

PROPER SCORING RULES, CALIBRATION AND SHARPNESS: ASSESSING PREDICTIONS FOR AN UNCERTAIN WORLD

Tilmann Gneiting

Department of Statistics, University of Washington, Seattle, WA

One of the major purposes of statistical analysis is to make forecasts for the future, and to provide suitable measures of the uncertainty associated with them. Consequently, predictions ought to be probabilistic in nature, taking the form of probability distributions over future quantities or events.

The goal of probabilistic forecasting is to maximize the sharpness of the predictive distributions subject to calibration. Calibration refers to the statistical consistency between the distributional forecasts and the observations. Sharpness relates to the concentration of the predictive distributions: The sharper the distributional forecasts, the less the uncertainty; and the sharper, the better; subject to calibration.

Scoring rules are omnibus performance measures that assign a numerical reward or penalty based on the predictive distribution and the event or value that materializes. A proper scoring rule addresses calibration and sharpness simultaneously, and encourages the forecaster to reveal her true beliefs.

I will review diagnostic tools for the assessment of calibration and sharpness, and discuss the use of proper scoring rules in comparing and ranking probabilistic prediction techniques in environmental, epidemiological and biomedical applications.