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

Piotr Sulewski[1]

1 Akademia Pomorska w Slupsku

Potential competing interests: No potential competing interests to declare.

A New Index for Measuring the Difference Between Two Probability Distributions

Review

The paper proposes a new index for measuring the difference between two probability distributions, named the distribution discrepancy index (DDI). The DDI is derived based on the concepts of informity, cross-informity, and informity divergence in the recently proposed informity theory. It is defined as the ratio between the informity divergence of two probability distributions and the sum of the two informities.

1. In the introduction, the author lists existing measures. I suggest presenting them in a table along with the name, definition, and possible values (see, e.g., Sulewski, P. (2021). "Equal-bin-width histogram versus equal-bin-count histogram." Journal of Applied Statistics, 48(12), 2092-2111.). In my opinion, the family of measures can be further extended (see also Sahler, W.A. "A survey of distribution.- free statistics based on distances between distribution functions." Metrika Vol. 13, (1968) 149 - 169.)

2. Related to:

$$\beta\left(Y_1 \cap Y_2\right) = \int p_1(y)p_2(y)dy = E_{p_1}\left[p_2(y)\right] = E_{p_2}\left[p_1(y)\right].$$

This formula is not new; it is only the simplified version of the Bhattacharyya formula. (See: "On a measure of divergence between two statistical populations..." Bull. Calcutta Math. Soc. Vol. 35, (1943), pp 99-109.)

3. Related to:

$$D(Y_1, Y_2) = \int [p_1(y) - p_2(y)]^2 dy = \beta\left(Y_1\right) - 2\beta\left(Y_1 \cap Y_2\right) + \beta\left(Y_2\right)$$

This formula is not new; it is the simplified version of the Cramér and Mises formula. (See: Sahler W.A. "A survey of distribution.- free statistics based on distances between distribution functions." Metrika Vol. 13, (1968) 149 - 169.) On the other hand, the integral formula is sufficient.

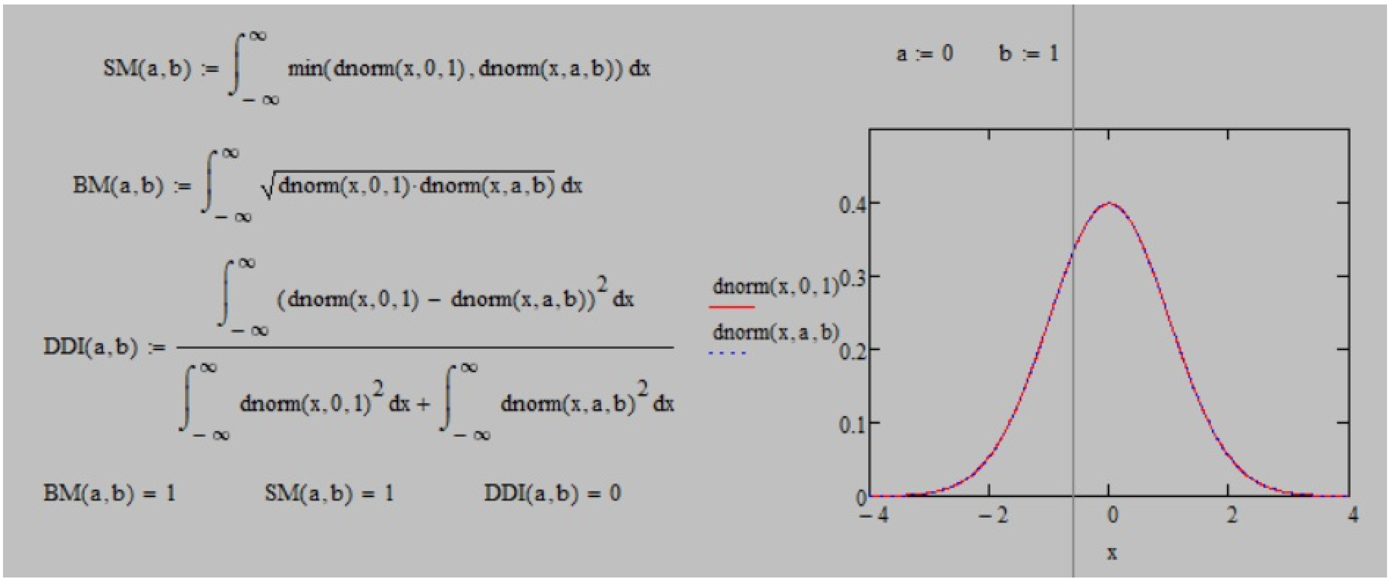
Section 3.1 certainly doesn't introduce anything new. Section 3.2 proposes some modifications to the existing similarity measure.

4. It is more intuitive that if the distributions are the same, then the measure = 1 (see Bhattacharyya and Sulewski formulas).

5. It would be useful to compare the measures in the application examples section.

I did some comparison in Mathcad.

1) If the distributions are the same, everything is OK.



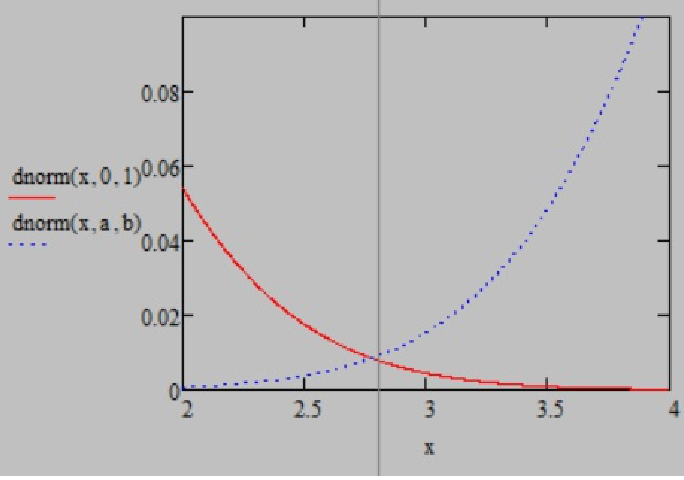
2) The DDI measure does not notice the similarity of very different distributions, unlike the SM and BM measures.

$$SM(a,b) := \int_{-\infty}^{\infty} \min(\mathrm{dnorm}(x,0,1), \mathrm{dnorm}(x,a,b))\, dx$$

$$a := 5.55 \quad b := 1$$

$$BM(a,b) := \int_{-\infty}^{\infty} \sqrt{\mathrm{dnorm}(x,0,1)\cdot \mathrm{dnorm}(x,a,b)}\, dx$$

$$DDI(a,b) := \frac{\int_{-\infty}^{\infty} (\mathrm{dnorm}(x,0,1) - \mathrm{dnorm}(x,a,b))^2\, dx}{\int_{-\infty}^{\infty} \mathrm{dnorm}(x,0,1)^2\, dx + \int_{-\infty}^{\infty} \mathrm{dnorm}(x,a,b)^2\, dx}$$



$$BM(a,b) = 0.021 \qquad SM(a,b) = 0.006 \qquad DDI(a,b) = 1$$

Summary

In my opinion, the article requires very significant changes. The article certainly does not propose a new measure (title to be changed), but a certain modification of the existing measure. The author refers to his other article, which, however, has not been published. The numerical example presented by the Reviewer shows that the DDI measure is worse than two already existing measures.