1 Introduction

2 Two Sample Problem

Let $\mathbf{X} = (X_1, \ldots, X_{n_x})$ and $\mathbf{Y} = \left(Y_1, \ldots, Y_{n_y}\right)$ be two mutually independent samples of independent and identically distributed random variables with $X_i \stackrel{d}{=} F_X$ and $Y_j \stackrel{d}{=} F_Y$. Consider testing $H_0 : F_X = F_Y$ versus a composite alternative $H_A : F_X \neq F_Y$, where both F_X and F_Y are continues and differentiable, so that their densities f_X and f_Y exist.

This manuscript focuses location-scale families with μ_x and μ_y are location parameters and σ_x and σ_y are scale parameters of F_X and F_Y, respectively. Consider the use of test statistics with the following form

$$T = \frac{1}{n_x n_y} \sum_{i=1}^{n_x} \sum_{j=n_x+1}^{n_x+n_y} g\left(Z_{ij}\right),$$

where Z_{ij} is a distance between X_i and Y_{j-n_x}. For example, $Z_{ij} = |X_i - Y_{j-n_x}|$, $i = 1, \ldots, n_x$, $j = n_x + 1, \ldots, n$, $n = n_x + n_y$. It is convenient consider a pooled sample,

$$\mathbf{D} = (\mathbf{X}, \mathbf{Y}) = \left(X_1, \ldots, X_{n_x}, Y_1, \ldots, Y_{n_y}\right) = \left(D_1, \ldots, D_{n_x}, D_{n_x+1}, \ldots, D_n\right).$$

The finite sample and asymptotic distributions of T can very complex, and the permutations tests represent a viable alternative to analytic derivation of such distributions. Realizations of random variables $X_\cdot$, $Y_\cdot$, $Z_{\cdot\cdot}$, and $D_\cdot$ will be denoted by $x_\cdot$, $y_\cdot$, $z_{\cdot\cdot}$, and $d_\cdot$, respectively.

Let $\hat{\mu}_x$ and $\hat{\mu}_y$ be the maximum likelihood estimators (MLEs) of μ_x and μ_y, respectively, estimated on samples $\mathbf{X}$ and $\mathbf{Y}$ under H_A. Similarly, $\hat{\sigma}_x$ and $\hat{\sigma}_y$ be the MLEs of σ_x and σ_y under the alternative. Under H_0, the $\hat{\mu}$ and $\hat{\sigma}$ are the MLEs of $\mu = \mu_x = \mu_y$ and $\sigma = \sigma_x = \sigma_y$ estimated on the pooled sample $\mathbf{D}$.

Then, under H_A, the joint density is estimated by

$$\hat{f}_{\mathbf{D}}\left(\mathbf{d}|H_A\right) := \prod_{i=1}^{n_x} f\left(d_i|\hat{\mu}_x, \hat{\sigma}_x\right) \prod_{i=n_x+1}^{n} f\left(d_i|\hat{\mu}_y, \hat{\sigma}_y\right),$$

and under H_0, by

$$\hat{f}_{\mathbf{D}}\left(\mathbf{d}|H_0\right) := \prod_{i=1}^{n_x} f\left(d_i|\hat{\mu}, \hat{\sigma}\right) \prod_{i=n_x+1}^{n} f\left(d_i|\hat{\mu}, \hat{\sigma}\right).$$

The log-likelihood ratio test statistic becomes

$$l = -2\log\frac{\hat{f}_{\mathbf{D}}\left(\mathbf{d}|H_0\right)}{\hat{f}_{\mathbf{D}}\left(\mathbf{d}|H_A\right)} = -2\sum_{i=1}^{n_x}\log\frac{f\left(d_i|\hat{\mu},\hat{\sigma}\right)}{f\left(d_i|\hat{\mu}_x,\hat{\sigma}_x\right)} - 2\sum_{i=n_x+1}^{n}\log\frac{f\left(d_i|\hat{\mu},\hat{\sigma}\right)}{f\left(d_i|\hat{\mu}_y,\hat{\sigma}_y\right)}. \tag{1}$$

When label permutations are applied to evaluate the distribution of l under H_0, the values of $\hat{\mu}$ and $\hat{\sigma}^2$ do not change and the test statistics l can be simplified to

$$l^* = \sum_{i=1}^{n_x}\log f\left(x_i,\hat{\mu}_x,\hat{\sigma}_x\right) + \sum_{i=n_x+1}^{n}\log f\left(y_i,\mu_y,\hat{\sigma}_y\right). \tag{2}$$

Special cases:

- (comparing location parameters with a shared scale parameter) If $\sigma_x = \sigma_y = \sigma$, then $\hat{\sigma}_x = \hat{\sigma}_y$ estimated under H_A, which can still be different from $\hat{\sigma}$ estimated under H_0. Since permutation tests based on l and l^* are equivalent, without loss of generality $\hat{\sigma}$ will be used to denote the MLE of σ estimated under H_A.

- (comparing scale parameters with a shared location parameter) If $\mu_x = \mu_y = \mu$, then $\hat{\mu}_x = \hat{\mu}_y$ under H_A, may be not the same as $\hat{\mu}$ estimated under H_0. Since permutation tests based on l and l^* are equivalent, $\hat{\mu}$ will be used to denote the MLE of μ based on a pooled sample $\mathbf{D}$.

3 Asymptotic optimality of label permutations for comparing Gaussian means with a shared variance

In Gaussian case all sampling information from $\mathbf{X}$ is fully absorbed by its sample mean $\bar{x} = n_x^{-1}\sum_{i=1}^{n_x} x_i$ and its sample variance

$$S_x^2 = \frac{1}{n_x - 1}\sum_{i=1}^{n_x}\left(x_i - \bar{x}\right)^2.$$

Using U-statistic form of the sample variance,

$$S_x^2 = \frac{2}{n_x\left(n_x - 1\right)}\sum_{1\le i_1 < i_2 \le n}\left(x_{i_1} - x_{i_2}\right)^2.$$

Similar expressions are applicable to S_y^2.

Theorem 1 *In Gaussian case with shared variance, permutation tests based on l^* and*

$$N_1 := -\left(n_x - 1\right) S_x^2 - \left(n_y - 1\right) S_y^2 = -\sum_{i=1}^{n_x} \left(d_i - \bar{x}\right)^2 - \sum_{i=n_x+1}^{n} \left(d_i - \bar{y}\right)^2$$

are equivalent for comparing means.

Proof: The following equality

$$\begin{aligned} l^* &= \sum_{i=1}^{n_x} \left[\log \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} - \frac{1}{2}\left(\frac{d_i - \bar{x}}{\hat{\sigma}^2}\right)^2\right] + \sum_{i=n_x+1}^{n} \left[\log \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} - \frac{1}{2}\left(\frac{d_i - \bar{y}}{\hat{\sigma}^2}\right)^2\right] \\ &= -\frac{1}{2\hat{\sigma}^2}\left[\left(n_x - 1\right) S_x^2 + \left(n_y - 1\right) S_y^2\right] + \tau\left(\hat{\sigma}^2\right) = \frac{N_1}{2\hat{\sigma}^2} + \tau\left(\hat{\sigma}^2\right), \end{aligned}$$

shows that l^* is a linear combination of $\hat{\sigma}^2$ and N_1. Since $\hat{\sigma}^2$ is invariant to label permutations, permutation tests based on l^* and N_1 are equivalent. **Q.E.D.**

Theorem 2 *In a Gaussian case with a shared variance and $n_x = n_y$, permutation tests based N_1, $N_2 := \left(\bar{y} - \bar{x}\right)^2$ and $N_3 := \sum_{i=1}^{n_x} \sum_{j=n_x+1}^{n} \left(d_i - d_j\right)^2$ are equivalent for comparing means.*

Proof.

$$\begin{aligned} N_3 &= \sum_{1 \le i < j \le n} (d_i - d_j)^2 - \sum_{1 \le i < j \le n_x} (d_i - d_j)^2 - \sum_{n_x+1 \le i < j \le n} (d_i - d_j)^2 \\ &= \sum_{1 \le i < j \le n} (d_i - d_j)^2 - \frac{n_x(n_x - 1)}{2} S_x^2 - \frac{n_y(n_y - 1)}{2} S_y^2 \\ &= \sum_{1 \le i < j \le n} (d_i - d_j)^2 - \frac{n_x}{2} \sum_{i=1}^{n_x} (d_i - \bar{x})^2 - \frac{n_y}{2} \sum_{i=n_x+1}^{n} (d_i - \bar{y})^2 \end{aligned} \tag{3}$$

where the first term in the right hand side does not depend on permutations.

When $n_x = n_y$, $N_3 = \sum_{1 \le i < j \le n} (d_i - d_j)^2 + 0.5 n_x N_1$, which makes N_3 and N_1 permutation equivalent. Since

$$\begin{aligned} (n-1)\hat{\sigma}^2 &= \sum_{i=1}^{n_x} \left(d_i - (\bar{x} + \bar{y})/2\right)^2 + \sum_{i=n_x+1}^{n} \left(d_i - (\bar{x} + \bar{y})/2\right)^2 \\ &= \sum_{i=1}^{n_x} [(d_i - \bar{x}) + (\bar{y} - \bar{x})/2]^2 + \sum_{i=n_x+1}^{n} [(d_i - \bar{y}) + (\bar{y} - \bar{x})/2]^2 \\ &= \sum_{i=1}^{n_x} [(d_i - \bar{x})]^2 + \sum_{i=1}^{n_x} [(\bar{y} - \bar{x})/2]^2 + \sum_{i=n_x+1}^{n} [(d_i - \bar{y})]^2 + \sum_{i=n_x+1}^{n} [(\bar{y} - \bar{x})/2]^2 \\ &= (n_x - 1)\left[S_x^2 + S_y^2\right] + \frac{n_x}{2}[\bar{y} - \bar{x}]^2 = -(n_x - 1)N_1 + \frac{n}{2} N_2 \end{aligned}$$

Choice	Number of Elements	$Cov\left(g\left(Z_{i_1j_1}\right), g\left(Z_{i_2j_2}\right)\right)$	i.i.d. case
$i_1 \neq i_2$ and $j_1 \neq j_2$	$n_x^2 n_y^2 - n_x^2 n_y - n_x n_y^2 + n_x n_y$	0	0
$i_1 = i_2$ and $j_1 \neq j_2$	$n_y(n_x^2 - n_x)$	$Cov\left(g\left(Z_{i_1j_1}\right), g\left(Z_{i_2j_2}\right)\right)$	C_1
$i_1 \neq i_2$ and $j_1 = j_2$	$n_x(n_y^2 - n_y)$	$Cov\left(g\left(Z_{i_1j_1}\right), g\left(Z_{i_2j_2}\right)\right)$	C_2
$i_1 = i_2$ and $j_1 = j_2$	$n_x n_y$	$Var\left(g\left(Z_{i_1j_1}\right)\right)$	V

Table 1: Summary of covariance matrix elements with an i.i.d. case (X_i are i.i.d. and Y_j are i.i.d.)

does not change with permutations, N_2 and N_1 are permutation equivalent test statistics. **Q.E.D.**

4 Asymptotics

Let the first two moments of X_i and Y_i exist and be finite, and $g(\cdot)$ be a monotone function. The objective is to consider asymptotic behavior of the random variable

$$T = \frac{1}{n_x n_y} \sum_{i=1}^{n_x} \sum_{j=n_x+1}^{n} g\left(Z_{ij}\right).$$

In this double sum there are $n_x n_y$ elements and there are $n_x n_y(n_x n_y - 1)/2$ distinct pairs $(Z_{i_1j_1}, Z_{i_2j_2})$. Then, $Cov\left(g\left(Z_{i_1j_1}\right), g\left(Z_{i_2j_2}\right)\right)$ takes different values for different choices of $i_1, i_2 \in \{1, \cdots, n_x\}$ and $j_1, j_2 \in \{n_x+1, \cdots, n\}$. For every finite n_x and n_y,

$$ET = \frac{1}{n_x n_y} \sum_{i=1}^{n_x} \sum_{j=n_x+1}^{n} E\left(g\left(Z_{ij}\right)\right)$$

and

$$\begin{aligned}
Var\left(T\right) = \frac{1}{n_x^2 n_y^2} Var\left(\sum_{i=1}^{n_x} \sum_{j=n_x+1}^{n} g\left(Z_{ij}\right)\right) &= \frac{1}{n_x^2 n_y^2} \sum_{i=1}^{n_x} \sum_{j=n_x+1}^{n} Var\left(g\left(Z_{ij}\right)\right) \\
&+ \frac{1}{n_x^2 n_y^2} \sum_{i=1}^{n_x} \sum_{j_1=n_x+1}^{n} \sum_{j_2=n_x+1, j_1\neq j_2}^{n} Cov\left(g\left(Z_{ij_1}\right), g\left(Z_{ij_2}\right)\right) \\
&+ \frac{1}{n_x^2 n_y^2} \sum_{i_1=1}^{n_x} \sum_{i_2=1, i_1\neq i_2}^{n_x} \sum_{j=n_x+1}^{n} Cov\left(g\left(Z_{i_1j}\right), g\left(Z_{i_2j}\right)\right).
\end{aligned}$$

In the case when X_i are i.i.d. and Y_j are i.i.d., $E\left(g\left(Z_{ij}\right)\right)=E_{iid}$ is the same for all combinations of i and j, with

$$ET=E\left(\frac{1}{n_x n_y}\sum_{i=1}^{n_x}\sum_{j=n_x+1}^{n}g\left(Z_{ij}\right)\right)=E_{iid}$$

and

$$\begin{aligned} Var\left(T\right) &= V\frac{n_x n_y}{n_x^2 n_y^2}+C_1\frac{n_x n_y(n_y-1)}{n_x^2 n_y^2}+C_2\frac{n_y n_x(n_x-1)}{n_x^2 n_y^2} \\ &= V\frac{1}{n_x n_y}+C_1\frac{n_y-1}{n_x n_y}+C_2\frac{n_x-1}{n_x n_y}. \end{aligned}$$

Asymptotic properties are considered under the following situations (1) $n_x\to\infty$ and n_y is fixed, (2) $n_y\to\infty$ and n_x is fixed, and (3) both n_x and n_y go to infinity. The first two cases are of lesser interest to us as when one of two sample sizes goes to ∞, the long term behavior is fully determined by either C_1 or C_2 defined in Table 1. The situation (3) is more interesting and will be considered with three distinct sub-cases:

Sub-case (3.1): If $n_y/n_x\to 0$ and both n_x and n_y go to ∞, then $n_x Var(T)\to C_1$. Considering T statistic as

$$T=\frac{1}{n_x}\sum_{i=1}^{n_x}\left(\frac{1}{n_y}\sum_{j=n_x+1}^{n}g\left(Z_{ij}\right)\right)=\frac{1}{n_x}\sum_{i=1}^{n_x}G_i$$

at a sufficiently large n_y, where $\frac{1}{n_y}\sum_{j=n_x+1}^{n}g\left(Z_{ij}\right)\approx G_i$ (close to its limiting case) we see that EG_i and $Var(G_i)$ are finite. Then CLT is applied and we prove that T is asymptotically normal.

Sub-case (3.2): if n_x goes to ∞ faster than n_y, meaning $n_x/n_y\to 0$, then $n_y Var(T)\to C_2$. The T is asymptotically normal by analogy with (3.1).

Sub-case (3.3): if the rates of convergence are the same for n_x and n_y, and $n_x/n_y\to k\in(0,\infty)$, then $n_x Var(T)\to C_1+kC_2$ and $n_y Var(T)\to k^{-1}C_1+C_2$. The test statistic T can be expressed as

$$T=\frac{1}{n_x}\sum_{i=1}^{n_x}\left(\frac{1}{n_y}\sum_{j=n_x+1}^{n}g\left(Z_{ij}\right)\right).$$

For a sufficiently large n_y, central limit theorem applies to conditional on i random variables $g\left(Z_{ij}\right)$. For these random variables, we have the same mean and variance: $E_i:=E\left(g\left(Z_{ij}\right)\right)$

and $V_i := Var\left(g\left(Z_{ij}\right)\right)$. Then, $\sqrt{n_y}\left(g\left(Z_{ij}\right) - E_i\right) \to N(0, V_i)$ and

$$\frac{1}{n_y}\sum_{j=n_x+1}^{n} g\left(Z_{ij}\right) = E_i + \frac{\sqrt{V_i}}{\sqrt{n_y}}\xi_i + o\left(\frac{1}{\sqrt{n_y}}\right),$$

where ξ_i is a standard normal r.v. Under $n_x(= kn_y)$,

$$\begin{aligned} T &= \frac{1}{kn_y}\sum_{i=1}^{kn_y}\left(E_i + \frac{\sqrt{V_i}}{\sqrt{n_y}}\xi_i + o\left(\frac{1}{\sqrt{n_y}}\right)\right) \\ &= E_{iid} + \frac{1}{kn_y}\sum_{i=1}^{kn_y}\frac{\sqrt{V_i}}{\sqrt{n_y}}\xi_i + o\left(\frac{1}{\sqrt{n_y}}\right) \end{aligned}$$

is asymptotically a linear combination of normal random variables, where $E_{iid} = \frac{1}{kn_y}\sum_{i=1}^{kn_y} E_i$. Then, $\sqrt{n_y}(T - E_{iid}) = \frac{1}{kn_y}\sum_{i=1}^{kn_y}\sqrt{V_i}\xi_i + o\left(1\right)$. The we apply CLT to $\sqrt{V_i}\xi_i$ (some assumptions) and get asymptotic normality of $\sqrt{n_y}(T - E_{iid})$. Similarly we can get asymptotic normality of $\sqrt{n_x}(T - E_{iid})$.

In the above X_i and Y_i follow a set of central limit regularity conditions which allows to rely on central limit theorem for analyzing asymptotic properties. This covers many distributions, but Cauchy and many other heavy-tailed distributions are not covered.

5 Permutation tests for heavy-tailed distributions

5.1 Likelihood ratio

When a parametric family is known, the maximum likelihood ratio test can be applied. The likelihood ratio test statistic (l) often depends on unknown nuisance parameters. These parameters are often estimated by the maximum likelihood method under H_0 and under H_A. It is not unusual that there is no explicit analytic solution for MLEs. For example, if data follow a distribution with unknown location and shift parameters, the MLEs have to be found numerically under H_A

$$\left(\hat{\mu}_x, \hat{\sigma}_x, \hat{\mu}_y, \hat{\sigma}_y\right) = \arg\max_{\mu_x,\sigma_x,\mu_y,\sigma_y}\prod_{i=1}^{n_x} f\left(d_i|\mu_x, \sigma_x\right)\prod_{i=n_x+1}^{n} f\left(d_i|\mu_y, \sigma_y\right), \tag{4}$$

and under H_0

$$(\hat{\mu}, \hat{\sigma}) = \arg\max_{\mu,\sigma} \prod_{i=1}^{n_x} f\left(d_i|\mu,\sigma\right) \prod_{i=n_x+1}^{n} f\left(d_i|\mu,\sigma\right). \tag{5}$$

An unknown analytic form of the MLE does not allow to analytically find a critical value. This problem is solved by re-sampling under H_0 with $X_i \sim f(x|\hat{\mu},\hat{\sigma})$ and $Y_i \sim f(y|\hat{\mu},\hat{\sigma})$, which approximates the distribution of l under H_0, and consequently allows researchers to estimate critical values.

Example (likelihood ratio test for Cauchy distribution): When X_i and Y_i follow a Cauchy distribution with location parameters μ_X and μ_Y, and scale parameters σ_X and σ_Y, the loglikelihood ratio test statistic is

$$l = -2\sum_{i=1}^{n_x} \log \frac{\hat{\sigma}_x \pi^{-1}}{\left(d_i - \hat{\mu}_x\right)^2 + \hat{\sigma}_x^2} - 2\sum_{i=n_x+1}^{n} \log \frac{\hat{\sigma}_y \pi^{-1}}{\left(d_i - \hat{\mu}_y\right)^2 + \hat{\sigma}_y^2} + 2\sum_{i=1}^{n} \log \frac{\hat{\sigma} \pi^{-1}}{\left(d_i - \hat{\mu}\right)^2 + \hat{\sigma}^2} \tag{6}$$

where the MLEs under H_A and H_0 are found as solutions to maximization tasks (4) and (5) respectively. The solutions were found numerically using box constrained maximization quasi-Newton optimization [1]. The domain for location parameter was constrained to the $(-3, 3)$ range and the domain of scale was $(0, 2)$. To reduces chances of local maximums, the maximization procedure was performed 10 times with starting values chosen randomly from uniform distributions: $\mathbf{Unif}(-3, 3)$ for the location and $\mathbf{Unif}(1.25, 2.75)$. The solution with the highest maximum likelihood value was selected as the MLE.

The critical values were also found numerically using $10,000$ simulations under H_0. For example, in Cauchy case, $10,000$ pairs of samples $\mathbf{X} \sim C(0, 1)$ and $\mathbf{Y} \sim C(0, 1)$ were generated and $10,000$ realizations of l were calculated. A 0.05-level quantile based on these $10,000$ realization was used as a critical value. This approach is a close approximation of real life scenarios when under H_0 the distributions are approximated by $\mathbf{X} \sim C(\hat{\mu}, \hat{\sigma})$ and $\mathbf{Y} \sim C(\hat{\mu}, \hat{\sigma})$. When these two approaches were compared numerically (not shown), the power properties of $C(\hat{\mu}, \hat{\sigma})$ scenarios were about 1% lower than in scenarios where critical values were found directly from $C(0, 1)$.

Example (permutation likelihood ratio for Cauchy distribution): When permutation tests are considered the loglikelihood ratio test is simplified to

$$l^* \;=\; -\frac{1}{n}\sum_{i=1}^{n_x}\log\frac{\hat{\sigma}_x}{\left(d_i-\hat{\mu}_x\right)^2+\hat{\sigma}_x^2}-\frac{1}{n}\sum_{i=n_x+1}^{n}\log\frac{\hat{\sigma}_y}{\left(d_i-\hat{\mu}_y\right)^2+\hat{\sigma}_y^2} \tag{7}$$

5.2 Other test statistics

Even though the central limit theorem cannot be applied to heavy-tailed distributions, the test statistic T can still be very useful for heavy-tailed distributions showing comparable power properties with the likelihood ratio test (see Table 2 for comparing l versus permutation tests). Consider an analog of N_3 as

$$L_\gamma = \sum_{i=1}^{n_x}\sum_{j=n_x+1}^{n}\ln(1+(d_i-d_j)^\gamma),$$

$\gamma \in \{0.5, 1, 2, \infty\}$. If $\gamma = \infty$, we define

$$L_\infty = \sum_{i=1}^{n_x}\sum_{j=n_x+1}^{n}\ln(|d_i-d_j|).$$

The test statistic L_γ is a special case of T when $g(z) = \ln(1+z^\gamma)$ for finite γ and $g(z) = \ln(|z|)$ when γ is infinite.

If $g(z) = |z|$, then $K_6 = \sum_{i=1}^{n_x}\sum_{j=n_x+1}^{n}|d_i - d_j|$.

If an underlying parametric family is known, the distributions of test statistics (L_γ, N_2, etc.) and the critical values can be estimated by the resampling approach described in Subsection 5.1. If, however, the distribution is not known permutations suggest a viable alternative solution.

5.3 Permutation test

Let

$$\mathbf{D} = \mathbf{D}(\pi_0) = (X_1, \ldots, X_{n_x}, Y_1, \ldots, Y_{n_y}) = (D_1, \ldots, D_n),$$

where π_0 defines an initial order of random variables D_i, and define all possible label permutations of $\mathbf{D}$ through the set

$$\big\{\mathbf{D}(\pi_k) = (\underbrace{D_1(\pi_k), \ldots, D_{n_x}(\pi_k)}_{X(\pi_k)}, \underbrace{D_{n_x+1}(\pi_k), \ldots, D_n(\pi_k)}_{Y(\pi_k)})\big\}_{k=1}^{r_2}, \tag{8}$$

where $r_2 = \frac{n!}{n_x! n_y!}$ is the total number of label permutations. A permutation test P-value is the ratio of the total number of permutations (r_1) satisfying $T(\mathbf{D}(\pi_k)) > T(\mathbf{D}(\pi_0))$ to r_2. If r_1/r_2 is less than a pre-determined significance level α, then H_0 is rejected.

Statistical power of N_2 was explored in Sturino et al (2010, [3]) and K_6 was investigated in Sirsky (2012, [4]). Melas et al considered both of these tests in (2016, [5]). These authors found that a premutation test based on the test statistic K_6 showed good power properties for a wide class of parametric models. Test statistics L_γ were not introduced before.

Section 6 compares statistical power of permutation tests based on L_γ, N_2 and K_6 with widely available two sample t-, Kolmogorov-Smirnov, and Wilcoxon-Mann-Whitney tests, and benchmark likelihood ratio tests.

Table 2: Statistical power for comparing permutation test statistics with likelihood ratio tests

F_1	F_2	n_x, n_y	$L_1(perm)$	l	$NC(perm)$	KS
$N(0,1)$	$N(0,1.5)$	$50, 50$		0.730	0.700	0.119
$N(0,1)$	$N(0.4,1)$	$50, 50$		0.405	0.395	0.343
$N(0,1)$	$N(0.25,1.25)$	$50, 50$		0.356	0.341	0.154
$C(0,1)$	$C(0,2)$	$50, 50$	0.514	0.579		0.179
$C(0,1)$	$C(1,1)$	$50, 50$	0.828	0.869		0.797
$C(0,1)$	$C(0.5,1.5)$	$50, 50$	0.371	0.408		0.269
$C(0,1)$	$C(0,1.5)$	$200, 200$	0.675	0.730		0.248

6 Power comparisons: Monte-Carlo simulations

Monte-Carlo studies covered simulation scenarios based on normal $N(\mu, \sigma)$, Cauchy $C(x_0, \gamma)$, log-normal $LN(\mu, \sigma)$, Pareto $P(x_m, k)$, Fisher $F(d_1, d_2)$, Weibull $W(k, \lambda)$, Beta $B(\alpha, \beta)$, and Gamma $G(k, \theta)$ distributions. Experimental settings mostly covered $n_x = n_y = 5$ (small) and $n_x = n_y = 50$ (moderate) sample sizes. Significance level was always equal to 5%.

Each Monte-Carlo experiment used $m = 2500$ resamples to evaluate power properties, leading to the Monte-Carlo error $\hat{\sigma}_m = \sqrt{\frac{\hat{p}(1-\hat{p})}{m}}$ and asymptotic 95% confidence intervals $(\hat{p} - 2\hat{\sigma}_m,\ \hat{p} + 2\hat{\sigma}_m)$, where $\hat{p}$ is a Monte-Carlo power estimate. In a most conservative case of $p = 0.5$, $\hat{\sigma}_m = 0.01$ and the confidence interval is approximately equal to $(\hat{p} - 0.02,\ \hat{p} + 0.02)$.

Since the number of permutations quickly grows with sample size, at $n_x > 20$ or $n_y > 20$, we will rely on randomized permutation test with 1600 resamples. This number was chosen following (Keller-Mcnulty, 1987 [6] and Marozzi, 2004 [7]).

Table 3: Statistical power at $n_X = n_Y = 20$

F_1	F_2	L_1	L_2	$L_{0.5}$	L_∞
$N(0,1)$	$N(1,1)$	0.802	0.83	0.76	0.685
$N(0,1)$	$N(0,3)$	0.86	0.871	0.846	0.812
$N(0,1)$	$N(1,2)$	0.734	0.748	0.706	0.662
$C(0,1)$	$C(2,1)$	0.906	0.902	0.911	0.903
$C(0,1)$	$C(0,6)$	0.912	0.91	0.916	0.914
$LN(0,1)$	$LN(1,1)$	0.776	0.774	0.764	0.725
$LN(0,1)$	$LN(0,4)$	0.925	0.882	0.952	0.956
$P(1,1)$	$P(3,1)$	0.991	0.989	0.994	0.997
$P(1,2)$	$P(1,6)$	0.86	0.878	0.814	0.727
$F(40,2)$	$F(40,20)$	0.894	0.881	0.868	0.801
$F(2,40)$	$F(20,40)$	0.796	0.609	0.819	0.809
$W(2,2)$	$W(2,4)$	0.949	0.962	0.937	0.902
$W(2,2)$	$W(8,2)$	0.959	0.882	0.961	0.944
$G(3,1)$	$G(3,2)$	0.908	0.915	0.891	0.844
$G(1,1)$	$G(2,1)$	0.741	0.745	0.723	0.680
$B(2,2)$	$B(5,2)$	0.904	0.924	0.867	0.804
$B(1,1)$	$B(8,8)$	0.810	0.103	0.869	0.857

References

[1] Byrd, R. H., Lu, P., Nocedal, J. and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. SIAM Journal on Scientific Computing, 16, 11901208. doi: 10.1137/0916069.

[2] Leman E. Testing Statistical Hypotheses. — 1979.

[3] Statistical methods for comparative phenomics using high-throughput phenotype microarrays / J. Sturino, I. Zorych, B. Mallick et al. // The International Journal of Biostatistics. — 2010. — Vol. 6. — P. 3-4.

[4] Sirsky M. On the Statistical Analysis of Functional Data Arasing from Designed Experiments : Ph. D. thesis / M. Sirsky ; University of Manitoba. — 2012.

[5] Melas V.B., Salnikov D.I., Gudulina A.O. The numerical compearing ofclassical and permutation methods of statistical hypothesis testing. Vestnik of Saint Petersburg University. Series 1. Mathematics. Mechanics. Astronomy, 2016, vol. 3(61), issue 3, pp. 415423.DOI: 10.21638/11701/spbu01.2016.309

[6] Keller-McNulty S., Higgins J. Effect of tail weight and outliers on power and type-I error of robust permutation tests for location // Communications in Statistics — Simulation and Computation. — 1987. — Vol. 16. — P. 17-35.

[7] Marozzi M. Some remarks about the number of permutations one should consider to perform a permutation test // Statistica — 2004.

[8] Ludbrook J., Dudley H. Why Permutation Tests Are Superior to t and F Tests in Biomedical Research // American Statistician — 1998 — Vol. 52. — P. 127-132.

[9] Permutation Tests in the Educational and Behavioral Sciences: A Systematic Review / M. Huo, M. Heyvaert, W. Van den Noortgate, P. Onghena. // Methodology European Journal of Research Methods for the Behavioral and Social Sciences. — 2014. — Vol. 10. — P. 43-59.