

Tim B.

The importance of negative controls in conditional independence testing

Conditional independence (CI) tests are tests of association between two variables X and Y (e.g., a genetic variant and a phenotype) while controlling for a vector of confounders Z (e.g., population structure). CI tests are among the most fundamental and widely-used hypothesis tests in statistics and the sciences. Despite their importance and popularity, CI tests are hard to conduct: recent theoretical results have revealed that assumption-free CI testing is fundamentally impossible (except in the special case where Z is discrete and takes a small number of values; Kim et al. 2021; Shah and Peters 2020). Unfortunately, the assumptions that underly CI tests typically are not checked in practice. This is not due to negligence; as we will see, checking the assumptions of CI tests is fraught with difficulties, especially in high dimensions. We therefore find ourselves faced with an apparent paradox: (i) CI testing is of fundamental importance; (ii) we must check the assumptions of CI tests in applying CI tests to data (else, we are not justified in believing the results); and (iii) checking the assumptions of CI tests is challenging.

We aim to address this challenge in the contemporary “high-multiplicity” setting in which we test thousands or tens of thousands (or more) of hypotheses with the objective of producing a discovery set that controls the false discovery rate (FDR; Benjamini and Hochberg 1995; Li and E. J. Candès 2021). Our core thesis is that “negative control data” – example data for which the null hypothesis is known to be true, roughly – are crucially important (and in some cases *required*) for CI testing with FDR control.

After briefly summarizing the vast landscape of CI testing procedures, we argue that negative controls play (or ought to play) a crucial role in FDR-controlled CI testing. We describe two broad types of negative controls — “experimental” and “in silico” — and argue that, although the former are superior statistically, the latter are more widely available and can be constructed directly from the data in many applications. Next, we introduce several ideas for working effectively with negative control data for the purposes of controlling FDR. In particular, we propose to calibrate the testing procedure against *both* the empirical negative control distribution *and* the theoretical null distribution that would obtain if the assumptions of the testing procedure were satisfied; this approach satisfies an appealing double-robustness property. Also, we introduce a practical method for assessing

whether a given testing procedure is robust to inflation observed in the tails of the null distribution. Finally, to facilitate these analyses, we describe the “symmetry plot,” a nonparametric analogue of the commonly-used quantile-quantile plot for assessing calibration.

As an auxiliary contribution, we introduce a new class of fast and powerful Gaussian test statistics compatible with a broad range of existing CI testing algorithms, including the conditional randomization test, the conditional permutation test, and the local permutation test (E. Candès et al. 2018; Berrett et al. 2020; Kim et al. 2021). These test statistics involve repeatedly fitting OLS, ridge regression, or additive spline models to the permuted (or resampled) data via an online QR decomposition-based procedure.

References

- Benjamini, Yoav and Yosef Hochberg (1995). “Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing”. In: *Journal of the Royal Statistical Society* 57.1, pp. 289–300.
- Berrett, Thomas B. et al. (2020). “The conditional permutation test for independence while controlling for confounders”. In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 82.1, pp. 175–197.
- Candès, Emmanuel et al. (2018). “Panning for gold: ‘model-X’ knockoffs for high dimensional controlled variable selection”. In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 80.3, pp. 551–577.
- Kim, Ilmun et al. (2021). “Local permutation tests for conditional independence”. In:
- Li, Shuangning and Emmanuel J. Candès (2021). “Deploying the Conditional Randomization Test in High Multiplicity Problems”. In: pp. 1–43.
- Shah, Rajen D. and Jonas Peters (2020). “The hardness of conditional independence testing and the generalised covariance measure”. In: *Annals of Statistics* 48.3, pp. 1514–1538.