Rényi entropy yields artificial biases not in the data and incorrect updating due to the finite-size data

Thomas Oikonomou[1,*] and G. Baris Bagci[2]

[1]*Department of Physics, School of Science and Technology, Nazarbayev University, Astana 010000, Kazakhstan*
[2]*Department of Physics, Mersin University, 33110 Mersin, Turkey*

We show that the Rényi entropy implies artificial biases not warranted by the data and incorrect updating information due to the finite size of the data despite being additive. It is demonstrated that this is so because it does not conform to the system and subset independence axioms of Shore and Johnson [J. IEEE Trans. Inf. Theory **26**, 26 (1980)]. We finally show that the escort averaged constraints do not remedy the situation.

I. INTRODUCTION

Boltzmann-Gibbs-Shannon (BGS) entropy and its quantum mechanical counterpart von Neumann entropy have found applications in many diverse fields of science [1–16]. Its use is rigorously rationalized through, e.g., the coding theorems [17] or the Khincin axioms [18]. On the other hand, the entropic functional does not suffice for many practical issues and one also needs the concomitant equilibrium probability distribution. This distribution is indeed known in statistical mechanics for closed systems such as the canonical one. However, it was Jaynes who obtained the same distribution through a procedure named entropy maximization procedure (MaxEnt) by relying on information theory [19]. From then on, obtaining the equilibrium entropy became an inference problem in statistical estimation theory.

Whether this inference is consistent or not is warranted by the Shore-Johnson (SJ) criteria [20]. These criteria consist of four axioms which ensure choosing the suitable entropic functional if one aims for self-consistent inferences from the data. It is only afterwards that this functional can be safely maximized. In this sense, SJ axioms are related to the premaximization scheme. If a functional violates one of the SJ axioms, then it cannot be securely used for the MaxEnt procedure [21].

On the other hand, many entropies have been recently introduced and applied in diverse fields of research [22–27]. The main and common claim of these entropies is that they are suitable for the cases where the number of accessible states does not grow exponentially [28–31]. One of the mostly used entropy measures is the Rényi measure. It is used extensively for quantum information, generalized statistical mechanics, and quantum gravity. Then this begs the question whether the Rényi entropies conform to some or all of the axioms, thereby allowing consistent inferences.

There are four axioms forming the SJ criteria [20,32–35]. The first one is the axiom of uniqueness whose requirement is the concavity (convexity) of the (relative) entropy measure to warrant a unique maximum. The second one is called the invariance axiom and requires the inferences to be independent of the adopted coordinate system. The third axiom

is the subset independence. It states that treatments of an independent subset of system states in terms of separate conditional probabilities and in terms of joint probabilities are equivalent. From a different point of view, it implies that the MaxEnt procedure should yield the same results whether the number of states is fully accounted for or not. The last axiom, i.e., system independence warrants the absence of biases when the systems are independent if the data do not provide evidence otherwise. In particular, for systems A and B with probabilities $\{u_i\}$ and $\{v_j\}$, respectively, their combination creates new bins with the joint probability $p_{ij} = u_i v_j$.

It is worth emphasis that SJ axioms have precedence to the MaxEnt, since it is interested in consistent and unbiased inference, i.e., premaximization [34]. These axioms are about deriving a functional which is only later to be maximized. Skipping this vital step implies the failure of both MaxEnt and postmaximization-related arguments such as extensivity but not vice versa.

It has recently been shown that the nonadditive q-entropies violate only the probability independence axiom [32]. It might be argued that the same result should also hold for the Rényi entropy, since the nonadditive q- and additive Rényi entropies are monotonic functions of each other and therefore should result in the same probability distributions after the maximization procedure. However, since the SJ axioms are related to the premaximization scheme and choose the functional which is later to be maximized, one must consider the Rényi entropy as a different and important case study compared to the nonadditive q-entropies.

II. BIASES, UPDATING AND RÉNYI ENTROPY

In order to check the SJ axioms for the Rényi entropy, we begin with the following functional:

$$H_q(\{p\}) - \lambda \left(\sum_k p_k a_k - \bar{a} \right), \tag{1}$$

where λ and $\bar{a}$ are Lagrange multiplier and the measured average of the quantity a, respectively. The normalization constraint is omitted for simplicity. The term H_q is the Rényi

---

*thomas.oikonomou@nu.edu.kz

entropy given as

$$H_q(\{p\}) = \frac{1}{1-q} \ln\left(\sum_{i=1}^{n} p_i^q\right), \qquad (2)$$

where q is the generalization parameter.

The first axiom can be easily checked, since the Rényi entropy H_q is concave for $q \in (0, 1)$. Therefore, the uniqueness axiom is satisfied by the Rényi entropy in the interval $q \in (0, 1)$.

The second axiom, i.e., the invariance axiom, requires some clarification. The discrete Rényi entropy (or discrete BGS entropy for that matter) is not coordinate invariant. Therefore, one generalizes them to the continuous case in the form of relative entropy so that one safely optimizes the associated relative entropy. This relative entropy expression, if it exists, can be seen to be coordinate invariant. For the BGS entropy, this is given by the Kullback-Leibler relative entropy, for example. Therefore, the crux of this axiom is to obtain the continuous version of the discrete entropy measure so that it will be coordinate invariant (see Ref. [36] for details). To this aim, consider the discrete Rényi entropy as the point of departure and following the Jaynes' approach, substitute $p_i = \rho(x_i)\Delta x_i = \rho(x_i)/[nm(x_i)]$ into the discrete expression to obtain

$$\lim_{n\to\infty} H_q(\{p\}) = \lim_{n\to\infty} \frac{1}{1-q} \ln\left\{\sum_{i=1}^{n}\left[\frac{\rho(x_i)}{m(x_i)}\right]^q m(x_i)\Delta x_i\right\}$$
$$+ \lim_{n\to\infty} \ln(n)$$

so that the continuous Rényi (relative) entropy $H_q[\rho\|m]$ reads

$$H_q[\rho\|m] = \frac{1}{1-q} \ln\left\{\int_a^b \left[\frac{\rho(x)}{m(x)}\right]^q m(x)\,dx\right\},$$

where the additive term is omitted exactly as in the BGS case, since the entropy change is what matters. Note that $m(x)$ is now a measure on the continuous version of the state space.

The third axiom in the SJ set of criteria is the subset independence. For the Rényi entropy to conform to this axiom, the following expression should be equal to zero [21]:

$$\frac{\partial}{\partial p_\ell}\left(\frac{\partial}{\partial p_k} - \frac{\partial}{\partial p_j}\right)\left(H_q - \lambda_a \sum_{i=1}^{n} p_i a_i\right). \qquad (3)$$

Substituting the Rényi entropy H_q, the expression above reads

$$\frac{\partial}{\partial p_\ell}\left[\frac{q\left(p_k^{q-1} - p_j^{q-1}\right)}{(1-q)\sum_{i=1}^{n} p_i^q} - \lambda_a a_k + \lambda_a a_j\right], \qquad (4)$$

which, after a little algebra, yields

$$\frac{q\left(p_k^{q-1} - p_j^{q-1}\right)}{1-q} \frac{\partial}{\partial p_\ell}\left(\sum_{i=1}^{n} p_i^q\right)^{-1}. \qquad (5)$$

Since this expression is not zero in general, we conclude that the Rényi entropy violates the third SJ axiom, i.e., the subset independence axiom. The implication of this violation is that the inference drawn from the updating procedure strongly depends on the set $S \subset D$ of limited data currently available to us, where D is the maximum state set of the system under scrutiny.

To check the last axiom, namely, the system independence axiom, we consider the following functional:

$$\Lambda = \frac{1}{1-q} \ln\left[\sum_{i,j} f(p_{ij})\right] - \lambda_a\left(\sum_{i,j} p_{ij}a_i - \overline{a}\right)$$
$$- \lambda_b\left(\sum_{i,j} p_{ij}b_j - \overline{b}\right). \qquad (6)$$

The maximization condition $\delta\Lambda = 0$ yields

$$\frac{f'(p_{ij})}{(1-q)\sum_{k,l} f(p_{kl})} - K_{ij} = 0, \qquad (7)$$

where $K_{ij} \equiv \lambda_a a_i - \lambda_b b_j$. Summing over all indices, we obtain the relation

$$Kf'(p_{ij}) = K_{ij}\sum_{k,l} f'(p_{kl}) \qquad (8)$$

with $K \equiv \sum_{i,j} K_{ij}$. Taking first the partial derivative $\frac{\partial}{\partial u_i}$ of the equation above,

$$Kf''(p_{ij})\frac{\partial p_{ij}}{\partial u_i} = K_{ij}\sum_l f''(p_{il})\frac{\partial p_{il}}{\partial u_i}, \qquad (9)$$

and then the partial derivative $\frac{\partial}{\partial v_j}$, we obtain

$$(K - K_{ij})\left[f'''(p_{ij})\frac{\partial p_{ij}}{\partial u_i}\frac{\partial p_{ij}}{\partial v_j} + f''(p_{ij})\frac{\partial^2 p_{ij}}{\partial u_i \partial v_j}\right] = 0. \qquad (10)$$

Since the first term in the parentheses is nonzero, the second one in the brackets must be equal to zero. Using the multiplicative joint probability $p_{ij} = u_i v_j$, the expression within the brackets yields

$$p_{ij} f'''(p_{ij}) + f''(p_{ij}) = 0. \qquad (11)$$

This differential equation is known to yield the BGS entropy [see Eq. (5) in Ref. [32]]. Therefore, we conclude that Rényi entropy also violates the system independence axiom.

A crucial point is that the Rényi entropy is often maximized through the so-called escort distributions to obtain the averaged quantities. These distributions are of the form $\frac{\sum_{i=1}^{n} p_i^q a_i}{\sum_{k=1}^{n} p_k^q}$. As we can show, the use of the escort distribution can not provide conformity of the Rényi entropy to the SJ axioms neither. The first two axioms, i.e., the uniqueness and coordinate invariance are still satisfied by the Rényi entropy, since the changes in the averaging scheme leave the entropy expression unchanged. However, the subset independence axiom relies on how we average the constraints. In fact, with the escort averaging, Eq. (3) becomes

$$\frac{\partial}{\partial p_\ell}\left(\frac{\partial}{\partial p_j} - \frac{\partial}{\partial p_k}\right)\left[H_q - \lambda\frac{\sum_i p_i^q a_i}{\sum_i p_i^q}\right]. \qquad (12)$$

In order to check whether the subset independence axiom is satisfied, the above relation should be equal to zero. After a little algebra, it reads

$$\left(p_j^{q-1} - p_k^{q-1}\right)\left[\frac{1}{q-1} + \lambda\left(a_\ell - 2\frac{\sum_i p_i^q a_i}{\sum_i p_i^q}\right)\right]$$
$$+ \lambda\left(p_j^{q-1}a_j - p_k^{q-1}a_k\right), \qquad (13)$$

which is zero only for $q \to 1$, i.e., in the BGS entropy. As a result, we conclude that the escort averaged constraints do not prevent the Rényi entropy from violating the subset independence axiom.

The next and last axiom to be checked is the system independence axiom. Similar to Eq. (6) but now with escort averages, we write

$$\Lambda = \frac{1}{1-q} \ln \left[\sum_{i,j} f(p_{ij}) \right] - \lambda_a \left[\frac{\sum_{i,j}(p_{ij})^q a_i}{\sum_{k,l}(p_{kl})^q} - \bar{a} \right]$$
$$- \lambda_b \left[\frac{\sum_{i,j}(p_{ij})^q b_j}{\sum_{k,l}(p_{kl})^q} - \bar{b} \right]. \quad (14)$$

Through $\delta \Lambda = 0$, we have

$$0 = \frac{q}{1-q} \frac{f'(p_{ij})}{\sum_{k,l} f(p_{kl})} - \lambda_a q \frac{(p_{ij})^{q-1} \left[a_i - \frac{\sum_{i,j}(p_{ij})^q a_i}{\sum_{k,l}(p_{kl})^q} \right]}{\sum_{k,l}(p_{kl})^q}$$
$$- \lambda_b q \frac{(p_{ij})^{q-1} \left[b_j - \frac{\sum_{i,j}(p_{ij})^q b_j}{\sum_{k,l}(p_{kl})^q} \right]}{\sum_{k,l}(p_{kl})^q}. \quad (15)$$

Multiplying this equation with p_{ij} and summing over i, j we see

$$0 = \sum_{i,j} p_{ij} f'(p_{ij}), \quad (16)$$

where again $q \in (0, 1)$. Applying now the derivative $\frac{\partial}{\partial u_i}$,

$$0 = \sum_{k,l} [f'(p_{kl}) + p_{kl} f''(p_{kl})] \frac{\partial p_{kl}}{\partial u_i}$$
$$= \sum_l [f'(p_{il}) + p_{il} f''(p_{il})] \frac{\partial p_{il}}{\partial u_i}, \quad (17)$$

and then the derivative $\frac{\partial}{\partial v_j}$, we have

$$0 = [2f''(p_{ij}) + p_{ij} f'''(p_{ij})] \frac{\partial p_{ij}}{\partial v_j} \frac{\partial p_{ij}}{\partial u_i}$$
$$+ [f'(p_{ij}) + p_{ij} f''(p_{ij})] \frac{\partial^2 p_{ij}}{\partial v_j \partial u_i}. \quad (18)$$

Substituting the probability multiplication $p_{ij} = u_i v_j$ we finally have

$$0 = f'(p_{ij}) + 3 p_{ij} f''(p_{ij}) + p_{ij}^2 f'''(p_{ij}), \quad (19)$$

whose solution reads

$$f(p_{ij}) = c_1 + c_2 \ln(p_{ij}) + \frac{c_3}{2} \ln^2(p_{ij}). \quad (20)$$

Apparently, the Rényi entropy violates the probability independence axiom, since the measure $\frac{1}{1-q} \ln \left[\sum_{i,j} f(p_{ij}) \right]$ with the above expression, i.e., $f(p_{ij}) = c_1 + c_2 \ln(p_{ij}) + \frac{c_3}{2} \ln^2(p_{ij})$, is not the Rényi entropy.

Before concluding, one can also consider a related entropy expression, i.e., the homogeneous entropy [37,38] in terms of the SJ axioms

$$S_q^H = \frac{\left[\sum_i p_i^{1/q} \right]^q - 1}{q - 1}. \quad (21)$$

The homogeneous entropy can easily be checked to satisfy the first axiom due to its concavity. However, it fails the second

one: retracing the aforementioned steps regarding the second SJ axiom, we obtain

$$S_q^H = \frac{1}{q-1} \left(n^{q-1} \left\{ \sum_i m(x_i) \left[\frac{\rho(x_i)}{m(x_i)} \right]^{1/q} \Delta x_i \right\}^q - 1 \right). \quad (22)$$

As can be seen, the expression does not converge as $n \to \infty$. In other words, the homogeneous entropy does not have a continuous expression and is bound to be used only for the discrete cases. Regarding the third axiom, i.e., subset independence, we find

$$\frac{\partial}{\partial p_\ell} \left(\frac{\partial}{\partial p_k} - \frac{\partial}{\partial p_j} \right) \left(S_q^H - \lambda_a \sum_i p_i a_i \right)$$
$$= \frac{q \left(p_k^{\frac{1}{q}-1} - p_j^{\frac{1}{q}-1} \right)}{q-1} \left(\sum_i p_i^{1/q} \right)^{q-1} \frac{\partial}{\partial p_\ell} \left(\sum_i p_i^{1/q} \right), \quad (23)$$

which is generally nonzero apart from the well-known limit $q \to 1$. Thus, the subset independence axiom is also violated by the homogeneous entropy. The last axiom, i.e., system independence, yields the relation $(1/q - 1) p_{ij} = 0$ when one follows the above procedure. Apparently, this relation is satisfied only for $q = 1$, therefore finally showing that the homogeneous entropy satisfies only the first, i.e., uniqueness, axiom of the SJ criteria, violating the other three axioms explicitly. Following similar steps, one can see that S_q^H does not conform to the SJ axioms even with the escort averaging.

III. CONCLUSIONS

To sum up, SJ criteria should be satisfied by any entropy measure if the criteria are to be used for consistent inferences from the data. The Rényi entropy violates both system and subset independence axioms, thereby resulting in the artificial biases not present in the data and updating information dictated by finite-size effects, respectively. As a result, it cannot be used to draw inferences in a consistent manner. On the other hand, the nonadditive q-entropies violate only the system independence axiom [32]. This might be surprising at first, since these two entropies are monotonic functions of one another. After all, they yield the same type of probability distributions after the entropy maximization procedure is carried out. The resolution of this apparent conflict is at the heart of the SJ axioms, since these axioms, even before the maximization procedure, determine whether the functional under scrutiny is worth maximizing at all. Despite the violation of the SJ criteria, one might nevertheless choose to maximize an entropy measure and obtain the concomitant optimum distribution. However, any such violation is bound to cast doubt on the practical use of the entropy expression such as yielding artificial biases or inapplicability for continuous physical systems even before one hopes for a consistent thermodynamics [39,40]. Since SJ criteria are related to the premaximization scheme [34], the Rényi entropy is an important case by itself whose features cannot be deduced from those of the nonadditive q-entropies. Note that the Rényi entropy is sometimes used together with a different averaging scheme, i.e., so-called escort averaging. However, this change in averaging constraints neither prevents the Rényi entropy from violating these two important axioms. Finally,

one might try to extend the scope of one of the axioms, for example, the probability independence axiom [41]. However, this does not suffice, since the Rényi entropy would still violate the subset independence axiom. In short, the Rényi entropy can be used for various purposes but is inadequate for consistent inferences to be drawn. Having considered also the homogeneous entropy and shown that it also fails the SJ criteria, we emphasize that the SJ criteria single out the Shannon entropy as the unique measure to draw consistent inferences if one adopts the linear averaging scheme.

[1] C.-S. Wong, J. Goree, Z. Haralson, and B. Liu, Nat. Phys. **14**, 21 (2018).

[2] E. Cocchi, L. A. Miller, J. H. Drewes, C. F. Chan, D. Pertot, F. Brennecke, and M. Köhl, Phys. Rev. X **7**, 031025 (2017).

[3] D. Mandal, K. Klymko, and M. R. DeWeese, Phys. Rev. Lett. **119**, 258001 (2017).

[4] J. D. Ramshaw, Phys. Rev. E **96**, 052122 (2017).

[5] C. Nardini, É. Fodor, E. Tjhung, F. van Wijland, J. Tailleur, and M. E. Cates, Phys. Rev. X **7**, 021007 (2017).

[6] L. Piroli, E. Vernier, P. Calabrese, and M. Rigol, Phys. Rev. B **95**, 054308 (2017).

[7] F. Binder, S. Vinjanampathy, K. Modi, and J. Goold, Phys. Rev. E **91**, 032119 (2015).

[8] B. Gardas and S. Deffner, Phys. Rev. E **92**, 042126 (2015).

[9] P. Skrzypczyk, A. J. Short, and S. Popescu, Nat. Commun. **5**, 4185 (2014).

[10] Z. Xi, M.-L. Hu, Y. Li, and H. Fan, Quant. Info. Proc. **17**, 34 (2018).

[11] E. Chitambar, A. Streltsov, S. Rana, M. N. Bera, G. Adesso, and M. Lewenstein, Phys. Rev. Lett. **116**, 070402 (2016).

[12] M. Uhl, P. Pietzonka, and U. Seifert, J. Stat. Mech. (2018) 023203.

[13] F. Saracco, M. J. Straka, R. Di Clemente, A. Gabrielli, G. Caldarelli, and T. Squartini, New J. Phys. **19**, 053022 (2017).

[14] M. Stella and M. De Domenico, Entropy **20**, 268 (2018).

[15] M. Nogueira, PLoS ONE **12**, e0173994 (2017).

[16] F. Arizmendi, M. Barreiro, and C. Masoller, Sci. Rep. **7**, 45676 (2017).

[17] P. Nath, Inf. Control **29**, 234 (1975).

[18] A. I. Khinchin, *Mathematical Foundations of Information Theory* (Dover, New York, 1957).

[19] E. T. Jaynes, Phys. Rev. **106**, 620 (1957); **108**, 171 (1957).

[20] J. E. Shore and R. W. Johnson, J. IEEE Trans. Inf. Theory **26**, 26 (1980).

[21] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, Rev. Mod. Phys. **85**, 1115 (2013).

[22] P. Jizba and J. Korbel, Physica A **444**, 808 (2016).

[23] R. Hanel and S. Thurner, Europhys. Lett. **93**, 20006 (2011).

[24] G. Kaniadakis, Physica A **365**, 17 (2006).

[25] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).

[26] A. Rényi, On measures of entropy and information, in *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, edited by J. Neyman (University of California Press, Berkeley, 1961), Vol. 1, pp. 547–561.

[27] G. B. Bagci, R. Sever, and C. Tezcan, Mod. Phys. Lett. B **18**, 467 (2004).

[28] N. Canosa and R. Rossignoli, Phys. Rev. Lett. **88**, 170401 (2002).

[29] A. Taruya and M. A. Sakagami, Phys. Rev. Lett. **90**, 181101 (2003).

[30] G. B. Bagci, A. Arda, and R. Sever, Int. J. Mod. Phys. B **20**, 2085 (2006).

[31] G. B. Bagci, Physica A **386**, 79 (2007).

[32] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, Phys. Rev. Lett. **111**, 180604 (2013).

[33] S. Pressé, Phys. Rev. E **90**, 052149 (2014).

[34] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, Entropy **17**, 5043 (2015).

[35] C. Tsallis, Entropy **17**, 2853 (2015).

[36] T. Oikonomou and G. B. Bagci, Phys. Rev. E **97**, 012104 (2018).

[37] S. Arimoto, Inf. Control **19**, 181 (1971).

[38] J. F. Lutsko, J. P. Boon, and P. Grosfils, Europhys. Lett. **86**, 40005 (2009).

[39] G. B. Bagci and T. Oikonomou, Phys. Rev. E **93**, 022112 (2016).

[40] A. Plastino, M. C. Rocca, and F. Pennini, Phys. Rev. E **94**, 012145 (2016).

[41] P. Jizba and J. Korbel, arXiv:1808.01172 (2018).