Discussion contribution

Judith ter Schure's contribution to the Discussion of 'Testing by betting: A strategy for statistical and scientific communication' by Glenn Shafer

Judith ter Schure
CWI, Amsterdam, The Netherlands

Correspondence: Judith ter Schure, CWI, Amsterdam, The Netherlands.
Email: Judith.ter.schure@cwi.nl

Professor Shafer has given a wonderful introduction into the usefulness of 'testing by betting' for interpreting statistical inferences. His talk provided examples of testing weather forecasters and political pundits, showcasing an approach that is more general than conventional statistical tests. Full generality, however, always leaves the applied statistician wondering how to use the approach, which reflects much of the deliberation in the chat.

Applied statisticians, like myself, might find it difficult to specify a bet on an outcome space. On the one hand, we think in types of data (e.g. time-to-event) and accompanying test statistics (e.g. logrank statistic) that are defined in terms of certain model parameters (e.g. hazard ratio). On the other hand, we are used to improving our designs by evaluating the rejection region in a power analysis. Yet instead of thinking of our bets as having power for an *assumed* alternative, we think of them as having power for parameter values specifying effects *we are not willing to miss*.

Bets in terms of these parameters provide a clear instruction to develop new statistical methodology (e.g. the t-test, contingency table test and logrank test, which we were able to design [1,2]): bets should make a profit for parameters of minimal interest, but not for any distribution in the null hypothesis. Nevertheless, staying close to statistical practice does call for a slight relaxation of 'opportunistic betting'. Unlike the Bayesian, we might not wish to use prior knowledge, or optimize our 'prequential' forecasts. Instead of focusing on the parameter values we assume most likely to be true – based on prior knowledge – we might wish to bet on the values for which, if they are true, we most dearly want this to show in our betting profit.

Our bets might deliberately miss out on profit for effects smaller than our smallest effect size of interest. Hence the choice of bets poses a challenge for retrospective meta-analysis (which we try to address in the ALL-IN meta-analysis philosophy [3]): each study can have its own protocol and effect of minimal interest.

I am already convinced that it will serve statistics well to replace p-values by bets, and power analyses by implied targets. But how do we know whether practitioners actually find this more intuitive? As an applied statistician, I will keep that question from the chat in mind as well, and see if I can design an experiment to test it.

REFERENCES

[1] Grünwald, P., de Heide, R., & Koolen, W. (2019). Safe testing. *arXiv preprint arXiv:1906.07801*.

[2] ter Schure, J., Pérez-Ortiz, M.F., Ly, A. & Grünwald, P.D. (2021) The Safe Logrank Test: Error Control under Continuous Monitoring with Unlimited Horizon. *arXiv preprint arXiv:2011.06931.*

[3] ter Schure, J. & Grünwald, P. (2021) ALL-IN meta-analysis: soon to appear as a preprint.