Review of: "A Proposed Heuristic for Guessing Distributions"

Robert Veszteg

Potential competing interests: No potential competing interests to declare.

The proposed method for making a guess tries to find a possible probability distribution by matching the observed median to the theoretical median of all probability distributions under consideration.

In the two examples described in the paper, the “layman” decision-maker is looking at a very specific family of distributions. In example 1, they are all symmetric. In example 2, it is the family of Pareto distributions.

So who exactly chooses that family of distributions for the layman who then needs to match medians? In other words, what exactly do wee need to assume about the decision-maker's statsitical sophistication for this model to work?

Later, for example 1, the authors claim that the chosen distribution is the best, becasue it seems to be in line with the observed income distribution from the United States. The trouble is that showing a single example where the method works is not a conclusive evidence. Using the data from the the very same link (3) that the authors are using, I have confirmed that for the real/observed US income distribution in 2021, distribution B performs better than disbribution A (the multinomial probabilities are 0.0974 vs 0.0799). But for 1971 that would be the exact opposite: distribution A fits better than B (0.1242 vs 0.0584). There is nothing in the proposed model that would help the decision-maker the correct guess for different problems, different years, different countries, etc.

Relying on a simple criterion, e.g. the median, will not make wonders. The article in its second version is still very much misleading as it hides important assumptions. I suggest the authors to make major changes and take a much more rigorous approach.