Consider a permutation test (the ideas here apply broadly to all sampling-based tests of independence or conditional independence, including marginal permutation test, conditional permutation test, marginal randomization test, and conditional randomization test). Let B be the number of resamples. Let $T_1, \dots, T_B$ be the ordered resampled test statistics. Let T^* be the test statistic computed on the raw data. Let I_i be the indicator that T^* is less than T_i, i.e.

$$I_i = \mathbb{I}\left(T^* \leq T_i\right).$$

We have the basic fact under the null hypothesis that

$$\mathbb{E}(I_i) = \mathbb{P}(T^* \leq T_i) = i/B.$$

We want to combine $I_1, \dots, I_B$ in such a way so as to produce an e-value E (i.e., a non-negative variable E such that $\mathbb{E}(E) \leq 1$). To test the hypothesis, we check if $E > \beta$ for some threshold β.

We call $f : \mathbb{R}^B \to \mathbb{R}^{\geq 0}$ a "combining function" if f combines $I_1, \dots, I_B$ in such a way so as to produce a valid E-value, i.e.,

$$f(I_1, \dots, I_B) \geq 0; \quad \mathbb{E}\left[f(I_1, \dots, I_B)\right] \leq 1.$$

Let $\mathcal{F}$ be the class of combining functions. An interesting subset of $\mathcal{F}$ is the set of linear combining functions. A function $g \in \mathcal{F}$ is a linear combining function if

$$g(I_1, \dots, I_B) = c + \sum_{i=1}^{B} a_i I_i$$

for given scalars $a_1, \dots, a_B$, and c. The linearity of expectation implies that

$$\mathbb{E}\left[g(I_1, \dots, I_B)\right] = c + \sum_{i=1}^{n} a_i \mathbb{E}(I_i) = c + \frac{1}{B}\sum_{i=1}^{B}(a_i)(i).$$

Therefore, under the stricter requirement that $\mathbb{E}(E) = 1$, we have the constraints

i $\sum_{i=1}^{B}(a_i)i = B(1-c)$

ii $c + \sum_{i=1}^{B} a_i I_i \geq 0.$

We give a couple examples of linear combining functions.

Example 1. Let $c = 0$ and $a_i = 1/i$ for all i. Then

$$\sum_{i=1}^{B} (1/i)i = B(1-0),$$

confirming constraint (i). Next,

$$0 + \sum_{i=1}^{B} \frac{I_i}{i} \geq 0,$$

confirming constraint (ii). Therefore,

$$E = \sum_{i=1}^{B} \frac{I_i}{i}$$

is a valid e-value. This is a left-tailed test: if T^* is small, then E is big, leading us to reject the null.

Example 2. The standard permutation test is a special case of the proposed framework. Let $c = 3/2 + 1/(2B)$ and $a_i = -1/B$ for all a_i. Verifying condition (i),

$$\begin{aligned}\sum_{i=1}^{B} (a_i)i = -\frac{1}{B}\sum_{i=1}^{B} i = -\frac{1}{B}\left(\frac{B^2+B}{2}\right) &= -(1/2)B - 1/2 \\ &= B(1-(3/2)-1/(2B)) = B(1-c).\end{aligned}$$

Next, verifying condition (ii),

$$(3/2) + 1/(2B) - (1/B)\sum_{i=1}^{B} I_i \geq (3/2) + 1/(2B) - 1 \geq 0.$$

Therefore,

$$E = 3/2 + 1/(2B) - \frac{1}{B}\sum_{i=1}^{B} I_i$$

is a valid e-value. Note that $p_B := \frac{1}{B}\sum_{i=1}^{B} I_i$ is simply the p-value corresponding to the right-sided permutation test. For given $\alpha \in (0,1)$, we reject the null hypothesis if and only if $p_B < \alpha$, which is equivalent to rejecting the null hypothesis if and only if

$$E > 3/2 + 1/(2B) - \alpha.$$

Therefore, the p-permutation test is a special case of the e-permutation test. We typically think of e-values as controlling FDR, but here e-values control type-I error.

We can extend this idea to control family-wise error rate.

Example 3. Suppose we are using CRT or CPT to test conditional independence. If we have misspecified the model for $X|Z$, then the resampled test statistics $T_1, \ldots, T_B$ will be off. In particular, qq-plots often reveal that in such settings the tail of the distribution is incorrect, leading to p-value inflation. We can use the proposed framework to downweight re-sampled test statistics in the tail of the distribution, possibly leading to better calibration. For example, consider again example 1, where previously we set $a_i = 1/i$ for all i. Suppose that we use (e.g.) a Gaussian kernel to assign the weights. Let

$$\hat{\mu} = \frac{1}{B}\sum_{i=1}^{B} T_i$$

be the average of the resampled test statistics. Next, let

$$K(t) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)t^2}$$

be the Gaussian kernel. Define a_i as

$$a_i := (B/i)\frac{K(|T_i - \hat{\mu}|/h)}{\sum_{i=1}^{B} K(|T_i - \hat{\mu}|/h)}.$$

Again, set $c = 0$. We see that

$$\sum_{i=1}^{B}(a_i)i = \frac{\sum_{i=1}^{B} K(|T_i - \hat{\mu}|/h)}{\sum_{i=1}^{B} K(|T_i - \hat{\mu}|/h)} = B = B(1-c),$$

verifying the first condition. Moreover, $A_i \geq 0$ for all i, satisfying the second condition. It follows that

$$E := \sum_{i=1}^{B} a_i I_i$$

is a valid e-value. This e-value is less sensitive to the tails of the sampling distribution than the e-value from example 1.

Advantages to this approach. The proposed framework has several advantages (I think). First, because the outputs are e-values, they can be combined under arbitrary dependence to control FDR using the e-BH procedure. Second, the framework is flexible: we can choose among many different "combining functions" (including the easy-to-use linear variants) to achieve different objectives. Certain combining functions might have robustness properties, for example. Third, we easily can combine e-values across data splits, enabling us to run multiple train-test splits without worrying about combining the resulting p-values.

Open questions. What is the "best" combining function that we could use?

Additional properties of the I_is. Recall that $I_i = \mathbb{I}(T^* \leq T_i)$. We state several additional properties of the I_is. Let $k_1, k_d, \ldots, k_p \in \mathbb{N}$ be indexes such that

$$1 \leq k_1 \leq k_2 \leq \cdots \leq k_p \leq B.$$

Furthermore, let $r_1, r_2, \ldots, r_p \in \mathbb{N}$ be non-negative integers. We have that

$$I_{k_1}^{r_1} I_{k_2}^{r_2} \ldots I_{k_p}^{r_p} = I_{k_1}.$$

The reason is as follows. First, because I_{k_i} is Bernoulli,

$$I_{k_1}^{r_1} I_{k_2}^{r_2} \ldots I_{k_p}^{r_p} = I_{k_1} I_{k_2} \ldots, I_{k_p}.$$

Next, we take cases on T^*. First, suppose that $T^* \leq k_1$. Then $T^* \leq k_2, \ldots, k_p$. Therefore,

$$I_{k_1} I_{k_2} \ldots I_{k_p} = 1 \cdot 1 \ldots 1 = I_{k_1}.$$

Next, suppose that $T^* > k_1$. Then $I_{k_1} = 0$, and so

$$I_{k_1} I_{k_2} \ldots I_{k_p} = 0 \cdot I_{k_2} \ldots I_{k_p} = 0 = I_{k_1}.$$

This property is useful because it allows us to more easily evaluate expressions of the form $\left(\sum_{i=1}^{n} I_i\right)^r$ for $r \in \mathbb{N}$. As a warmup, we consider $r = 1$:

$$\sum_{i=1}^{B} I_i = a_i^{(1)} I_i,$$

where $a_1^{(1)} = 1$. Next, consider $r = 2$:

$$\left(\sum_{i=1}^{B} I_i\right)^2 = \sum_{i=1}^{B}\sum_{j=1}^{B} I_i I_j = \sum_{i=1}^{B} I_i^2 + 2\sum_{i=1}^{B-1}\sum_{j=i+1}^{B} I_i I_j = \sum_{i=1}^{B} I_i + 2\sum_{i=1}^{B-1}\sum_{j=i+1}^{B} I_i$$
$$= \sum_{i=1}^{B} I_i + \sum_{i=1}^{B-1} 2(B-i) I_i = \sum_{i=1}^{B} I_i + \sum_{i=1}^{B} 2(B-i) I_i = \sum_{i=1}^{B} (2B-2i+1) I_i = \sum_{i=1}^{B} a_i^{(2)} I_i,$$

where $a_i^{(2)} = 2B - 2i + 1$.

Consider the multinomial theorem:

$$\left(\sum_{i=1}^{B} I_i\right)^r = \sum_{k_1+k_2+\cdots+k_B=r} \binom{r}{k_1, k_2, \ldots, k_B} \prod_{i=1}^{B} I_i^{k_i},$$

where

$$\binom{r}{k_1, k_2, \ldots, k_B} = \frac{r!}{k_1! k_2! \ldots k_B!}.$$

Let

$$C(r, B) := \left\{ (k_1, k_2, \ldots, k_B) \in \{1, \ldots, r\}^B : \sum_{i=1}^{B} k_i = r \right\},$$

i.e., the set of tuples of length B of integers from 1 to r such that the elements of the tuple sum to r. Let $\tau : C(r, B) \to B$ be defined by

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

i.e., τ is the minimal nonzero index of a given configuration. The inverse τ^{-1} of τ is the set of configurations with a given minimal nonzero index. That is, for $s \in \{1, \ldots, B\}$,

$$\tau^{-1}(s) = \left\{ (0, \ldots, 0, k_s, k_{s+1}, \ldots, k_B) : \sum_{i=s}^{B} k_i = r \right\}.$$

Therefore, the multinomial theorem reduces to

$$\left(\sum_{i=1}^{B} I_i\right)^r = \sum_{(k_1,\dots,k_B):k_1+\dots+k_B=r} \binom{r}{k_1, k_2, \dots, k_B} I_{\tau(k_1,\dots,k_B)}$$
$$= \sum_{s=1}^{B} \sum_{(0,\dots,0,k_s,\dots,k_B):k_s\geq 1,k_s+\dots+k_B=r} \binom{r}{k_1, k_2, \dots, k_B} I_s.$$

We consider the above sum:

$$\sum_{(0,\dots,0,k_s,\dots,k_B):k_s\geq 1,k_s+\dots+k_B=r} \binom{r}{k_1, k_2, \dots, k_B}$$
$$= \sum_{(k_1,\dots,k_{B-s+1}):k_1\geq 1,k_1+\dots+k_{B-s+1}=r} \binom{r}{k_1, k_2, \dots, k_B}$$
$$= \sum_{j=1}^{r} \sum_{l_1+\dots+l_{B-s}=r-j} \binom{r}{j, l_1, \dots, l_{B-s}}$$
$$= \sum_{j=1}^{r} \sum_{l_1+\dots+l_{B-s}=r-j} \frac{r!}{j!l_1!\dots l_{B-s}!}$$
$$= \sum_{j=1}^{r} \sum_{l_1+\dots+l_{B-s}=r-j} \frac{r(r-1)\dots(r-j+1)(r-j)!}{j!l_1!\dots l_{B-s}!}$$
$$= \sum_{j=1}^{r} \sum_{l_1+\dots+l_{B-s}=r-j} \frac{r!}{(r-j)!j!} \binom{r-j}{l_1, \dots, l_{B-s}}$$
$$= \sum_{j=1}^{r} \frac{r!}{(r-j)!j!} \sum_{l_1,\dots,l_{B-s}} \binom{r-j}{l_1, \dots, l_{B-s}}$$
$$= \sum_{j=1}^{r} \binom{r}{j} (B-s)^{r-j} = \sum_{j=1}^{r} \binom{r}{j} (B-s)^{r-j} 1^j$$
$$= \sum_{j=0}^{r} \binom{r}{j} (B-s)^{r-j} 1^j - 1(B-s)^r = (B-s+1)^r - (B-s)^r := M(B,r,s).$$

Therefore,

$$\left(\sum_{i=1}^{B} I_i\right)^r = \sum_{s=1}^{B} M(B,r,s) I_s.$$

The number $M(B, r, s)$ can be quite large. However, normalizing the sum by B neutralizes this problem:

$$\begin{aligned}
\left(\frac{1}{B}\sum_{i=1}^{B} I_i\right)^r &= (1/B)^r \sum_{i=1}^{B} \left[(B-i+1)^r - (B-i)^r\right] I_i \\
&= \sum_{i=1}^{B} \left[(1-i/B+1/B)^r - (1-i/B)^r\right] I_i.
\end{aligned}$$

Let us check this formula for specific values of r. First, setting $r = 1$, we have $M(B, 1, s) = (B - s + 1) - (B - s) = 1$. Therefore,

$$\left(\sum_{i=1}^{B} I_s\right)^1 = \sum_{i=1}^{B} I_s.$$

Next, setting $r = 2$, we have

$$M(B, 1, s) = (B + s - 1)^2 - (B - s)^2 = 2B - 2i + 1.$$

Therefore,

$$\left(\sum_{i=1}^{B} I_i\right)^2 = \sum_{i=1}^{B} [2B - 2i + 1] I_i,$$

which matches what we have above.

Overall, assuming the above is correct, we have two key properties of the I_is. First, $\mathbb{E}(I_i) = i/B$. Second, for an analytic function $f : \mathbb{R} \to \mathbb{R}$, we have that

$$f\left(\frac{1}{B}\sum_{i=1}^{B} I_i\right) \approx C + \sum_{i=1}^{B} a_i I_i,$$

that is, we can approximate f arbitrarily well with a linear combination of the a_is. It follows from these properties that

$$\mathbb{E}\left[f\left(\frac{1}{B}\sum_{i=1}^{B} I_i\right)\right] \approx C + \sum_{i=1}^{B} \frac{(a_i)i}{B}.$$

In other words, we can put the empirical p-value through any analytic function to obtain an (asymptotically exact) linear combining function. (Need simulations to demonstrate empirically). As a corollary, the binomial theorem implies that

$$f\left(\frac{1}{B}\sum_{i=1}^{B} I_i - x_0\right)$$

is also approximately a linear combination.

We have that

$$\sum_{i=1}^{r} a_r \left(\sum_{i=1}^{B} I_i\right)^r = \sum_{i=1}^{r} \left(\sum_{i=1}^{B} a_r M(B,r,i) I_i\right) = \sum_{i=1}^{B} \left(I_i \sum_{i=1}^{r} a_r M(B,r,i)\right).$$