

## Paper proposal

Tim B

In this document I chart out several possible directions for a next statistics paper. The following keywords are relevant: randomization tests, conditional randomization/permutation tests,  $e$ -values, multiple hypothesis testing, testing/training on the same data, and sample splitting. I propose two main directions and a few tertiary directions.

## Direction 1: A more general framework for randomization tests

### Motivation

Randomization tests are statistical tests in which a test statistic is recomputed over permuted, resampled, rotated, or otherwise transformed versions of the data to produce an empirical null distribution against which a statistic computed on the raw data is compared. Randomization tests are ubiquitous throughout all of statistics and science. For example, randomization tests commonly are used in neuroscience to perform nonparametric inference on GLMs (Winkler et al. 2014) and in genetics to control for population stratification in GWAS (). See Dobriban 2021 for a recent review and theoretical analysis of randomization tests. Despite their widespread popularity, randomization tests pose several practical challenges, especially in high-multiplicity settings: miscalibrated  $p$ -values can cause type-I error inflation; combining  $p$ -values from many hypotheses can result in excessively conservative or liberal discovery sets, especially when tests are dependent; and sample splitting – a procedure required by certain randomization tests, such as the recently-proposed holdout randomization test (CITE) – results in non-reproducible  $p$ -values.

Building closely on the work of several authors (e.g., Wang and Ramdas 2020; Vovk 2020) we propose a simple new framework for randomization tests that helps to resolve these challenges. The framework leverages  $e$ -values, test statistics that by definition have unit expectation under the null. We recover the standard,  $p$ -value based approach to randomization tests a special case of our framework. Although our theory applies broadly, we focus mostly on (marginal) permutation tests, conditional randomization tests, and conditional permutation tests as illustrative examples.

Let  $T^*$  be the test statistic computed on the raw data, and let  $T_1, T_2, \dots, T_B$  be the *ordered* test statistics recomputed on the permuted (or resampled, etc.) data. Define  $I_i = \mathbb{I}(T^* \leq T_i)$ . For given constants  $a_0, a_1, \dots, a_B \in \mathbb{R}$ , define  $e$  by

$$e = a_0 + \sum_{i=1}^B a_i I_i,$$

where  $e$  is subject to the constraint that  $\mathbb{E}[e] = 1$ . In other words,  $e$  is an  $e$ -value that is a linear combination of the  $I_i$ s. We call  $e$  a “randomization test  $e$ -value (RT  $e$ -value)” so as to distinguish it from other, more general  $e$ -values.

## Constructing test statistics $e$

We consider methods for constructing RT  $e$ -values, leveraging two key properties of the  $I_i$ s. First, because  $T^*$  is i.i.d. with the  $T_i$ s, we have that  $\mathbb{P}(T^* \leq T_i) = i/B$ . Therefore,  $\mathbb{E}[I_i] = i/B$ . The second key property we state as a proposition.

**Proposition 1** *For  $r, B \in \mathbb{N}$ , we have that*

$$\left( \sum_{i=1}^B I_i \right)^r = \sum_{i=1}^B [(B - i + 1)^r - (B - i)^r] I_i.$$

*Equivalently, the  $r$ th power of the empirical  $p$ -value  $p_B := \frac{1}{B} \sum_{i=1}^B I_i$  is*

$$p_B^r = \sum_{i=1}^B [(1 - i/B + 1/B)^r - (1 - i/B)^r] I_i := \sum_{i=1}^B M(B, r, i) I_i.$$

*Finally, for given a  $x_0 \in \mathbb{R}$  and coefficients  $c_0, c_1, \dots, c_B$ , the polynomial  $\sum_{j=0}^r a_j (p_B - x_0)^j$*

In other words, the  $r$ th power of the sum of the  $I_i$ s is an extremely simple linear combination of the  $I_i$ s, not some complicated multinomial expression. The proof of Proposition 1 is a combinatorial argument (see appendix); I have not seen this result – although it is simple – derived elsewhere.

## Direction 2: A class of fast and powerful test statistics for the conditional randomization (permutation) test

## References

- Dobriban, Edgar (2021). “Consistency of invariance-based randomization tests”. In: pp. 1–35 (cit. on p. 1).
- Vovk, Vladimir (2020). “A note on data splitting with  $e$ -values”. In: pp. 1–11 (cit. on p. 1).
- Wang, Ruodu and Aaditya Ramdas (2020). “False discovery rate control with  $e$ -values”. In: pp. 1–32 (cit. on p. 1).
- Winkler, Anderson M., Gerard R. Ridgway, Matthew A. Webster, Stephen M. Smith, and Thomas E. Nichols (2014). “Permutation inference for the general linear model”. In: *NeuroImage* 92, pp. 381–397 (cit. on p. 1).