Statistical Reproducibility of Hypothesis Tests

Frank Coolen[1][*]

[1]Department of Mathematical Sciences, Durham University, Durham, DH1 3LE, UK
frank.coolen@durham.ac.uk

Abstract

In recent years there has been much attention to reproducibility of important test outcomes in many areas of applied statistics. Many reported results in the literature turned out not to hold up when experiments were repeated. There are many reasons for this, including the publication practices, and these have been widely discussed. Surprisingly, perhaps, there has been little attention to implicit variability of the statistical test methods themselves: if a test were repeated under the same circumstances, would it lead to the same result for a hypothesis test? We consider this topic of statistical reproducibility as a prediction problem in a frequentist statistics setting, using nonparametric methods to add rather minimal assumptions so that the inferences on reproducibility are directly based on the actual test observations.

In this presentation, I will briefly discuss some general issues of test reproducibility, followed by an introduction to nonparametric predictive inference for statistical hypothesis testing. I will discuss examples of several statistical tests and address practical and computational issues. The presentation will be partly based on the papers below, and will also include results and insights from ongoing research and applications.

References

[1] Coolen, F.P.A. and Bin Himd, S. (2014). *Nonparametric predictive inference for reproducibility of basic nonparametric tests.* Journal of Statistical Theory and Practice 8, 591-618.

[2] Coolen, F.P.A. and Alqifari, H.N. (2018). *Nonparametric predictive inference for reproducibility of two basic tests based on order statistics.* REVstat - Statistical Journal 16, 167-185.

[3] Marques, F.J., Coolen, F.P.A. and Coolen-Maturi, T. (2019). *Introducing nonparametric predictive inference methods for reproducibility of likelihood ratio tests.* Journal of Statistical Theory and Practice 13, article 15.

[4] Coolen, F.P.A. and Bin Himd, S. (2020). *Nonparametric predictive inference bootstrap with application to reproducibility of the two-sample Kolmogorov-Smirnov test.* Journal of Statistical Theory and Practice 14, article 26.

[5] Coolen, F.P.A. and Marques, F.J. (2020). *Nonparametric predictive inference for test reproducibility by sampling future data orderings.* Journal of Statistical Theory and Practice 14, article 62.

---

[*]Corresponding author.