

Reply

Reply to C. Tsallis' "Conceptual Inadequacy of the Shore and Johnson Axioms for Wide Classes of Complex Systems"

Steve Pressé ^{1,*}, Kingshuk Ghosh ², Julian Lee ³ and Ken A. Dill ⁴

¹ Department of Physics, Indiana Univ.-Purdue Univ. Indianapolis, Indianapolis, IN 46202, USA

² Department of Physics and Astronomy, University of Denver, Denver, CO 80208, USA;

E-Mail: kingshuk.ghosh@du.edu

³ Department of Bioinformatics, Soongsil University, Seoul 156-743, Korea;

E-Mail: suineg21@gmail.com

⁴ Laufer Center for Physical and Quantitative Biology and Departments of Physics and Chemistry, Stony Brook University, Stony Brook, NY 11794, USA; E-Mail: dill@laufercenter.org

* Author to whom correspondence should be addressed; E-Mail: stevenpresse@gmail.com.

Academic Editor: Kevin H. Knuth

Received: 7 June 2015 / Accepted: 14 July 2015 / Published: 17 July 2015

Abstract: In a recent PRL (2013, 111, 180604), we invoked the Shore and Johnson axioms which demonstrate that the least-biased way to infer probability distributions $\{p_i\}$ from data is to maximize the Boltzmann-Gibbs entropy. We then showed which biases are introduced in models obtained by maximizing nonadditive entropies. A rebuttal of our work appears in *entropy* (2015, 17, 2853) and argues that the Shore and Johnson axioms are inapplicable to a wide class of complex systems. Here we highlight the errors in this reasoning.

Keywords: nonadditive entropies; nonextensive statistical mechanics; strongly correlated random variables; Shore and Johnson axioms

In [1], we invoked the powerful results of Shore and Johnson (SJ) [2], who showed that, under quite general circumstances, the least-biased way to infer a probability distribution $\{p_i\}$ is to maximize the Boltzmann-Gibbs relative entropy

$$H = - \sum_i p_i \log(p_i/q_i) \quad (1)$$

(or any function that is monotonic with this entropy), under constraints where q_i is the prior distribution on p_i that contains any foreknowledge of the system. In [1], we showed that mathematical forms of H

that are not monotonic with Equation (1)—which we call noncanonical entropies—lead to unwarranted biases. In his comment on our [1], Tsallis objects to this argument on multiple grounds [3]. We show here the flaws in his objections.

First, Tsallis contends that “nonadditive entropies emerge from strong correlations (among the random variables of the system) which are definitively out of the SJ hypothesis” adding that the “SJ axiomatic framework addresses essentially systems for which it is possible to provide information about their subsystems without providing information about the interactions between the subsystems, which clearly is not the case where nonadditive entropies are to be used”. These statements are incorrect. SJ is not limited in any way to small interactions or weak correlations; it can handle interactions of any strength.

Standard methods of statistical physics—which are grounded in Equation (1)—provide a clear recipe for treating correlated systems while assuming no prior subsystem correlations. For example, we do not need a special form of Equation (1) in order to build spin-spin correlations into an Ising model. On the contrary, the Boltzmann weights for a spin-spin-correlated Ising model are constructed by assuming that spin correlations originate from the data (which is used to constrain Equation (1)). In the absence of spin-spin correlations, standard statistical physics returns an Ising model with decoupled spins, as it should. On the other hand, if we choose to assume *a priori* some particular coupling between the spins that do not originate from data, then SJ prove that it should be introduced through $\{q_i\}$ exactly as it appears in Equation (1). In particular, models of power laws can only arise in a principled way either from data constraints or from $\{q_i\}$ in conjunction with Equation (1). There is no principled basis for power laws that can be obtained by re-assigning the meaning of H and changing its form [1].

Second, Tsallis asserts that “We see therefore that the SJ set of axioms, demanding, as they do, system independence, are not applicable unless we have indisputable reasons to believe that the data that we are facing correspond to a case belonging to the exponential class, and by no means correspond to strongly correlated cases such as those belonging to the power-law or stretched-exponential classes, or even others”. This is not correct. The SJ inference procedure is not limited to independent systems. Rather, SJ asserts that *when systems are independent of each other, then* the joint outcome for two independent systems must be the product of marginal probabilities if data are provided for systems independently [4]. SJ is otherwise perfectly applicable broadly across situations not involving independent systems. In the language of Bayesian inference, the SJ axioms are used to derive a form for the noninformative prior over the model (the probability distribution $\{p_i\}$) [5]. The SJ axioms have no input on the likelihood function that otherwise determines how data updates the prior. What is more, because priors and data are confounded in the Tsallis entropy, the parametrization of the fitting parameter q in the Tsallis entropy relies on a misapplication of Bayes’ theorem as has been shown in [5].

Third, Tsallis asserts that “This is the deep reason why, in the presence of strong correlations, the BG entropy is generically not physically appropriate anymore since it violates the thermodynamical requirement of entropic extensivity.” In this statement, Tsallis is incorrectly asserting that extensivity must be a foundational property from which the functional form of the entropy follows. In fact, the logic is just the opposite. Extensivity—or not—of an entropy is the outcome of an inference problem at hand, not its input. Certainly throughout much of equilibrium thermodynamics, extensivity happens to hold. But that is a matter of the particulars of that (macroscopic) class of applications.

Said differently, this argument confuses the distinction between *pre-maximization* and *post-maximization* entropies. SJ focus on pre-maximization. SJ seek a functional H that, upon maximization, achieves certain properties required for drawing consistent unbiased inferences. At this stage, no system property (such as how entropy scales with system size) is yet relevant. This is just establishing a very general inference principle. However, once maximization has been performed, $H(\{p_i = p_i^*\}) = -\sum_i p_i^* \ln p_i^*$ is an entropy function that may be used to make predictions about physical systems including how properties scale with system size. The SJ argument is agnostic about whether extensivity holds or not for the post-maximization entropy $H(\{p_i = p_i^*\})$.

Conclusions

In short, the great power of the SJ arguments is in showing that Equation (1) is an extremely broad and deep result, applicable across all matters of inference of probability distributions, given only priors (q_i) on p_i and given new information. The power of the SJ arguments is that they apply upstream of any particular application, whether it should involve equilibria or dynamics, materials or informational channel capacities or other, weak or strong correlations, extensivity or not, or any other particularity. We are assured by [2] that no other form of entropy function—beyond those monotonic with H —can generate unbiased inferences.

Acknowledgments

Steve Pressé acknowledges support from the NSF (MCB Award No. 1412259) and the Purdue Research Foundation. Ken A. Dill acknowledges the support of NIH grant 5R01GM090205-02 and the Laufer Center. Kingshuk Ghosh acknowledges support from the Research Corporation for Science Advancement as a Cottrell Scholar.

Author Contributions

All four authors contributed to the research and have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Pressé, S.; Ghosh, K.; Lee, J.; Dill, K. Nonadditive entropies yield probability distributions with biases not warranted by the data. *Phys. Rev. Lett.* **2013**, *111*, 180604.
2. Shore, J.E.; Johnson, R.W. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans. Inf. Theory* **1980**, *26*, 26–37.
3. Tsallis, C. Conceptual inadequacy of the Shore and Johnson axioms for wide classes of complex systems. *Entropy* **2015**, *17*, 2853–2861.
4. Pressé, S.; Ghosh, K.; Lee, J.; Dill, K. Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.* **2013**, *85*, 1115–1141.

5. Pressé, S. Nonadditive entropy maximization is inconsistent with Bayesian updating. *Phys. Rev. E.* **2015**, *90*, 052149.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).