Physica A 388 (2009) 2337–2353

Contents lists available at ScienceDirect

Physica A

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

Generalized statistics framework for rate distortion theory

R.C. Venkatesan [a,*], A. Plastino [b]

[a] Systems Research Corporation, Aundh, Pune 411007, India

[b] IFLP, National University La Plata & National Research Council (CONICET), C. C., 727 1900, La Plata, Argentina

ARTICLE INFO

Article history:
Received 1 February 2008
Received in revised form 5 February 2009
Available online 15 February 2009

PACS:
05.40.-a
05.20.Gg
89.70.-a
89.70.Cf

Keywords:
Generalized Tsallis statistics
Rate distortion
Additive duality
Compression information
Nonextensive alternating minimization algorithm
Distortion–compression plane

ABSTRACT

Variational principles for the rate distortion (RD) theory in lossy compression are formulated within the ambit of the generalized nonextensive statistics of Tsallis, for values of the nonextensivity parameter satisfying $0 < q < 1$ and $q > 1$. Alternating minimization numerical schemes to evaluate the nonextensive RD function, are derived. Numerical simulations demonstrate the efficacy of generalized statistics RD models.

1. Introduction

The generalized (nonadditive) statistics of Tsallis' [1,2] has recently been the focus of much attention in statistical physics, and allied disciplines. Nonadditive statistics,[1] which generalizes the extensive Boltzmann–Gibbs–Shannon statistics, has much utility in a wide spectrum of disciplines ranging from complex systems and condensed matter physics to financial mathematics.[2] This paper investigates nonadditive statistics within the context of Rate Distortion (RD) theory in lossy data compression.

RD theory constitutes one of the cornerstones of contemporary information theory [3,4], and is a prominent example of *source coding*. It addresses the problem of determining the minimal amount of entropy (or information) R that should be communicated over a channel, so that a compressed (reduced) representation of the source (input signal) can be approximately reconstructed at the receiver (output signal) without exceeding a given distortion D.

For a thorough exposition of RD theory Section 13 of [3] should be consulted. Consider a discrete random variable $X \in \mathcal{X}$[3] called the *source alphabet* or the *codebook*, and, another discrete random variable $\tilde{X} \in \tilde{\mathcal{X}}$ which is a compressed

* Corresponding address: Systems Research Corporation, I.T.I. Road, Aundh, Maharashtra, Pune 411007, India.
E-mail addresses: ravi@systemsresearchcorp.com, ravicv@eth.net (R.C. Venkatesan), plastino@venus.fisica.unlp.edu.ar (A. Plastino).

1 The terms generalized statistics, nonadditive statistics, and nonextensive statistics are used interchangeably.

2 A continually updated bibliography of works related to nonextensive statistics may be found at http://tsallis.cat.cbpf.br/biblio.htm.

3 Calligraphic fonts denote sets.

0378-4371/$ – see front matter
doi:10.1016/j.physa.2009.02.003

representation of X. The compressed representation $\tilde{X}$ is sometimes referred to as the *reproduction alphabet* or the *quantized codebook*. By definition, quantization is the process of approximating a continuous range of values (or a very large set of possible discrete values) by a relatively small set of discrete symbols or integer values.

The mapping of $x \in \mathcal{X}$ to $\tilde{x} \in \tilde{\mathcal{X}}$ is characterized by a conditional (transition) probability $p(\tilde{x}|x)$. The information rate distortion function is obtained by minimizing the generalized mutual entropy $I_q(X;\tilde{X})$ (defined in Section 2)[4] over all normalized $p\left(\tilde{x}|x\right)$. Note that in RD theory $I_q(X;\tilde{X})$ is known as the *compression information* (see Section 4). Here, q is the nonextensivity parameter [1,2] defined in Section 2.

RD theory has found applications in diverse disciplines, which include data compression and machine learning. Deterministic annealing [5,6] and the information bottleneck method [7] are two influential paradigms in machine learning, that are closely related to RD theory. The representation of RD theory in the form of a variational principle, expressed within the framework of the Shannon information theory, has been established [3]. The computational implementation of the RD problem is achieved by application of the Blahut–Arimoto alternating minimization algorithm [3,8], derived from the celebrated Csiszár–Tusnády theory [9].

Since the work on nonextensive source coding by Landsberg and Vedral [10], a number of studies on the information theoretic aspects of generalized statistics pertinent to coding related problems have been performed by Yamano [11], Furuichi [12,13], and Suyari [14], amongst others. The source coding theorem, central to the RD problem, has been derived by Yamano [15] using generalized statistics. A preliminary work by Venkatesan [16] has investigated into the re-formulation of RD theory and the information bottleneck method, within the framework of nonextensive statistics.

Generalized statistics has utilized a number of constraints to define expectation values. The linear constraints originally employed by Tsallis of the form $\langle A\rangle = \sum_i p_i A_i$ [1], were convenient owing to their similarity to the maximum entropy constraints. The linear constraints were abandoned because of difficulties encountered in obtaining an acceptable form for the partition function. These were subsequently replaced by the Curado–Tsallis (C–T) [17] constraints $\langle A\rangle_q = \sum_i p_i^q A_i$. The C–T constraints were later discarded on physics related grounds, $\langle 1\rangle_q \neq 1$, and replaced by the normalized Tsallis–Mendes–Plastino (T–M–P) constraints [18] $\langle\langle A\rangle\rangle_q = \sum_i \frac{p_i^q}{\sum_i p_i^q} A_i$. The dependence of the expectation value on the normalized *pdf* renders the canonical probability distributions obtained using the T–M–P constraints to be *self-referential*. A fourth form of constraint, prominent in nonextensive statistics, is the optimal Lagrange multiplier (OLM) constraint [19,20]. The OLM constraint removes the self-referentiality caused by the T–M–P constraints by introducing centered mean values.

A recent formulation by Ferri, Martinez, and Plastino [21] has demonstrated a methodology to "rescue" the linear constraints in maximum (Tsallis) entropy models, and, has related solutions obtained using the linear, C–T, and, T–M–P constraints. This formulation [21] has commonality with the studies of Wada and Scarfone [22], Bashkirov [23], and, Di Sisto et al. [24]. This paper extends the work in Ref. [16], by employing the Ferri–Martinez–Plastino formulation [21] to formulate self-consistent nonextensive RD models for $0 < q < 1$ and $q > 1$.

Tsallis statistics is described by two separate ranges of the nonextensivity parameter, i.e. $0 < q < 1$ and $q > 1$. Within the context of coding theory and learning theory, each range of q has its own specific utility. Un-normalized Tsallis entropies take different forms for $0 < q < 1$ and $q > 1$, respectively. For example, as defined in Section 2, for $0 < q < 1$, the generalized mutual entropy is of the form $I_{0<q<1}\left(X;\tilde{X}\right) = -\sum_{x,\tilde{x}} p\left(x,\tilde{x}\right)\ln_q\left(\frac{p(x)p(\tilde{x})}{p\left(x,\tilde{x}\right)}\right)$.

For $q > 1$, as described in Section 2, the generalized mutual entropy is defined by $I_{q>1}(X;\tilde{X}) = S_q(X) + S_q(\tilde{X}) - S_q(X,\tilde{X})$, where $S_q(X)$ and $S_q(\tilde{X})$ are the marginal Tsallis entropies for the random variables X and $\tilde{X}$, and, $S_q(X,\tilde{X})$ is the joint Tsallis entropy. Unlike the Boltzmann–Gibbs–Shannon case, $I_{0<q<1}(X;\tilde{X})$ can never acquire the form of $I_{q>1}$, and vice versa. While the form of $I_{0<q<1}(X;\tilde{X})$ is important in a number of applications of practical interest in coding theory and learning theory, un-normalized Tsallis entropies for $q > 1$ demonstrate a number of important properties such as the *generalized data processing inequality* and the *generalized Fano inequality* [12].

It may be noted that normalized Tsallis entropies do exhibit the *generalized data processing inequality* and the *generalized Fano inequality* [11]. As pointed out by Abe [25], normalized Tsallis entropies do not possess Lesche stability. However, for applications in communications theory and learning theory, the local stability criterion of Yamano [26] may be evoked to justify the use of normalized Tsallis entropies described in terms of *escort probabilities*. Ongoing studies, which will be reported elsewhere, have established the relation between the solutions of generalized RD theory for un-normalized Tsallis entropies using linear constraints that are reported in this paper, and, normalized Tsallis entropies using T–M–P constraints defined in terms of *escort probabilities*, in a manner similar to that employed by Wada and Scarfone [27].

To reconcile the different forms of the generalized mutual entropy for $0 < q < 1$ and $q > 1$, the *additive duality of nonextensive statistics* [28] is evoked in Section 3. This results in dual Tsallis entropies characterized by re-parameterization of the nonextensivity parameter $q^* = 2-q$, results in a *dual generalized RD theory*. An important feature of dual Tsallis entropies is the similarity of the forms of the Tsallis entropies with their counterparts in Boltzmann–Gibbs–Shannon statistics, the difference being $\log(\bullet) \rightarrow \ln_{q^*}(\bullet)$ [27].

[4] The absence of a principled nonextensive channel coding theorem prompts the use of the term mutual entropy instead of mutual information.