# Comment on Glenn Shafer's *Testing by betting: A strategy for statistical and scientific communication*

Dr. Ryan Martin
North Carolina State University
Raleigh, North Carolina, USA
rgmarti3@ncsu.edu

September 22, 2020

Statistics plays a central role in science, so it's imperative that scientists can communicate statistical ideas clearly. I commend Shafer for his efforts in this direction, but my opinions differ somewhat. Coincidentally, my views have been heavily influenced by Glenn's work in a different direction, including a paper that was read before The Royal Statistical Society almost 40 years ago (Shafer 1982).

Shafer's betting score proposal is largely motivated by concerns that p-values are "too complicated for effective communication" (Sec. 1). But the use of p-values is statistically well-founded: "the p-value function provides the full statistical story for the particular data value relative to the model" (Fraser 2013, p. 43). So, if p-values are right and "will never completely disappear" (Sec. 3), then wouldn't a new betting language make scientific communication even more complicated? Wouldn't it be better to understand what p-values are and how they should be interpreted?

Fortunately, p-values can be understood using Shafer's plausibility theory (e.g., Shafer 1976). Martin and Liu (2014) showed that, under certain conditions, every p-value corresponds to the *plausibility* of the hypothesis, based on the given data, relative to a valid *inferential model* (Martin and Liu 2013, 2015); similar characterizations of confidence and conformal prediction are given in Martin (2017) and Cella and Martin (2020), respectively. Consequently, the colloquial definition of p-values we teach our students, i.e., *a p-value measures the plausibility of the hypothesis*, is mathematically justified. Moreover, interpretation is straightforward: a small p-value indicates the hypothesis is implausible, while a large p-value means the hypothesis is plausible which, e.g., does not imply that the data provides evidence supporting the hypothesis.

Concerning efficiency, statistical procedures based on p-values are often best. Since Shafer's betting scores must be smaller than $(\text{p-value})^{-1}$, his betting score-based procedures (Sec. 4.2) are less efficient than the p-valued-based procedures. Conservative solutions might be warranted under certain circumstances, but should we make across-the-board sacrifices in efficiency for the betting score interpretation?

Finally, I wonder about the use of betting scores when scientists aren't actually bet-

ting. Shafer's proposed language must lose some of its meaning if those using it are only going through the motions. There are good reasons to make betting a real part of the scientific process (e.g., Crane 2018) and Shafer's proposal would be much more convincing in such a framework.

# References


Cella, L. and Martin, R. (2020). Valid distribution-free inferential models for prediction. `https://www.researchers.one/article/2020-01-12`.

Crane, H. (2018). The fundamental principle of probability: Resolving the replication crisis with skin in the game. `www.researchers.one/article/2018-08-16`.

Fraser, D. A. S. (2013). Discussion: "Confidence distribution, the frequentist distribution estimator of a parameter: a review" [mr3047496]. *Int. Stat. Rev.*, 81(1):42–48.

Martin, R. (2017). A mathematical characterization of confidence as valid belief. `https://www.researchers.one/article/2018-07-3`.

Martin, R. and Liu, C. (2013). Inferential models: a framework for prior-free posterior probabilistic inference. *J. Amer. Statist. Assoc.*, 108(501):301–313.

Martin, R. and Liu, C. (2014). A note on p-values interpreted as plausibilities. *Statist. Sinica*, 24(4):1703–1716.

Martin, R. and Liu, C. (2015). *Inferential Models: Reasoning with Uncertainty*, volume 147 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, N.J.

Shafer, G. (1982). Belief functions and parametric models. *J. Roy. Statist. Soc. Ser. B*, 44(3):322–352. With discussion.