

1 Introduction

The null hypothesis of no difference between two populations without reliable assumptions of parametric distributions has long been known (ref,ref) and permutation tests suggest viable solutions. Permutation tests evaluate distributions of statistics of interests under the null hypothesis, and consequently, secure the desired type 1 error Lehmann (1986) (Chapter 5, Section 10, Theorem 6). Power properties of the test depend on a test statistic which is chosen to be permuted. In Lehmann (1986) (Section 10 of Chapter 5) sample mean based permutation test is a most powerful test for comparing normal distributions with a shared variance. When data are not normal, Wilcoxon-Mann-Whitney U-statistic is often used instead. Chung and Romano (2016) generalized the two-sample U-statistics to a larger class of U-statistics including sample means, sample variances, and the Wilcoxon-Mann-Whitney U-statistic.

When U-statistics are asymptotically normal, permutation tests result in high power properties. However, the application of U-test statistics for comparing *heavy-tailed distributions* results in low statistical power. To extend the application of U-statistics, this manuscript proposes the use of U-statistics with asymmetric-logarithm-based kernels. As it will be shown in Section XXX, the use of such asymmetric test statistics for distributions with Paretian tails creates asymptotically normal asymmetric test statistics. Power properties of these tests are investigated through various Monte-Carlo simulation studies in Section YYY. The manuscript conclude with a short summary in Section ZZZ.

2 Two Sample Problem

Let $\mathbf{X} = (X_1, \dots, X_{n_x})$ and $\mathbf{Y} = (Y_1, \dots, Y_{n_y})$ 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 continuous and differentiable, so that their densities f_X and f_Y exist.

This manuscript focuses distributions with location and scale parameters, where 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

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

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

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

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 ., Y ., Z ., and D . will be denoted by x ., y ., z ., and d ., 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

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

and under H_0 , by

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

The log-likelihood ratio test statistic becomes

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

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

$$l^* = \sum_{i=1}^{n_x} \log f(d_i|\hat{\mu}_x, \hat{\sigma}_x) + \sum_{i=n_x+1}^n \log f(d_i|\hat{\mu}_y, \hat{\sigma}_y). \quad (2)$$

Special cases:

| Choice | Number of Elements | $Cov(g(Z_{i_1j_1}), g(Z_{i_2j_2}))$ | 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 $j_1 \neq j_2$ | $n_y(n_x^2 - n_x)$ | $Cov(g(Z_{i_1j_1}), g(Z_{i_2j_2}))$ | C_1 |
| $i_1 \neq i_2$ and $j_1 = j_2$ | $n_x(n_y^2 - n_y)$ | $Cov(g(Z_{i_1j_1}), g(Z_{i_2j_2}))$ | C_2 |
| $i_1 = i_2$ and $j_1 = j_2$ | $n_x n_y$ | $Var(g(Z_{i_1j_1}))$ | 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.)

- (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 D .

3 Finite mean and variance and large sample distributions

Let the first two moments of $g(Z_{ij})$ exist and be finite, $Z_{ij} = |X_i - Y_j|$ and $g(\cdot)$ be a non-negative loss function. The objective is to consider asymptotic behavior of the random variable

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

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 $(Z_{i_1 j_1}, Z_{i_2 j_2})$. Asymptotic distribution of U -statistics with asymmetric kernels such as T is shown to be normal in XXX. Below we derive the large sample mean and variance of T under difference asymptotic scenarios.

3.1 Finite n_x and n_y : independent observations

If $Cov(g(Z_{i_1 j_1}), g(Z_{i_2 j_2}))$ takes different values for different choices of $i_1, i_2 \in \{1, \dots, n_x\}$ and $j_1, j_2 \in \{n_x + 1, \dots, n\}$, then for every finite n_x and n_y ,

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

and

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

3.2 Finite n_x and n_y : independent and identically distribution observations

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

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

and

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

3.3 Large sample properties

Asymptotic properties are considered under the following situations (1) $n_x \rightarrow \infty$ and n_y is fixed, (2) $n_y \rightarrow \infty$ 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 $n_y/n_x \rightarrow 0$ and both n_x and n_y go to ∞ , then $n_x \text{Var}(T) \rightarrow C_1$. Considering T statistic as

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

at a sufficiently large n_y , where $\frac{1}{n_y} \sum_{j=n_x+1}^n g(Z_{ij}) \approx G_i$ (close to its limiting case) we see that EG_i and $\text{Var}(G_i)$ 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 $n_x/n_y \rightarrow 0$, then $n_y \text{Var}(T) \rightarrow C_2$. 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 $n_x/n_y \rightarrow k \in (0, \infty)$, then $n_x \text{Var}(T) \rightarrow C_1 + kC_2$ and $n_y \text{Var}(T) \rightarrow k^{-1}C_1 + C_2$. The test statistic T can be expressed as

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

For a sufficiently large n_y , central limit theorem applies to conditional on i random variables $g(Z_{ij})$. For these random variables, we have the same mean and variance: $E_i := E(g(Z_{ij}))$ and $V_i := \text{Var}(g(Z_{ij}))$. Then, $\sqrt{n_y}(g(Z_{ij}) - E_i) \rightarrow N(0, V_i)$ and

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

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, $\sqrt{n_y}(T - E_{iid}) = \frac{1}{kn_y} \sum_{i=1}^{kn_y} \sqrt{V_i} \xi_i + o(1)$. Then 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 $g(Z_{ij})$ follow a set of central limit regularity conditions which allows to rely on central limit theorem for analyzing asymptotic properties.

3.4 Heavy tailed distributions (Paretian tails)

Moreover, if X_i and Y_j belong to the family of stable distributions (Normal, Cauchy, Levy, etc.) even if their moments do not exist (the stability parameter < 2), for some choices of $g(\cdot)$ moments of $g(Z_{ij})$ do exist and asymptotic normality holds. For example, for stable Z_{ij} with $g(Z_{ij}) = \log(1 + Z_{ij}^p)$ ($p > 0$) moments exist as follows from Theorem 1. Hence, asymptotic normality of T holds and researchers can apply normal approximation of the distribution of T when sample size becomes large and permutation test is cost prohibitive.

For an absolutely continuous random variable Z to have *Pareto tails*, we require $\forall \epsilon > 0$ there exists $M > 0$ so that $\forall x > M$,

$$\left| \frac{\partial}{\partial x} \Pr(|Z| > x) - \frac{\alpha x_m^\alpha}{x^{\alpha+1}} \right| < \epsilon$$

with some shape parameter α , also called a *tail index*, and a scale parameter x_m .

Theorem 1 *If Z is a random variable with Pareto tails, then $\forall \gamma, p > 0$, $E[\log(1 + |Z|^\gamma)]^p < \infty$.*

Proof: To show that $E[\log(1 + |Z|^\gamma)]^p$ is finite we use

$$E[\log(1 + |Z|^\gamma)]^p = E[I_{|Z| < M} \log(1 + |Z|^\gamma)]^p + E[I_{|Z| > M} \log(1 + |Z|^\gamma)]^p. \quad (3)$$

Further we consider only upper tail with $Z > M$. The first term in (3) is finite, the second term $E[I_{|Z| > M} \log(1 + Z^\gamma)]^p$ is dominated above by

$$\begin{aligned} & \int_{I_{|z| > M}} [\log(1 + z^\gamma)]^p \frac{\alpha x_m^\alpha}{z^{\alpha+1}} dz + \epsilon \Pr(|z| > M) \\ &= \int_{I_{|z| > M}} \left[\frac{\log(1 + z^\gamma)}{\log(z^\gamma)} \right]^p [\log(z^\gamma)]^p z^{-\alpha-1} \alpha x_m^\alpha dz + \epsilon \Pr(|z| > M) \\ &= \int_{I_{|z| > M}} \left[\frac{\log(1 + z^\gamma)}{\log(z^\gamma)} \right]^p \gamma^p [\log(z)]^p z^{-\alpha-1} \alpha x_m^\alpha dz + \epsilon \Pr(|z| > M) \end{aligned} \quad (4)$$

Let $G_p(z) = [\log(z)]^p z^{-\alpha-1}$. Sequential application of L'Hospital's rule

$$\begin{aligned} \lim_{z \rightarrow \infty} G_p(z) &= \lim_{z \rightarrow \infty} \frac{\frac{\partial}{\partial z} [\log(z)]^p}{\frac{\partial}{\partial z} z^{\alpha+1}} = \lim_{z \rightarrow \infty} \frac{p [\log(z)]^{p-1} z^{-1}}{\alpha z^\alpha} \\ &= \lim_{z \rightarrow \infty} \frac{p [\log(z)]^{p-1}}{\alpha z^{\alpha+1}} = \lim_{z \rightarrow \infty} \frac{p [\log(z)]^{p-1}}{\alpha z^{\alpha+1}} = \lim_{z \rightarrow \infty} \frac{p}{\alpha} G_{p-1}(z) \end{aligned}$$

gives a recursive relationship which ensures that $G_p(z)$ converges down to zero as $z \rightarrow \infty$, as

$$\lim_{z \rightarrow \infty} G_p(z) = \lim_{z \rightarrow \infty} \frac{p^p}{\alpha^p} G_0(z) = \frac{p^p}{\alpha^p} \lim_{z \rightarrow \infty} \frac{1}{z^{\alpha+1}} = 0.$$

Other terms in the integrand of the integral of (4) also converge either to zero or to a constant.

Thus, for a sufficiently large M there exists ϵ' so that

$$\left[\frac{\log(1 + z^\gamma)}{\log(z^\gamma)} \right]^p < \epsilon'$$

and, going back to two tails with $Z > M$ (upper) and $Z < -M$ (lower),

$$\begin{aligned} \mathbb{E} [I_{|Z|>M} \log(1 + Z^\gamma)]^p &< \int_{I_{|z|>M}} [\log(1 + z^\gamma)]^p \frac{\alpha x_m^\alpha}{z^{\alpha+1}} dz + \epsilon \Pr(|z| > M) \\ &< \epsilon' \int_{I_{|z|>M}} \frac{\alpha x_m^\alpha}{z^{\alpha+1}} dz + \epsilon \Pr(|z| > M) < \epsilon' + \epsilon \end{aligned} \quad (5)$$

which ensures that the integral (4) is finite when $\alpha > 0$. **Q.E.D.**

4 Permutation tests for heavy-tailed distributions

4.1 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

$$(\hat{\mu}_x, \hat{\sigma}_x, \hat{\mu}_y, \hat{\sigma}_y) = \arg \max_{\mu_x, \sigma_x, \mu_y, \sigma_y} \prod_{i=1}^{n_x} f(d_i | \mu_x, \sigma_x) \prod_{i=n_x+1}^n f(d_i | \mu_y, \sigma_y), \quad (6)$$

and under H_0

$$(\hat{\mu}, \hat{\sigma}) = \arg \max_{\mu, \sigma} \prod_{i=1}^{n_x} f(d_i | \mu, \sigma) \prod_{i=n_x+1}^n f(d_i | \mu, \sigma). \quad (7)$$

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

$$l = -2 \sum_{i=1}^{n_x} \log \frac{\hat{\sigma}_x \pi^{-1}}{(d_i - \hat{\mu}_x)^2 + \hat{\sigma}_x^2} - 2 \sum_{i=n_x+1}^n \log \frac{\hat{\sigma}_y \pi^{-1}}{(d_i - \hat{\mu}_y)^2 + \hat{\sigma}_y^2} + 2 \sum_{i=1}^n \log \frac{\hat{\sigma} \pi^{-1}}{(d_i - \hat{\mu})^2 + \hat{\sigma}^2} \quad (8)$$

where the MLEs under H_A and H_0 are found as solutions to maximization tasks (6) and (7) respectively. The solutions were found numerically using box constrained maximization quasi-Newton optimization Zhu et al. (1995). The domain for location parameter was constrained to the $(-3, 3)$ range and the domain of scale was $(0, 2)$. To reduce chances of local maximums, the maximization procedure was performed 10 times with starting values chosen randomly from uniform distributions: $\text{Unif}(-3, 3)$ for the location and $\text{Unif}(1.25, 2.75)$. 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 realizations 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

$$l^* = -\frac{1}{n} \sum_{i=1}^{n_x} \log \frac{\hat{\sigma}_x}{(d_i - \hat{\mu}_x)^2 + \hat{\sigma}_x^2} - \frac{1}{n} \sum_{i=n_x+1}^n \log \frac{\hat{\sigma}_y}{(d_i - \hat{\mu}_y)^2 + \hat{\sigma}_y^2} \quad (9)$$

4.2 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 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 4.1. If, however, the distribution is not known permutations suggest a viable alternative solution.

4.3 Permutation test

Let

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

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

$$\{\mathbf{D}(\pi_k) = (\underbrace{D_1(\pi_k), \dots, D_{n_x}(\pi_k)}_{X(\pi_k)}, \underbrace{D_{n_x+1}(\pi_k), \dots, D_n(\pi_k)}_{Y(\pi_k)})\}_{k=1}^{r_2}, \quad (10)$$

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.

Section 5 compares statistical power of permutation tests based on L_γ 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 | $L_{0.5}$ | L_1 | L_2 | LL_{norm} | LL_{cauchy} | LL_{levy} | wilcox.test | ks.test |
|---------|-------------|-----------|-------|-------|-------------|---------------|-------------|-------------|---------|
| N(0, 1) | N(0, 1) | 4.9 | 5 | 4.3 | 5.3 | 5.1 | 5.2 | 4.1 | 4.1 |
| N(0, 1) | N(0.6, 1) | 71.8 | 75.3 | 78.9 | 75 | 62.2 | 19.2 | 81.7 | 68.4 |
| N(0, 1) | N(0, 2) | 88.6 | 89.3 | 90.3 | 99.1 | 69.8 | 75.5 | 5.5 | 37.7 |
| N(0, 1) | N(0.5, 1.5) | 67.5 | 71 | 72.5 | 88.2 | 51.7 | 16.8 | 45.9 | 50.7 |
| N(0, 1) | N(0.5, 2) | 94.3 | 95.1 | 95.4 | 99.6 | 81.4 | 56.1 | 33.3 | 69.1 |
| C(0, 1) | C(0, 1) | 4.7 | 4.5 | 4.5 | 4.7 | 4.2 | 6.1 | 5.5 | 4.1 |
| C(0, 1) | C(1, 1) | 81.3 | 80.8 | 79.1 | 5.9 | 85.3 | 5.2 | 72.2 | 80.2 |
| C(0, 1) | C(0, 3) | 91.1 | 91.2 | 90.9 | 33.8 | 93.7 | 24.5 | 6.4 | 45.9 |
| C(0, 1) | C(1, 2) | 82.2 | 81.9 | 80.7 | 17.2 | 85.3 | 11.8 | 46.1 | 65.6 |
| C(0, 1) | C(1, 3) | 96.2 | 96.4 | 96.4 | 31 | 97.8 | 21.8 | 28.8 | 70.4 |
| L(0, 1) | L(0, 1) | 5.1 | 5.3 | 5.3 | 5.6 | 5 | 4.3 | 4.6 | 3.4 |
| L(0, 1) | L(1, 1) | 81.2 | 66.1 | 50.9 | 4.4 | 39.6 | 100 | 52.9 | 92.3 |
| L(0, 1) | L(0, 3) | 74.6 | 72.3 | 70.8 | 11.3 | 69.3 | 92.7 | 85.5 | 76.2 |
| L(0, 1) | L(0.5, 1.5) | 43.4 | 34.3 | 29.6 | 5.9 | 27.5 | 98.9 | 50.5 | 58.6 |
| L(0, 1) | L(0.5, 2) | 65.5 | 56.4 | 51.6 | 6.1 | 48.8 | 99.5 | 73.1 | 78 |

5 Power comparisons: Monte-Carlo simulations

Monte-Carlo studies covered simulation scenarios generated from stable distributions: normal $N(\mu, \sigma)$, Cauchy $C(\mu, \sigma)$, and Levy $L(\mu, \sigma)$. Experimental settings included $n_x = n_y = 5$ (small) and $n_x = n_y = 50$ (moderate) sample sizes. Significance level was set 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 suggested by Keller-Mcnulty and Higgins (1987) and used by Marozzi (2004)).

Table 3: Statistical power at $n_X = n_Y = 50$

| F_1 | F_2 | $L_{0.5}$ | L_1 | L_2 | LL_{cauchy} | $LL_{logcauchy}$ | wilcox.test | ks.test |
|----------|----------|-----------|-------|-------|---------------|------------------|-------------|---------|
| LC(0, 1) | LC(0, 1) | 4.4 | 4.3 | 4.2 | 6 | 6.1 | 6.5 | 5.1 |
| LC(0, 1) | LC(1, 1) | 65.9 | 65 | 64.8 | 87.7 | 86.7 | 71.9 | 80.4 |
| LC(0, 1) | LC(0, 3) | 61.5 | 55.9 | 52.6 | 41.4 | 94.4 | 5.9 | 45.8 |
| LC(0, 1) | LC(1, 2) | 65.1 | 64.5 | 64 | 72.3 | 86 | 44.2 | 66.9 |
| LC(0, 1) | LC(1, 3) | 79.9 | 79.1 | 78.1 | 67.9 | 97.2 | 28.5 | 70.8 |

6 Illustrative example

7 Discussion

Permutation tests based on L_γ test statistics (U-statistics with logarithmic kernel) show high power properties in the class of distributions with Paretian tails. The power properties are slightly inferior to the likelihood ratio tests (LRT), but in contrast to LRTs no parametric assumptions are needed except for an assumption of Paretian tails in the class of absolutely continuous distributions. Monte-Carlo simulations showed the robustness of L_γ -test for stable distributions.

References

- Chung, E. and J. P. Romano (2016). Asymptotically valid and exact permutation tests based on two-sample u-statistics. *Journal of Statistical Planning and Inference* 168, 97–105.
- Keller-Mcnulty, A. and J. J. Higgins (1987). Effect of tail weight and outliers on power and type-i error of robust permutation tests for location: Effect of tail weight and outliers on power and type-i error. *Communications in Statistics-Simulation and Computation* 16(1), 17–35.
- Lehmann, E. (1986). *Testing Statistical Hypotheses*. Probability and Statistics Series. Wiley.
- Marozzi, M. (2004). Some remarks about the number of permutations one should consider to perform a permutation test. *Statistica* 64(1), 193–201.
- Zhu, C., R. Byrd, P. Lu, and J. Nocedal (1995). A limited memory algorithm for bound constrained optimisation. *SIAM J. Sci. Stat. Comput* 16(5), 1190–1208.