Quaderni di Statistica
Vol. 10, 2008

A permutation solution for two-sample location-scale test

Rosa Arboretti Giancristofaro
Department of Mathematics, University of Ferrara, Italy
E-mail: rbrrso@unife.it

Livio Corain
Department of Management and Engineering, University of Padova, Italy
E-mail: livio.corain@unipd.it

Summary: The aim of this work is to present and discuss a permutation solution for the two-sample location-scale testing problem by means of a simulation study. As suggested by the simulation results, we can confirm that the proposed solution is a good alternative to traditional procedures, such as the Lepage test. One of the greatest advantages of our permutation solution is that it has a good behaviour both under the null hypothesis and in power with small sample sizes. Hence, in each situation where the normality assumption may be hard to justify, this nonparametric procedure can be considered a valid solution.

Keywords: Location-scale problem, Nonparametric combination, Permutation tests.

1. Description of the problem

Let Z be a random variable with zero mean, variance equal to one and unknown continuous distribution P, and let Z_{ij}, $i=1,\ldots,n_j$, $j=1,2$, be the i.i.d elements of two independent random samples $\boldsymbol{Z}_j$, $j=1,2$, from Z.

With reference to the following model:

$$X_{ij} = \mu_j + \sigma_j \cdot Z_{ij},\ i=1,\ldots,n_j,\ \ j=1,2,$$

we wish to test the following hypotheses of interest:

$$\begin{cases} H_0 : F(X_1) = F(X_2) \\ H_1 : F(X_1) \neq F(X_2) \end{cases} \tag{1}$$

where $F(X_1)$ and $F(X_2)$ are the unknown cumulative distribution functions of the random variables X_1 and X_2. If hypothesis (1) can be rewritten as:

$$\begin{cases} H_0 : \{(m_1 = m_2) \cap (s_1 = s_2)\} \\ H_1 : \{(m_1 <\neq> m_2) \cup (s_1 <\neq> s_2)\} \end{cases} \tag{2}$$

we are dealing with the so-called location-scale testing problem.

Several solutions for this problem have been proposed in the literature following either a parametric or a nonparametric approach. The best known nonparametric solution was addressed within the rank framework and is the result of work carried out by Lepage (1971). It is based on a combination of the Wilcoxon and Mann-Whitney statistic W for location alternatives and the Ansari-Bradley statistic A for scale alternatives:

$$L = \frac{[W - E(W)]^2}{Var(W)} + \frac{[A - E(A)]^2}{Var(A)},$$

where $E(W)$, $E(A)$, $VAR(W)$ and $VAR(A)$ are the expected values and variances of W and A under H_0. Note that since W and A are not correlated under H_0, then the L statistic has a limiting chi-square distribution with 2 degrees of freedom. The Lepage test is considered to be robust in case of non-normal distributions.

Several other solutions have been proposed in the literature to increase the performance of the Lepage test (Podgor and Gastwirth, 1994). When replacing both components W and A by arbitrary linear rank tests it is possible to obtain so-called Lepage-type tests that were introduced by Büning and Thadewal (2000). Kössler (2006) computed their asymptotic efficacies and proposed an adaptive test as well.

Cucconi (1968) proposed a different rank solution to the same problem which, unfortunately, is not much known in the literature as the original paper was written in Italian and published by a national journal. However, the solution is interesting and it is based on the following test statistic:

$$C = \frac{U^2 + V^2 - 2rUV}{2(1 - r^2)},$$

where:

$$U = \frac{6\sum_{i=1}^{n_1} W_{i1}^2 - n_1(n+1)(2n+1)}{\sqrt{n_1 n_2 (n+1)(2n+1)(8n+11)/5}},$$

$$V = \frac{6\sum_{i=1}^{n_1} (n+1-W_{i1})^2 - n_1(n+1)(2n+1)}{\sqrt{n_1 n_2 (n+1)(2n+1)(8n+11)/5}},$$

$$r = \frac{2(n^2-4)}{(2n+1)(8n+11)} - 1,$$

and where W_{i1} is the rank of the first sample X_{i1} and $n=n_1+n_2$ is the sum of the sizes of the two samples. It is worth noting that U is based on the squares of the ranks W_{1i}, while V is based on the squares of the quantity $(n+1-W_{i1})$ of the first sample.

2. A permutation solution

In this section we present two permutation solutions for the two-sample joint location-scale problem, described in (2).

With reference to the theory of nonparametric combination of dependent permutation tests (Pesarin, 2001), this problem can be solved following a two-stage procedure. At first we perform two separate tests, one for location and one for scale, where both tests are approximately unbiased and consistent, then we combine them into a global test which is appropriate for testing (2).

As far as the first stage of our proposed solution is concerned, the two separate tests are indicated as T_m and T_s, where T_m is related to testing H_{0m}: $m_1=m_2$, i.e. the sub-hypothesis of equality of location parameters independently of scale, while the second test T_s consists in testing H_{0s}: $s_1=s_2$, i.e. the sub-hypothesis of equality of scale parameters independently of location.

With the goal of deriving an appropriate test T_m, let us assume that response X_j, is symmetrically distributed with unknown distribution P around the location parameter m_j, with scale parameter s_j, j=1,2. Let $\tilde{X}$ be the sample median of the pooled sample $\boldsymbol{X}=[\boldsymbol{X}_1,\boldsymbol{X}_2]$ and consider the transformation ${}_m\boldsymbol{Y}_j=(\boldsymbol{X}_j-\tilde{X})$, j=1,2. Note that the sub-hypothesis H_{0m}: $m_1=m_2$ is true if and only if the distribution of ${}_m\boldsymbol{Y}_j$, j=1,2, is symmetrical around zero. In this way, in order to test H_{0m}, it is appropriate to define two separate test statistics for symmetry, T_1 and T_2, one for each individual sample, then combine them with an appropriate combining function. More specifically, we define:

$$T_m = j\left(T_1 - T_2\right), T_j = \sum_{i=1}^{n_j} {}_mY_{ij} \cdot S_{ij} / n_j \,,\ j = 1,2 \tag{3}$$

where $j(\cdot)$ corresponds to $+(\cdot)$ if the alternative is '<', and to $-(\cdot)$ if the alternative is '>', and finally to absolute value $|(\cdot)|$ if the alternative is '?', and where S=$\{S_{ij},\ i=1,\ldots,n_j,\ j=1,2\}$ is a random sample from the random variable S which takes the values +1,−1 with probability ½.

Since m_1 and m_2 are unknown parameters, there is no exact solution for testing the sub-hypothesis H_{0s}: $s_1=s_2$, but we can obtain an approximate permutation solution by using the following test statistic:

$$T_s = j\left(\sum_{i=1}^{n_1} {}_sY_{1i}^2 / n_1 - \sum_{i=1}^{n_2} {}_sY_{2i}^2 / n_2\right). \tag{4}$$

where $j(\cdot)$ has the same meaning as above and where ${}_sY_{ij}=(X_{ij}-\bar{X}_j)$, $i=1,\ldots,n_j$, j=1,2, is the deviation of the X_{ij} elements from their own sample mean $\bar{X}_j$, j=1,2.

In order to obtain an appropriate permutation solution for H_{0h}, $h=m,s$, let us consider $T_{obs,h}$, $h=m,s$, i.e. the observed value of T_m and T_s, and let us independently and randomly exchange the elements ${}_hY_{ij}$ $i=1,\ldots,n_j$, j=1,2, $h=m,s$, of the pooled samples ${}_m\boldsymbol{Y}=[\boldsymbol{Y}_1,\boldsymbol{Y}_2]$ and ${}_s\boldsymbol{Y}=[\boldsymbol{Y}_1,\boldsymbol{Y}_2]$, with respect to the original two samples 1 and 2. Hence, we perform a random permutation between the two samples and we independently repeat a random permutation B times, obtaining an estimate of the permutation distribution of T_m and T_s under the null hypothesis.

Hence, we reject H_{0h}, h=m,s, when the observed p-value l_{obs} is lower than the selected significance a-level:

$$l_{obs,h} = \frac{\cdot \left[T_h^* \geq T_{obs,h}\right]}{B},$$

where h=m,s, and B is the number of random permutations.

An alternative solution that can be used to perform two separate tests for location and scale can be based on the approximation of permutation tests on known distributions. If we assume that s_j, j=1,2, is finite and if the sample size is large enough, we can refer to the Central Limit Theorem to approximate the standardized distribution of T_m to the standard normal distribution:

$$\frac{T_m}{\sum_{i=1}^{n_1} {}_mY_{1i}^2 \big/ n^2{}_1 + \sum_{i=1}^{n_2} {}_mY_{2i}^2 \big/ n^2{}_2} \xrightarrow{d} N(0,1).$$

Moreover, by using the approximation of the sample variance to the chi-square distribution, we can approximate the distribution of T_s to Fisher's F distribution, i.e. :

$$T_s = \frac{\left[\sum_{i=1}^{n_1} {}_sY_{1i}^2 \big/ (n_1 - 1)\right] / (n_1 - 1)}{\left[\sum_{i=1}^{n_2} {}_sY_{2i}^2 \big/ (n_2 - 1)\right] / (n_2 - 1)} \approx F_{n_1 - 1, n_2 - 1}.$$

In order to obtain a final test, which is appropriate for testing hypotheses (2), we have to consider a global test T'' as a combination of the two aspect under testing, i.e. location and scale. For this goal, let us consider the nonparametric combination methodology (Pesarin, 2001), to be applied on the two previously obtained tests T_m and T_s. When considering an appropriate combining function y, the derivation of T'' consists in the following steps:

2.a the combined observed value of the second-order test is evaluated using the same random permutations as in the first stage, and is given by:

$$T''_{obs} = y\,(l_{obs,m}, l_{obs,s})\,;$$

2.b the *r*-th combined value of two statistics is then calculated by:

$$T''^{*}_{r} = y\,(l^{*}_{r,m}, l^{*}_{r,s})\,,$$

where $l^{*}_{r,h} = .\,\left[T^{*}_{h} \geq T^{*}_{r,h}\right]\big/ B$, h=m,s, r =1,…,B;

2.c the p-value of combined test T'' is thus estimated as:

$$l''_{obs,y} = \frac{.\,\left[T''^{*} \geq T''_{obs}\right]}{B}\,;$$

2.d if $l'' \leq a$, the global null hypothesis H_0 is rejected at significance level a;

A general characterization of the class of combining functions is given by the following three main features for combining function y (Pesarin, 2001):

a) it must be non-increasing in each argument:

$$y\,(..., l_h, ...) \geq y\,(..., l'_h, ...)\ if\ l_h < l'_h,\ h \in \{1,\ldots,k\};$$

b) it must attain its supremum value $\bar{y}$, possibly non finite, even when only one argument reaches zero:

$$y\,(..., l_h, ...) \rightarrow \bar{y}\ \text{if}\ l_h \rightarrow 0,\ h \in \{1,\ldots,k\};$$

c) $\forall a > 0$, the critical value of every y is assumed to be finite and strictly smaller than the supremum value:

$$T''_a < \bar{y}\,.$$

The above properties define the class C of combining functions. Some of the functions most often used to combine independent tests (Fisher, Lancaster, Liptak, Tippett, Mahalanobis, etc.) are included in this class. We have considered the following ones:

– Fisher combination: $T''_F = -2 \cdot \sum_h \log(l_h)$,

– Liptak combination: $T''_L = -\sum_h \Phi^{-1}(1 - l_h)$,

– Tippett combination: $T''_T = -\min_{1 \leq h \leq k}(l_h)$.

When using the approximation of permutation tests T_m and T_s on standard normal and F distribution, it is possible to approximate the above combining functions on known distributions as well, more specifically:

– Fisher combination: $T''_F \xrightarrow{d} c_4^2$,

– Liptak combination: $T''_F \xrightarrow{d} N(0,2)$,

– Tippett combination: $T''_T \xrightarrow{d} Triangular(0,1,0)$.

3. Simulation study

In this section we evaluate the appropriateness of the proposed permutation solutions for two-sample location-scale testing by means of a set of three Monte Carlo simulation studies. For each simulation study we estimated the global permutation *p*-value (with 1000 random permutations) for joint location-scale testing (against bilateral alternatives), and we applied Fisher, Liptak and Tippett combining functions. For the sake of comparison, we also used the approximation of permutation tests T_m and T_s on known distributions.

Depending on the sample sizes, we consider two main simulation settings:

1. the first setting with several combinations of small/moderate and balanced/unbalanced sample sizes, where data are generated from normal distribution;
2. the second setting with balanced sample size, where data are generated from some non-normal distributions;

The first considered simulation setting consists of 1000 Monte Carlo simulations for the generation of two samples, with the following sample size:

– n_1=5, n_2=10,
– n_1=10, n_2=10,
– n_1=20, n_2=40,

where data are generated from normal distribution with parameters:

- $m_1 = 0$, $s_1 = 1$; $m_2 = 0$, $s_2 = 1$;
- $m_1 = 0$, $s_1 = 1$; $m_2 = 0.2$, $s_2 = 1.5$;
- $m_1 = 0$, $s_1 = 1$; $m_2 = 0.5$, $s_2 = 2$;
- $m_1 = 0$, $s_1 = 1$; $m_2 = 0.8$, $s_2 = 4$.

The second considered simulation setting consists of 1000 Monte Carlo simulations for the generation of two samples with balanced sample size $n_1 = n_2 = 25$, where data are generated from the following distributions:

1. normal,
2. Laplace,
3. log-normal,

with location and scale parameters:

- $m_1 = 0$, $s_1 = 1$; $m_2 = 0$, $s_2 = 1$;
- $m_1 = 0$, $s_1 = 1$; $m_2 = 0.5$, $s_2 = 1.5$;
- $m_1 = 0$, $s_1 = 1$; $m_2 = 0.5$, $s_2 = 2$;
- $m_1 = 0$, $s_1 = 1$; $m_2 = 1$, $s_2 = 1.5$;
- $m_1 = 0$, $s_1 = 1$; $m_2 = 1$, $s_2 = 2$.

With the aim at comparing our proposed permutation solutions with other location-scale testing procedures suggested in the literature, we perform a third final simulation study by considering the following competitors: Lepage test, modified Lepage test, adaptive Lepage test, rank-sum test, Grambsch - O'Brien generalized *t*-test and Grambsch - O'Brien generalized rank-sum test. For details on such competitors we refer the reader to Hollander and Wolfe (1999) and Grambsch - O'Brien (1991). It consists of 1000 Monte Carlo simulations of two samples with balanced sample size $n_1 = 25$, $n_2 = 25$, drawn from normal distribution with the following parameter settings:

- $m_1 = 0$, $s_1 = 1$; $m_2 = 0$, $s_2 = 1$;
- $m_1 = 0$, $s_1 = 1$; $m_2 = 1$, $s_2 = 2$.

Results from the three simulation studies are shown in Tables 1, 2.a, 2.b and 3.

Table 1. Rejection rates (a=0.05) for the first simulation study

Combining function	permutation tests			approximated permutation tests on known distributions		
	n_1=5 n_2=10	n_1=10 n_2=10	n_1=20 n_2=40	n_1=5 n_2=10	n_1=10 n_2=10	n_1=20 n_2=40
	$N_1(0,1)$ $N_2(0,1)$			$N_1(0,1)$ $N_2(0,1)$		
Fisher	5.5	4.9	5.8	7.5	4.5	9.2
Liptak	5.3	5.0	5.4	6.6	5.1	7.1
Tippett	4.9	5.3	5.4	8.5	5.7	7.3
	$N_1(0,1)$ $N_2(0.2,1.5)$			$N_1(0,1)$ $N_2(0.2,1.5)$		
Fisher	4.2	9.6	16.8	12.6	13.1	16.7
Liptak	4.8	9.2	17.4	11.2	12.1	16.1
Tippett	4.0	9.4	10.7	9.5	10.2	19.3
	$N_1(0,1)$ $N_2(0.5,2)$			$N_1(0,1)$ $N_2(0.5,2)$		
Fisher	10.8	23.2	53.2	14.2	19.3	29.0
Liptak	13.6	23.8	56.8	13.5	18.6	26.4
Tippett	7.5	17.3	39.9	14.1	20.0	24.8
	$N_1(0,1)$ $N_2(0.8,4)$			$N_1(0,1)$ $N_2(0.8,4)$		
Fisher	21.8	53.6	90.0	14.5	51.3	85.8
Liptak	22.6	52.6	88.2	12.7	51.8	84.3
Tippett	11.4	46.2	81.7	15.1	34.5	81.3

Table 2.a Rejection rates (a=0.05) for the second simulation study (permutation tests)

Random Distribution	Combining function	m_1=0,s_1=1 m_2=0,s_2=1	m_1=0,s_1=1 m_2=0.5,s_2=1.5	m_1=0,s_1=1 m_2=0.5,s_2=2	m_1=0,s_1=1 m_2=1,s_2=1.5	m_1=0,s_1=1 m_2=1,s_2=2
		permutation tests				
Normal	Fisher	4.4	58.7	86.6	92.5	96.4
	Liptak	4.6	57.8	74.5	93.3	94.4
	Tippett	4.7	53.5	87.8	86.1	93.8
Laplace	Fisher	5.2	33.5	54.2	65.9	75.1
	Liptak	5.2	32.7	46.2	65.8	72.1
	Tippett	5.2	29.6	52.6	59.1	70.4
Log-Normal	Fisher	8.0	62.8	80.2	91.9	95.3
	Liptak	9.2	68.2	85.4	93.1	97.1
	Tippett	3.0	42.8	63.2	81.3	87.4

Table 2.b Rejection rates (a=0.05) for the second simulation study (approximated permutation tests)

Random Distribution	Combining function	$m_1=0, s_1=1$ $m_2=0, s_2=1$	$m_1=0, s_1=1$ $m_2=0.5, s_2=1.5$	$m_1=0, s_1=1$ $m_2=0.5, s_2=2$	$m_1=0, s_1=1$ $m_2=1, s_2=1.5$	$m_1=0, s_1=1$ $m_2=1, s_2=2$
		approximated permutation tests on known distributions				
	Fisher	4.5	46.0	82.7	81.1	93.7
Normal	Liptak	4.7	44.2	73.0	80.6	93.3
	Tippett	4.2	38.8	82.6	70.7	87.1
	Fisher	12.3	43.6	74.4	62.2	81.4
Laplace	Liptak	10.9	39.2	65.2	62.0	77.9
	Tippett	12.0	42.1	73.2	55.9	78.9
	Fisher	37.8	91.3	97.3	97.6	99.8
Log-Normal	Liptak	35.4	90.7	97.2	97.8	99.8
	Tippett	39.3	90.1	97.3	96.7	99.8

Table 3. Rejection rates (a=0.05) for the third simulation study

Location-scale test	$m_1=0, s_1=1$ $m_1=0, s_1=1$	$m_1=0, s_1=1$ $m_2=1, s_2=2$
Permutation test	4.4	96.4
Lepage	4.8	84.3
Modified Lepage	4.6	86.7
Adaptive Lepage	4.6	87.5
Rank sum	9.8	67.9
Grambsch-O'Brien gen. t	9.3	92.2
Grambsch-O'Brien gen. rank-sum	9.7	91.8

4. Discussion and conclusion

Table 1 shows that when either the sample size and the true difference between parameters of two populations increases, the rejection rates also increase considerably and this is confirmation that the proposed permutation solution is a valid solution for the two-sample location-scale testing problem. This is true whether starting from permutation or approximated permutation tests with respect to the first-stage separated location and scale tests. As far as the combining function is concerned, Fisher and Liptak seem to be more 'stable' than Tippett combining function.

When considering several types of random distributions for data generation (Tables 2.a and 2.b), it is worth noting that:

- under normality, permutation tests have a slightly better power performance than approximated permutation tests on known distributions, most likely because despite data being normal, the sample size of n_1=25, n_2=25 is never too large;
- under non normal distributions, the power of approximated permutation tests is only apparently higher than the power of permutation tests since under H_0, approximated permutation tests present inflated nominal alpha levels.

With reference to the comparison with solutions proposed in the literature (Table 3), first of all it is worth noting that our proposed permutation test does respect the nominal a-level under H_0 and this is also true for Lepage-type tests. Moreover, in term of rejection rates, permutation solution seems to be more powerful than Lepage-type tests. Under H_0 Rank sum and Grambsch-O'Brien-type tests have an anticonservative behaviour, hence under H_1 the related rejection rates are inflated. In conclusion, we can say that the proposed permutation solution for the two-sample location-scale testing problem is a good alternative compared to traditional procedures such as the Lepage test.

As final remark, it is possible to expect a future valuable analytical perspective to be developed from this permutation solution we proposed in this work. In fact, since permutation tests are particularly suitable to derive multivariate tests via nonparametric combination, it would be possible to develop a solution for the multivariate two sample location-scale problem. Such solution might be compared with recent ones proposed by the literature (Zu and He, 2006; Rousson, 2002).

References

Bürning H., Thandewald T. (2000), An adaptive two-sample location-scale test of Lepage type for symmetric distributions, *Journal of Statistical Computation and Simulation*, 65, 287-310.

Cucconi O. (1968), Un nuovo test non parametrico per il confronto tra due gruppi campionari, *Giornale degli Economisti*, XXVII, 225-248.

Grambsch, P., O'Brien, P. (1991), The effects of transformations and preliminary tests for non-linearity in regression, *Statistics in Medicine*, *10*, 697-709.

Hollander M., Wolfe D.A. (1999), *Nonparametric Statistical Methods*, 2nd Edition, Wiley, New York.

Kössler W. (2006), Asymptotic power and efficiency of Lepage-type test for the treatment of combined location-scale alternative, *Working paper of Humboldt University in Berlin*, http://edoc.hu-berlin.de/series/informatik-berichte/200/PDF/200.pdf .

Lepage Y. (1971), A combination of Wilcoxon's and Ansary-Bradley's statistics, *Biometrika*, 58 , 213-217.

Pesarin F. (2001), *Multivariate Permutation Tests: With Applications in Biostatistics*, John Wiley & Sons, Ltd., Chichester.

Podgor M.J., Gastwirth J.L. (1994), On non-parametric and generalized tests for the two-sample problem with location and scale change alternatives, *Statistics in Medicine*, 13, 747-758.

Rousson V. (2002), On distribution-free tests for the multivariate two-sample location-scale model, *Journal of Multivariate Analysis*, 80, 1, 43-57.

Zu Y.J., He X.M. (2006), On the limiting distributions of multivariate depth-based rank sum statistics and related tests, *Annals of Statistics*, 34, 6, 2879-2896.