Review of: "Two New Indices for Measuring the Difference Between Two Probability Distributions"

Karoly Heberger[1]

1 Research Centre for Natural Sciences

Potential competing interests: No potential competing interests to declare.

I will not comment on this work further, as the novelty is questionable and my comments were not answered properly:

“If one uses a well-known quantity, as e.g., the Pearson correlation coefficient (r), s/he cannot claim to introduce a novel quantity as its square (r^2), though r and r^2 have different ranges
[1 to +1 and 0 to +1, respectively] and also different meanings [“closeness” to linear relationships (direction included) and explained variance by the model, respectively]. A simple algebraic transformation cannot form the basis of a novelty claim."

Response: **While I agree with your example, I don’t think it is relevant to the distribution discrepancy index (DDI) I proposed.**

Why? Is DDI not a linear combination of DSI?

“It is especially problematic here because the distribution similarity index (DSI), denoted by φ(X1,X2), (X1 and X2 are random variables), has already been defined earlier [Huang 2023]. ”Putting the same quantity on a reverse scale does not amount to defining a new measure; the distribution discrepancy index (DDI) = 1 – φ(X1,X2)."

Response: **The distribution discrepancy index (DDI) was derived based on the concept of infirmity (SIC!) introduced in my previous paper (Huang 2023) and the concept of informity divergence that is introduced in this paper.**

R^2 is also based on a different concept than *r*, i.e., explained variance, and may have a different direction.

Response (cont): **The DDI, which measures the discrepancy, happens to be the complement of the distribution similarity index (DSI) presented in Huang (2023) that measures the similarity.**

Yes, it complements, but it is the same index just in the opposite direction. No novelty is buried in linear combinations of earlier defined terms/indices/concepts. Imagine what would happen to science if all linear combinations of already defined indices would be accepted as new ones.

Further on, Huang (2023) has not been peer reviewed and accepted yet. My suspect is that “informity” is just a new name for old term(s). No refutation or justification has been provided yet. As I understood, DSI has only the advantage to

Kullback–Leibler (KL) *divergence* of being symmetric. Apart from a simple scaling factor (transferring asymmetric indices to symmetric ones), the advantages of symmetric indices are at least dubious, as revealed earlier.

My reservations are still valid:

"It should be evaluated by independent experts:

1. Whether the “informity concept” is new.
2. Is the “informity” a better measure of “informativeness” (whatever it is) than its competitors?
3. Whether the above terms and “cross-informity” and “joint informity” are useful magnitudes and whether they are not linear combinations of earlier, well-known measures.
4. The difference between entropy, hypervolume, *etc.*, and the suggested concepts should be determined, eventually with empirical examinations and comparisons."

A declaration by other experts is not sufficient that “DSI is novel”; proofs and comparisons are needed.

Response (cont): **I think there is nothing wrong with introducing the DDI after introducing the DSI. The DDI and DSI allow us to compare the two distributions from different perspectives, and it is a nice property to be complementary to each other.**

Oh, yes, nothing wrong, except that it adds noise to the existing knowledge. It may confuse superficial scientists, who may use it without considering the direction.

The present version simply hides the linear combination of DSI.

Let us return to discussion only if renowned experts (of not Asian origin) clear the above detailed (and earlier suggested) problems. NOT BEFORE.