

A Location-Scale Permutation Test

Marco Marozzi

Abstract A permutation test for the location-scale problem is proposed within the nonparametric combination framework. The test is based on the combination of the permutation version of the Student t test and the permutation version of Pan $L50$ test for scale. Type-one error rate and power of this test are compared with those of Lepage and Cucconi tests. It is shown that the proposed test is preferable under normal- and light-tailed distributions.

Keywords: testing statistical hypotheses, resampling and permutation methods, computer intensive methods

1 Introduction

Two samples, the $\underline{X}_1$ sample and the $\underline{X}_2$ sample of n_1 and n_2 independent observations from populations with continuous distribution functions F_1 and F_2 respectively, are considered. The problem is to test the null hypothesis

$$F_1(t)=F_2(t) \quad \forall t \in]-\infty, +\infty[.$$

In particular, we consider the location-scale problem, and therefore we assume that

$$F_1(t) = G\left(\frac{t - \mu_1}{\sigma_1}\right) \text{ and that } F_2(t) = G\left(\frac{t - \mu_2}{\sigma_2}\right),$$

where $G(\cdot)$ denotes the distribution function for a continuous variable with location 0 and scale 1, μ_1 and μ_2 (σ_1 and σ_2) the locations (scales) of population 1 and 2, respectively. Since the normal assumption is often not fulfilled in practice and sample

sizes are often not large, the problem is addressed within the nonparametric permutation framework.

The better known nonparametric method for the location-scale problem is the test of Lepage (1971) which is a rank test based on the direct combination (i.e. the sum) of the squares of the standardized Wilcoxon-Mann-Whitney statistic, and the standardized Ansari-Bradley statistic. In the literature, nonparametric tests for jointly detecting location and scale changes are based on the combination of two tests, one for location and another for scale. Generally, as in the case of the Lepage test, the combination is achieved through the sum of the squared standardized test statistics. Cucconi (1968) proposed a different solution based on the squares of ranks and contrary-ranks. This test has been recently studied by Marozzi (2009a) that performed a detailed power simulation study of the Cucconi test and computed a table of critical values that were missing in the literature. Moreover he showed that the test of Cucconi maintains the size very close to α and is more powerful than the Lepage test.

Pesarin (2001) proposed a very general framework concerning the nonparametric combination of tests. The idea of this paper is to propose a permutation test for the location-scale problem within the nonparametric combination framework. More precisely, we propose in section 2 a test based on the combination of a test for location and a test for scale. In section 3 a simulation study is performed to compare type-one error rate and power of the proposed test with those of the Lepage and Cucconi tests. Some concluding remarks ends the paper in section 4.

2 The Permutation Test

Let $\underline{X} = (\underline{X}_1, \underline{X}_2) = (X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2}) = (X_1, \dots, X_{n_1}, X_{n_1+1}, \dots, X_n)$ denote the pooled sample. To obtain the PLS permutation test for the location-scale problem, first we separately address the location problem and the scale problem. Let

$$\underline{Y} = (\underline{X}_1 / \text{MAD}(\underline{X}_1), \underline{X}_2 / \text{MAD}(\underline{X}_2)) = (X_1 / \text{MAD}(\underline{X}_1), \dots, X_n / \text{MAD}(\underline{X}_2)) = (Y_1, \dots, Y_n)$$

be the median absolute deviation rescaled pooled sample, and let

$$\underline{Z} = (\underline{X}_1 - \text{MED}(\underline{X}_1), \underline{X}_2 - \text{MED}(\underline{X}_2)) = (X_1 - \text{MED}(\underline{X}_1), \dots, X_n - \text{MED}(\underline{X}_2)) = (Z_1, \dots, Z_n)$$

be the median aligned pooled sample. Let $(u_1^*, \dots, u_n^*)$ be a random permutation of unit labels $(1, \dots, n)$, then the corresponding permutation of $\underline{Y}$ and $\underline{Z}$ are respectively $\underline{Y}^* = (Y_{u_1^*}, \dots, Y_{u_n^*}) = (Y_1^*, \dots, Y_n^*)$ and $\underline{Z}^* = (Z_{u_1^*}, \dots, Z_{u_n^*}) = (Z_1^*, \dots, Z_n^*)$.

To compare locations μ_1 and μ_2 of the two populations we use the test based on

$$PT(\underline{Y}^*) = \frac{\left| \sum_{i=1}^{n_1} Y_i^* / n_1 - \sum_{i=n_1+1}^n Y_i^* / n_2 \right|}{\sqrt{\left((n_1^{-1} + n_2^{-1}) \left(\sum_{i=1}^{n_1} (Y_i^* - \sum_{i=1}^{n_1} Y_i^* / n_1)^2 + \sum_{i=n_1+1}^n (Y_i^* - \sum_{i=n_1+1}^n Y_i^* / n_2)^2 \right) \right) / (n-2)}},$$

which is the permutation version of the Student statistic computed on the median absolute deviation rescaled elements.

To compare scales σ_1 and σ_2 of the two populations we use the *PL50* test, which is the permutation version (Marozzi, 2009b) of the *L50* test due to Pan (1999) and based on

$$PL50(\underline{Z}^*) = \frac{|\ln \bar{Z}_1^* - \ln \bar{Z}_2^*|}{\sqrt{\frac{1}{n_1} \frac{(V_1^*)^2}{(\bar{Z}_1^*)^2} + \frac{1}{n_2} \frac{(V_2^*)^2}{(\bar{Z}_2^*)^2}}},$$

where $\bar{Z}_1^* = \sum_{i=1}^{n_1} |Z_i^*| / n_1$, $\bar{Z}_2^* = \sum_{i=n_1+1}^n |Z_i^*| / n_2$, $(V_1^*)^2 = \sum_{i=1}^{n_1} (|Z_i^*| - \bar{Z}_1^*)^2 / (n_1 - 1)$ and

$(V_2^*)^2 = \sum_{i=n_1+1}^n (|Z_i^*| - \bar{Z}_2^*)^2 / (n_2 - 1)$. In detecting differences in scale, the *PL50* test behaves better than the original version under a range of distributions from symmetric to skewed and from low to high kurtosis (Marozzi, 2009b). The original *L50* test was proposed as a modification of the traditional Levene test to fix a structural problem of this latter.

Resampling from $\underline{Y}$ and $\underline{Z}$ respectively for the *PT* test and the *PL50* test, instead of resampling directly from $\underline{X}$, is necessary in order that pooled sample elements were at least approximately exchangeable under the corresponding partial null hypothesis H_{0a} : $\mu_1 = \mu_2$ in the first case and H_{0b} : $\sigma_1 = \sigma_2$ in the second one. To compute the *p*-values of *PT* and *PL50* we compute the permutation null distribution of *PT* and *PL50* statistics as the distribution function of the permutation values of them: ${}_1PT, \dots, {}_KPT$ where ${}_kPT = PT(\underline{Y}_k^*)$, $\underline{Y}_k^*$ is the *k*th permutation of $\underline{Y}$ and $k=1, \dots, K=n!/n_1!/n_2!$; and ${}_1PL50, \dots, {}_KPL50$ where ${}_kPL50 = PL50(\underline{Z}_k^*)$, $\underline{Z}_k^*$ is the *k*th permutation of $\underline{Z}$ and $k=1, \dots, K$. Therefore the *p*-values are respectively

$$M_{PT}({}_0PT) = \frac{n_1!n_2!}{n!} \sum_{k=1}^{\frac{n!}{n_1!n_2!}} I({}_kPT \geq {}_0PT) \text{ and } M_{PL50}({}_0PL50) = \frac{n_1!n_2!}{n!} \sum_{k=1}^{\frac{n!}{n_1!n_2!}} I({}_kPL50 \geq {}_0PL50),$$

where ${}_0PT$ and ${}_0PL50$ are the observed values ${}_0PT = PT(\underline{Y})$ and ${}_0PL50 = PL50(\underline{Z})$ of the statistics and $I(\cdot)$ denotes the indicator function.

To obtain the *PLS* test we combine the *p*-values of the *PT* and *PL50* tests. This is equivalent to combine the test statistics since *p*-values and statistics are one to one decreasingly related. We use the Fisher combining function obtaining the following test statistic for the null hypothesis H_0 : $H_{0a} \cap H_{0b}$

$$PLS = \ln(1/M_{PT}) + \ln(1/M_{PL50}).$$

Pesarin (2001) reports several combining functions and states that there exist situations in which one combining function is preferable to the others. For example, the Tippett combining function may have a good power behavior when only one among the partial null hypotheses is true; that of Liptak is generally good when the partial null hypotheses are jointly true. The Fisher combining function is suggested when nothing is expected about the partial null hypotheses because has an intermediate behaviour, and for this reason it is adopted here. See also Zaykin et al. (2002) and Won et al. (2009).

The observed value of PLS is ${}_0PLS = \ln(1/M_{PT}({}_0PT)) + \ln(1/M_{PL50}({}_0PL50))$. The null distribution of PLS is the distribution function of ${}_1PLS, \dots, {}_KPLS$ where ${}_kPLS = \ln(1/M_{PT}({}_kPT)) + \ln(1/M_{PL50}({}_kPL50))$. Large values of ${}_0PLS$ are evidence against H_0 , that should be rejected if $M_{PLS}({}_0PLS) \leq \alpha$ where $M_{PLS}({}_0PLS) = \sum_{k=1}^K I({}_kPLS \geq {}_0PLS) / K$.

3 A Comparative Simulation Study

A simulation study was designed to evaluate the size and the power of the proposed test, the Cucconi test and the Lepage test. We used five distributions, that range from low to high kurtosis, in the simulations to generate sample observations:

- standard normal distribution $N(0,1)$;
- uniform distribution;
- bimodal obtained as a mixture of a $N(-1.5,1)$ with probability 0.5 and a $N(1.5,1)$ with probability 0.5;
- Laplace double exponential distribution;
- Student's t with 2 df.

Note that we consider only symmetric distributions because if one considers skewed distributions a change in location is not qualitatively different with respect to a change in scale and therefore it is not well specified the location-scale alternative. We considered $(n_1, n_2) = (10, 10), (10, 30), (30, 10)$ and $(30, 30)$. The null hypothesis tested was that the populations have the same distribution. The alternative hypothesis is that $\mu_1 - \mu_2 \neq 0$ or $\sigma_1 / \sigma_2 \neq 1$. We considered three cases for examining the power: (i) $\mu_1 - \mu_2 = 0$ and $\sigma_1 / \sigma_2 > 1$; (ii) $\mu_1 - \mu_2 > 0$ and $\sigma_1 / \sigma_2 = 1$; (iii) $\mu_1 - \mu_2 > 0$ and $\sigma_1 / \sigma_2 > 1$. To speed up computations, 1000 permutations were computed. For each simulation setting 5000 Monte-Carlo simulations were considered. Note that for the Glivenko-Cantelli theorem, 1000 permutations are enough for permutation p -value estimation (see among others Pesarin (2001) and Marozzi (2004a)).

Tables 1 and 2 display the results, the size of the tests appears in the first column of inner cells, which corresponds to the null hypothesis. The Lepage and the Cucconi test maintain their size very close to the nominal 5% level. The PLS test has a slightly greater size than that of them, but for the larger samples its size is very similar. The PLS test is more powerful than the L and the C test for the normal, uniform, and bimodal distributions, with the exceptions of the situations in which the scale but not the

location is shifted for $(n_1, n_2) = (10, 10)$ and $(10, 30)$. For the Laplace double exponential distribution the *PLS* is practically as powerful as the *L* and the *C* test. For the Student *t* with 2 df the *PLS* test is less powerful. Therefore the combined location-scale *PLS* test works better than the *L* and the *C* test under normal- and light-tailed distributions, whereas the *L* and the *C* test are preferable for heavier than normal tailed distributions. It should be noted that the Lepage and the Cucconi test perform quite similarly, but the Cucconi test is on the whole more powerful than the Lepage test, in accordance with the results found by Marozzi (2009a). The results of *PLS* under the unbalanced sample size settings worth a comment. The reason is that the unbalancement is very high: a size is three times than the other one. This affects in particular the power of *PLS* under H_{1b} (and therefore under H_1) but not under H_{1a} . The power is higher when the scale shift is in the larger sample.

Table 1. Power estimates with $\alpha=0.05$ and $(n_1, n_2) = (10, 10)$ and $(10, 30)$.

$(n_1, n_2) = (10, 10)$					$(n_1, n_2) = (10, 30)$			
Normal					Normal			
$\mu_1 - \mu_2$	0	0	1	1	0	0	0.75	0.75
σ_1 / σ_2	1	2	1	2	1	1.5	1	1.5
<i>PLS</i>	0.053	0.260	0.538	0.513	0.057	0.130	0.522	0.470
<i>L</i>	0.050	0.239	0.416	0.390	0.053	0.198	0.392	0.433
<i>C</i>	0.049	0.269	0.414	0.419	0.050	0.246	0.395	0.478
Uniform					Uniform			
$\mu_1 - \mu_2$	0	0	1	1	0	0	0.75	0.75
σ_1 / σ_2	1	2	1	2	1	1.5	1	1.5
<i>PLS</i>	0.063	0.352	0.494	0.596	0.074	0.190	0.536	0.541
<i>L</i>	0.053	0.381	0.351	0.432	0.046	0.333	0.350	0.434
<i>C</i>	0.056	0.458	0.334	0.479	0.048	0.447	0.356	0.507
Bimodal					Bimodal			
$\mu_1 - \mu_2$	0	0	1.5	2.5	0	0	1	2
σ_1 / σ_2	1	1.5	1	1.5	1	1.5	1	1.5
<i>PLS</i>	0.062	0.151	0.397	0.624	0.074	0.178	0.348	0.723
<i>L</i>	0.049	0.178	0.266	0.443	0.053	0.312	0.211	0.541
<i>C</i>	0.051	0.204	0.256	0.433	0.049	0.399	0.213	0.607
Laplace					Laplace			
$\mu_1 - \mu_2$	0	0	1	1	0	0	0.75	0.75
σ_1 / σ_2	1	2	1	2	1	1.5	1	1.5
<i>PLS</i>	0.055	0.189	0.579	0.446	0.060	0.091	0.537	0.438
<i>L</i>	0.053	0.160	0.542	0.436	0.057	0.133	0.556	0.531
<i>C</i>	0.050	0.172	0.551	0.454	0.055	0.162	0.553	0.548
t_2					t_2			
$\mu_1 - \mu_2$	0	0	1	2	0	0	0.8	1.1
σ_1 / σ_2	1	2.4	1	2.4	1	1.8	1	1.8
<i>PLS</i>	0.053	0.236	0.291	0.456	0.060	0.155	0.268	0.306
<i>L</i>	0.050	0.241	0.246	0.499	0.049	0.235	0.253	0.470
<i>C</i>	0.051	0.253	0.253	0.518	0.049	0.275	0.255	0.491

Table 2. Power estimates with $\alpha=0.05$ and $(n_1, n_2)=(30, 10)$ and $(30, 30)$.

$(n_1, n_2)=(30, 10)$					$(n_1, n_2)=(30, 30)$				
Normal					Normal				
$\mu_1 - \mu_2$	0	0	0.75	1	0	0	0.5	0.5	
σ_1 / σ_2	1	1.8	1	1.8	1	1.3	1	1.3	
<i>PLS</i>	0.060	0.395	0.484	0.711	0.047	0.186	0.473	0.523	
<i>L</i>	0.046	0.246	0.386	0.443	0.047	0.148	0.356	0.387	
<i>C</i>	0.049	0.246	0.389	0.435	0.045	0.165	0.356	0.405	
Uniform					Uniform				
$\mu_1 - \mu_2$	0	0	0.75	1	0	0	0.5	0.5	
σ_1 / σ_2	1	1.8	1	1.8	1	1.3	1	1.3	
<i>PLS</i>	0.066	0.539	0.457	0.813	0.054	0.285	0.463	0.622	
<i>L</i>	0.046	0.378	0.337	0.487	0.046	0.251	0.335	0.440	
<i>C</i>	0.048	0.456	0.341	0.462	0.048	0.353	0.322	0.483	
Bimodal					Bimodal				
$\mu_1 - \mu_2$	0	0	1	2	0	0	0.75	1.1	
σ_1 / σ_2	1	1.5	1	1.5	1	1.3	1	1.3	
<i>PLS</i>	0.077	0.295	0.314	0.741	0.051	0.268	0.352	0.690	
<i>L</i>	0.052	0.208	0.219	0.503	0.049	0.249	0.257	0.515	
<i>C</i>	0.053	0.219	0.221	0.442	0.047	0.304	0.250	0.545	
Laplace					Laplace				
$\mu_1 - \mu_2$	0	0	0.75	1	0	0	0.5	0.5	
σ_1 / σ_2	1	1.8	1	1.8	1	1.3	1	1.3	
<i>PLS</i>	0.060	0.273	0.527	0.601	0.059	0.129	0.489	0.457	
<i>L</i>	0.054	0.153	0.556	0.521	0.055	0.107	0.513	0.484	
<i>C</i>	0.050	0.138	0.553	0.544	0.056	0.117	0.506	0.486	
t_2					t_2				
$\mu_1 - \mu_2$	0	0	0.8	1.7	0	0	0.6	0.8	
σ_1 / σ_2	1	2.2	1	2.2	1	1.6	1	1.6	
<i>PLS</i>	0.054	0.316	0.245	0.517	0.058	0.216	0.251	0.394	
<i>L</i>	0.055	0.274	0.268	0.495	0.057	0.250	0.289	0.513	
<i>C</i>	0.049	0.249	0.268	0.515	0.054	0.260	0.289	0.522	

4 Conclusion

A permutation test for the location-scale problem has been proposed within the nonparametric combination framework. More precisely, the *PLS* test is based on the combination of the permutation version of the Student *t* test and of the permutation version of Pan *L50* test for scale. Type-one error rate and power of the test have been studied and compared with those of the test of Lepage and of the test of Cucconi. It has been shown that the *PLS* test is more powerful, with few exceptions, than the *L* and the

C test for the normal, uniform, and bimodal distributions. For the Laplace double exponential distribution the PLS is practically as powerful as the L and the C test. For the Student t with 2 df the PLS test is less powerful. Therefore the combined location-scale PLS test works better than L and C under normal- and light-tailed distributions, whereas the L and the C test are preferable for heavier than normal tail distributions. We caution that our results were obtained via simulation experiments.

References

1. Cucconi, O.: Un nuovo test non parametrico per il confronto tra due gruppi campionari. *Giornale degli Economisti* **XXVII**, 225-248 (1968)
2. Lepage, Y.: A combination of Wilcoxon's and Ansari-Bradley's statistics. *Biometrika* **58**, 213-217 (1971)
3. Marozzi, M.: Some remarks about the number of permutations one should consider to perform a permutation test. *Statistica* **LXIV**, 193-201 (2004a)
4. Marozzi, M.: A bi-aspect nonparametric test for the two-sample location problem. *Computational Statistics and Data Analysis* **44**, 639-648 (2004b)
5. Marozzi, M.: Multivariate tri-aspect non-parametric testing. *Journal of Nonparametric Statistics* **19**, 269-282 (2007)
6. Marozzi, M.: Some notes on the location-scale Cucconi test. *Journal of Nonparametric Statistics* **21**, 629-647 (2009a)
7. Marozzi, M.: A combined resampling test for scale in two-sample designs. *Proceedings of the 6th St. Petersburg Workshop on Simulation*, 563-568 (2009b)
8. Pan, G.: On a Levene type test for equality of two variances. *Journal of Statistical Computation and Simulation* **63**, 59-71 (1999)
9. Pesarin, F.: *Multivariate Permutation Tests with Applications in Biostatistics*. Wiley, Chichester (2001)
10. Won, S, Morris, N, Lu, Q. and Elston, R. C.: Choosing an optimal method to combine p -values. *Statistics in Medicine* **28**, 1537-1553 (2009)
11. Zaykin, D. V., Zhivotovsky, L. A., Westfall, P. H. and Weir, B. S.: Truncated product method for combining p -values. *Genetic Epidemiology* **22**, 170-185 (2002)