Reply to "Comment on '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, Nazarbayev University, Nur-Sultan 010000, Kazakhstan*
[2]*Department of Physics, Mersin University, Mersin 33110, Turkey*

We reply to the preceding Comment by Jizba and Korbel [Jizba and Korbel, Phys. Rev. E **100**, 026101 (2019)] by first pointing out that the Schur concavity proposed by them falls short of identifying the correct intervals of normalization for the optimum probability distribution even though normalization is a necessary ingredient in the entropy maximization procedure. Second, their treatment of the subset independence axiom requires a modification of the Lagrange multipliers one begins with, thereby rendering the optimization less trustworthy. We also explicitly demonstrate that the Rényi entropy violates the subset independence axiom and compare it with the Shannon entropy. Third, the composition rule offered by Jizba and Korbel is shown to yield probability distributions even without a need for the entropy maximization procedure at the expense of creating artificial bias in the data.

We have recently shown that the Rényi entropy violates the Shore-Johnson (SJ) subset and system independence axioms [1,2].[1] Jizba and Korbel (JK) hold different views in this regard [3]. Therefore, we consider each SJ axiom below except for the second one, i.e., the invariance axiom, on which we have a consensus with JK.

Uniqueness axiom. In our assessment of this axiom, we used the concavity as a criterion which led us to the q interval being $(0,1)$. Inspecting the Rényi maximum distribution, one can see that for a variable interval $x \in (x_{\min}, x_{\max})$ with a proper $x_{\min}$ the distribution is normalizable for $0 < q < 1 + (x_{\max})^{-1}$. Apparently, in this inequality we identify also q values greater than the unity. The Schur concavity that JK used as the uniqueness criterion indeed allows these values. However, there is a major drawback in this consideration, namely, JK, in their entire derivation, assumed that q and x are mathematically independent quantities. As can be seen from the inequality above, this only holds when $q \in (0, 1)$, since for $q > 1$ the deformation parameter carries information about x. Thus, the $q > 1$ values are irrelevant and must be discarded. Accordingly, the Schur-concavity criterion, either at the premaximization or at the postmaximization state, will be reduced to the ordinary concavity criterion. From a different point of view, the inequality above reveals that when $q > 1$, the q value is dictated by the finite size of the data, which agrees with the title of our work. Since this happens anyway, the interval of interest to be explored is for $q \in (0, 1)$.

---

*thomas.oikonomou@nu.edu.kz
[1]Our result is valid whenever one optimizes the Rényi entropy. Hence, it does not impose any restrictions on the quantum information-theoretic applications of this entropy as long as entropy optimization is not included.

Subset independence axiom. Contrary to our opinion, JK argue that the Rényi entropy satisfies this axiom. The crux of their argument can be traced back to the idea that the maximization of $f(\sum_i g(p_i))$ should yield the same result as $\sum_i g(p_i)$. This certainly looks correct *prima facie* when one considers the entropy maximization procedure as already carried out, i.e., at a postmaximization stage. However, as explicitly pointed out in Ref. [4], the SJ axioms are concerned with the premaximization stage and they are about choosing the functional which is only later to be maximized to draw consistent inferences from the data. Since SJ axioms are directly related to the premaximization in order to exclude the opportunity for a possible erroneous inference procedure, these axioms yield different judgements facing the entropies that are closely related to one another. A case in point is to consider the Tsallis and Rényi entropies. Although they are monotonically related to one another, the use of SJ axioms shows that the Tsallis entropy only violates the system independence axiom whereas the Rényi entropy violates both the subset independence and system independence axioms with ordinary linear averaged constraints [1,5]. This difference between the two occurs not because of additivity versus non-additivity, since SJ axioms are general enough. The reason is that SJ axioms consider the premaximization stage and check what can be consistently maximized or not (and in which interval of admissible parameters) beforehand. Otherwise, one can surely maximize the Rényi entropy and obtain the concomitant probability distribution. The SJ axioms just warn us that we should not trust this distribution for consistent inferences.

The next related issue can be explicitly observed from the treatment of the subset independence by JK. Note that for their discussion in Eqs. (7)–(14), JK are forced to change the Lagrange multipliers in order to achieve their statement about the equivalence $f(\sum_i g(p_i)) \sim \sum_i g(p_i)$ to confirm the subset

independence for the Rényi entropy. This is erroneous, since one can easily verify that the maximization of $f(\sum_i g(p_i))$ characterizes a different maximum state rather than the maximization of $\sum_i g(p_i)$, i.e., $\beta = \frac{\partial \sum_i g(p_i)}{\partial U}$, while $\beta \neq \frac{\partial f(\sum_i g(p_i))}{\partial U}$ in their notation [6,7]. It is also worth emphasizing that this move already shows that there is something wrong with the maximization procedure carried out in this manner, since it yields the infamous physical temperature problem plaguing the field. In other words, due to this manipulation by JK, the Lagrange multiplier β is not inverse temperature and even worse does not yield the same value as the one obtained from the Clausius relation [8]. Note that one also loses the connection with thermodynamics as a result of such an unjustified entropy maximization procedure (see Ref. [9] for a nice exposition of such a case).

In addition, one can easily show from their Eq. (14), by eliminating the Lagrange multiplier α invoking the maximization conditions, that the maximum distribution contains a cross term, i.e., U, whose values depend on the particular subset under scrutiny, therefore violating the subset independence. We also provide a simple example, based on the original SJ criterion for continuous variables describing the subset independence axiom, to numerically demonstrate this violation in the Rényi case and compare it with the Shannon case. According to Shore and Johnson, for the total set $D = S_1 \cup S_2$ with $S_1 \cap S_2 = \emptyset$, the subset independence axiom is satisfied when [see Eqs. (13) and (14) in Ref. [2]]

$$q_D(x) = m(S_1)q_{S_1}(x) + m(S_2)q_{S_2}(x), \tag{1}$$

where $q_D(x)$, $q_{S_1}(x)$, and $q_{S_2}(x)$ are the optimized distributions in the respective set and

$$m(S_i) = \int_{x \in S_i} q_D(x)dx \Rightarrow m(S_1) + m(S_2) = 1. \tag{2}$$

Consider then $D = [0, \infty) = S_1 \cup S_2 = [0, 1) \cup [1, \infty)$. The Shannon posterior distributions are calculated from the entropy maximization procedure to be

$$q_D(x) = \beta e^{-\beta x},$$

$$q_{S_1}(x) = \frac{\beta}{1 - e^{-\beta}} e^{-\beta x}, \tag{3}$$

$$q_{S_2}(x) = \beta e^{\beta(1-x)},$$

with the coefficients determined from Eq. (2) as

$$m(S_1) = 1 - e^{-\beta}, \quad m(S_2) = e^{-\beta}. \tag{4}$$

Substituting then Eqs. (3) and (4) into Eq. (1), we verify the fulfillment of the subset independence axiom for the Shannon case. The Rényi posterior distributions, on the other hand, are given by

$$q_D(x) = \frac{\beta e_q^{-\beta(x-U)}}{[e_q^{\beta U}]^q},$$

$$q_{S_1}(x) = \frac{\beta e_q^{-\beta(x-U_1)}}{[e_q^{\beta U_1}]^q - [e_q^{\beta(U_1-1)}]^q}, \tag{5}$$

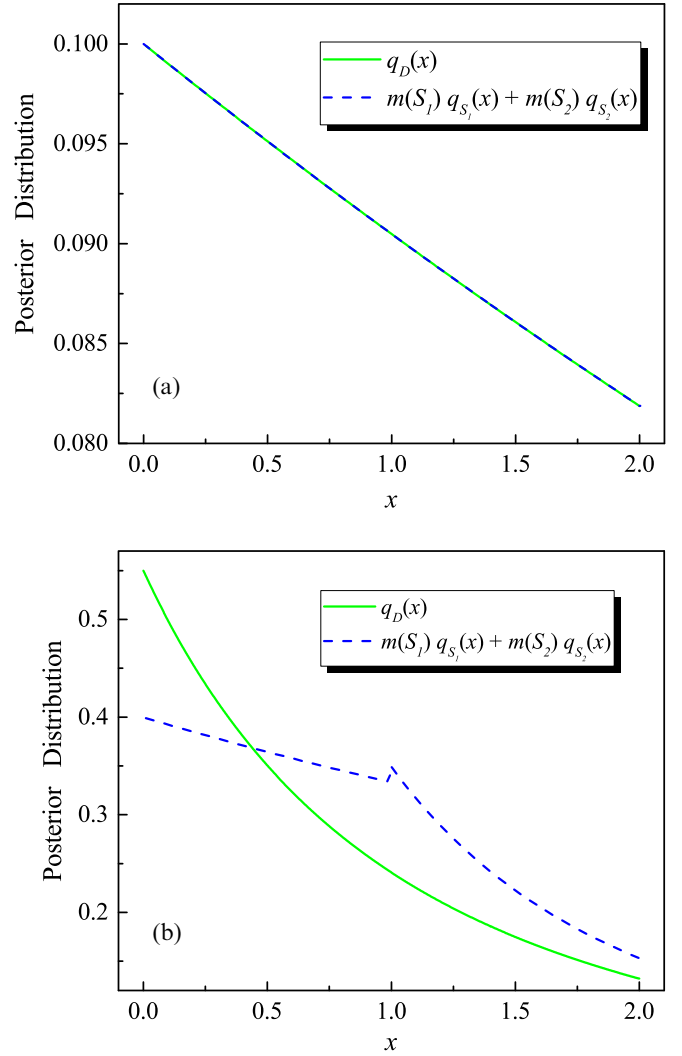$$q_{S_2}(x) = \frac{\beta e_q^{-\beta(x-U_2)}}{[e_q^{\beta(U_2-1)}]^q},$$

FIG. 1. Posterior (maximum) probability distributions for the (a) Shannon and (b) Rényi entropies. In both graphs the green solid lines and the blue dashed lines are the posterior distributions on the left- and right-hand sides of Eq. (1), respectively. Fulfillment of the subset independence for the Shannon entropy is shown in (a), while the violation of the former axiom for the Rényi entropy is in shown in (b).

where we introduced the notation $e_q^x := [1 + (q-1)q^{-1}x]^{1/(q-1)}$ and $\{U, U_1, U_2\}$ are the mean values in the respective set $\{D, S_1, S_2\}$. Also the probability normalization in each set requires $q < 1$ with $q > \beta U(1 + \beta U)^{-1}$, $q > \beta U_1(1 + \beta U_1)^{-1}$, and $q > \beta U_2(1 + \beta U_2)^{-1}$. The coefficients $m(S_i)$ are calculated again by means of Eq. (2) as

$$m(S_1) = 1 - \left[\frac{e_q^{\beta(U-1)}}{e_q^{\beta U}}\right]^q, \quad m(S_2) = \left[\frac{e_q^{\beta(U-1)}}{e_q^{\beta U}}\right]^q. \tag{6}$$

As can be immediately seen here, the coefficients depend on the mean value of the entire set D instead of S_i, which causes the violation of the subset independence. Indeed, substituting Eqs. (5) and (6) into Eq. (1), we verify that the latter is not satisfied. For the reader's convenience, we demonstrate graphically our results. In Figs. 1(a) and 1(b) we plot the

left- and right-hand sides of Eq. (1) for the Shannon and Rényi entropies, respectively, for the randomly chosen values $\beta = 0.1$ and $q = 0.55$ (with normalization conditions taken into account).

Correctly then, the Livesey-Skilling criterion reveals that only trace-form entropies satisfy the subset independence axiom. We note here that the Livesey-Skilling criterion for capturing the subset independence is more accurate than the one originally presented by Shore and Johnson. Showing this, however, lies beyond the scope of this Reply.

System independence axiom. First of all, JK in this part of their Comment compare our discussion with their modified SJ axiom in Ref. [3] ("... strong system independence is added to the SJ desiderata") and not with the original SJ axiom in Ref. [2]. In this sense their comment is irrelevant for our discussion in Ref. [1].

However, let us closely inspect the modified axiom introduced by JK. The main move of JK is to introduce a probability composition rule different from the one adopted by Shore and Johnson. As a result, instead of the original composition rule $p_{ij} = u_i v_j$, one now has $g(p_{ij}) = g(u_i v_j) = g(u_i)g(v_j)$ (Supplemental Material of [3], p. 3). However, once one introduces this novel composition rule, then the solution is given in advance and independently of the maximization procedure, namely, $g(x) \sim x^q$ [10]. In other words, through the modification proposed by JK, the entropy maximization procedure becomes redundant. Then one naturally asks what the use of the SJ axioms could be in the first place if they would be deemed redundant anyway by modifying only one of the axioms. Note also that one can obtain any entropic structure one wishes by such modifications of this axiom completely bypassing the entropy maximization procedure.

Finally, as already noted in Ref. [5], different composition rules, e.g., generalized q products, introduce biases not warranted in the data at all. We can understand this in the following simple way. At the maximum state we must have $\sum_{i,j} p_{ij} a_i = \bar{a} = \sum_i u_i a_i$, or equivalently $\sum_{i,j} a_i(p_{ij} - u_i v_j) = 0$ (the same of course holds for the second system too). Apparently, the fulfillment of this equation is always met only when $p_{ij} = u_i v_j$. Therefore, the use of the multiplicative joint probability composition rule at the premaximization state ensures the accuracy of the SJ inference procedure.

Before concluding, JK correctly emphasize that the scope of the original SJ axioms is linear constraints and the use of the escort averages (or any other averaging procedure) remains an open problem. Since JK did not present any explicit inconsistency in our treatment, we provide our view on this issue here. Shore-Johnson axioms are conceptually general enough in the sense that the violation of the second axiom, for example, dooms an entropy measure to the discrete uses only independently of the averaging scheme. In fact, it is important to remember that the entropy measures themselves are obtained from specific averaging procedures. In this regard, note that the Rényi entropy itself is obtained from the exponential averaging of the information gain whereas the Shannon entropy is formed by the linear average of the very same information gain. Therefore, the main issue is not the averaging procedure inherent in the entropy expressions but the constraints as duly noted by JK. However, constraints only enter into the SJ axioms when the functional is under scrutiny. For example, when one checks the second axiom above, we do not look for any averaging procedure in general, but try to find a consistent continuous generalization of the discrete entropy expression [11]. In other words, the conceptual message of the SJ axioms allows us to choose the appropriate averaging procedure for a new entropy measure so that the entropy measure can be used to draw consistent inferences from the data. For linearly averaged constraints, however, we uniquely have the Shannon entropy.

To sum up, contrary to the criticisms of JK, we showed that (i) invoking Schur concavity as a criterion for the uniqueness of the maximization solution, in contrast to the ordinary entropy concavity, may violate the probability normalization condition; (ii) the Livesey-Skilling criterion correctly describes the subset independence axiom, allowing only trace form entropies for a self-consistent inference procedure; and (iii) the modified JK subsystem independence axiom renders the maximization procedure redundant.

[1] T. Oikonomou and G. B. Bagci, Phys. Rev. E **99**, 032134 (2019).

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

[3] P. Jizba and J. Korbel, Phys. Rev. Lett. **122**, 120601 (2019).

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

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

[6] H. Karabulut, Eur. J. Phys. **27**, 709 (2006).

[7] T. Oikonomou and G. B. Bagci, Phys. Lett. A **381**, 207 (2017).

[8] S. Abe, S. Martínez, F. Pennini, and A. Plastino, Phys. Lett. A **281**, 126 (2001); S. Abe, Physica A **368**, 430 (2006).

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

[10] R. Hanel and S. Thurner, Eur. Phys. Lett. **93**, 20006 (2011).

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