The Lepage location-scale test revisited

Article *in* Far East Journal of Theoretical Statistics · January 2008

1 author:

Marco Marozzi
Università Ca' Foscari Venezia

Far East Journal of Theoretical Statistics
Volume 24, Number 2, 2008, Pages 137-155
Published online: April 2, 2008
This paper is available online at http://www.pphmj.com

THE LEPAGE LOCATION-SCALE TEST REVISITED

MARCO MAROZZI

Dipartimento di Economia e Statistica
Università della Calabria
via P. Bucci Cubo 0c
87036 Arcavacata di Rende (CS), Italy
e-mail: marozzi@unical.it

Abstract

Generally, to make inferences about possible difference between two populations, a test for location is considered. Sometimes, there is more interest in scale differences rather than in location ones (e.g., industrial quality control). This paper is instead focused on joint nonparametric testing for location and scale. A test by Lepage [4] is considered. This is a rank test based on a combination of the Wilcoxon-Mann-Whitney test for location and the Ansari-Bradley test for scale. It is shown that the Lepage idea may be developed and extended within the nonparametric combination framework, for example by considering different combining functions or different test statistics. Moreover, it is easy to adopt a weighting testing strategy. The classical version of the Lepage test has been compared via simulation with various permutation versions of it, partly developed within the nonparametric combination framework. The rank test of Cucconi [1] for location-scale problem has been considered as well. It is shown that the tests maintain the type-one error rate close to the nominal level. The classical Lepage test works well and like the nonparametric combination test based on the Fisher combining function which is more flexible. A multivariate version of the classical Lepage test is presented.

2000 Mathematics Subject Classification: 62G10, 62G35, 62G09.

Keywords and phrases: the location-scale problem, nonparametric methods, rank testing, permutation testing, nonparametric combination of tests.

Received October 12, 2007

1. Introduction

Very often, when one would like to make inferences about possible difference between two populations, considers a test for location. Sometimes, the interest is on scale differences rather than on location ones (e.g., industrial quality control). Problems in which a treatment not only affects the location or scale of a distribution but also can change both simultaneously are common in toxicological, medical and epidemiological studies. The paper is focused on location-scale differences and in particular on a test proposed by Lepage [4]. We discuss this test within the nonparametric combination framework (Pesarin [12]), generalize Lepage idea for location-scale testing and propose different versions of the original Lepage's test. A simulation study has been performed to compare the tests. The Cucconi [1] rank test has been considered as well. In this paper we do not consider location-scale testing based on adaptive designs; for this approach see for example Wassmer [13] and Neuhäuser [11].

Let observations $X_{11}, \ldots, X_{1n_1}$ and $X_{21}, \ldots, X_{2n_2}$ be random samples from continuous populations X_1 and X_2, and F_1 and F_2 be their respective cumulative distribution functions. The general null hypothesis of interest in comparing two populations is that X_1 and X_2 variables have the same distribution, formally

$$H_0 : F_1(t) = F_2(t), \quad \text{for every } -\infty < t < +\infty.$$

In the paper, the location-scale model is considered. This model corresponds to take $F_1(t) = G\left(\frac{t-\delta_1}{\sigma_1}\right)$ and $F_2(t) = G\left(\frac{t-\delta_2}{\sigma_2}\right)$, where $G(.)$ is the distribution function for a continuous variable with location 0 and scales 1, δ_1 and δ_2 $(\sigma_1$ and $\sigma_2)$ are the locations (scales) of populations 1 and 2, respectively. According to this model, we formalize the system of hypotheses as

$$H_0 : (\delta_1 = \delta_2) \cap (\sigma_1 = \sigma_2) \quad \text{versus} \quad H_1 : (\delta_1 \neq \delta_2) \cup (\sigma_1 \neq \sigma_2).$$

In Section 2, the Lepage idea for jointly testing location and scale is reviewed. In Section 3, the test is extended in various directions within the nonparametric combination framework. In Section 4, the classical Lepage test is compared via simulation with different permutation tests based on it and developed in Section 3, and with the Cucconi test which is based on a quite different rank solution. The test is extended for the multivariate problem in Section 5. Section 6 concludes with discussion. Being not much familiar in the literature, the Cucconi test is briefly reviewed in Appendix.

2. The Lepage Test

Lepage [4] proposed a test statistic for the two sample location-scale problem which is a combination of the Wilcoxon-Mann-Whitney and Ansari-Bradley statistics. The Lepage statistic L for the two sample location-scale problem is

$$L = \frac{(W - E(W))^2}{\mathrm{VAR}(W)} + \frac{(A - E(A))^2}{\mathrm{VAR}(A)},$$

where W is the Wilcoxon-Mann-Whitney statistic and A is the Ansari-Bradley one. To compute W and A statistics, the combined sample $Y = (X_{11}, \ldots, X_{1n_1}, X_{21}, \ldots, X_{2n_2}) = (X_1, \ldots, X_{n_1}, X_{n_1+1}, \ldots, X_n)$ elements should be ordered from least to greatest. Let W_{ji} denote the rank of X_{ji} in the combined sample. Then $W = \sum_{i=1}^{n_2} W_{2i}$. To compute the A statistic assign the score 1 to both the smallest and the largest observations in Y, the score 2 to the second smallest and second largest, and so on. Let A_{ji} denote the score of X_{ji} in the combined sample. Then $A = \sum_{i=1}^{n_2} A_{2i}$. $E(.)$ and $\mathrm{VAR}(.)$ denote the expected value and variance of W and A under H_0, see, e.g., Hollander and Wolfe [3] for the corresponding formulae.

For testing at the α level of significance, reject H_0 if $L \geq l_\alpha$, where the constant l_α is chosen so that the type-one error rate is α. Tables for the Lepage test can be found in Lepage [5]. Lepage [4] showed that since

W and A statistics are not correlated under H_0, the L statistic has a limiting chi-squared distribution with 2 degrees of freedom. In the presence of ties among sample observations, the L statistic should be computed using the tie-modified versions of W and A statistics (see Hollander and Wolfe [3] for details). Lepage [7] and Duran et al. [2] studied the asymptotic relative efficiency of the test.

The idea of the Lepage test is that of tackling the location-scale problem through a combination of a test for location and a test for scale. Since that the W statistic gives information on location shifts and A on scale shifts, W and A jointly give information on H_0. However, it should be emphasized that the W test is biased for testing location shifts when $\sigma_1 \neq \sigma_2$ and so is A for testing scale shifts when $\delta_1 \neq \delta_2$. Lepage uses the standardized version $\widetilde{W} = \frac{W - E(W)}{\sqrt{\mathrm{VAR}(W)}}$ and $\widetilde{A} = \frac{A - E(A)}{\sqrt{\mathrm{VAR}(A)}}$ of W and A, and, since the interest is on the general alternative hypothesis $H_1 = \overline{H}_0$, he adopts the square transformation. To combine $\widetilde{W}$ and $\widetilde{A}$ he simply uses the sum. We emphasize this idea writing

$$L = {}_1L = {}_1\psi(\widetilde{W}, \widetilde{A}) = \widetilde{W}^2 + \widetilde{A}^2.$$

Lepage [4] mentioned two other ways for combining $\widetilde{W}$ and $\widetilde{A}$

$${}_2L = {}_2\psi(\widetilde{W}, \widetilde{A}) = \max(|\widetilde{W}|, |\widetilde{A}|)$$

and

$${}_3L = {}_3\psi(\widetilde{W}, \widetilde{A}) = |\widetilde{W}| + |\widetilde{A}|.$$

Let us now consider the permutation version of ${}_1L$, ${}_2L$ and ${}_3L$ tests. Let Y^* denote a random permutation of the combined sample $Y^* = (X_{u_1^*}, \ldots, X_{u_n^*}) = (X_1^*, \ldots, X_n^*)$, where $(u_1^*, \ldots, u_n^*)$ is a permutation of $(1, \ldots, n)$. The permutation versions of ${}_1L$, ${}_2L$ and ${}_3L$ statistics are, respectively,

$${}_1L^* = (\widetilde{W}^*)^2 + (\widetilde{A}^*)^2,\ {}_2L^* = \max(|\widetilde{W}^*|, |\widetilde{A}^*|) \text{ and } {}_3L^* = |\widetilde{W}^*| + |\widetilde{A}^*|,$$

where

$$\widetilde{W}^* = \frac{W^* - E(W)}{\sqrt{\mathrm{VAR}(W)}}, \quad \widetilde{A}^* = \frac{A^* - E(A)}{\sqrt{\mathrm{VAR}(A)}},$$

$$W^* = \sum\nolimits_{i=n_1+1}^{n} W_i^*, \quad A^* = \sum\nolimits_{i=n_1+1}^{n} A_i^*,$$

W_i^* and A_i^* are, respectively, the rank and the Ansari-Bradley score of X_i^*. The observed values of ${}_1L^*$, ${}_2L^*$ and ${}_3L^*$ are, respectively, ${}_1L_0 = \widetilde{W}^2 + \widetilde{A}^2 = L$, ${}_2L_0 = \max(|\widetilde{W}|, |\widetilde{A}|)$ and ${}_3L_0 = |\widetilde{W}| + |\widetilde{A}|$. The p-value of the test is estimated by

$$\hat{M}_{{}_iL^*} = \frac{1}{B}\sum_{b=1}^{B} I({}_iL_b^* \geq {}_iL_0),$$

where ${}_iL_b^*$ denotes the value of ${}_iL^*$, $i = 1, 2, 3$ in the bth $(b = 1, \ldots, B)$ permutation of Y and $I(.)$ denotes the indicator function. Efficient algorithms to compute exact permutation p-values for some univariate problems are available (see, e.g., Mehta et al. [10]). Such algorithms compute exactly the permutation test statistic distribution in the tails, but to perform a combined test (as we are going to do in Sections 3 and 5) the whole distribution should be studied. Since algorithms for exact combined test p-value computing are not available in the literature, we estimate them. $\hat{M}_{{}_iL^*}$ is an unbiased and (for the Glivenko-Cantelli theorem) strong-consistent estimator of the p-value of the test. If $\hat{M}_{{}_iL^*} \leq \alpha$, then H_0 is rejected, otherwise is accepted.

3. Extension of the Lepage Idea to Location-scale Testing

The Lepage idea to location-scale testing may be developed and extended in various directions within the nonparametric combination framework (Pesarin [12]). The framework is based on a natural idea, that of breaking down a (complex) problem into a set of easier to solve sub-problems, each of which is related to a particular aspect of the original problem. To address a complex testing problem, it is suggested to emphasize $K \geq 2$ partial aspects so that the null hypothesis may be broken down into K sub-hypotheses as $H_0 = \bigcap_{k=1}^{K} H_{0k}$ (H_0 is true if all H_{0k} are jointly true). The general alternative is

$$H_1 = \overline{H}_0 = \overline{\bigcap_{k=1}^{K} H_{0k}} = \bigcup_{k=1}^{K} \overline{H}_{0k} = \bigcup_{k=1}^{K} H_{1k}.$$

The procedure for testing H_0 versus H_1 consists of two steps. First, each of the K partial null hypotheses is tested. Then, the results of the first step are jointly managed to solve the general problem. The practitioner has only to take care of partial problems (which are easier than the general one) through proper tests. In fact, a test for the general problem is developed via the nonparametric combination of partial tests.

Regarding location-scale testing, it is clear that two partial aspects may be emphasized ($K = 2$), the location and scale aspect. The corresponding decomposition of $H_0 : (\delta_1 = \delta_2) \cap (\sigma_1 = \sigma_2)$ is $H_0 = H_{0l} \cap H_{0s}$, where $H_{0l} : \delta_1 = \delta_2$ and $H_{0s} : \sigma_1 = \sigma_2$. $H_1 = \overline{H}_0$ is represented as $H_1 = H_{1l} \cup H_{1s}$, where $H_{1l} : \delta_1 \neq \delta_2$ and $H_{1s} : \sigma_1 \neq \sigma_2$. In the first step, H_{0l} and H_{0s} are tested by means of $T_l^* = |\widetilde{W}^*|$ and $T_s^* = |\widetilde{A}^*|$, respectively. The corresponding p-values are estimated as

$$\hat{M}_{T_l^*}(T_{l0}) = \frac{0.5 + \sum_{b=1}^{B} I(T_{lb}^* \geq T_{l0})}{B+1}$$

and

$$\hat{M}_{T_s^*}(T_{s0}) = \frac{0.5 + \sum_{b=1}^{B} I(T_{sb}^* \geq T_{s0})}{B+1},$$

where $T_{l0} = |\widetilde{W}|$, $T_{s0} = |\widetilde{A}|$, $T_{lb}^* = |\widetilde{W}_b|$ and $T_{sb}^* = |\widetilde{A}_b|$. With respect to standard permutation p-value estimation, 0.5 and 1 are added to the numerator and denominator of the fraction, respectively. The reason is to obtain estimated p-values in the open interval $]0, 1[$ avoiding computational problems which may arise in the second step of the procedure. However, since large B is used, this correction is practically irrelevant.

In the second step, the test statistic for the location-scale problem is obtained by nonparametric combination of the p-values associated with T_l^* and T_s^* partial tests. We combine partial p-values rather than partial test statistics without loss of generality because partial p-values are permutationally equivalent to partial test statistics. Let φ be a proper combining function, the test statistic for jointly testing location and scale parameters is

$$_{\varphi}T_{ls}^* = \varphi(M_{T_l^*}, M_{T_s^*}).$$

The observed value of $_{\varphi}T_{ls}^*$ is estimated as

$$_{\varphi}\hat{T}_{ls0} = \varphi(\hat{M}_{T_l^*}(T_{l0}), \hat{M}_{T_s^*}(T_{s0}))$$

and its distribution simulated by means of the same B permutations of the first step. For example, the bth permutation value of $_{\varphi}T_{ls}^*$ is computed as $_{\varphi}\hat{T}_{lsb} = \varphi(\hat{M}_{T_l^*}(T_{lb}), \hat{M}_{T_s^*}(T_{sb}))$. Large values of the observed test statistic are evidence against H_0. The p-value of the test is estimated by

$$\hat{M}_{_{\varphi}T_{ls}^*} = \frac{1}{B}\sum_{b=1}^{B} I(_{\varphi}\hat{T}_{lsb}^* \geq {_{\varphi}\hat{T}_{ls0}})$$

and H_0 is rejected if $\hat{M}_{_{\varphi}T_{ls}^*} \leq \alpha$. The most often used combining functions are

(i) the Fisher combining function $\varphi_F = \ln\left(\frac{1}{M_{T_l^*}}\right) + \ln\left(\frac{1}{M_{T_s^*}}\right)$;

(ii) the normal (Liptak) combining function

$$\varphi_N = \Phi^{-1}(1 - M_{T_l^*}) + \Phi^{-1}(1 - M_{T_s^*}),$$

where $\Phi^{-1}(.)$ denotes the quantile function of the standard normal distribution function;

(iii) the Tippett combining function $\varphi_T = \max(1 - M_{T_l^*}, 1 - M_{T_s^*})$.

It is very interesting to note that the Lepage statistic can be seen as an example of a direct combination of the square of the standardized Wilcoxon-Mann-Whitney and Ansari-Bradley statistics ${}_1L = \widetilde{W}^2 + \widetilde{A}^2$. Direct combination functions can be used when all partial test statistics are homogeneous, in the sense that they have the same limiting distribution. They are based on the sum function which acts on test statistics rather than p-values. ${}_1L$ is a proper combined statistic because the combined elements have the same chi-squared with one degree of freedom limiting distribution. The other two statistics suggested by Lepage [4] can be seen as particular examples of the nonparametric combination procedure as well. ${}_3L = |\widetilde{W}| + |\widetilde{A}|$ is similar to ${}_1L : |.|$ function is used instead of $(.)^2$ function in order to assess the global alternative hypothesis. ${}_2L = \max(|\widetilde{W}|, |\widetilde{A}|)$ is analogous to the Tippett statistic (the first one follows the max statistic rule, the other one the min p-value rule). Lepage prefers ${}_1L$ because its limiting distribution is the well-known chi-squared distribution.

The nonparametric combination framework for location-scale testing allows to generalize the Lepage idea. The generalization is expressed in various ways. Other combining functions can be used and not only those based on direct combinations or on the maximum statistic rule. Pesarin [12] states that this possibility may be very useful, since 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 H_{1l} and H_{1s} is true; that of Liptak is generally good when H_{1l} and H_{1s} are jointly true. The Fisher combining function is suggested when nothing is expected about H_{1l} and H_{1s} because it has an intermediate behavior. It is worth noting that it is easy to develop a weighted testing procedure. In certain contexts, for example in industrial quality control, the scale aspect may be more important than the location one. Therefore, it is very useful to assign different weighting factors to the two aspects: $w_l, w_s > 0$ with $w_l > w_s$. For example, the weighted location-scale test statistic is $\varphi_W = w_l \ln(1/M_{T_l^*}) + w_s \ln(1/M_{T_s^*})$ using the Fisher combining function. Note that it is difficult to perform a weighted testing procedure within a different approach than the permutation/rank one. It should be also emphasized that one can easily take advantage of the flexibility of the nonparametric combination framework by assessing the location and scale problems using different statistics than W and A. For example, the location aspect may be addressed using the Marozzi [9] test, which does not use ranks and has a good power behavior under heavy-tailed and/or highly-skewed distributions.

4. A Comparative Simulation Study

To evaluate how the classical version of the Lepage test performs in detecting location-scale shifts, it has been compared via simulation with various permutation versions of it (partly developed within the nonparametric combination framework) ${}_1L^*$, ${}_2L^*$, ${}_3L^*$, ${}_{\varphi_F}T_{ls}^*$, ${}_{\varphi_N}T_{ls}^*$ and ${}_{\varphi_T}T_{ls}^*$. The rank test C of Cucconi [1] has been considered as well. Since it is not much familiar in the literature, is briefly reviewed in Appendix.

Six distributions have been considered: a standard normal distribution; a chi-squared with eight degrees of freedom; a double-exponential; an exponential; a standard Cauchy; a standard half-Cauchy distribution. Three sample-size settings (n_1, n_2) are considered (10, 10), (10, 30) and (30, 30). We consider three different alternative hypotheses (i) $H_1 : (\delta_1 - \delta_2 > 0) \cap (\sigma_1/\sigma_2 = 1)$; (ii) $H_1 : (\delta_1 - \delta_2 = 0) \cap (\sigma_1/\sigma_2 > 1)$ and (iii) $H_1 : (\delta_1 - \delta_2 > 0) \cap (\sigma_1/\sigma_2 > 1)$. Shifts values are specified so that the power of the original Lepage test is about 50% at $\alpha = 5\%$. For each simulation setting, we considered 10,000 Monte Carlo simulations and 2,000 random permutations of Y.

Table 1 reports type-one error rate estimates. As it can be seen, the type-one error rate of the tests is close to the nominal level. Moreover, the tests tend to be conservative, in particular under normal, chi-squared and double exponential distributions. The Cucconi test is the most conservative one.

Table 1. Size estimates (percent), $\alpha = 5\%$

n_1	10	10	30	n_1	10	10	30
n_2	10	30	30	n_2	10	30	30
test				test			
	Normal				Chi-squared		
L	5.08	4.80	4.55	L	5.33	4.79	4.73
${}_1L^*$	4.95	4.76	4.81	${}_1L^*$	5.21	4.80	5.02
${}_2L^*$	4.55	4.76	4.84	${}_2L^*$	4.62	4.50	4.78
${}_3L^*$	4.90	5.01	4.58	${}_3L^*$	5.21	5.02	5.00
$\varphi_F\, T_{ls}^*$	4.93	4.85	4.84	$\varphi_F\, T_{ls}^*$	5.22	4.81	5.09
$\varphi_T\, T_{ls}^*$	4.61	4.84	4.92	$\varphi_T\, T_{ls}^*$	4.68	4.52	4.80
$\varphi_N\, T_{ls}^*$	4.98	5.08	4.60	$\varphi_N\, T_{ls}^*$	5.40	5.10	5.14
C	4.04	4.36	4.53	C	4.26	4.19	4.66
	Double-exponential				Exponential		
L	5.43	4.92	4.72	L	5.04	4.81	4.98
${}_1L^*$	5.34	4.99	4.96	${}_1L^*$	4.95	4.80	5.21
${}_2L^*$	4.69	4.81	5.12	${}_2L^*$	4.79	4.42	5.34
${}_3L^*$	5.18	4.90	5.03	${}_3L^*$	4.88	4.87	5.32
$\varphi_F\, T_{ls}^*$	5.35	4.95	4.95	$\varphi_F\, T_{ls}^*$	5.01	4.89	5.25
$\varphi_T\, T_{ls}^*$	4.79	4.78	5.14	$\varphi_T\, T_{ls}^*$	4.79	4.42	5.50
$\varphi_N\, T_{ls}^*$	5.18	4.92	4.90	$\varphi_N\, T_{ls}^*$	4.93	5.02	5.45
C	4.11	4.30	4.75	C	4.19	4.22	5.06
	Cauchy				Half-Cauchy		
L	5.13	5.12	4.80	L	5.03	5.08	4.89
${}_1L^*$	4.93	5.11	4.95	${}_1L^*$	4.91	4.91	5.16
${}_2L^*$	4.81	4.94	5.03	${}_2L^*$	4.62	4.76	5.08
${}_3L^*$	5.23	5.01	5.04	${}_3L^*$	4.88	4.94	5.11
$\varphi_F\, T_{ls}^*$	5.15	5.23	5.06	$\varphi_F\, T_{ls}^*$	5.00	4.95	5.30
$\varphi_T\, T_{ls}^*$	4.68	4.87	4.96	$\varphi_T\, T_{ls}^*$	4.56	4.74	5.18
$\varphi_N\, T_{ls}^*$	5.18	5.15	5.13	$\varphi_N\, T_{ls}^*$	4.99	4.97	5.09
C	4.03	4.49	4.94	C	4.16	4.45	4.86

Table 2 reports power estimates when $\delta_1 - \delta_2 > 0$ and $\sigma_1/\sigma_2 = 1$. It is shown (without surprise) that the L test performs very similarly to its permutation version ${}_1L^*$. Among the straight permutation versions of the Lepage test $({}_1L^*,\ {}_2L^*,\ {}_3L^*)$, the test based on the sum of absolute value of $\widetilde{W}$ and $\widetilde{A}$ statistics is the worst one. Among tests obtained through nonparametric combination of $\widetilde{W}$ and $\widetilde{A}$ tests, the best one is the test based on the Fisher combining function, while the worst is the one based on the Liptak function. The latter is the worst test. It should be noted that the C test has intermediate performance.

Table 2. Power estimates (percent) when $\delta_1 - \delta_2 > 0$ and $\sigma_1/\sigma_2 = 1$, $\alpha = 5\%$

n_1	10	10	30	n_1	10	10	30
n_2	10	30	30	n_2	10	30	30
test				test			
	Normal				Chi-squared		
L	48.6	53.6	49.0	L	47.8	48.6	47.2
${}_1L^*$	48.2	53.4	49.8	${}_1L^*$	47.1	48.5	47.8
${}_2L^*$	48.4	53.5	50.3	${}_2L^*$	44.3	50.0	45.3
${}_3L^*$	37.8	46.7	43.4	${}_3L^*$	40.4	37.3	44.9
$\varphi_F\, T_{ls}^*$	48.1	53.0	49.1	$\varphi_F\, T_{ls}^*$	47.6	47.2	48.0
$\varphi_T\, T_{ls}^*$	50.1	54.1	50.5	$\varphi_T\, T_{ls}^*$	45.8	50.7	45.7
$\varphi_N\, T_{ls}^*$	31.8	44.0	38.0	$\varphi_N\, T_{ls}^*$	35.3	33.0	41.9
C	43.9	51.7	48.7	C	43.1	44.5	48.1
	Double-exponential				Exponential		
L	46.7	50.2	51.4	L	47.8	49.3	53.1
${}_1L^*$	46.2	49.8	51.9	${}_1L^*$	47.1	48.8	53.8
${}_2L^*$	44.3	47.9	51.8	${}_2L^*$	38.2	42.2	46.7
${}_3L^*$	39.7	45.6	46.3	${}_3L^*$	47.4	44.0	54.5
$\varphi_F\, T_{ls}^*$	46.3	49.8	51.3	$\varphi_F\, T_{ls}^*$	48.5	49.6	54.7
$\varphi_T\, T_{ls}^*$	45.6	48.5	52.0	$\varphi_T\, T_{ls}^*$	38.8	42.5	46.7
$\varphi_N\, T_{ls}^*$	34.3	43.2	41.1	$\varphi_N\, T_{ls}^*$	45.1	40.9	53.5
C	43.7	47.4	50.9	C	42.5	39.4	58.1
	Cauchy				Half-Cauchy		
L	51.3	50.1	49.6	L	50.6	52.0	52.7
${}_1L^*$	50.8	49.9	50.1	${}_1L^*$	49.6	51.6	53.4
${}_2L^*$	47.7	46.5	49.4	${}_2L^*$	39.5	40.4	46.8
${}_3L^*$	46.8	47.8	45.9	${}_3L^*$	50.9	51.5	54.0
$\varphi_F\, T_{ls}^*$	50.5	50.0	49.8	$\varphi_F\, T_{ls}^*$	51.2	52.9	54.1
$\varphi_T\, T_{ls}^*$	48.7	47.0	49.6	$\varphi_T\, T_{ls}^*$	39.2	40.4	46.9
$\varphi_N\, T_{ls}^*$	41.4	45.3	41.6	$\varphi_N\, T_{ls}^*$	48.9	49.1	53.4
C	49.0	46.7	49.2	C	45.4	40.8	58.3

As it is shown in Table 3, similar claims apply when $\delta_1 - \delta_2 = 0$ and $\sigma_1/\sigma_1 > 1$: L and ${}_1L^*$ perform similarly, ${}_3L^*$ is the worst one among straight permutation versions of the L test idea and ${}_{\varphi_N}T^*_{ls}$ is the worst test.

Table 3. Power estimates (percent) when $\delta_1 - \delta_2 = 0$ and $\sigma_1/\sigma_2 > 1$, $\alpha = 5\%$

n_1	10	10	30	n_1	10	10	30
n_2	10	30	30	n_2	10	30	30
test				test			
	Normal				Chi-squared		
L	51.5	49.8	50.3	L	52.9	52.3	48.5
${}_1L^*$	50.9	49.5	51.0	${}_1L^*$	52.4	52.3	49.1
${}_2L^*$	48.3	46.8	50.9	${}_2L^*$	51.5	49.7	49.0
${}_3L^*$	46.2	47.3	44.6	${}_3L^*$	42.0	49.0	43.6
${}_{\varphi_F}T^*_{ls}$	50.1	49.5	50.1	${}_{\varphi_F}T^*_{ls}$	52.4	52.3	48.4
${}_{\varphi_T}T^*_{ls}$	46.5	46.4	50.7	${}_{\varphi_T}T^*_{ls}$	53.2	50.3	49.3
${}_{\varphi_N}T^*_{ls}$	40.7	44.1	39.8	${}_{\varphi_N}T^*_{ls}$	34.6	47.6	38.5
C	53.6	58.7	59.5	C	48.4	51.4	48.5
	Double-exponential				Exponential		
L	52.2	50.2	51.1	L	50.2	49.8	52.9
${}_1L^*$	51.3	50.1	51.8	${}_1L^*$	49.4	49.5	53.4
${}_2L^*$	48.8	48.3	52.5	${}_2L^*$	46.7	44.7	51.0
${}_3L^*$	46.5	46.7	45.8	${}_3L^*$	43.2	48.9	49.6
${}_{\varphi_F}T^*_{ls}$	50.7	49.6	51.1	${}_{\varphi_F}T^*_{ls}$	50.0	50.1	53.5
${}_{\varphi_T}T^*_{ls}$	46.9	47.9	52.3	${}_{\varphi_T}T^*_{ls}$	48.0	45.1	51.2
${}_{\varphi_N}T^*_{ls}$	40.8	43.3	40.1	${}_{\varphi_N}T^*_{ls}$	37.5	48.4	46.3
C	52.8	57.2	57.6	C	45.4	50.0	53.6
	Cauchy				Half-Cauchy		
L	50.2	49.4	50.3	L	49.9	49.5	50.7
${}_1L^*$	49.7	49.1	50.7	${}_1L^*$	49.4	49.3	51.1
${}_2L^*$	47.5	47.4	50.9	${}_2L^*$	48.4	48.5	51.5
${}_3L^*$	44.9	45.0	46.3	${}_3L^*$	41.5	44.1	44.8
${}_{\varphi_F}T^*_{ls}$	48.6	48.6	50.2	${}_{\varphi_F}T^*_{ls}$	49.3	49.2	50.4
${}_{\varphi_T}T^*_{ls}$	45.7	46.9	50.8	${}_{\varphi_T}T^*_{ls}$	49.7	49.1	51.7
${}_{\varphi_N}T^*_{ls}$	39.5	41.6	41.2	${}_{\varphi_N}T^*_{ls}$	35.0	41.4	39.0
C	46.0	50.7	48.7	C	46.3	47.4	50.6

Table 4 shows that when $\delta_1 - \delta_2 > 0$ and $\sigma_1/\sigma_2 > 1$, ${}_3L^*$ is no more the worst test among ${}_1L^*$, ${}_2L^*$ and ${}_3L^*$, and that, as expected (note that the Tippett function is generally suggested when only one among partial null hypotheses is expected to be true), ${}_{\varphi_N}T^*_{ls}$ is no more the worst one among ${}_{\varphi}T^*_{ls}$ type tests. It is worth noting that the classical Lepage test works well in this case as in the other ones, and so does its permutation version ${}_1L^*$.

Table 4. Power estimates (percent) when $\delta_1 - \delta_2 > 0$ and $\sigma_1/\sigma_2 = 1$, $\alpha = 5\%$

n_1	10	10	30	n_1	10	10	30
n_2	10	30	30	n_2	10	30	30
test				test			
	Normal				Chi-squared		
L	50.2	52.3	49.0	L	52.3	49.9	47.2
${}_1L^*$	49.5	52.0	49.6	${}_1L^*$	51.7	49.6	48.0
${}_2L^*$	44.7	46.4	46.4	${}_2L^*$	52.0	51.9	48.6
${}_3L^*$	44.6	52.2	46.8	${}_3L^*$	40.5	40.0	41.3
${}_{\varphi_F}T^*_{ls}$	50.2	53.0	49.7	${}_{\varphi_F}T^*_{ls}$	51.5	48.4	47.2
${}_{\varphi_T}T^*_{ls}$	46.2	46.9	46.7	${}_{\varphi_T}T^*_{ls}$	53.5	52.5	49.2
${}_{\varphi_N}T^*_{ls}$	39.5	51.5	44.2	${}_{\varphi_N}T^*_{ls}$	32.7	36.8	36.1
C	46.7	53.4	50.1	C	47.2	46.6	46.8
	Double-exponential				Exponential		
L	49.9	51.5	51.2	L	48.6	49.5	51.3
${}_1L^*$	49.2	51.1	51.8	${}_1L^*$	48.1	49.3	52.0
${}_2L^*$	41.2	45.2	47.2	${}_2L^*$	43.4	48.3	48.6
${}_3L^*$	48.5	51.4	50.5	${}_3L^*$	43.7	39.0	48.7
${}_{\varphi_F}T^*_{ls}$	49.9	52.0	52.2	${}_{\varphi_F}T^*_{ls}$	49.1	48.5	51.9
${}_{\varphi_T}T^*_{ls}$	41.8	45.4	47.2	${}_{\varphi_T}T^*_{ls}$	45.0	49.0	48.7
${}_{\varphi_N}T^*_{ls}$	45.6	50.6	48.5	${}_{\varphi_N}T^*_{ls}$	39.7	34.6	45.9
C	47.2	50.4	52.5	C	43.0	42.2	53.2
	Cauchy				Half-Cauchy		
L	47.3	48.3	50.0	L	49.7	49.9	49.9
${}_1L^*$	46.7	48.2	50.6	${}_1L^*$	49.0	49.6	50.7
${}_2L^*$	41.2	41.8	44.8	${}_2L^*$	40.8	44.7	45.2
${}_3L^*$	44.7	48.6	50.5	${}_3L^*$	47.5	42.2	50.2
${}_{\varphi_F}T^*_{ls}$	46.9	48.7	51.3	${}_{\varphi_F}T^*_{ls}$	50.1	49.9	51.6
${}_{\varphi_T}T^*_{ls}$	41.7	41.9	44.9	${}_{\varphi_T}T^*_{ls}$	41.6	45.3	45.3
${}_{\varphi_N}T^*_{ls}$	42.3	47.6	49.1	${}_{\varphi_N}T^*_{ls}$	44.1	38.5	49.0
C	44.8	46.2	48.5	C	44.9	42.9	53.6

On the whole, the simulation study shows that the best tests are L, ${}_1L^*$ and ${}_{\varphi_F}T^*_{ls}$. The classical Lepage test L (and its permutation version ${}_1L^*$) works like the nonparametric combination based ${}_{\varphi_F}T^*_{ls}$. It is important to emphasize that using a test like ${}_{\varphi_F}T^*_{ls}$, the practitioner may take advantage not only of the Lepage idea for jointly testing location and scale but also of the flexibility of the nonparametric combination framework. Acting within this framework, as discussed in Section 3, the practitioner, for example, may replace one of the partial tests (or all of them) with a test which performs better in certain situations (an example is the Marozzi [9] location test for heavy-tailed and highly-skewed distributions).

5. The Multivariate Lepage Test

As shown in the previous section, the classical Lepage test performs well as compared with various versions of it, e.g., ${}_{\varphi_F}T^*_{ls}$, which is however more flexible. Therefore it is of some interest to propose a multivariate version of the classical Lepage test. Let populations 1 and 2 be P-variate with $P \geq 2$. The multivariate dataset can be represented as

$$\underline{\underline{X}} = \begin{bmatrix} {}^1\underline{X}_1 & {}^1\underline{X}_2 \\ \vdots & \vdots \\ {}^p\underline{X}_1 & {}^p\underline{X}_2 \\ \vdots & \vdots \\ {}^P\underline{X}_1 & {}^P\underline{X}_2 \end{bmatrix}$$

$$= \begin{bmatrix} {}^1X_{11} & \cdots & {}^1X_{1n_1} & {}^1X_{21} & \cdots & {}^1X_{2n_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ {}^pX_{11} & \cdots & {}^pX_{1n_1} & {}^pX_{21} & \cdots & {}^pX_{2n_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ {}^PX_{11} & \cdots & {}^PX_{1n_1} & {}^PX_{21} & \cdots & {}^PX_{2n_2} \end{bmatrix}$$

$$= \begin{bmatrix} {}^{1}X_1 & \cdots & {}^{1}X_{n_1} & {}^{1}X_{n_1+1} & \cdots & {}^{1}X_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ {}^{p}X_1 & \cdots & {}^{p}X_{n_1} & {}^{p}X_{n_1+1} & \cdots & {}^{p}X_n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ {}^{P}X_1 & \cdots & {}^{P}X_{n_1} & {}^{P}X_{n_1+1} & \cdots & {}^{P}X_n \end{bmatrix}$$

$$= \begin{bmatrix} {}^{1}\underline{X} \\ \vdots \\ {}^{p}\underline{X} \\ \vdots \\ {}^{P}\underline{X} \end{bmatrix},$$

where ${}^{p}X_{ij}$ denotes the ith $(i = 1, ..., n_j)$ sample element from the pth $(p = 1, ..., P)$ component of population j $(j = 1, 2)$ and ${}^{p}\underline{X}$ is the pth pooled-sample. We are interested in testing

$$H_0 : \bigcap_{p=1}^{P} [({}^{p}\delta_1 = {}^{p}\delta_2) \cap ({}^{p}\sigma_1 = {}^{p}\sigma_2)]$$

against

$$H_1 : \bigcup_{p=1}^{P} [({}^{p}\delta_1 \neq {}^{p}\delta_2) \cup ({}^{p}\sigma_1 \neq {}^{p}\sigma_2)],$$

where ${}^{p}\delta_j$ $({}^{p}\sigma_j)$ is the location (scale) of the pth component of population j $(j = 1, 2)$. Within the multivariate problem, permutation of individual data vectors should be considered. For each component p we obtain a partial Lepage ${}^{p}L$ test, based on

$${}^{p}L^* = \frac{[{}^{p}W^* - E(W)]^2}{\mathrm{VAR}(W)} + \frac{[{}^{p}A^* - E(A)]^2}{\mathrm{VAR}(A)},$$

where

$$ {}^{p}W^{*} = \sum_{i=n_1+1}^{n} {}^{p}W_{i}^{*}, $$

$$ {}^{p}A^{*} = \sum_{i=n_1+1}^{n} {}^{p}A_{i}^{*}, $$

${}^{p}W_{i}^{*}$ and ${}^{p}A_{i}^{*}$ are, respectively, the rank and the Ansari-Bradley score of ${}^{p}X_{i}^{*}$. Then we combine nonparametric ${}^{1}L, \ldots, {}^{p}L, \ldots, {}^{P}L$ tests. We use the Fisher combining function, which corresponds to take as observed value of the multivariate test statistic

$$ \underline{L}_0 = \sum_{p=1}^{P} \ln(1/M_{{}^{p}L^{*}}({}^{p}L_0)), $$

where ${}^{p}L_0$ is the observed value of the Lepage statistic on ${}^{p}\underline{X}$. The p-value is estimated by

$$ M_{\underline{L}^{*}} = \sum_{b=1}^{B} I(\underline{L}_b^{*} \geq \underline{L}_0)/B, $$

where

$$ \underline{L}_b^{*} = \sum_{p=1}^{P} \ln(1/M_{{}^{p}L^{*}}({}^{p}L_b^{*})) $$

and ${}^{p}L_b^{*}$ is the bth permutation value of the Lepage statistic on ${}^{p}\underline{X}$.

6. Conclusion

Jointly nonparametric testing for location and scale has been discussed. In particular, a rank test proposed by Lepage [4] has been reviewed and re-proposed within the nonparametric combination framework. Special ways of generalization of the Lepage idea has been presented and discussed. They are mainly based on the great flexibility of the nonparametric combination. For example, we discuss the possibility to replace one of the partial tests (or all of them) with a test which performs better in certain situations. To evaluate how the classical version of the Lepage test performs in detecting location-scale shifts, it has been compared via simulation with various permutation versions of it, partly developed within the nonparametric combination framework. The Cucconi test has been considered as well. The simulation study shows that the tests control the type-one error rate (they are often conservative) and that the classical Lepage test L (and its permutation version ${}_1L^*$) works well and like the nonparametric combination based ${}_{\varphi F}T^*_{ls}$. L, ${}_1L^*$ and ${}_{\varphi F}T^*_{ls}$ are the best tests. It is worth noting that the practitioner, by using ${}_{\varphi F}T^*_{ls}$ may take advantage not only of the Lepage idea to jointly testing location and scale but also of the flexibility of the nonparametric combination framework. A multivariate version of the classical Lepage test has been proposed as well.

Appendix

The Cucconi Test. On the same location-scale problem discussed in the paper, Cucconi [1] proposed a quite different rank solution based on

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

where

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

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

and

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

Note that U is based on the squares of the ranks W_{1i}, while V is based on the squares of the counter-ranks $(n + 1 - W_{1i})$ of the first sample. Let U' and V' be U and V computed on the second sample. Then

$$U' = -U \quad \text{and} \quad V' = -V$$

and so it does not matter if one acts on the first or second sample to compute C. Cucconi [1] proved that under

$$H_0 E(U) = E(V) = 0$$

and

$$\mathrm{VAR}(U) = \mathrm{VAR}(V) = 1.$$

Of course U and V are (negative) dependent. More precisely,

$$\begin{aligned}\rho &= \mathrm{CORR}(U, V)\\ &= \mathrm{COVAR}(U, V)\end{aligned}$$

and if $n > 2$, then $-1 < \rho < -7/8$. Under H_0, the asymptotic density function of (U, V) is a bivariate normal

$$f(u, v) = \frac{1}{2\pi\sqrt{1 - \rho_0^2}} \exp\left(-\frac{u^2 + v^2 - 2\rho_0 uv}{2(1 - \rho_0^2)}\right),$$

where

$$\rho_0 = -7/8.$$

When n_1 is not too different to n_2, the normal approximation is already good when $n_1, n_2 > 6$. The null hypothesis should be rejected when $C > -\ln\alpha$.

The Cucconi test is unbiased and consistent. Geometrically, the rejection region of the test is the set of points (u, v) outside the ellipse with equation

$$u^2 + v^2 - 2\rho_0 uv = -2(1 - \rho_0^2)\ln\alpha.$$

Note that when $\delta_1 > \delta_2$ and $\sigma_1 = \sigma_2$, $E(U) > 0$ and $E(V) < 0$; when $\delta_1 = \delta_2$ and $\sigma_1 < \sigma_2$, $E(U) < 0$ and $E(V) < 0$; when $\delta_1 > \delta_2$ and $\sigma_1 < \sigma_2$, $E(U)$ may be closed to 0 but $E(V) < 0$. Therefore, under H_0 (U, V) is centered on (0, 0) while under H_1 is not.

References

[1] O. Cucconi, Un nuovo test non parametrico per il confronto tra due gruppi campionari, Giornale degli Economisti 27 (1968), 225-248.

[2] B. S. Duran, W. S. Tsai and T. O. Lewis, A class of location-scale nonparametric tests, Biometrika 63 (1976), 173-176.

[3] M. Hollander and D. A. Wolfe, Nonparametric Statistical Methods, 2nd ed., Wiley, New York, 1999.

[4] Y. Lepage, A combination of Wilcoxon's and Ansari-Bradley's statistics, Biometrika 58 (1971), 213-217.

[5] Y. Lepage, A table for a combined Wilcoxon Ansari-Bradley statistic, Biometrika 60 (1973), 113-116.

[6] Y. Lepage, Asymptotically optimum rank tests for contiguous location and scale alternatives, Comm. Statist. Theory Methods 4 (1975), 671-687.

[7] Y. Lepage, Asymptotic power efficiency for a location and scale problem, Comm. Statist. Theory Methods 5 (1976), 1257-1274.

[8] Y. Lepage, A class of nonparametric tests for location and scale parameters, Comm. Statist. Theory Methods 6 (1977), 649-659.

[9] M. Marozzi, A bi-aspect nonparametric test for the two-sample location problem, Comput. Statist. Data Anal. 44 (2004), 639-648.

[10] C. R. Mehta, N. R. Patel and P. Senchauduri, Importance sampling for estimating exact probabilities in permutational inference, J. Amer. Statist. Assoc. 83 (1988), 999-1005.

[11] M. Neuhäuser, An adaptive location-scale test, Biom. J. 43 (2001), 809-819.

[12] F. Pesarin, Multivariate Permutation Tests with Applications in Biostatistics, Wiley, Chichester, 2001.

[13] G. Wassmer, Basic concepts of group sequential and adaptive group sequential test procedures, Statist. Papers 41 (2000), 253-279.