Valid sequential inference on probability forecast performance

Alexander Henzi[1*]; Johanna F. Ziegel[1]

alexander.henzi@stat.unibe.ch   johanna.ziegel@stat.unibe.ch

[1]   University of Bern, Bern, Switzerland

Abstract:
Probability forecasts for binary events play a central role in many applications. Their quality is commonly assessed with proper scoring rules, which assign forecasts a numerical score such that a correct forecast achieves a minimal expected score. In this paper, we construct e-values for testing the statistical significance of score differences of competing forecasts in sequential settings. E-values have been proposed as an alternative to p-values for hypothesis testing, and they can easily be transformed into conservative p-values by taking the multiplicative inverse. The e-values proposed in this article are valid in finite samples without any assumptions on the data generating processes. They also allow optional stopping, so a forecast user may decide to interrupt evaluation taking into account the available data at any time and still draw statistically valid inference, which is generally not true for classical p-value based tests. In a case study on postprocessing of precipitation forecasts, state-of-the-art forecasts dominance tests and e-values lead to the same conclusions.

Keywords:
Probability forecast; forecast dominance; proper scoring rule; consistent scoring function; e-value