1 = P_2$ almost everywhere. Property (2.8) can be easily verified by using the Jensen inequality: If $\sum_j \alpha_j = 1$ and $f''(x) > 0$ then $\sum_j \alpha_j f(x_j) \geq f(\sum_j \alpha_j x_j)$. The symmetry of $JS_1(P_1, P_2)$ in its arguments P_1 and P_2 is clear.

3. Non-logarithmic JSD

Expression (1.1) can be interpreted as the unnormalized q -average of the generalized information content of the probability distribution $P = (p_1, \dots, p_w)$ given by $\sigma_j^q \equiv -(p_j^{1-q} - 1)/(1 - q)$. Thus for two probability distribution P_1 and P_2 , the change in information is given by

$$\Delta \sigma_j^q = \frac{(1 - p_{2j}^{1-q}) - (1 - p_{1j}^{1-q})}{1 - q}. \quad (3.1)$$

Taking the unnormalized q -average of $\Delta \sigma_j^q$ with respect to P_1 yields

$$K_q(P_1, P_2) = \sum_j \frac{p_{1j}^q (p_{1j}^{1-q} - p_{2j}^{1-q})}{1 - q}, \quad (3.2)$$

which is a non-logarithmic generalized Kullback divergence measure. In term of the q -logarithmic function, $\ln_q(p) = (p^{1-q} - 1)/(1 - q)$, definition (3.2) can be written as

$$K_q(P_1, P_2) = - \sum_j p_{1j} \ln_q \frac{p_{2j}}{p_{1j}}. \quad (3.3)$$

It can be shown that $K_q \geq 0$ for $q > 0$ and $K_q = 0$ for $q = 0^{10}$. In this context the direct generalization of L_1 for $q \neq 1$ is

$$L_q(P_1, P_2) = K_q \left(P_1, \frac{P_1 + P_2}{2} \right) + K_q \left(P_2, \frac{P_1 + P_2}{2} \right). \quad (3.4)$$

Obviously, $L_q \geq 0$ for $q > 0$ and $L_q = 0$ for $q = 0$. After a little algebra, we can rewrite L_q in the form

$$\begin{aligned} L_q(P_1, P_2) = & - \sum_j (p_{1j}^q + p_{2j}^q) \ln_q \left(\frac{p_{1j} + p_{2j}}{2} \right) \\ & + \sum_j p_{1j}^q \ln_q p_{1j} + \sum_j p_{2j}^q \ln_q p_{2j} \end{aligned} \quad (3.5)$$

or equivalently

$$\begin{aligned} L_q(P_1, P_2) = & 2 \left\{ - \sum_j \left(\frac{1}{2} p_{1j}^q + \frac{1}{2} p_{2j}^q \right) \ln_q \left(\frac{p_{1j} + p_{2j}}{2} \right) \right. \\ & \left. + \frac{1}{2} \sum_j p_{1j}^q \ln_q p_{1j} + \frac{1}{2} \sum_j p_{2j}^q \ln_q p_{2j} \right\}. \end{aligned} \quad (3.6)$$

This last equation reveals the fact that we have assumed in obtaining expression (3.5) equal weights for the probability distributions P_1 and P_2 ($\frac{1}{2}$). Thus the natural generalization of (3.6) for arbitrary weights π_1 and π_2 is

$$\begin{aligned} JS_q^{(\pi_1, \pi_2)}(P_1, P_2) = & - \sum_j (\pi_1 p_{1j}^q + \pi_2 p_{2j}^q) \\ & \times \ln_q (\pi_1 p_{1j} + \pi_2 p_{2j}) + \pi_1 \sum_j p_{1j}^q \ln_q p_{1j} \\ & + \pi_2 \sum_j p_{2j}^q \ln_q p_{2j}. \end{aligned} \quad (3.7)$$

This is our definition of the non-logarithmic Jensen–Shannon divergence (NLJSD). The limit $q \rightarrow 1$ recovers JS_1 . The extension of definition (3.7) to an arbitrary number of probability distributions P_1, \dots, P_n is

$$\begin{aligned} JS_q^{(\pi_1, \dots, \pi_n)}(P_1, \dots, P_n) = & - \sum_j (\pi_1 p_{1j}^q + \dots + \pi_n p_{nj}^q) \\ & \times \ln_q (\pi_1 p_{1j} + \dots + \pi_n p_{nj}) + \pi_1 \sum_j p_{1j}^q \ln_q p_{1j} + \dots \\ & + \pi_n \sum_j p_{nj}^q \ln_q p_{nj}. \end{aligned} \quad (3.8)$$

All the properties mentioned above for the standard case, are valid for the non-logarithmic ones. It is direct to verify that $JS_q \geq 0$ for $q > 0$ and $JS_q = 0$ for $q = 0$. For $q > 0$ we have $JS_q = 0$ if and only if $P_1 = P_2$.

It should be noted that our generalization of the JSD does not consist in replacing H_1 by H_q in expression (2.6). Incidentally, by using the Jensen inequality it is possible

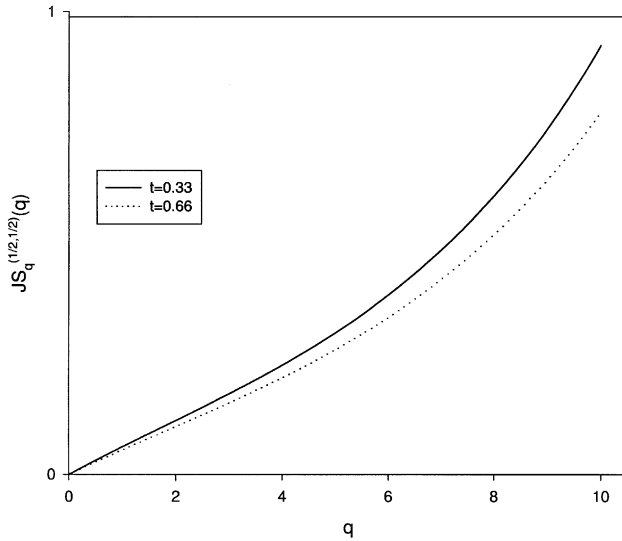


Fig. 1. Plot of the non-logarithmic JSD as a function of q . Here the probability distributions for a binary alphabet are $P_1 = (t, 1 - t)$ and $P_2 = (1 - t, t)$.

to show that

$$JS_q^{(\pi_1, \pi_2)}(P_1, P_2) > H_q(\pi_1 P_1 + \pi_2 P_2) - \pi_1 H_q(P_1) - \pi_2 H_q(P_2) \quad (3.9)$$

for $q > 1$. Thus, identity (2.6) does not satisfy anymore for $JS_q^{(\pi_1, \pi_2)}$. Besides this fact the divergence here introduced share with JS_1 its fundamental statistical properties. For example in some particular context JS_q can be interpreted as the (non-extensive) mutual information (see Ref. [2, Section III.B] for details).

As a function of q , the divergence JS_q is a monotonally increasing one. To illustrate this fact, Fig. 1 shows a plot of JS_q versus q . In this case we assume a binary alphabet and the probability distributions $P_1 = (t, 1 - t)$ and $P_2 = (1 - t, t)$, $0 \leq t \leq 1$.

4. NLJSD in the context of sequences analysis

4.1. Segmenting sequences

Let $\mathcal{A} = \{a_1, \dots, a_N\}$ an alphabet of N symbols and let $S = a_{n_1} a_{n_2} \dots a_{n_L}$ be a sequence of characters from \mathcal{A} which we think as partitioned in two subsequences S_1 and S_2 each one built up from different probability distributions. Let us denote by α_i (β_i) the probability of occurrence of symbol a_i in S_1 (S_2). The lengths of S_1 and S_2 are L_1 and L_2 , respectively. Throughout this section we will assume the weights $\pi_1 = L_1/L$ and $\pi_2 = L_2/L$ for the probabilities α_i and β_i , respectively. In this way, for the sequence S ,

we can evaluate the NLJSD as

$$\begin{aligned}
 JS_q^{(\pi_1, \pi_2)}(\mathcal{A}, S_1, S_2) = & - \sum_{j=1}^N (\pi_1 \alpha_j^q + \pi_2 \beta_j^q) \ln_q (\pi_1 \alpha_j + \pi_2 \beta_j) \\
 & + \pi_1 \sum_{j=1}^N \alpha_j^q \ln_q \alpha_j + \pi_2 \sum_{j=1}^N \beta_j^q \ln_q \beta_j .
 \end{aligned} \quad (4.1)$$

The extension to an arbitrary number of subsequences is obvious.

For a given (non-stationary) sequence which is divided in two disjoint subsequences, the quantities α_i and β_i are not directly observables. The observables are the frequencies of occurrence of symbols a_i , that is, the number of symbols a_i in subsequence S_k ($k = 1, 2$), $N_i^{(k)}$ over the length L_k :

$$f_i^{(k)} = \frac{N_i^{(k)}}{L_k} .$$

Thus in evaluating the divergence JS_q for a real sequence we first must calculate the frequencies $f_i^{(k)}$.

In practice the segmentation procedure consists in separating a non-stationary sequence in two or more stationary subsequences. In Ref. [2] a recursive segmentation algorithm based on the JSD has been developed. We refer the reader there for details. We just mention that in this algorithm the maximum value of the JSD sampled along the sequence plays a central role. To study the behavior of this quantity for the NLJSD, we did some Monte Carlo simulations. In this point we follow Grosse et al. [2]. We evaluated the average of $JS_q^{(\pi_1, \pi_2)}$ over an ensemble of 2000 binary sequences of length $L = 1500$. Each sequence is made up by joining two subsequences of length $L_1 = 500$ and $L_2 = L - L_1 = 1000$. The subsequence of length L_1 is generated from a probability distribution $(t, 1 - t)$ and the other one from a probability distribution $(s, 1 - s)$. The results obtained are presented in Fig. 2. Here we put $t = 0.33$ and $s = 0.66$. In this graph we have plotted $\langle JS_q^{(\pi_1, \pi_2)} \rangle$ versus the cursor position along the sequence. We observe that for different values of q a global maximum in $\langle JS_q^{(\pi_1, \pi_2)} \rangle$ occurs at the fusion point. It should be also noted that the global maximum increases as q increases. This behavior is made evident in the inset of Fig. 2 where we plotted the maximum value of $\langle JS_q^{(\pi_1, \pi_2)} \rangle$ that takes place at the fusion point, $\langle JS_q^{(\pi_1, \pi_2)} \rangle^{\max}$, as a function of q for certain range of q . This two facts suggest that $JS_q^{(\pi_1, \pi_2)}$ might be a more adequate tool, than the standard one, to segment non-stationary sequences into stationary ones. The most adequate value of q will depend on the particular problem under study.

In every recursive procedure it is necessary to define a criterion to stop the process. In the above-mentioned segmentation algorithm the process is halted when no significant difference between potential subsequences are found. To quantify the “significant difference” a quantity called the *significance value* is introduced. For an observed value x of the JSD, the significance value is defined as the probability of obtaining this value or a lower value under the hypothesis that all subsequences are generated from the same probability distribution:

$$s(x) \equiv \text{Prob}\{JS \leq x\} .$$

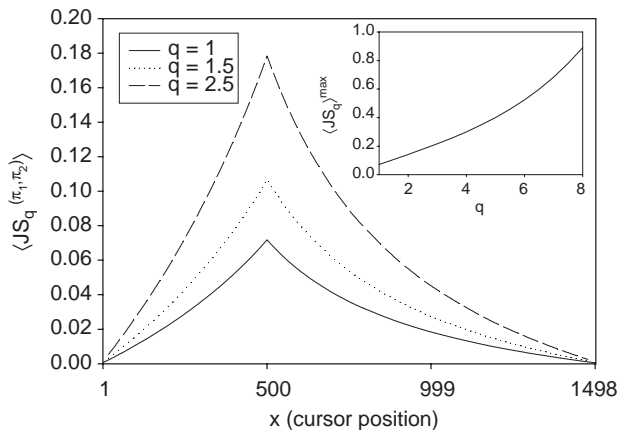


Fig. 2. Graphs of the non-logarithmic JSD for different values of q as a function of the cursor position along simulated sequences. The inset shows the maximum value of $\langle JS_q \rangle$ as a function of q (see text for details).

In this context the significance value works as a threshold value. For a given significance value, let us say s_0 , the segmentation process stops when none of the possible cutting points (i.e., where the JSD reaches a maximum) has a significance value greater than s_0 . In that case it is said that the sequence is segmented at significance threshold s_0 .

For the NLJSD, and after some algebra, it is possible to estimate

$$s_q(x) \sim \frac{\gamma(v/2, (L/q)x)}{\Gamma(v/2)}, \quad (4.2)$$

where $v = (N-1)/(m-1)$ and m is the number of subsequences. $\gamma(a, x)$ and $\Gamma(a)$ are the incomplete and complete gamma function, respectively. For $q=1$ the last expression reduces to those obtained for the standard JSD [2].

4.2. Branching property

Some statistical properties of a sequence are strongly dependent on the alphabet used to describe the sequence. This fact is particularly relevant when a measure of compositional complexity of a sequence is investigated [5].

Let Γ_j , $j = 1 \dots r$, $r \leq N$ be a partition of the alphabet \mathcal{A} . To each member of this partition we assign a character γ_j (in general from another alphabet). Let us denote by $\Gamma = \{\gamma_1, \dots, \gamma_r\}$ this new alphabet (grouped alphabet). For a given sequence S of characters of \mathcal{A} , we can construct a new sequence of characters belonging to Γ by substituting each symbol a_i by γ_j , such that $a_i \in \Gamma_j$. The branching property of the Tsallis entropy [13] allows us to relate the entropies evaluated by considering different alphabets. This property is written in the form

$$H_q(S, \mathcal{A}) = H_q(S, \Gamma) + \sum_{j=1}^r \Theta_j^q H_q(S, \Gamma_j), \quad (4.3)$$

where $\Theta_j = N_j/L$ and N_j is the number of symbols of S belonging to Γ_j . Incidentally we note that property (4.3) generalizes the branching property for the Boltzmann–Gibbs–Shannon entropy [16].

From definition (2.7) and property (4.3) for $q = 1$, it is possible to derive the branching property for the SJSD, which reads as

$$JS_1^{(\pi_1, \pi_2)}(\mathcal{A}) = JS_1^{(\pi_1, \pi_2)}(\Gamma) + \sum_j \Theta_j JS_1^{(\tilde{\pi}_{1j}, \tilde{\pi}_{2j})}(\Gamma_j), \quad (4.4)$$

where we have explicitly written the alphabet used in evaluating each term and where $\tilde{\pi}_{1j} = \pi_1 \theta_j^{(\alpha)} / \Theta_j$, $\tilde{\pi}_{2j} = \pi_2 \theta_j^{(\beta)} / \Theta_j$, $\theta_j^{(\alpha)} = N_j^{(\alpha)} / L_1$ and $\theta_j^{(\beta)} = N_j^{(\beta)} / L_2$ with $N_j^{(\alpha)}$ ($N_j^{(\beta)}$) the number of symbols belonging to Γ_j in subsequence S_1 (S_2). $JS_1^{(\tilde{\pi}_{1j}, \tilde{\pi}_{2j})}(\Gamma_j)$ stands for the divergence evaluated by considering only the symbols belonging to the alphabet Γ_j . Expression (4.4) gives the divergence between two subsequences in terms of partial divergences between the same two subsequences associated with the grouped alphabet Γ and the partial ones Γ_j .

The deduction of the corresponding branching property for the NLJSD is much more involved. After some algebra we arrive to

$$\begin{aligned} JS_q^{(\pi_1, \pi_2)}(\mathcal{A}) = & JS_q^{(\pi_1, \pi_2)}(\Gamma) + \sum_j \Theta_j \frac{[(\theta_j^{(\alpha)})^{q-1} + (\theta_j^{(\beta)})^{q-1}]}{2} JS_q^{(\tilde{\pi}_{1j}, \tilde{\pi}_{2j})}(\Gamma_j) \\ & + \frac{(1-q)}{2} \left\{ \sum_j \Theta_j \left[\tilde{\pi}_{1j} (\theta_j^{(\alpha)})^{q-1} K_q(\alpha_i^{(j)}, p_i^{(j)}) \right. \right. \\ & \times \left. \left\{ 2 \ln_q \theta_j - \ln_q \frac{\theta_j^{(\alpha)}}{\theta_j^{(\beta)}} \right\} \right. \\ & \left. \left. + \tilde{\pi}_{2j} (\theta_j^{(\beta)})^{q-1} K_q(\beta_i^{(j)}, p_i^{(j)}) \left\{ 2 \ln_q \theta_j - \ln_q \frac{\theta_j^{(\beta)}}{\theta_j^{(\alpha)}} \right\} \right] \right\}, \quad (4.5) \end{aligned}$$

where $\alpha_i^{(j)}$ and $\beta_i^{(j)}$ are the probabilities of occurrence of symbol a_i in subsequence S_1 and S_2 , respectively, just considering the alphabet Γ_j and $p_i^{(j)}$ is the probability of occurrence of the symbol a_i along all the sequence S considering only symbols belonging to Γ_j . K_q is the Kullback divergence given by (3.2). For $q = 1$ we reobtain identity (4.4). The third term in (4.5) have to do with the non-extensive character of the formalism. It should be noted that this term only involve the alphabets Γ and Γ_j .

5. Conclusions

The Jensen–Shannon divergence has shown to be an excellent statistical tools in segmentation procedures. Furthermore this type of divergence measure appears suitable for a metric in the functional space of probability distributions. In this paper we have studied the basic properties of a non-logarithmic version of the Jensen–Shannon divergence, which we propose as a potentially more adequate divergence measure between

two or more probability distributions. Our expectation is that, for every specific use, better discrimination in the segmentation procedure, will be achieved with appropriate ranges of values of q . This is the case in several applications of Tsallis statistics, from the study of turbulent flow [17] to the study of electroencephalographic recordings [18]. At the present the study of some applications of the non-logarithmic JSD is in progress. Several interesting effects emerge when we study the NLJSD as a function of q which will be presented in a future paper.

Acknowledgements

We acknowledge valuable discussions with D. Prato and A. Plastino and we are grateful to Secretaria de Ciencia y Tecnica de la Universidad Nacional de Córdoba for financial assistance. PWL is a member of CONICET and AM is a fellowship holder of SECYT-UNC. This work was partially supported by Grant BIO2002-04014-C03-03 from Spanish Government.

References

- [1] J. Lin, *IEEE Trans. Inform. Theory* 37 (1) (1991) 145.
- [2] I. Grosse, P. Bernaola-Galvan, P. Carpena, R. Román-Roldán, J. Oliver, H.E. Stanley, *Phys. Rev. E* 65 (2002) 041905.
- [3] P. Bernaola-Galvan, R. Román-Roldán, J.L. Oliver, *Phys. Rev. E* 53 (1996) 5181.
- [4] R.K. Azad, P. Bernaola-Galvan, R. Ramaswamy, J.S. Rao, *Phys. Rev. E* 65 (2002) 051909.
- [5] P. Bernaola-Galvan, J.L. Oliver, R. Román-Roldán, *Phys. Rev. Lett.* 83 (1999) 3336.
- [6] P. Bernaola-Galvan, I. Grosse, P. Carpena, J.L. Oliver, R. Román-Roldán, H.E. Stanley, *Phys. Rev. Lett.* 85 (2000) 1342.
- [7] J.F. Gómez-Lopera, J. Martínez-Aroza, A.M. Robles-Pérez, R. Román-Roldán, *J. Math. Imaging Vision* 13 (2000) 35.
- [8] F. Topsøe, Preprint University of Copenhagen, 2002.
- [9] C. Tsallis, *J. Stat. Phys.* 52 (1988) 479.
- [10] C. Tsallis, *Phys. Rev. E* 58 (1998) 1442.
- [11] P. Graben, *Phys. Rev. E* 64 (2001) 051104.
- [12] R.S. Johal, *Phys. Rev. E* 58 (1998) 4147.
- [13] S. Abe, Y. Okamoto (Eds.), *Nonextensive Statistical Mechanics and Its Applications*, Springer, Berlin, 2001.
- [14] C. Tsallis, S.V.F. Levy, A.M.C. Souza, R. Maynard, *Phys. Rev. Lett.* 27 (1995) 3589.
- [15] S. Buldyrev, A. Goldberger, S. Havlin, Chung-Kang Peng, M. Simons, H.E. Stanley, *Phys. Rev. E* 47 (1993) 4514.
- [16] T. Cover, J. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.
- [17] C. Beck, *Phys. Rev. Lett.* 87 (2001) 180601.
- [18] A. Capurro, L. Diambra, D. Lorenzo, O. Macadar, M.T. Martín, C. Mostaccio, A. Plastino, E. Rofman, M.E. Torres, J. Velutti, *Physica A* 257 (1998) 149.