Tim B.

The need for negative controls in conditional independence testing

- Conditional independence tests (CI tests) assess the association between two variables (e.g., a genetic variant and a phenotype) while controlling for one or more confounding variables (e.g., population structure). CI tests are among the most widely-used and fundamental hypothesis tests in applied statistics and science.

- Recent results have indicated that assumption-free conditional independence testing is impossible: all valid CI tests must make an assumption or set of assumptions about the data-generating mechanism (except for in the trivial case where the confounding vector Z is discrete and takes a small number of values).

- In practice, these assumptions are rarely checked, especially in so-called "high-multiplicity" settings in which there are tens of thousands (or more) of hypotheses.

- We argue that negative controls — — are required in virtually applications of conditional independence tests.

- Conceptual thoughts.

  - **Breaking the limits of black box-based inference with negative controls**. Black boxes can be leveraged for predictive inference (via conformal prediction) and independence testing (via permutation tests) without assumptions beyond i.i.d. However, CI testing is more difficult. We must inject additional information into the system — in the form of negative controls — to solve the problem with black boxes. We must supplement the black boxes with external, negative control information to conduct reliable conditional independence tests; the black boxes alone are unsatisfactory.

  - **The blessings of multiplicity**. It is unclear how one would test the assumptions of model-free CI methods (e.g., local permutation test) when there are only a few hypotheses to test. However, when there are many hypotheses (including many negative control hypotheses), we check assumptions by fitting the model to negative controls and checking for calibration.

- Now that we have demonstrated the importance of negative controls, we explore how to use them effectively. We propose to use a doubly robust strategy for calibrating the test statistic. Let $\hat{\mathcal{L}}_n$ be the estimated empirical null distribution. Let $\mathcal{L}_t$ denote the theoretical null distribution that would result if the assumptions were satisfied. We want our procedure to control FDR for *both* $\hat{\mathcal{L}}_n$ and $\mathcal{L}_t$. This strategy avoids calibration against the empirical null distribution, which is fraught with difficulties.

- We commonly use qq-plots to assess the distribution of the test statistic under the null hypothesis (typically, this is a Gaussian distribution or uniform distribution over the integers $\{1, \ldots, B\}$). QQ-plot checking is necessary for methods that operate on p-values (e.g., BH). We define an analogous kind of plot called a "symmetry plot" (or s-plot) for methods that operate more generally on symmetric test statistics (e.g., BC). Basically, the variable $X \sim \mathcal{L}_n$ should satisfy the property $-|X| = |X|$. Therefore, let $x^-_{(1)}, x^-_{(2)}, \ldots, x^-_{(r_1)}$ be the ordered test statistics that are less than zero (so that $x^-_{(1)} \geq x^-_{(2)} \geq \cdots \geq x^-_{(r_1)}$), and let $x^+_{(1)}, x^+_{(2)}, \ldots, x^+_{(r_2)}$ be the ordered test statistics that are greater than zero $x^+_{(1)} \leq x^+_{(2)} \leq \ldots x^+_{(r_2)}$. We plot these ordered quantities against one another.

- We can use symmetry-preserving transformations on $N(0, 1)$ to improve the calibration of the empirical null distribution (e.g., binning). For example, there is not much mass in the tails. Thus, we can bin all observations less than $-s$ or greater than s for some $s \in \mathbb{R}$ and get the tails looking more symmetric. We even can verify the improvement using an s-plot.