

Statistics paper proposal

Tim B

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

Direction 1: A more general framework for randomization tests

Motivation

Randomization tests are tests in which a 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 routinely are used in neuroscience to perform nonparametric inference on GLMs (Winkler et al. 2014) and in genetics to test the association of SNPs (Johnson et al. 2010). See Dobriban 2021 for a recent review and unifying 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 across many hypothesis tests 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 (Tansey et al. 2021) – can result in non-reproducible p -values.

Building closely on the work of several authors (Wang and Ramdas 2020; Vovk 2020; Vovk and Wang 2021) we propose a simple new framework for randomization testing that helps to resolve these challenges. The framework leverages e -values, test statistics that by definition have (at most) unit expectation under the null. We capture the standard, p -value based approach to randomization testing a special case of our framework. Although the 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. We assume for simplicity that there are no ties among the T_i s, although this assumption does not impact the theory much. Define $I_i = \mathbb{I}(T^* \leq T_i)$. For given constants $a_0, a_1, \dots, a_B \in \mathbb{R}$, define the test statistic e by

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

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

Two key properties of the I_i s

We state two key properties of the I_i s that enable us to construct linear e -values. We defer the proofs – which are combinatorial in nature – to the appendix. The first proposition gives the expectation of I_i .

Proposition 1 *Let $T^*, X_1, \dots, X_B \sim F$ for some distribution F . Let T_1, \dots, T_B be the order statistics of the X_i s. For $i \in \{1, \dots, B\}$,*

$$\mathbb{P}(T^* \leq T_i) = \frac{i}{B+1}.$$

Therefore, setting $I_i = \mathbb{I}(T^ \leq T_i)$, it follows $E[I_i] = i/(B+1)$.*

Intuitively, the T_i s define $B+1$ “slots”: the slot before T_1 , the slot after T_1 and before T_2 , and so on, up until the slot after T_B . The probability that T^* is less than T_i is equal to the probability that T^* lands in one of the first i slots, which is $i/(B+1)$. The second proposition derives a simple expression for the r th power of the sum of the I_i s.

Proposition 2 *For $r, B \in \mathbb{N}$,*

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

Equivalently, the r th power of the empirical right-sided 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 $x_0 \in \mathbb{R}$ and coefficients $c_0, c_1, \dots, c_r \in \mathbb{R}$, the r th degree polynomial $\sum_{j=0}^r c_j (p_B - x_0)^j$ is given by

$$\sum_{j=0}^r c_j (p_B - x_0)^j = \left(a_0 + \sum_{j=1}^r a_j x_0^j \right) + \sum_{i=1}^B \left[\sum_{j=1}^r \sum_{k=1}^j a_j x_0^{j-k} \binom{j}{k} M(B, k, i) \right] I_i.$$

In other words, the r th power of the sum of the I_i s is a simple linear combination of the I_i s, not a messy multinomial expression as one might initially expect. Equivalently, the r th power of the empirical right-sided p -value (i.e., $p_B = (1/B) \sum_{i=1}^B I_i$) is a simple linear combination of the I_i s. Finally, the r th degree polynomial of p_B centered at x_0 with coefficients c_0, \dots, c_r is a straightforward extension of the above.

Constructing linear e -values

We give several examples of constructing linear e -values.

Example 1: Warmup.

Example 2: Recovering p -values as a special case.

Example 3: Powerful linear e -values via p -to- e calibrators. We outline a general and flexible strategy for producing powerful linear e -values. Vovk and Wang introduced the notion of a p -to- e calibrator, which is a function that transforms a p -value into an e -value (Vovk and Wang 2021). A p -to- e calibrator f satisfies the following properties: (i) f is defined on $[0, 1]$; (ii) f is nonnegative; (iii) f is decreasing; and (iv) f integrates to (at most) 1. Vovk and Wang furthermore defined the notion of an “admissible” p -to- e calibrator, which roughly speaking is a calibrator that is “good” (or at least “not bad”), i.e. it yields powerful e -values. An admissible calibrator f satisfies the following additional properties: (i) f is left-continuous; (ii) $f(0) = \infty$; and (iii) f integrates to *exactly* 1. To illustrate the above definitions, let f be an admissible p -to- e calibrator, and let $P \sim U(0, 1)$ be a uniformly distributed random variable with cumulative density function $G(p) = p$. We have by LOTUS that $\mathbb{E}[f(P)] = \int_0^1 f(p) dG(p) = \int_0^1 f = 1$.

Vovk, Wang, and Ramdas introduced several examples of p -to- e calibrators. First, for given $\kappa \in (0, 1)$, define the function $f_\kappa : [0, 1] \rightarrow \mathbb{R}^{\geq 0}$ by $f_\kappa(p) = \kappa p^{\kappa-1}$. It is easy to check that f_κ satisfies the properties listed above and is therefore an admissible p -to- e calibrator. Next, define $g : [0, 1] \rightarrow \mathbb{R}^{\geq 0}$ by

$$g(p) = \frac{1 - p + p \ln(p)}{p \log^2(p)}.$$

The function g also is an admissible p -to- e calibrator, although this is a bit harder to check.

We propose a general procedure for constructing linear e -values using p -to- e calibrators. The j th derivative of the p -to- e calibrator f_κ is

$$f_\kappa^{(j)}(p) = \prod_{i=0}^j (i - j) p^{\kappa-j-1}.$$

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

1 Other considerations

References

- Dobriban, Edgar (2021). “Consistency of invariance-based randomization tests”. In: pp. 1–35 (cit. on p. 1).
- Johnson, Randall C. et al. (2010). “Accounting for multiple comparisons in a genome-wide association study (GWAS)”. In: *BMC Genomics* 11.1, pp. 2–7 (cit. on p. 1).
- Sury, B., Tianming Wang, and Feng-Zhen Zhao (2004). “Identities Involving Reciprocals of Binomial Coefficients”. In: *Journal of Integer Sequences* 7.2, p. 3 (cit. on p. 5).
- Tansey, Wesley, Victor Veitch, Haoran Zhang, Raul Rabadan, and David M. Blei (2021). “The Holdout Randomization Test for Feature Selection in Black Box Models Wesley”. In: *Journal of Computational and Graphical Statistics*, pp. 1–12 (cit. on p. 1).
- Vovk, Vladimir (2020). “A note on data splitting with e -values”. In: pp. 1–11 (cit. on p. 1).

Vovk, Vladimir and Ruodu Wang (2021). “E-values: Calibration, combination and applications”. In: *Annals of Statistics* 49.3, pp. 1736–1754 (cit. on pp. 1, 3).

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).

Appendices

Proposition 1 proof

Suppose without loss of generality that F is supported on the real line (the proof is similar in the case of bounded F). For notational simplicity, let $Y = T^*$. Denote the density of F by f (i.e., $f = F'$). Denote the density of Y , T_i , and (Y, T_i) by $f_Y = f$, f_{T_i} , and f_{Y, T_i} respectively. Standard distributional theory on order statistics indicates that the CDF F_{T_i} of T_i is given by

$$F_{T_i}(t) = \sum_{j=i}^B \binom{B}{j} [F(t)]^j [1 - F(t)]^{B-j}.$$

Meanwhile, the CDF F_Y of Y is given by $F_Y = F$. We seek to show that $\mathbb{P}(Y \leq T_i) = i/(B+1)$. We have that

$$\begin{aligned} \mathbb{P}(Y \leq T_i) &= \int_{-\infty}^{\infty} \int_y^{\infty} f_{Y, T_i}(y, t) dt dy = \int_{-\infty}^{\infty} \int_y^{\infty} f_Y(y) f_{T_i}(t) dt dy \\ &= \int_{-\infty}^{\infty} f(y) \left[\int_y^{\infty} f_{T_i}(t) dt \right] dy = \int_{-\infty}^{\infty} f(y) [F_{T_i}(\infty) - F_{T_i}(y)] dy = 1 - \int_{-\infty}^{\infty} f(y) F_{T_i}(y) dy. \end{aligned} \tag{1}$$

The problem reduces to evaluating the integral in the final equality of (1). We have that

$$\begin{aligned} \int_{-\infty}^{\infty} f(y) F_{T_i}(y) dy &= \int_{-\infty}^{\infty} f(y) \sum_{j=i}^B \binom{B}{j} F^j(y) [1 - F(y)]^{B-j} dy \\ &= \int_{-\infty}^{\infty} f(y) \sum_{j=i}^B \binom{B}{j} F^j(y) \sum_{k=0}^{B-j} \binom{B-j}{k} (-1)^k F^k(y) dy \\ &= \sum_{j=i}^B \binom{B}{j} \sum_{k=0}^{B-j} \binom{B-j}{k} (-1)^k \int_{-\infty}^{\infty} f(y) F^{j+k}(y) dy \\ &= \sum_{j=i}^B \binom{B}{j} \sum_{k=0}^{B-j} \binom{B-j}{k} (-1)^k \left[\frac{1}{1+j+k} F^{j+k+1}(y) \Big|_{-\infty}^{\infty} \right] \end{aligned}$$

$$= \sum_{j=i}^B \binom{B}{j} \sum_{k=0}^{B-j} \binom{B-j}{k} (-1)^k \left[\frac{1}{1+j+k} \right]. \quad (2)$$

Sury et al. 2004 (Corollary 2.2) showed that

$$\sum_{k=0}^{B-j} (-1)^k \binom{B-j}{k} \frac{1}{1+j+k} = \frac{1}{(B+1) \binom{B}{j}}.$$

Substituting the above into (2), we obtain.

$$\begin{aligned} \sum_{j=i}^B \binom{B}{j} \sum_{k=0}^{B-j} \binom{B-j}{k} (-1)^k \left[\frac{1}{1+j+k} \right] = \\ \sum_{j=i}^B \binom{B}{j} \frac{1}{(B+1) \binom{B}{j}} = \sum_{j=i}^B \frac{1}{B+1} = \frac{B-i+1}{B+1}. \end{aligned} \quad (3)$$

Finally, combining (1), (2), and (3), we conclude that

$$\mathbb{P}(Y \leq T_i) = \mathbb{P}(T^* \leq T_i) = \frac{B+1}{B+1} - \frac{B-i+1}{B+1} = \frac{i}{B+1}.$$

Proposition 2 proof

Define the set $C(r, B)$ by

$$C(r, B) := \left\{ (k_1, \dots, k_B) \in \{0, \dots, r\}^B : \sum_{i=1}^B k_i = r \right\},$$

i.e., $C(r, B)$ is the set of length- B tuples of integers from 0 to r such that the elements of the tuple sum to r . Next, let the function $\tau : C(r, B) \rightarrow \{1, \dots, B\}$ be defined by

$$\tau(k_1, \dots, k_B) = \min \{i \in \{1, \dots, B\} : k_i \geq 1\},$$

i.e., $\tau(k_1, \dots, k_B)$ is the position of the minimal nonzero element of a (k_1, \dots, k_B) . Finally, let τ^{-1} be the pre-image of τ . It is easy to see that, for $i \in \{0, \dots, B\}$,

$$\tau^{-1}(i) = \{(0, \dots, 0, k_i, k_{i+1}, \dots, k_B) \in C(r, B) : k_i \geq 1\}.$$

In other words, $\tau^{-1}(i)$ is the set of tuples whose first nonzero entry is the i th entry.

Before proceeding, we establish an important property of the I_i s. If at least one of the k_i s is nonzero, we have that

$$\prod_{i=1}^B I_i^{k_i} = I_{\tau(k_1, \dots, k_B)}. \quad (4)$$

This equality holds for the following reason. Assume without loss of generality that there are $N \in \{1, \dots, B\}$ nonzero k_i s. Let $\sigma : \{1, \dots, N\} \rightarrow \{1, \dots, B\}$ give the position of the

i th nonzero k_i (so that $\sigma(1)$ is the position of the first nonzero k_i , $\sigma(2)$ is the position of the second, etc.). We can write

$$\prod_{i=1}^B I_i^{k_i} = \prod_{i=1}^N I_{\sigma(i)}^{k_{\sigma(i)}}, \quad (5)$$

i.e., we can remove all I_i s that are raised to the power of zero. Because the I_i s are Bernoulli random variables, we have that $I_{\sigma(i)}^{k_{\sigma(i)}} = I_{\sigma(i)}$. Next, recall that $I_{\sigma(i)} = \mathbb{I}(T^* \leq T_{\sigma(i)})$, where $T_{\sigma(1)} < T_{\sigma(2)} < \dots < T_{\sigma(n)}$. If $T^* \leq T_{\sigma(i)}$, then by transitivity, $T^* \leq T_{\sigma(2)} < \dots < T_{\sigma(n)}$, implying $I_{\sigma(i)} = 1$ for all $i \in \{1, \dots, N\}$. Therefore, $I_{\sigma(1)} = 1 = I_{\sigma(1)} \dots I_{\sigma(n)}$. On the other hand, if $T^* > T_{\sigma(i)}$ then $T_{\sigma(1)} = 0$, implying $I_{\sigma(1)} = 0 = I_{\sigma(1)} \dots I_{\sigma(n)}$. Combining these cases, we conclude that $I_{\sigma(1)} = I_{\sigma(1)} \dots I_{\sigma(B)}$. Equation 5 therefore reduces to

$$\prod_{i=1}^N I_{\sigma(i)}^{k_{\sigma(i)}} = I_{\sigma(1)}. \quad (6)$$

Finally, because $\sigma(1)$ is the position of the first nonzero k_i , we have that $\sigma(1) = \tau(k_1, \dots, k_B)$. Combining this fact with (5) and (6) yields the conclusion (4).

Having established this lemma, we can evaluate the r th power of the sum of the I_i s. The multinomial theorem states that

$$\begin{aligned} \left(\sum_{i=1}^B I_i \right)^r &= \sum_{(k_1, \dots, k_B): k_1 + \dots + k_B = r} \binom{r}{k_1, k_2, \dots, k_B} \prod_{i=1}^B I_i^{k_i} \\ &= \sum_{(k_1, \dots, k_B): k_1 + \dots + k_B = r} \binom{r}{k_1, k_2, \dots, k_B} I_{\tau(k_1, \dots, k_B)}, \end{aligned} \quad (7)$$

where the second equality follows from (4). The final term in (7) is simply a linear combination of the I_i s. We therefore can factor out the terms in the sum corresponding to I_i for each i , yielding

$$\begin{aligned} \sum_{(k_1, \dots, k_B): k_1 + \dots + k_B = r} \binom{r}{k_1, k_2, \dots, k_B} I_{\tau(k_1, \dots, k_B)} &= \sum_{i=1}^B \sum_{(k_1, \dots, k_B) \in \tau^{-1}(i)} \binom{r}{k_1, k_2, \dots, k_B} I_i \\ &= \sum_{i=1}^B I_i \sum_{(k_1, \dots, k_B) \in \tau^{-1}(i)} \binom{r}{k_1, k_2, \dots, k_B}. \end{aligned} \quad (8)$$

We evaluate the inner sum (8), which is the coefficient corresponding to I_i in the linear combination. We have that

$$\begin{aligned} \sum_{(k_1, \dots, k_B) \in \tau^{-1}(i)} \binom{r}{k_1, k_2, \dots, k_B} &= \sum_{(0, \dots, 0, k_i, \dots, k_B): k_i \geq 1, k_i + \dots + k_B = r} \binom{r}{k_1, k_2, \dots, k_B} \\ &= \sum_{(k_1, \dots, k_{B-i+1}): k_1 \geq 1, k_1 + \dots + k_{B-i+1} = r} \binom{r}{k_1, k_2, \dots, k_{B-i+1}} = \sum_{j=1}^r \sum_{l_1 + \dots + l_{B-i} = r-j} \binom{r}{j, l_1, \dots, l_{B-i}} \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^r \sum_{l_1+\dots+l_{B-i}=r-j} \frac{r!}{j!l_1!\dots l_{B-i}!} = \sum_{j=1}^r \sum_{l_1+\dots+l_{B-i}=r-j} \frac{r(r-1)\dots(r-j+1)(r-j)!}{j!l_1!\dots l_{B-i}!} \\
&= \sum_{j=1}^r \sum_{l_1+\dots+l_{B-i}=r-j} \frac{r!}{(r-j)!j!} \binom{r-j}{l_1, \dots, l_{B-i}} = \sum_{j=1}^r \frac{r!}{(r-j)!j!} \sum_{l_1, \dots, l_{B-i}} \binom{r-j}{l_1, \dots, l_{B-i}} \\
&= \sum_{j=1}^r \binom{r}{j} (B-i)^{r-j} = \sum_{j=1}^r \binom{r}{j} (B-i)^{r-j} 1^j = \sum_{j=0}^r \binom{r}{j} (B-i)^{r-j} 1^j - 1(B-i)^r \\
&= (B-i+1)^r - (B-i)^r. \quad (9)
\end{aligned}$$

Combining (7), (8), and (9), we conclude that

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

Next, setting $p_B = \frac{1}{B} \sum_{i=1}^B I_i$, we obtain

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

We next consider the r th degree polynomial of p_B^r . First, for $j \in \mathbb{N}$, and $x_0 \in \mathbb{R}$, we have that

$$\begin{aligned}
(p_B - x_0)^r &= \sum_{k=0}^j \binom{j}{k} p_B^k x_0^{j-k} = x_0^j + \sum_{k=1}^j \binom{j}{k} p_B^k x_0^{j-k} \\
&= x_0^j + \sum_{k=1}^j \binom{j}{k} \left[\sum_{i=1}^B M(B, k, i) I_i \right] x_0^{j-k} = x_0^j + \sum_{i=1}^B \left[\sum_{k=1}^j x_0^{j-k} \binom{j}{k} M(B, k, i) \right] I_i.
\end{aligned}$$

Finally, let $c_0, c_1, \dots, c_r \in \mathbb{R}$ be polynomial coefficients. We have that

$$\begin{aligned}
\sum_{j=0}^r c_j (p_B - x_0)^j &= c_0 + \sum_{j=1}^r c_j (p_B - x_0)^j = c_0 + \sum_{l=1}^r c_j \left[x_0^j + \sum_{i=1}^B \left[\sum_{k=1}^j x_0^{j-k} \binom{j}{k} M(B, k, i) \right] I_i \right] \\
&= \left(c_0 + \sum_{j=1}^r c_j x_0^j \right) + \sum_{i=1}^B \left[\sum_{j=1}^r \sum_{k=1}^j c_j x_0^{j-k} \binom{j}{k} M(B, k, i) \right] I_i,
\end{aligned}$$

completing the proof.