Jensen-Shannon Divergence: A Multipurpose Distance for Statistical and Quantum Mechanics

Pedro W. Lamberti, Ana P. Majtey, Marcos Madrid and María E. Pereyra

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

Abstract. Many problems of statistical and quantum mechanics can be established in terms of a distance; in the first case the distance is usually defined between probability distributions; in the second one, between quantum states. The present work is devoted to review the main properties of a distance known as the Jensen-Shannon divergence (JSD) in its classical and quantum version. We present two examples of application of this distance: in the first one we use it as a quantifiers of the stochastic resonance phenomenon in ion channels; in the second one we use the JSD to propose a geometrical view of entanglement for two qubits states.

Keywords: Metrics for probability spaces; Distances in Quantum Theory; Jensen-Shannon Divergence
PACS: 02.50.-r - 03.65.-w, 89.70.+c

INTRODUCTION

Geometry and physics have lived together in close association for centuries. The relationship has had its ups and downs, but for the last century it has blossomed in a remarkable manner. This is a particularly noticeable after the formulation of the General Relativity theory by A. Einstein and of the Quantum Mechanics by Schrödinger, Dirac, Heisenberg et al. In the first case, the theory is established in terms of a four dimensional manifold, the space-time, with geometrical properties given by the Universe's matter-energy distribution; in the second one the states of a physical system are represented as vectors in a Hilbert space [1, 2].

As was pointed out by Gauss and Riemann in their pioneer works, one of the basic geometrical notion is the concept of distance between points of a given space (differentiable manifold, in modern terminology). According to their conceptions, the distance function (metric) determines all the geometrical properties of the space.

To be precise, let us recall that a distance over an arbitrary set X (a topological space, a Hilbert space, etc.), is a non-negative function D from the cartesian product $X \times X$ into the real numbers set that verifies:

$$D(x,y) = 0 \; iff \; x = y \tag{1}$$
$$D(x,y) = D(y,x) \; \forall \; x,y \in X \; (symmetric) \tag{2}$$

If the function D verifies the *triangle inequality*

$$D(x,z) \leq D(x,y) + D(y,z), \forall x,y,z \in X \tag{3}$$

it is said that D is a metric for the space X.

When distances between (discrete) probability distributions are investigated, the set X is given by a subset of $\Re^N$:

$$X = M_+^1(N) \equiv \{(p_1, \ldots, p_N), p_i \in \Re, p_i \geq 0, \sum_i p_i = 1\}.$$

In the probability theory context, several distances between two probability distributions are found in the literature. Even though the most frequently used are the Minkowskian metrics

$$d_\nu(P,P') \equiv \left(\sum_j |p_j - p'_j|^\nu \right)^{\frac{1}{\nu}} \quad \nu \geq 1 \tag{4}$$

CP913, *Nonequilibrium Statistical Mechanics and Nonlinear Physics, XV Conference*
edited by O. Descalzi, O. A. Rosso, and H. A. Larrondo

several important results are expressed in terms of other metrics [3]. In the expression (4) p_j and p'_j are the event j's probabilities of occurrence for each of the probability distributions P and P', respectively. The case $\nu = 2$ corresponds to the euclidean one.

A frequently used quantity to compare two probability distributions, arose in information theory, is the Kullback-Leibler divergence [4]:

$$K(P,P') \equiv \sum_j p_j \log\left(\frac{p_j}{p'_j}\right), \tag{5}$$

which does not verify properties (2) and (3) of the metric definition. More important even, this "distance" is not always well defined. Throughout this paper, log represents the logarithm in basis two.

Another example is provided by the Wootters's distance, which has the expression:

$$d_W(P,P') = \arccos\left(\sum_j \sqrt{p_j p'_j}\right) \tag{6}$$

This distance was introduced by Wootters in 1981 in the context of studying the problem of distinguishability between quantum states [5]. Roughly speaking, eq. (6) gives the greatest number of distinguishable probability distributions between P and P'. The Wootters's distance is symmetric, but it does not verify the triangle inequality (3).

In spite of these previous examples are very relevant, they do not exhaust the list of all the distances introduced in statistical and probabilistic studies (for a more complete list, see for example, reference [6]). At this point a question naturally emerges: why do we need distances? In answering this question, it should be observed that in several problems of statistical and quantum mechanics we need to compare "objects" (for example symbolic sequences, signals or classical and quantum states.).

Applications of distances range from the development of error detecting and error correcting codes [7], hypothesis testing in statistics [8], and philogenetic studies [9, 10], on the one hand, and to compare the amount of information carried by quantum states [11], on the other. In the realm of statistical physics many relevant problems are stated in terms of distances between probability distributions: to set up the notion of stability [12] and to define a complexity measure [13], just to mention two particularly interesting examples. Several outstanding results arose in classical and quantum information theory can be established in terms of distances too [14], [15].

During last years our research interest was focused in investigating an alternative to the previously mentioned distances: the Jensen Shannon divergence (JSD). Recently this quantity has been exhaustively applied in several problems of interest in theoretical physics and related topics.

Next section is devoted to review the main properties of the classical and quantum versions of this distance.

JENSEN-SHANNON DIVERGENCE

Classical JSD

The JSD was introduced in 1987 by C. Rao and generalized by J. Lin in 1991 as an alternative to the Kullback-Leibler divergence [16, 17]. To overcome the already mentioned limitations of this divergence, they proposed the quantity

$$\begin{aligned} JS(P,P') &\equiv \frac{1}{2}\left[K(P,\frac{P+P'}{2}) + K(P',\frac{P+P'}{2})\right] \\ &= -\frac{1}{2}\left[\sum_j p_j \log\left(\frac{2p_j}{p_j+p'_j}\right) + \sum_j p'_j \log\left(\frac{2p'_j}{p_j+p'_j}\right)\right], \end{aligned} \tag{7}$$

to evaluate the distance between the probability distributions P and P'. In terms of the Shannon entropy $H[P] = -\sum_j p_j \log p_j$, expression (7) can be rewritten in the form:

$$JS(P,P') = H\left(\frac{P+P'}{2}\right) - \frac{1}{2}H(P) - \frac{1}{2}H(P') \tag{8}$$

By using logarithm in basis two, this quantity has units of bits.

The JSD is positive definite, symmetric, bounded and, unlike the Kullback-Leibler divergence, always well defined (in the sense that it can be evaluated even when $p_j \neq 0$ and $p'_j = 0$ for some j). More important yet, its square root verifies the triangle inequality (3) (but JS does not!!). Therefore, it can be concluded that $\sqrt{JS(P,P')}$ is a true metric for the probability distribution space [18].

The JSD can be interpreted in the frame of information theory [19] and in a Bayesian probabilistic sense [17]. Incidentally, it is worth mentioning that for P close to P' the JSD coincides (up to numerical factors) with the Wootters and the Fubini-Study distances and, for the continuous case, is related to the Fisher information measure [22].

Lin has proposed a generalization of (7) to several probability distributions [17]. Let $P^{(k)}$, $k = 1..M$, a set of M probability distributions. The JSD among these probabilities distributions is given by:

$$JS[P^{(1)},..,P^{(M)}] = H[\sum_k \pi^{(k)} P^{(k)}] - \sum_k^M \pi^{(k)} H[P^{(k)}], \tag{9}$$

where the numbers $\pi^{(k)}$, $k = 1..M$, $\sum_k \pi^{(k)} = 1$, are weights properly chosen that allow to assign different relevance to each probability distribution.

The JSD has been extensively applied in different contexts, for example, as a tool for segmenting symbolic sequences [19, 20], for analyzing literary texts and musical score [21], as the disequilibrium for defining a non-triviality measure for chaotic maps [23], to study EEG records [24], as a complexity measure for genomic sequences [25], to compare electron distributions in atoms [26], to define an alternative distinguishability criterium between pure quantum states [22] and as a measure of discernibility between wave functions [27].

Quantum JSD

Distances between quantum states play a central role in quantum information theory (see, for example, chapter 9 of reference [11]). Trace distance, fidelity and the Bures's metric are frequently used as measures of distance between quantum states. Let us recall that a quantum state is mathematically represented by a density operator acting over the Hilbert space of the system, that is, by an hermitian operator, positive definite and with trace equal to one .

An extensively referenced distance between quantum states, is the relative entropy, which is the natural generalization of the Kullback-Leibler divergence, eq. (5), to the context of quantum theory [28, 29]:

$$K(\rho\|\sigma) = \mathrm{Tr}(\rho \log \rho - \rho \log \sigma) \tag{10}$$

with ρ and σ density operators, and where Tr stands for the trace operator. This measure has several inconveniences: it is not symmetric, not bounded and not always well defined (for example, $K(\rho\|\sigma)$ is not defined when σ represents a pure state). To be precise it should be said that in order that $K(\rho\|\sigma) < \infty$ it is necessary that $ker(\sigma) \subset ker(\rho)$, where ker denotes the kernel of the operator (the kernel of an operator is the subspace spanned by eigenvectors corresponding to null eigenvalues).

Following the steps that lead to the definition of the JSD from the Kullback-Leibler divergence (first line of eq. (7)), we can introduce the quantum JSD in the form [30]:

$$JS(\rho\|\sigma) \equiv \frac{1}{2}\left[K(\rho\|\frac{\rho+\sigma}{2}) + K(\sigma\|\frac{\rho+\sigma}{2})\right] \tag{11}$$

which, in terms of the von-Neumann entropy, $H_N(\rho) = -\mathrm{Tr}(\rho \log \rho)$, can be rewritten in the form

$$JS(\rho\|\sigma) = H_N\left(\frac{\rho+\sigma}{2}\right) - \frac{1}{2}H_N(\rho) - \frac{1}{2}H_N(\sigma) \tag{12}$$

This measure is positive definite, symmetric, bounded, and always well defined; in particular the condition over the kernels of the operators ρ and σ can be removed. It has all the adequate dynamical properties for a distance between quantum states . When the quantum JSD is generalized as a distance among several quantum states, according to the recipe (9), the resulting quantity has a very noticeable interpretation: it gives the upper bound of the mutual information that can be gathered between two quantum computing operators (Holevo bound) [30].

TWO EXAMPLES OF APPLICATION OF THE JSD

To conclude this paper, we present two applications of the JSD; in the first one we use the classical JSD as a quantifier of the stochastic resonance phenomenon in ion channels; in the second one we use the quantum JSD to give a geometrical characterization of entanglement for two qubits states.

Example 1

Stochastic resonance (SR) characterizes a cooperative phenomenon wherein the addition of a small amount of noise to an input driving signal can optimally amplify the output response. Generically this phenomenon has three basic ingredients: i) a nonlinear stochastic classical or quantum system; ii) certain kind of meta-stability, such as a potential barrier or a fixed threshold; and iii) a source of noise [31]. SR effect plays a prominent role in such diverse scientific areas as sensory biology, signal information and detection, or in conventional physical and chemical nonlinear noisy systems. Several quantifiers of this phenomenon have been proposed [32, 33] including some arose in information theory [34, 35].

In a very remarkable experiment, Bezrukov and Vodyanoy observed the SR phenomenon in (artificial) alameticyn -voltage gated- ion channels [36]. Motivated in this experimental result Goychuk and Hänggi have advanced in a characterization of the SR in ion channels by means of some information theory originated quantities [34]. According to their approach a ion channel plays the role of a transducing device that transmits an input information, a voltage, to the signal-modulated ion current response: $V \mapsto I$. The input is a time-dependent voltage which has three components:

$$V(t) = V_0 + V_s(t) + V_n(t),$$

where V_0 is a constant bias voltage; $V_s(t)$ is some time-dependent unbiased signal and $V_n(t)$ represents a stationary Gaussian Markovian noise of amplitude σ. Here the signal voltage is assumed of the form $V_s(t) = A\cos(\Omega t),\ A \ll V_0$ (weak signal).

The gating dynamics of a single ion channel is governed by the equations verified by the time-dependent probabilities for the channel to be open or closed, $P_o(t)$ and $P_c(t)$, respectively. On his part, the probabilities $P_o(t)$ and $P_c(t)$ can be expressed in terms of the joint multi-time probability distributions $Q_s^{c,(o)}(t, \tau_s, ...\tau_1)$ for switches between open and closed states to occur at time $\tau_1, \tau_2, ..., \tau_s$ and to end up a time t in either the open state "o" or closed state "c", respectively (see reference [34] for a complete description of these probability distributions):

$$P_o(t) = Q_0^o(t) + \sum_{s=1}^{\infty} \int_0^t d\tau_s \int_0^{\tau_s} d\tau_{s-1} ... \times \int_0^{\tau_2} d\tau_1 Q_s^o(t, \tau_s, ...\tau_1) \tag{13}$$

and a similar expression is valid for the probability of the closed conformation, $P_c(t)$.

The basic idea in quantifying the SR phenomenon by means of the JSD, is to compare the corresponding probability distributions when the signal $V_s(t)$ is present and absent (an upper index "(0)" in a quantity indicates that no voltage signal is applied):

$$JS(I|V_s;T) = \sum_{\alpha=o,c} JS^{\alpha}(I|V_s;T) \tag{14}$$

where

$$JS^{\alpha}(I|V_s;T) = \frac{1}{2}\left[K(Q^{\alpha}, \frac{Q^{\alpha}+Q^{(0)\alpha}}{2}) + K(Q^{(0)\alpha}, \frac{Q^{\alpha}+Q^{(0)\alpha}}{2}\right] \tag{15}$$

with

$$\begin{aligned} K(Q^{\alpha}, \frac{Q^{\alpha}+Q^{(0)\alpha}}{2}) &= Q_0^{\alpha} \log \frac{2Q_0^{\alpha}}{Q_0^{\alpha}+Q_0^{(0)\alpha}} + \sum_{s=1}^{\infty} \int_0^T d\tau_s \int_0^{\tau_s} d\tau_{s-1} \times ... \\ &\times \int_0^{\tau_2} d\tau_1 Q_s^{\alpha}(T, \tau_s, ..., \tau_1) \log \frac{2Q_s^{\alpha}(T, \tau_s, ...\tau_1)}{Q_s^{\alpha}(T, \tau_s, ...\tau_1) + Q_s^{(0)\alpha}(T, \tau_s, ..., \tau_1)} \end{aligned} \tag{16}$$

and an analogous expression for the second term in (15). All these quantities can be evaluated up to second order in A. In figure 1 we plotted the average over a period ($\frac{2\pi}{\Omega}$) of the JSD for different values of the bias voltage. All the

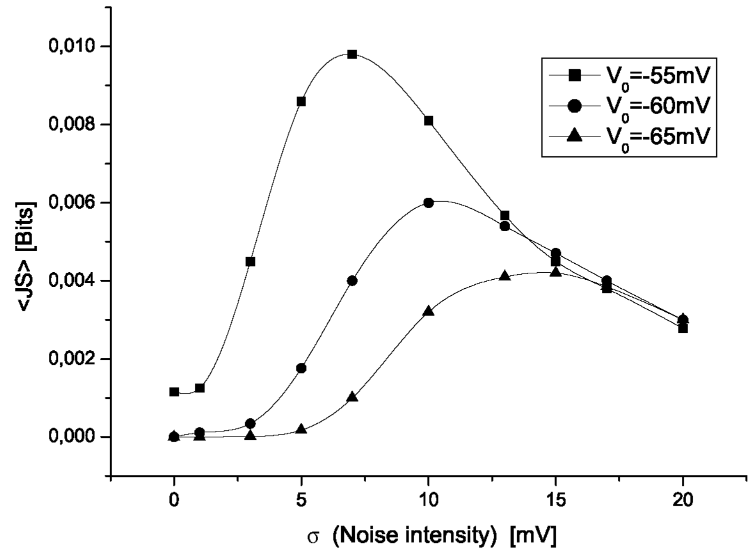

FIGURE 1. Time averaged JSD as a function of noise intensity for different bias voltages.

data required to do this plot has been taken from reference [34], and correspond to a potassium selective *Shaker IR* ion channel. The main result here is that the JSD reveal the typical SR dependence on the noise intensity. By comparing our results with those obtained by Goychuk and Hänggi, we can conclude that the JSD behaves, in this example, like the rate of information gain for the transduction process.

Example 2

Entanglement is the basic resource required to implement quantum information processes [11]. One of the most active issues in quantum information theory is the search of measures that allow to quantify the degree of entanglement of a system [37]. As shown by Vedral and collaborators, a measure of entanglement for bipartite systems can be built from a distance defined over the corresponding Hilbert space, whenever it verifies certain conditions [15]. Remarkably, the quantum JSD verifies these conditions [30]. In order to explore the JSD as a quantifier of entanglement for bipartite systems, we will evaluate it between a Werner state and a pure Bell state. These are states for a system of two 1/2-spin particles.

The Bell basis is expressed in terms of the canonical basis in the form: $|\Phi^{\pm}\rangle = \frac{1}{\sqrt{2}}[|++\rangle \pm |--\rangle]$ and $|\Psi^{\pm}\rangle = \frac{1}{\sqrt{2}}[|+-\rangle \pm |-+\rangle]$. Let us consider the density operator ρ_W corresponding to a Werner state:

$$\rho_W = F|\Psi^-\rangle\langle\Psi^-| + \frac{1-F}{3}\left(|\Psi^+\rangle\langle\Psi^+| + |\Phi^+\rangle\langle\Phi^+| + |\Phi^-\rangle\langle\Phi^-|\right)$$

where F is the fidelity of ρ_W with respect to the pure state $\sigma = |\Psi^-\rangle\langle\Psi^-|$. The Werner state is separable (unentangled) if $F \leq 1/2$; otherwise it is entangled.

The JSD between ρ_W and σ is:

$$d_{JS}(F) \equiv JS[\rho_W \| \sigma] = \frac{1}{2}\left[F\log F - (1+F)\log(\frac{1+F}{2})\right] + \frac{1-F}{2} \tag{17}$$

The JSD is a monotonically decreasing function of the fidelity F. Therefore, it is possible to invert the relationship (17) and express F as a function of the distance d_{JS}:

$$F = F(d_{JS}) \tag{18}$$

Thus the condition for the Werner state ρ_W of being entangled can be rewritten in terms of the distance d_{JS} in the form:

$$0 \leq d_{JS} \leq -\frac{3}{4}\log(\frac{3}{4}) = 0.311.$$

We can interpret this last inequality, from a geometrical point of view, as defining an hyper-sphere centered around the state σ in the 2-qubits space [38]. Outside this sphere only separable Werner states are found. A more detailed study of this geometrical view will be presented in a future work.

ACKNOWLEDGMENTS

The authors are grateful to Secretaria de Ciencia y Técnica de la Universidad Nacional de Córdoba, Argentina, for financial assistance. AM, MP are fellowship holders of CONICET and PWL is a member of CONICET. This work was partially supported by a Grant from CONICET. PWL is very grateful to the organizers of Medyfinol 2006 for their warm hospitality during his stay in Mar del Plata.

REFERENCES

1. R.M. Wald, *General Relativity*, The University of Chicago Press, Chicago, 1984.
2. P.A.M. Dirac, *Quantum Mechanics*, Oxford University Press, 1958.
3. A.N. Shyriaev, *Probability Theory*, Springer, 1996.
4. S. Kullback and R.A. Leibler, *Ann. Math. Stat.* **22**, 79-86 (1951).
5. W. Wootters, *Phys. Rev D* **23**, 357-362 (1981).
6. A. Majtey, PhD Thesis, Universidad Nacional de Córdoba, Argentina (in spanish).
7. R.V. Hamming, *Bell Sys. Tech. Journal* **29**, 147-160 (1950).
8. P.M. Robinson, *Rev. Econ. Stud* **88**, 437-453 (1991).
9. M. Li, J.H. Badger, X. Chen, S. Kwong, P. Kearney and H. Zhang, *Bioinformatics* **17**, 149-154 (2001).
10. M. Li, X. Chen, X. Li, B. Ma and P. M. N. Vitányi, *IEEE Trans. Inf. Theory* **50**, 3250-3264 (2004).
11. M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, 2000.
12. B. Lesche, *J. Stat. Phys.* **27**, 419-423 (1982).
13. R. Lopez-Ruiz, H. Mancini and X. Calbet, *Phys. Lett. A* **209**, 321-326 (1995).
14. A. Daffertshofer, A. R. Plastino and A. Plastino, *Phys. Rev. Lett.* **88**, 210601 (2002).
15. V. Vedral, M.B. Plenio, M.A. Rippin and P.L. Knight, *Phys. Rev. Lett* **78**, 2275-2278 (1997).
16. C.R. Rao, *Differential Geometry in Statistical Inference, IMS-Lectures Notes* **10** 217-225 (1987)
17. J. Lin, *IEEE Trans. Inf. Theory* **37**, 145-151 (1991).
18. J. Schindelin and D. Endres, *IEEE Trans. Inform. Theory* **49**, 1858-1860 (2003).
19. I. Grosse, P. Bernaola-Galvan, P. Carpena, R. Roman Roldan, J. Oliver & H.E. Stanley, *Phys. Rev. E* **65**, 041905 (2002).
20. P.W. Lamberti and A.Majtey, *Physica A*, **329**, 81-86 (2003).
21. D. Zanette, *Complex Systems*, To appear (2007).
22. A. Majtey, P.W. Lamberti, M.T. Martin and A. Plastino, *Eur. Phys. J. D* **32**, 413-418 (2005).
23. P.W. Lamberti, M.T. Martin, A. Plastino and O. Rosso, *Physica A* **334**, 119-126 (2004).
24. M.E. Pereyra, P.W. Lamberti and O. Rosso, in *Physica A* **379**, 122-132 (2007).
25. R. Román-Roldán, P. Bernaola-Galván and J. Oliver, Phys. Rev. Lett. **80**, 1344-1347 (1998).
26. K. Ch. Chatzisavvas, Ch. C. Moustakidis and C.P. Panos, *J. Chem. Phys.* **123**, 174111 (2005).
27. M. Casas, P.W. Lamberti, A. Plastino, A.R. Plastino and G. Roston, *Il Nuovo Cimento* **120B**, 521-532 (2005).
28. G. Lindblad *Comm. Math. Phys.* **33**, 305-322 (1973).
29. V. Vedral, *Rev. Mod. Phys.* **74**, 197-234 (2002).
30. A. Majtey, P.W. Lamberti and D. Prato, *Phys. Rev. A* **72** 052310 (2005)
31. L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni, *Rev. Mod. Phys.* **70**, 223-288 (1998).
32. B. McNamara and K. Wiesenfeld, *Phys. Rev. A* **39**, 4854-4869 (1989).
33. P. Jung and P. Hänggi, *Phys. Rev. A* **44**, 8032-8042 (1991).
34. I. Goychuk and P. Hänggi, *Phys. Rev E* **61**, 4272-4280 (2000).
35. C. Tessone, A Plastino and H. Wio, *Physica A* **326**, 37-54 (2003).
36. S.M. Bezrukov and I. Vodyanoy, *Nature* **378**, 362-364 (1995).
37. C.H. Bennett, H.J. Bernstein, S. Popescu and B. Schumacher, *Phys. Rev. A* **53**, 2046-2052 (1996).
38. A. Borras, A.R. Plastino, M. Casas and A. Plastino, *Advaces in Mathematical Physics* To appear (2007).