Accepted version of the article published in IEEE Signal Processing Letters, 2023. DOI: 10.1109/LSP.2023.3291308.
Available at: http://ieeexplore.ieee.org

On the Estimation of Tsallis Entropy and a Novel Information Measure Based on its Properties

Aniol Martí, Ferran de Cabrera and Jaume Riba, *Senior Member, IEEE*

***Abstract*—This article explores a plug-in estimator of second-order Tsallis entropy based on Kernel Density Estimation (KDE) and its implicit regularization process. First, it is shown that the expected value of the estimator corresponds to the entropy of an Additive White Gaussian Noise (AWGN) model. Then, we prove various relevant properties of the Tsallis entropy: it is monotonically non-decreasing under random variables addition, its derivative with respect to the Gaussian noise power is monotonically non-increasing, and it is concave in the additive noise power. From these, we derive an information metric that provides an alternative to the strategy of regularization.**

***Index Terms*—Tsallis Entropy, U-Statistics, Kernel Density Estimation, Entropy Estimation, Gaussian Convolutions**

I. Introduction

SINCE the formulation of Shannon's entropy and its impact on communication theory, there has been an increasing interest in defining more general entropy measures to facilitate the use and estimation of entropy in complex systems. Alfréd Rényi studied the most general formula to quantify information while keeping the additivity property of independent systems [1], [2], and he presented order-q Rényi entropies in 1961:

$$h_q(\mathsf{X}) = \frac{1}{1-q}\log\int_{\mathcal{D}} f_{\mathsf{X}}^q(x)\,\mathrm{d}x, \tag{1}$$

where X is a continuous random variable with density f_{X} supported in $\mathcal{D}$, and $q \in [0,\infty)\setminus\{1\}$. In 1988, Constantino Tsallis proposed a possible generalization to the Boltzmann-Gibbs entropy in statistical mechanics [3], [4]:

$$S_q(\mathsf{X}) = \frac{1}{q-1}\left(1-\int_{\mathcal{D}} f_{\mathsf{X}}^q(x)\,\mathrm{d}x\right). \tag{2}$$

Rényi entropy and Tsallis entropy are related, and its relation can be expressed as follows:

$$h_q(\mathsf{X}) = \frac{1}{1-q}\log((1-q)S_q(\mathsf{X})+1). \tag{3}$$

Observe that both entropies tend to Shannon entropy as $q \to 1$.

The case $q = 2$ is of special interest because it provides a very simple relation between both entropies. In addition, the expected value of the probability density function, usually referred as Information Potential (IP), arises [5, Ch. 2]:

$$\mathrm{IP}(\mathsf{X}) = \int_{\mathcal{D}} f_{\mathsf{X}}^2(x)\,\mathrm{d}x = 1 - S_2(\mathsf{X}). \tag{4}$$

This work was (partially) funded by the Spanish Ministry of Science and Innovation project RODIN (PID2019-105717RB-C22), grant 2021 SGR 01033, fellowship 2019 FI 00620 and fellowship 2023 FI "Joan Oró" 00050 by Generalitat de Catalunya and the European Social Fund, and fellowship 2022 FPI-UPC 028 by Universitat Politècnica de Catalunya and Banc de Santander.

The authors are with the Signal Processing and Communications Group (SPCOM), Departament de Teoria del Senyal i Comunicacions, Universitat Politècnica de Catalunya (UPC), 08034 Barcelona, Spain (e-mail: {aniol.marti, ferran.de.cabrera, jaume.riba}@upc.edu).

Information metrics derived from the second-order case are more suitable for data-driven methods and estimation, for instance, using plug-in methods based on Kernel Density Estimation (KDE). These entropic measures arise in, for instance, pattern recognition [6], [7], digital communications [8] and radar imaging [9] problems. Nevertheless, the estimation of the entropy measure itself is still a complex task. Many research has been performed in this regard (see, for example, [10]), with the conclusion that, in general, some sort of regularization is required. As a matter of fact, the estimator derived from KDE [11] is also intrinsically regularized by the kernel variance, which can be seen as a contamination of the original random variable that introduces an undesirable bias to the estimate. In this article, although we also use KDE, the focus shifts to the change of entropy induced by the bias, rather than to the estimation itself. From this analysis, an entropic measure based on the repercussion of the contamination process is presented.

The paper is structured as follows. First, in Section II, we determine the estimator of second-order Tsallis entropy and we derive its expected value. In Section III, we present and prove some properties that naturally arise from the estimator proposed in the previous section. Section IV addresses a novel information measure that leverages the properties presented in previous sections. Finally, in Section V, we present our conclusions and we state future research lines.

II. Estimation of Tsallis Entropy

The Tsallis entropy estimator presented in this section is based on the IP estimator proposed in [12]–[14] and then improved in [15], taking advantage of U-statistics [16]. Given a sequence $\mathcal{X} = \{x_1, x_2, \ldots, x_L\}$ of i.i.d. samples drawn from the distribution of X, if a Gaussian kernel with bandwidth $\sqrt{v}$ is used to estimate f_{X} from $\mathcal{X}$, the IP estimator becomes[1]:

$$\begin{aligned} \widehat{\mathrm{IP}}(\mathsf{X}) &= \frac{2}{L(L-1)}\mathbf{1}_L^{\mathrm{T}}(\mathbf{K}\odot\mathbf{U})\mathbf{1}_L, \\ [\mathbf{K}]_{ij} &= \frac{1}{\sqrt{4\pi v}}\mathrm{e}^{-\frac{(x_i-x_j)^2}{4v}}, \quad 0 < i,j \le L, \end{aligned} \tag{5}$$

where $\mathbf{1}_L$ is a vector with L ones, $\mathbf{U}$ is a strictly upper (or lower) triangular matrix with the non-zero elements equal to one, and $\odot$ is the Hadamard product [17]. This estimator can be easily extended to second-order Tsallis entropy thanks to its affine relation with IP. Therefore, the second-order Tsallis entropy estimator is:

$$\widehat{S}_2(\mathsf{X}) = 1 - \frac{2}{L(L-1)}\mathbf{1}_L^{\mathrm{T}}(\mathbf{K}\odot\mathbf{U})\mathbf{1}_L = 1 - \widehat{\mathrm{IP}}(\mathsf{X}). \tag{6}$$

[1]Note that the resulting kernel bandwidth corresponds to $\sqrt{2v}$ due to the convolution between two Gaussian kernels, each with kernel bandwidth $\sqrt{v}$.

Next, we determine the expected outcome of the Tsallis entropy estimate in a lemma. It is worth noting that in [5, Ch. 2.5] the bias of the original estimator is studied, albeit only an approximation through the Taylor series. Here, the expected outcome of the estimator is studied from the point of view of the original entropy measure without any approximation.

Lemma 1. *Let* $\mathsf{N} \sim \mathcal{N}(0,1)$*,* $v > 0$*. The estimator proposed in* (6) *is an unbiased estimator of* $S_2(\mathsf{X} + \sqrt{v}\mathsf{N})$*:*

$$\mathrm{E}\left(\widehat{S}_2(\mathsf{X})\right) = S_2\big(\mathsf{X} + \sqrt{v}\mathsf{N}\big). \tag{7}$$

Proof. Consider a random variable $\mathsf{N} \sim \mathcal{N}(0,1)$, a random variable X with density f_{X} and a scalar $v > 0$. The estimator proposed in Equation (6) is:

$$\widehat{S}_2(\mathsf{X}) = 1 - \widehat{\mathrm{IP}}(\mathsf{X}). \tag{8}$$

Taking into account that linear functions preserve unbiasedness [18, p. 150], we only need to prove that

$$\mathrm{E}\left(\widehat{\mathrm{IP}}(\mathsf{X})\right) = \mathrm{IP}\big(\mathsf{X} + \sqrt{v}\mathsf{N}\big). \tag{9}$$

By linearity,

$$\mathrm{E}\left(\widehat{\mathrm{IP}}(\mathsf{X})\right) = \frac{2}{L(L-1)} \mathbf{1}_L^{\mathrm{T}} \, \mathrm{E}(\mathbf{K} \odot \mathbf{U}) \mathbf{1}_L. \tag{10}$$

We proceed by calculating $\mathrm{E}([\mathbf{K}]_{ij})$. First, consider the sequence $\mathcal{Y}$ formed by the pairwise differences

$$\begin{aligned} y_{i,j} = x_i - x_j, \quad & 1 \le i \le L-1 \\ & i+1 \le j \le L \end{aligned} \tag{11}$$

drawn from a random variable Y. Taking into account that x_i and x_j are i.i.d., the density of Y is

$$f_{\mathsf{Y}}(y) = f_{\mathsf{X}}(y) * f_{\mathsf{X}}(-y), \tag{12}$$

and its characteristic function is given by:

$$\varphi_{\mathsf{Y}}(t) = \varphi_{\mathsf{X}}(t)\varphi_{\mathsf{X}}(-t) = |\varphi_{\mathsf{X}}(t)|^2, \tag{13}$$

where we have used the hermiticity property. We can now write the expected value of the kernel matrix elements as:

$$\mathrm{E}([\mathbf{K}]_{ij}) = \frac{1}{\sqrt{4\pi v}} \mathrm{E}\left(\mathrm{e}^{\frac{-\mathsf{Y}^2}{4v}}\right). \tag{14}$$

Using the inversion formula of the characteristic function:

$$\begin{aligned} \mathrm{E}\left(\mathrm{e}^{\frac{-\mathsf{Y}^2}{4v}}\right) &= \int_{-\infty}^{+\infty} \mathrm{e}^{\frac{-y^2}{4v}} f_{\mathsf{Y}}(y) \, \mathrm{d}y \\ &= \int_{-\infty}^{+\infty} \mathrm{e}^{\frac{-y^2}{4v}} \frac{1}{2\pi} \int_{-\infty}^{+\infty} \mathrm{e}^{-\mathrm{i}ty} |\varphi_{\mathsf{X}}(t)|^2 \, \mathrm{d}t \, \mathrm{d}y \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\varphi_{\mathsf{X}}(t)|^2 \, \mathrm{d}t \int_{-\infty}^{+\infty} \mathrm{e}^{-\frac{y^2}{4v} - \mathrm{i}ty} \, \mathrm{d}y. \end{aligned} \tag{15}$$

The second integral has a known solution [19, p. 337], thus:

$$\mathrm{E}\left(\mathrm{e}^{\frac{-\mathsf{Y}^2}{4v}}\right) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \sqrt{4\pi v} \, \mathrm{e}^{-vt^2} |\varphi_{\mathsf{X}}(t)|^2 \, \mathrm{d}t, \tag{16}$$

and substituting in Equation (14):

$$\mathrm{E}([\mathbf{K}]_{ij}) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \mathrm{e}^{-vt^2} |\varphi_{\mathsf{X}}(t)|^2 \, \mathrm{d}t. \tag{17}$$

Observe that the exponential in the integral is the characteristic function of a $\mathcal{N}(0, 2v)$ distribution. It can be expressed as:

$$\varphi_{\sqrt{2v}\mathsf{N}}(t) = \mathrm{e}^{-vt^2} = \left|\mathrm{e}^{\frac{-vt^2}{2}}\right|^2 = \left|\varphi_{\sqrt{v}\mathsf{N}}(t)\right|^2. \tag{18}$$

Therefore,

$$\begin{aligned} \mathrm{E}([\mathbf{K}]_{ij}) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left|\varphi_{\sqrt{v}\mathsf{N}}(t)\right|^2 |\varphi_{\mathsf{X}}(t)|^2 \, \mathrm{d}t \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left|\varphi_{\sqrt{v}\mathsf{N}}(t)\varphi_{\mathsf{X}}(t)\right|^2 \, \mathrm{d}t \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left|\varphi_{\mathsf{X}+\sqrt{v}\mathsf{N}}(t)\right|^2 \, \mathrm{d}t. \end{aligned} \tag{19}$$

Taking into account that the characteristic function of a random variable is the Fourier transform of its density, we can finally apply Parseval's theorem,

$$\mathrm{E}([\mathbf{K}]_{ij}) = \int_{-\infty}^{+\infty} f^2_{\mathsf{X}+\sqrt{v}\mathsf{N}}(x) \, \mathrm{d}x = \mathrm{IP}\big(\mathsf{X} + \sqrt{v}\mathsf{N}\big). \tag{20}$$

All the entries in $\mathbf{K}$ have the same expected value. Substituting it in (10), we obtain:

$$\mathrm{E}\left(\widehat{\mathrm{IP}}(\mathsf{X})\right) = \mathrm{IP}\big(\mathsf{X} + \sqrt{v}\mathsf{N}\big), \tag{21}$$

proving that the proposed estimator provides an unbiased estimate of $S_2(\mathsf{X} + \sqrt{v}\mathsf{N})$. ☐

III. Properties of Tsallis Entropy

In the previous section we have shown that the proposed estimator of Tsallis entropy offers an unbiased estimate of $S_2(\mathsf{X} + \sqrt{v}\mathsf{N})$. Now, we present and prove some properties that arise from the entropy of a random variable perturbed with Additive White Gaussian Noise (AWGN). In later sections we propose an application that leverages these properties.

A. Monotonicity under Sum of Random Variables

It is known that Shannon entropy is monotonically non-decreasing under i.i.d. random variables addition [20], [21]. We now show that Tsallis entropy (and Rényi entropy) is also weakly monotonic.

Lemma 2. *Let* $q \in [0, \infty) \setminus \{1\}$*. Let* X*,* Y *be two independent continuous random variables. Let* $S_q(\mathsf{X})$*,* $S_q(\mathsf{Y})$ *be the order-*q *Tsallis entropies of* X *and* Y*, respectively. Then,*

$$S_q(\mathsf{X}) \le S_q(\mathsf{X} + \mathsf{Y}). \tag{22}$$

Proof. Consider the *entropy power* of $\mathsf{X} + \mathsf{Y}$:

$$N_q(\mathsf{X} + \mathsf{Y}) = \mathrm{e}^{2h_q(\mathsf{X}+\mathsf{Y})}, \tag{23}$$

It is known that entropy power is non-decreasing [22]:

$$N_q(\mathsf{X} + \mathsf{Y}) \ge \max\{N_q(\mathsf{X}), N_q(\mathsf{Y})\}. \tag{24}$$

Thus, without loss of generality, we can assume that $N_q(\mathsf{X}) \le N_q(\mathsf{Y})$, and substituting with (3):

$$((1-q)S_q(\mathsf{X}) + 1)^{\frac{2}{1-q}} \le ((1-q)S_q(\mathsf{X}+\mathsf{Y}) + 1)^{\frac{2}{1-q}}. \tag{25}$$

For $q \in [0,1)$:

$$(1-q)S_q(\mathsf{X}) + 1 \le (1-q)S_q(\mathsf{X}+\mathsf{Y}) + 1 \implies S_q(\mathsf{X}) \le S_q(\mathsf{X}+\mathsf{Y}). \tag{26}$$

If $q > 1$:

$$(1-q)S_q(\mathsf{X}+\mathsf{Y}) + 1 \le (1-q)S_q(\mathsf{X}) + 1 \implies S_q(\mathsf{X}) \le S_q(\mathsf{X}+\mathsf{Y}), \tag{27}$$

which concludes the proof. □

B. Monotonicity of the Derivative under Gaussian Addition

This property also emerges from the sum of random variables, but for the particular case where one of them is AWGN. In the Shannon entropy context, de Bruijn identity [23], [24, p. 672] gives a connection between Fisher information and entropy, and it allows to prove that the former is monotonically non-increasing. In [25], a de Bruijn-like identity for the second-order Tsallis entropy is derived:

$$\frac{\partial}{\partial v} S_2(\mathsf{X} + \sqrt{v}\mathsf{N}) = -\mathrm{E}\left(\frac{\partial^2}{\partial \mathsf{Y}^2} f_{\mathsf{Y}}(\mathsf{Y})\right), \tag{28}$$

where $\mathsf{Y} = \mathsf{X} + \sqrt{v}\mathsf{N}$. This is a result of particularizing $q = 2$ and assuming Gaussian noise. Nevertheless, it is worth noting that (7) reveals that this assumption is valid for the presented estimator, provided that Gaussian kernels are used. The consequence of (28) is that the contamination process given by (6) can be evaluated in terms of the curvatures of the shape of the resulting distribution. For instance, distributions with fine details are more susceptible to the contamination, which is translated to the derivative of the Tsallis entropy with respect to v. This particular point of view will be used in Section IV to derive an information measure based on the sensitivity of the distribution of X in front of the estimation process itself. However, the absence of the logarithm in (28) makes it difficult to provide a prove of the non-increasing monotonicity following a similar procedure to the de Bruijn identity. Instead, we present a novel proof of this result using the characteristic function.

Lemma 3. *Let* $\mathsf{N} \sim \mathcal{N}(0,1)$, X *a random variable with density* f_{X} *and* $v > 0$. *Then,*

$$\frac{\partial}{\partial v} S_2(\mathsf{X} + \sqrt{v}\mathsf{N}) \le \frac{\partial}{\partial v} S_2(\sqrt{v}\mathsf{N}). \tag{29}$$

Proof. First observe that the previous statement is true iff:

$$\frac{\partial}{\partial v}\mathrm{IP}\big(\mathsf{X} + \sqrt{v}\mathsf{N}\big) \ge \frac{\partial}{\partial v}\mathrm{IP}\big(\sqrt{v}\mathsf{N}\big). \tag{30}$$

Now consider the IP of the Gaussian perturbation:

$$\mathrm{IP}\big(\sqrt{v}\mathsf{N}\big) = \int_{-\infty}^{+\infty} f^2_{\sqrt{v}\mathsf{N}}(x)\,\mathrm{d}x. \tag{31}$$

Applying Parseval's theorem,

$$\mathrm{IP}\big(\sqrt{v}\mathsf{N}\big) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathrm{e}^{-vt^2}\,\mathrm{d}t. \tag{32}$$

Its derivative with respect to v is

$$\frac{\partial}{\partial v}\mathrm{IP}\big(\sqrt{v}\mathsf{N}\big) = \frac{-1}{2\pi}\int_{-\infty}^{+\infty} t^2\mathrm{e}^{-vt^2}\,\mathrm{d}t. \tag{33}$$

Now consider $\mathsf{Y} = \mathsf{X} + \sqrt{v}\mathsf{N}$, with density f_{Y}. Its IP is

$$\mathrm{IP}(\mathsf{Y}) = \int_{-\infty}^{+\infty} f^2_{\mathsf{Y}}(y)\,\mathrm{d}y. \tag{34}$$

Again with Parseval's theorem,

$$\mathrm{IP}(\mathsf{Y}) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |\varphi_{\mathsf{X}}(t)|^2\mathrm{e}^{-vt^2}\,\mathrm{d}t, \tag{35}$$

where φ_{X} is the characteristic function of X. Assuming that it does not depend on v, the derivative of $\mathrm{IP}(\mathsf{Y})$ is given by

$$\frac{\partial}{\partial v}\mathrm{IP}(\mathsf{Y}) = \frac{-1}{2\pi}\int_{-\infty}^{+\infty} |\varphi_{\mathsf{X}}(t)|^2\, t^2\mathrm{e}^{-vt^2}\,\mathrm{d}t. \tag{36}$$

Substituting Equations (33) and (36) into (30) we obtain

$$\begin{aligned}
&\frac{-1}{2\pi}\int_{-\infty}^{+\infty} |\varphi_{\mathsf{X}}(t)|^2\, t^2\mathrm{e}^{-vt^2}\,\mathrm{d}t \ge \frac{-1}{2\pi}\int_{-\infty}^{+\infty} t^2\mathrm{e}^{-vt^2}\,\mathrm{d}t \\
&\implies \int_{-\infty}^{+\infty} |\varphi_{\mathsf{X}}(t)|^2\, t^2\mathrm{e}^{-vt^2} - t^2\mathrm{e}^{-vt^2}\,\mathrm{d}t \le 0 \\
&\implies \int_{-\infty}^{+\infty} t^2\mathrm{e}^{-vt^2}\big(|\varphi_{\mathsf{X}}(t)|^2 - 1\big)\,\mathrm{d}t \le 0.
\end{aligned} \tag{37}$$

Taking into account that $|\varphi_{\mathsf{X}}(t)|^2 \le 1$ and $t^2\mathrm{e}^{-vt^2} \ge 0$, we conclude that (30) is always true [26]. Thus, inequality (29) is also true. □

C. Concavity in Noise Power

This last property states that second-order Tsallis entropy is concave in v, the noise power.

Lemma 4. *Let* $\mathsf{N} \sim \mathcal{N}(0,1)$, X *a random variable with density* f_{X} *and* $v > 0$. *Then,* $S_2(\mathsf{X} + \sqrt{v}\mathsf{N})$ *is concave in* v.

Proof. Consider the Tsallis entropy of $\mathsf{Y} = \mathsf{X} + \sqrt{v}\mathsf{N}$:

$$\begin{aligned}
S_2(\mathsf{Y}) &= 1 - \int_{-\infty}^{+\infty} f^2_{\mathsf{Y}}(y)\,\mathrm{d}y \\
&= 1 - \frac{1}{2\pi}\int_{-\infty}^{+\infty} |\varphi_{\mathsf{X}}(t)|^2\mathrm{e}^{-vt^2}\,\mathrm{d}t.
\end{aligned} \tag{38}$$

Its second derivative with respect to v is:

$$\frac{\partial^2}{\partial v^2} S_2(\mathsf{Y}) = -\frac{1}{2\pi}\int_{-\infty}^{+\infty} |\varphi_{\mathsf{X}}(t)|^2\, t^4\,\mathrm{e}^{-vt^2}\,\mathrm{d}t \le 0, \tag{39}$$

proving the concavity in v of $S_2(\mathsf{X} + \sqrt{v}\mathsf{N})$. □

IV. Information Measure from Tsallis Entropy

As stated in the introduction, estimating entropy is a complex task that usually requires some kind of regularization [10]. In (7), we observe that the proposed estimator already provides an intrinsic regularization term driven by the kernel bandwidth. The choice of an appropiate value for v is therefore crucial. While the Silverman rule is often employed for such tasks [27], it is not a general approach to the problem and may give inaccurate results if data is non-Gaussian. In contrast, we now propose an information metric that considers the repercussion of the contamination variance v, following the observations from (28). The goal is to determine the sensitivity of the Tsallis

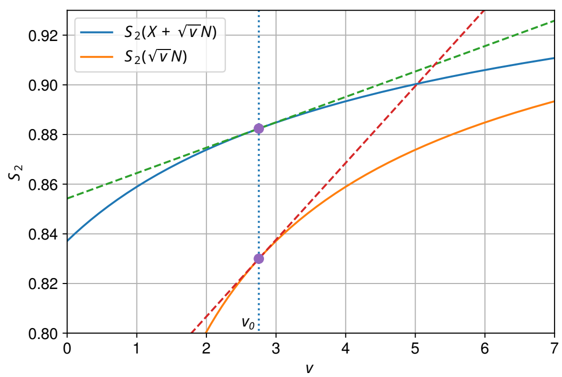

Fig. 1. Entropy of a random variable with Gaussian contamination versus the entropy of the Gaussian contamination itself, and their derivatives.

entropy in front of changes in v, and to compare it with the sensitivity of the Tsallis entropy of just the Gaussian process.

Consider $\mathsf{N} \sim \mathcal{N}(0,1)$, X a continuous random variable, and $v > 0$. Using Lemmas 2, 3 and 4, for any $k > 1$ there exists a unique v_0 such that

$$\frac{\partial}{\partial v} S_2(\mathsf{X} + \sqrt{v}\mathsf{N})_{|v=v_0} = \frac{1}{k}\frac{\partial}{\partial v} S_2(\sqrt{v}\mathsf{N})_{|v=v_0}. \tag{40}$$

A graphical illustration of this reasoning is shown in Fig. 1. The derivatives, represented in dashed lines, denote the sensitivity of the second-order Tsallis entropy in terms of v. The stronger is the contamination v, the closer to 1 becomes k, hence linking both parameters. Now, taking into account Lemma 1 and the linearity of differentiation and expectation,

$$\mathrm{E}\left(\frac{\partial}{\partial v}\widehat{S}_2(\mathsf{X})\right) = \frac{\partial}{\partial v} S_2(\mathsf{X} + \sqrt{v}\mathsf{N}). \tag{41}$$

The derivative of the Gaussian perturbation is given by

$$\frac{\partial}{\partial v} S_2(\sqrt{v}\mathsf{N}) = \frac{1}{4\sqrt{\pi} v^{3/2}}. \tag{42}$$

Thus, using (41) and letting $\mathsf{X} \sim \mathcal{N}(0, \beta)$ in (40),

$$\frac{1}{4\sqrt{\pi}(\beta + v_0)^{3/2}} = \frac{1}{k}\frac{1}{4\sqrt{\pi} v_0^{3/2}}. \tag{43}$$

After some manipulation, we obtain the information metric

$$\mathrm{V}(\mathsf{X}) = v_0 \cdot \left(k^{2/3} - 1\right) = \beta, \tag{44}$$

where k is the ratio of the slopes of the derivatives and it is inversely proportional to the perturbation power v_0. Note that $\mathrm{V}(\mathsf{X})$ is equal to the variance of X if it is normally distributed, but not in the general case.

A. Gaussian Mixture Model Cluster Variance

To conclude this section, we show that $\mathrm{V}(\mathsf{X})$ provides an upper bound of the cluster variance of a Gaussian Mixture Model (GMM). We assume that all clusters have the same variance β, which is relevant, for instance, when estimating the quality link of a digital communications channel [28].

Given a discrete random variable Y with probability mass function p_{Y}, a random variable $\mathsf{N}' \sim \mathcal{N}(0, 1)$ and scalars α, $\beta > 0$, $\mathsf{X} = \sqrt{\alpha}\mathsf{Y} + \sqrt{\beta}\mathsf{N}'$ follows a GMM with density:

$$f_{\mathsf{X}}(x) = \sum_{k=0}^{M-1} p_{\mathsf{Y}}(y_k) \frac{1}{\sqrt{2\pi\beta}} \mathrm{e}^{\frac{-(x-\sqrt{\alpha}y_k)^2}{2\beta}}. \tag{45}$$

Given that a GMM is a linear combination of Gaussian densities, its Tsallis entropy is

$$S_2(\mathsf{X}) = 1 - \frac{1}{\sqrt{4\pi\beta}} \sum_{j,k=0}^{M-1} p_{\mathsf{Y}}(y_j) p_{\mathsf{Y}}(y_k) \mathrm{e}^{\frac{-\alpha(y_j - y_k)^2}{4\beta}}. \tag{46}$$

Likewise, any AWGN model on X yields to the contamination of all clusters equally, and therefore, from the estimate $\widehat{S}_2(\mathsf{X})$ given in (6), we can write:

$$\mathrm{E}\left(\widehat{S}_2(\mathsf{X})\right) = S_2(\sqrt{\alpha}\mathsf{Y} + \sqrt{\beta + v}\mathsf{N}). \tag{47}$$

Finally, we announce the main result of this section.

Lemma 5. *Let $\mathsf{X} = \sqrt{\alpha}\mathsf{Y} + \sqrt{\beta}\mathsf{N}'$ be a random variable with density f_{X} given by* (45)*. Then,*

$$\beta \leq \mathrm{V}(\mathsf{X}). \tag{48}$$

Proof. Following (47) and Lemma 3 we have:

$$\frac{\partial S_2(\sqrt{\alpha}\mathsf{Y} + \sqrt{\beta + v}\mathsf{N})}{\partial v} \leq \frac{\partial S_2(\sqrt{\beta + v}\mathsf{N})}{\partial v} \leq \frac{\partial S_2(\sqrt{v}\mathsf{N})}{\partial v}, \tag{49}$$

and following the same procedure as in (40):

$$\begin{aligned}\frac{\partial S_2(\sqrt{\alpha}\mathsf{Y} + \sqrt{\beta + v}\mathsf{N})}{\partial v}\Big|_{v=v_0} &= \frac{1}{k_1}\frac{\partial S_2(\sqrt{\beta + v}\mathsf{N})}{\partial v}\Big|_{v=v_0} \\ &= \frac{1}{k_2}\frac{\partial S_2(\sqrt{v}\mathsf{N})}{\partial v}\Big|_{v=v_0}.\end{aligned} \tag{50}$$

From the second equality we obtain:

$$\beta = v_0\left((k_2/k_1)^{2/3} - 1\right). \tag{51}$$

Observing that $k_2 > k_1 > 1$,

$$\beta \leq v_0\left(k_2^{2/3} - 1\right) = \mathrm{V}(\mathsf{X}), \tag{52}$$

concluding the proof. □

The consequence of Lemma 5 is that for $\alpha \to 0$, then $\mathrm{V}(\mathsf{X})$ approximates β from above for any value of $k > 1$. The equality is also given by $\beta \to \infty$, which for a fixed α one can assume that $\beta \gg \alpha$. Then, f_{X} resembles a single Gaussian distribution, and the resulting $\mathrm{V}(\mathsf{X})$ also becomes β.

V. Conclusions

Inspired by the expected outcome of the estimate of the second-order Tsallis entropy based on KDE, we have shown the weakly increasing monotonicity under addition of independent random variable of Tsallis entropy for $q \in [0, \infty) \setminus \{1\}$, the concavity in the noise power v and the weakly decreasing monotonicity of its derivative also with respect to v. Then, we have proposed an information metric that leverages these properties, shifting the focus from the bias introduced by the regularization to the sensitivity of the distribution in front of it. We have shown that this information metric is equal to the variance in the Gaussian case, and it upper bounds the cluster variance of a GMM, when all the clusters have the same variance. These two properties of $\mathrm{V}(\mathsf{X})$ suggest the possibility of applying this measure to other common problems in the literature. For instance, the first property can be exploited to perform normality detection, whereas the second one, can be used to estimate the quality link of a communication channel.

References

[1] A. Rényi, "On measures of entropy and information," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, vol. 4.1, 1961.

[2] T. van Erven and P. Harremos, "Rényi Divergence and Kullback-Leibler Divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, 2014.

[3] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *Journal of Statistical Physics*, vol. 52, no. 1, 1988.

[4] S. Amari, *Information Geometry and Its Applications*. Tokyo: Springer Japan, 2016, vol. 194, ISBN: 978-4-431-55977-1.

[5] J. C. Principe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives* (Information Science and Statistics). New York, NY: Springer New York, 2010, ISBN: 978-1-4419-1569-6.

[6] M. Portes De Albuquerque, I. A. Esquef, and A. R. Gesualdi Mello, "Image thresholding using Tsallis entropy," *Pattern Recognition Letters*, vol. 25, no. 9, 2004.

[7] H. Yazid, S. N. Basah, S. A. Rahim, M. J. A. Safar, and K. S. Basaruddin, "Performance analysis of entropy thresholding for successful image segmentation," *Multimedia Tools and Applications*, vol. 81, no. 5, 2022.

[8] F. de Cabrera and J. Riba, "Entropy-based non-data-aided SNR estimation," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, 2019.

[9] M.-S. Kang, J.-H. Bae, S.-H. Lee, and K.-T. Kim, "Efficient isar autofocus via minimization of tsallis entropy," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 6, 2016.

[10] Q. Wang, S. R. Kulkarni, and S. Verdú, "Universal Estimation of Information Measures for Analog Sources," *Found. Trends Commun. Inf. Theory*, vol. 5, no. 3, 2009.

[11] E. Parzen, "On Estimation of a Probability Density Function and Mode," *The Annals of Mathematical Statistics*, vol. 33, no. 3, 1962.

[12] D. Erdogmus and J. C. Principe, "An error-entropy minimization algorithm for supervised training of nonlinear adaptive systems," *IEEE Transactions on Signal Processing*, vol. 50, no. 7, 2002.

[13] J. C. Aubuchon and T. P. Hettmansperger, "A note on the estimation of the integral of $f^2(x)$," *Journal of Statistical Planning and Inference*, vol. 9, no. 3, 1984.

[14] H. Joe, "Estimation of entropy and other functionals of a multivariate density," *Annals of the Institute of Statistical Mathematics*, vol. 41, no. 4, 1989.

[15] F. de Cabrera, J. Riba, and G. Vázquez, "Entropy-based covariance determinant estimation," in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2017.

[16] V. S. Koroliuk and Y. V. Borovskikh, *Theory of U-statistics* (Mathematics and its applications). Dordrecht, Boston: Kluwer Academic Publishers, 1994, ISBN: 978-0-7923-2608-3.

[17] R. A. Horn and C. R. Johnson, *Matrix analysis*, 2nd. New York, NY: Cambridge University Press, 2017, ISBN: 978-0-521-54823-6.

[18] H. Van Trees, K. Bell, and Z. Tian, *Detection Estimation and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory*. Wiley, 2013, ISBN: 978-1-118-53970-5.

[19] I. S. Gradshteĭn, I. M. Ryzhik, and A. Jeffrey, *Table of integrals, series, and products*, 7th ed. Academic Press, 2007, ISBN: 978-0-12-373637-6.

[20] T. A. Courtade, "Monotonicity of entropy and Fisher information: a quick proof via maximal correlation," *Communications in Information and Systems*, vol. 16, no. 2, 2016.

[21] S. Artstein, K. Ball, F. Barthe, and A. Naor, "Solution of Shannon's problem on the monotonicity of entropy," *Journal of the American Mathematical Society*, vol. 17, no. 4, 2004.

[22] E. Ram and I. Sason, "On Rényi Entropy Power Inequalities," *IEEE Transactions on Information Theory*, vol. 62, no. 12, 2016.

[23] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Information and Control*, vol. 2, no. 2, 1959.

[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. 2006, ISBN: 978-0-471-24195-9.

[25] J. Riba and F. de Cabrera, "A Proof of de Bruijn Identity based on Generalized Price's Theorem," in *2019 IEEE International Symposium on Information Theory (ISIT)*, Paris, France, 2019.

[26] G. Grimmett and D. Stirzaker, *Probability and random processes*. Oxford University Press, 2001, ISBN: 978-0-19-857223-7.

[27] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, First edition. Boca Raton, FL: CRC Press, 2018, ISBN: 978-1-351-45617-3.

[28] D. R. Pauluzzi and N. C. Beaulieu, "A comparison of SNR estimation techniques for the AWGN channel," *IEEE Transactions on Communications*, vol. 48, no. 10, 2000.