Geophysical Research Abstracts
Vol. 18, EGU2016-1613, 2016
EGU General Assembly 2016

`scoringRules` – A software package for probabilistic model evaluation

Sebastian Lerch (1,2), Alexander Jordan (2), and Fabian Krüger (2)

(1) Institute of Stochastics, Karlsruhe Institute of Technology, Karlsruhe, Germany (sebastian.lerch@h-its.org), (2) Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

Models in the geosciences are generally surrounded by uncertainty, and being able to quantify this uncertainty is key to good decision making. Accordingly, probabilistic forecasts in the form of predictive distributions have become popular over the last decades. With the proliferation of probabilistic models arises the need for decision theoretically principled tools to evaluate the appropriateness of models and forecasts in a generalized way.

Various scoring rules have been developed over the past decades to address this demand. Proper scoring rules are functions $S(F, y)$ which evaluate the accuracy of a forecast distribution F, given that an outcome y was observed. As such, they allow to compare alternative models, a crucial ability given the variety of theories, data sources and statistical specifications that is available in many situations.

This poster presents the software package `scoringRules` for the statistical programming language `R`, which contains functions to compute popular scoring rules such as the continuous ranked probability score for a variety of distributions F that come up in applied work. Two main classes are parametric distributions like normal, t, or gamma distributions, and distributions that are not known analytically, but are indirectly described through a sample of simulation draws. For example, Bayesian forecasts produced via Markov Chain Monte Carlo take this form. Thereby, the `scoringRules` package provides a framework for generalized model evaluation that both includes Bayesian as well as classical parametric models.

The `scoringRules` package aims to be a convenient dictionary-like reference for computing scoring rules. We offer state of the art implementations of several known (but not routinely applied) formulas, and implement closed-form expressions that were previously unavailable. Whenever more than one implementation variant exists, we offer statistically principled default choices.