ON THE ESTIMATION OF TSALLIS ENTROPY AND A NOVEL INFORMATION MEASURE BASED ON ITS PROPERTIES

Aniol Martí, Ferran de Cabrera, and Jaume Riba

Department of Signal Theory and Communications, Universitat Politècnica de Catalunya (UPC)

Motivation

Estimation of information measures in information-theoretical learning problems is a fundamental, yet very complex task [1]. Moreover, estimation methods usually require some kind of self-regularization process that can potentially hinder the accuracy. By assuming that this regularization is mandatory for some estimators, this work tries to answer the question of whether we can utilize the regularization for the benefit of estimating information. In short, we prove three relevant lemmas of the estimation of the second-order Tsallis entropy that revolve around the regularization process. Finally, we propose an information measure that leverages the previous properties.

Baseline estimation

Let X be a continuous random variable with density f_{X} supported in $\mathcal{D}$. The Tsallis entropy of order $q \geq 0$ is defined as follows:

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

A special case is given for $q = 2$:

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

The previous entropy measure results in a very interesting expression for the purpose of estimation [2], [3]. Given i.i.d. samples $\mathcal{X} = \{x_1, x_2, \dots, x_L\}$ drawn from X, if a Gaussian kernel with bandwidth $\sqrt{v}$ is used to estimate f_{X}, the estimator becomes:

$$\hat{S}_2(\mathsf{X}) = 1 - \frac{2}{L(L-1)}\mathbf{1}_L^{\mathrm{T}}(\mathbf{K} \odot \mathbf{U})\mathbf{1}_L, \qquad [\mathbf{K}]_{ij} = \frac{1}{\sqrt{4\pi v}}\exp\left(\frac{-(x_i - x_j)^2}{4v}\right). \quad (3)$$

Lemma 1. *Let* $\mathsf{N} \sim \mathcal{N}(0,1)$, $v > 0$. *The estimator in* (3) *is an unbiased estimator of* $S_2(\mathsf{X} + \sqrt{v}\mathsf{N})$:

$$\mathrm{E}\big(\hat{S}_2(\mathsf{X})\big) = S_2(\mathsf{X} + \sqrt{v}\mathsf{N}). \quad (4)$$

Monotonicity I

Following the monotonicity of Shannon's entropy under i.i.d. random variables addition, we prove that Tsallis 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,*

$$S_q(\mathsf{X}) \leq S_q(\mathsf{X} + \mathsf{Y}). \quad (5)$$

Monotonicity II

For the particular case where one of the random variables in Lemma 2 is normally distributed, we can determine a de Bruijn-like identity for the second-order Tsallis entropy:

$$\frac{\partial}{\partial v} S_2(\mathsf{X} + \sqrt{v}\mathsf{N}) = -\mathrm{E}\left(\frac{\partial^2}{\partial \mathsf{Y}^2} f_{\mathsf{Y}}(\mathsf{Y})\right), \quad (6)$$

where $\mathsf{Y} = \mathsf{X} + \sqrt{v}\mathsf{N}$. Thanks to the unbiasedness in Lemma 1, (6) also applies to the estimator in (3). However, the lack of logarithm (compared to Shannon's entropy) makes difficult to to develop the right-side of (6). Instead, we present a novel bound:

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

$$\frac{\partial}{\partial v} S_2(\mathsf{X} + \sqrt{v}\mathsf{N}) \leq \frac{\partial}{\partial v} S_2(\sqrt{v}\mathsf{N}). \quad (7)$$

Concavity

Finally, we also prove that the second-order Tsallis entropy is concave with respect to the noise power:

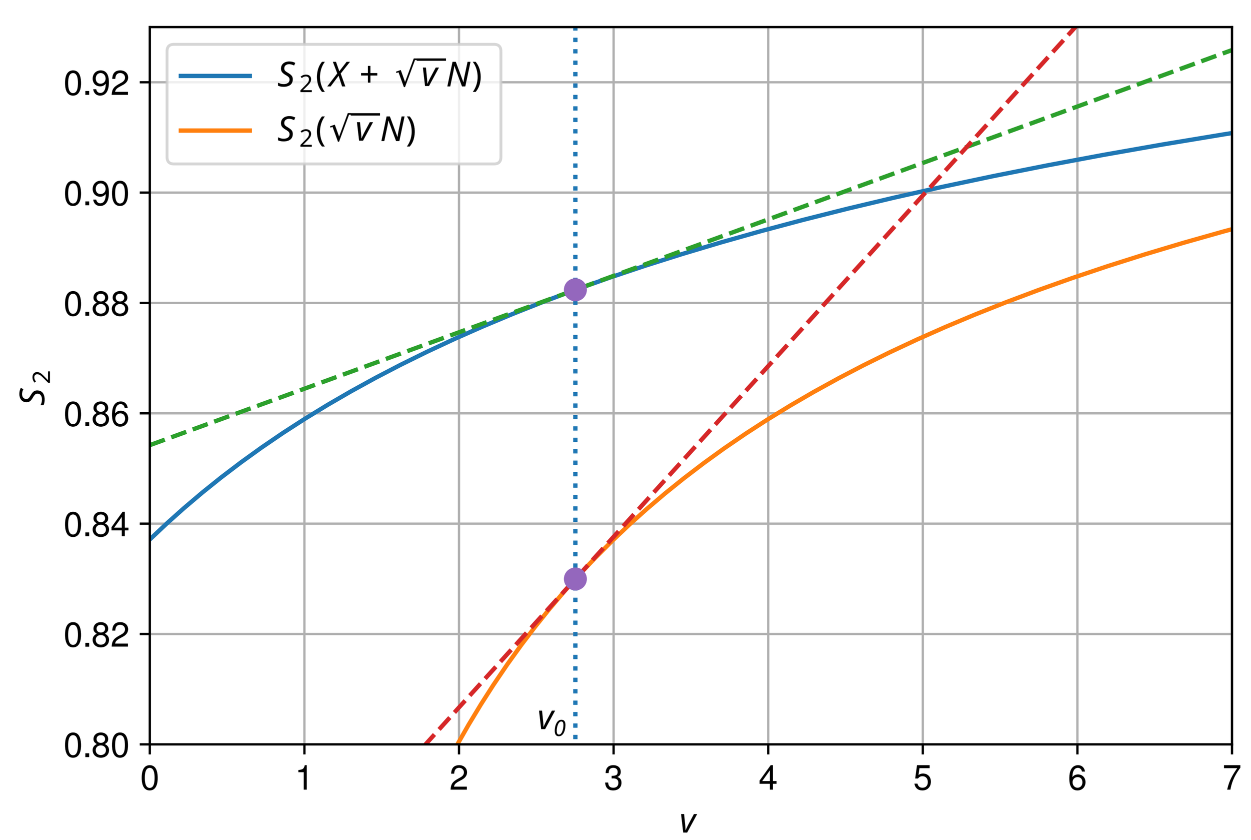
Lemma 4. *Let* $\mathsf{N} \sim \mathcal{N}(0,1)$, X *a random variable with density* f_{X} *and* $v > 0$. *Then,* $S_2(\mathsf{X} + \sqrt{v}\mathsf{N})$ *is concave in* v.

A novel information measure

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

$$\frac{\partial}{\partial v} S_2(\mathsf{X} + \sqrt{v}\mathsf{N})_{|v=v_0} = \frac{1}{k}\frac{\partial}{\partial v} S_2(\sqrt{v}\mathsf{N})_{|v=v_0}. \quad (8)$$

The relevance of the previous expression is the link between v and k. As the contamination becomes stronger, v increases and k becomes closer to one. This rationale is depicted in the following figure:



However, to produce an estimate we need to start from an estimator. Thus we join the the estimate $\hat{S}_2(\mathsf{X})$ with (8), and thanks to Lemma 1 and the linearity of differentiation and expectation we obtain the following expressions:

$$\mathrm{E}\left(\frac{\partial}{\partial v}\hat{S}_2(\mathsf{X})\right) = \frac{\partial}{\partial v} S_2(\mathsf{X} + \sqrt{v}\mathsf{N}). \quad (9)$$

Given $\frac{\partial}{\partial v} S_2(\sqrt{v}\mathsf{N}) = (4\sqrt{\pi}v^{3/2})^{-1}$ and letting $\mathsf{X} \sim \mathcal{N}(0, \beta)$ results in

$$\frac{1}{4\sqrt{\pi}(\beta + v_0)^{3/2}} = \frac{1}{k}\frac{1}{4\sqrt{\pi}v_0^{3/2}}, \quad (10)$$

which after some manipulations we can obtain the desired information measure:

$$\mathrm{V}(\mathsf{X}) = v_0 \cdot (k^{2/3} - 1) = \beta \quad (11)$$

GMM Cluster Variance

An application of (11) is to provide an upper bound of the cluster variance of a Gaussian Mixture Model (GMM) with the following density:

$$f_{\mathsf{X}}(x) = \sum_{k=0}^{M-1} \frac{p_{\mathsf{Y}}(y_k)}{\sqrt{2\pi\beta}}\exp\left(\frac{-(x - \sqrt{\alpha}y_k)^2}{2\beta}\right). \quad (12)$$

Then, the following lemma arises:

Lemma 5. *Let* $\mathsf{X} = \sqrt{\alpha}\mathsf{Y} + \sqrt{\beta}\mathsf{N}'$ *be a random variable with density* f_{X} *given by* (12). *Then,*

$$\beta \leq \mathrm{V}(\mathsf{X}). \quad (13)$$

References

[1] J. Príncipe, *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.

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

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

Acknowledgements

This work was funded by project MAYTE (PID2022-136512OB-C21) by MCIU/AEI/10.13039/501100011033 and ERDF "A way of making Europe", grant 2021 SGR 01033 and grant 2019 FI 00620 by Generalitat de Catalunya, grant 2023 FI "Joan Oró" 00050 by Generalitat de Catalunya and the ESF+, and grant 2022 FPI-UPC 028 by UPC and Banc de Santander.

Contact Information

- Email: aniol.marti@upc.edu
- UPC Campus Nord, C/ Jordi Girona 1-3, 08034 Barcelona