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

Neven Elezović

Potential competing interests: No potential competing interests to declare.

The paper must be rejected. The entire concept of DDI and DSI is incorrectly formulated. Specific comments are as follows:

1) The paper discusses the similarity of distributions, separately considering classes of discrete and continuous distributions. This setup alone indicates possible issues. For example, this concept cannot compare the similarity of a binomial distribution B(n, p) with its normal approximation N(np, npq).

2) The Distribution Discrepancy Index (DDI) is always within the interval [0, 1]. The authors propose categorization based on index values of 0.25, 0.5, 0.75, which also applies to the Distribution Similarity Index (DSI) since DSI = 1 - DDI. On what basis were these numbers chosen, and what do they actually represent? Which examples confirm that a similarity index of 0.8 will be suitable for distributions that we perceive as similar? The example provided in the paper does not address this. The authors should have considered, for instance, the approximation of a binomial distribution by a Poisson distribution to get an impression of the DSI coefficient.

3) In the consideration of discrete variables, the authors should limit themselves to variables with integer values for this concept to make any sense. Otherwise, unwanted results are obtained. For example, a discrete random variable X and its shift Y = X + epsilon for an infinitesimally small epsilon have similar distributions in all practical applications. However, according to the formulas in the article, the similarity of these distributions is 0.

4) The only example given for continuous distributions in section 5.2 is completely incorrect. I do not understand how the author can justify the written formulas, as no calculation is provided. What is clear to me is that if Y_1 has a normal distribution N(mu_1, sigma_1^2), then beta(Y_1) = E(p(Y_1)) must depend on the expectation mu_1. The author provides a value that depends only on the deviation sigma_1, which is incorrect. Also, the result for "cross-informity" is incorrect. The derivation of the formula is not provided, so I do not know how the authors arrived at it, but the formula's structure suggests that they considered the difference Y_1 - Y_2 and assumed that Y_1 and Y_2 are independent, which is completely unfounded in the context of this paper.

Conclusion: The paper must be rejected. No simple revision can fix it.