# Some notes on the location–scale Cucconi test

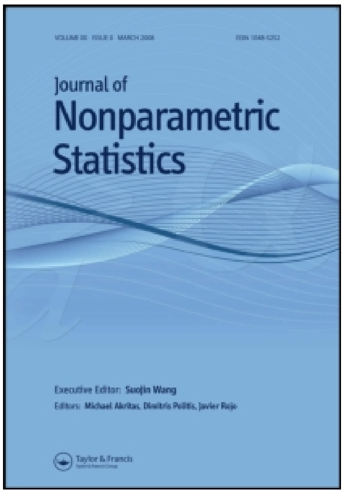
Article in Journal of Nonparametric Statistics · July 2009

DOI: 10.1080/10485250902952435



1 author:

Marco Marozzi
Università Ca' Foscari Venezia




**Journal of Nonparametric Statistics**
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713645758

## Some notes on the location-scale Cucconi test

Marco Marozzi [a]
[a] Dipartimento di Economia e Statistica, Università della Calabria, Italy

Online Publication Date: 01 July 2009

**To cite this Article** Marozzi, Marco(2009)'Some notes on the location-scale Cucconi test',Journal of Nonparametric Statistics,21:5,629 — 647

**To link to this Article: DOI:** 10.1080/10485250902952435
**URL:** http://dx.doi.org/10.1080/10485250902952435

*Journal of Nonparametric Statistics*
Vol. 21, No. 5, July 2009, 629–647

Taylor & Francis
Taylor & Francis Group

# Some notes on the location–scale Cucconi test

Marco Marozzi*

*Dipartimento di Economia e Statistica, Università della Calabria, Via P. Bucci Cubo 0c, 87036 Rende (CS), Italy*



(*Received 12 May 2008; final version received 5 April 2009*)

The best known and most used rank test for the location–scale problem is due to Lepage [Y. Lepage, *A combination of Wilcoxon's and Ansari–Bradley's statistics*, Biometrika 58 (1971), pp. 213–217.], but this paper is focused on the location–scale rank test of Cucconi [O. Cucconi, *Un nuovo test non parametrico per il confronto tra due gruppi campionari*, Giorn. Econom. XXVII (1968), pp. 225–248.], proposed earlier but not nearly as well-known. The test is of interest because, contrary to the other location–scale tests, it is not a quadratic form combining a test for location and a test for scale differences, and it is based on squared ranks and squared contrary-ranks. Moreover, it is easier to compute the test of Cucconi than those of Lepage, Manly–Francis, Büning–Thadewald, Neuhäuser, Büning and Murakami. Exact critical values for the test have been computed for the very first time. The power of the Cucconi test has been studied for the very first time and compared with that of the Lepage and other tests that include several Podgor–Gastwirth efficiency robust tests. Simulations show that the test of Cucconi maintains a size very close to $\alpha$ and is more powerful than the Lepage test, and therefore should be taken into account as a better alternative when it is not possible to develop an efficiency robust procedure for the problem at hand. The simulation study considers also the case of different shapes for the parent distributions, and the case of tied observations which is generally not considered in power studies. The presence of ties does not lower the performance of the Cucconi test, the contrary happens for the Lepage test. The tests are applied to real and fictitious biomedical data.

**Keywords:** the location-scale problem; rank testing; the Cucconi test; the Lepage test

## 1. Introduction

The comparison of two samples is one of the most important problems in statistical testing. If it is assumed that parent population distributions may differ only in location, there are many parametric and nonparametric tests at our disposal. There are many tests also for the scale problem. It is well known that under normal distributions the $t$ test and the $F$ test are the uniformly most powerful unbiased tests for the location and scale problem, respectively, at least for the one-sided alternative; and that the $t$ test is $\alpha$ robust for nonnormal distributions (except for very heavy-tailed ones), whereas the $F$ test is non $\alpha$ robust [1,2]. Even if the usual two-sample problem tests for

*Email: marozzi@unical.it

ISSN 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/10485250902952435
http://www.informaworld.com

a location change, in many biomedical situations the treatment can change location and scale simultaneously [3]. The two-sample location–scale problem arises also in other fields such as climate dynamics [4] and stock prices [5].

Since in practice, normal data as well as either pure location or pure scale alternatives are not common, we need a test that works well for nonnormal data in jointly testing for location and scale changes. Tukey [6] emphasises that a new test to be useful in practice should be quick and compact, in the sense that it should be easily used in practice without much effort by practitioners. If the procedure is quick and compact, then it can afford somewhat reduced power because it can be used so much often as to more than compensate for its possible loss of power. Let us consider the two-stage testing scheme proposed by Manly and Francis [7]. In the first stage, an exact permutation test is carried out to determine whether there is evidence that the distributions underlying the samples to be compared are different. If the test is not significant, the procedure stops and you conclude that there is no evidence of scale or location differences. If the test is significant, then a robust approximate permutation test for detecting location changes is carried out, together with a bootstrap procedure that indicates whether this test is reliable. A robust test for detecting scale changes is also performed. This testing scheme was found to control the type-one error rate even when the samples come from extreme distributions. The authors emphasised that this is achieved at the expense of often concluding that there is evidence for a difference between the populations under consideration without indicating the nature of the difference, and for this reason did not suggest a general use of the procedure. We emphasise another drawback of the procedure: it is not neither quick nor compact in the sense of Tukey, since it is quite difficult to be applied by the practitioners. For the same reason, in this paper we are not interested in location–scale tests based on adaptive designs [8,9], or the tests proposed by Neuhäuser [10], Büning [11] and Murakami [12]. Note that, for what concerns the tests of Neuhäuser and Murakami, as confirmed via mail by both authors, no formulae are available in the literature for the moments of some statistics necessary for applying their tests (therefore, one to apply these tests has to generate the permutation distribution of the statistics by simulation or complete enumeration, and compute the necessary moments).



In this paper, we address the location–scale problem within the rank framework in Section 2. The best known and most used rank test is due to Lepage [13]. We consider the Cucconi [14] test proposed earlier but not nearly well known. 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, the combination is achieved through the sum of the squared standardised test statistics, this is just the case of the Lepage test and the tests cited above. The solution presented by Cucconi [14] is different, as it is not a quadratic form combining a test for location and a test for scale differences, and addresses the location–scale problem by considering the squares of ranks and contrary ranks. It is important to note that to perform the test of Cucconi you need only to compute the ranks of the observations in the pooled sample, whereas to perform the test of Lepage you need also to compute the Ansari–Bradley scores. It is worth noting that a detailed power simulation study of the Cucconi test and a table of critical values are missing in the literature so far. On the contrary, the power of the Lepage test has been widely studied, even recently [8,10–12,15,16]. For this reason, the main aim of this paper is to study in Section 3 power and type-one error rate of the Cucconi rank test, and make comparisons with the much better known Lepage test. Four Podgor and Gastwirth [15] efficiency robust tests have been considered as well, along with the Cramer–Von Mises and Kolmogorov–Smirnov tests, since they are widely used for the general two-sample problem. The simulation study considers also the case of different shapes for the parent distributions and the case of tied observations. A table of critical values for the Cucconi test at $\alpha = 1\%$, 5% and 10% is displayed in Section 3. The tests are applied to real and fictitious biomedical data in Section 4. Section 5 concludes the paper with discussion.

## 2. The Cucconi and Lepage tests

Let $X_1$ and $X_2$ be the continuous random variables underlying two populations of interest, $F_1$ and $F_2$ be their respective distribution functions. The general system of hypotheses when one compares two populations is

$$H_0 : X_1 \stackrel{d}{=} X_2 \text{ against } H_1 : X_1 \stackrel{d}{\neq} X_2,$$

where $X_1 \stackrel{d}{=} X_2$ means that $F_1(t) = F_2(t) \forall t \in \Re$ and $X_1 \stackrel{d}{\neq} X_2$ that $\exists A \subset \Re$: $F_1(t) \neq F_2(t)$, $t \in A$ with $\Pr(A) > 0$. In particular, we would like to address the location–scale problem that corresponds to take

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

where $G(.)$ is the distribution function for a continuous variable with location 0 and scale 1, $\mu_1$ and $\mu_2$ ($\sigma_1$ and $\sigma_2$) are the locations (scales) of population 1 and 2, respectively. Let observations $X_{11}, \ldots, X_{1n_1}$ and $X_{21}, \ldots, X_{2n_2}$ be random samples from populations 1 and 2, respectively. For the location–scale problem, Cucconi [14] proposed a rank test based on

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

where

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

$n = n_1 + n_2$, $W_{ji}$ denotes the rank of $X_{ji}$ in the pooled sample $\underline{X} = (X_{11}, \ldots, X_{1n_1}, X_{21}, \ldots, X_{2n_2}) = (X_1, \ldots, X_{n_1}, X_{n_1+1}, \ldots, X_n)$ and $\rho = 2(n^2-4)/((2n+1)(8n+11)) - 1$. Note that $U$ is based on the squares of the ranks $W_{1i}$, while $V$ is based on the squares of the contrary-ranks $(n+1-W_{1i})$ of the first sample. Let $U'$ and $V'$ be $U$ and $V$ computed on the second sample, then $U' = -U$ and $V' = -V$, and so it does not matter if one acts on the first or second sample to compute $C$. Under $H_{0.} E(U) = E(V) = 0$ and $\text{VAR}(U) = \text{VAR}(V) = 1$ because $E\left(\sum_{i=1}^{n_1} W_{1i}^2\right) = n_1(n+1)(2n+1)/6$ and $\text{VAR}\left(\sum_{i=1}^{n_1} W_{1i}^2\right) = n_1 n_2 (n+1)(2n+1)(8n+11)/180$. Of course, it is $E\left(\sum_{i=1}^{n_1} (n+1-W_{1i})^2\right) = E\left(\sum_{i=1}^{n_1} W_{1i}^2\right)$, $\text{VAR}\left(\sum_{i=1}^{n_1} (n+1-W_{1i})^2\right) = \text{VAR}\left(\sum_{i=1}^{n_1} W_{1i}^2\right)$, and $U$ and $V$ are negatively dependent. More precisely, $\text{CORR}(U,V)$ takes values in the interval $[-1, -7/8[$, in fact after some calculations it follows that $\text{CORR}(U, V) = \text{COVAR}(U, V) = 2(n^2-4)/((2n+1)(8n+11)) - 1 = \rho$, the minimum $-1$ occurs when $n = 2$, and the supremum is reached when $n$ tends to infinity

$$\lim_{n\to\infty} \rho = \lim_{n\to\infty} \frac{2n^2-8}{16n^2+30n+11} - 1 = -\frac{7}{8} = \rho_0.$$

Under $H_0$ $(U,V)$ is centred on $(0, 0)$, whereas it is not under $H_1$. When $\mu_1 \neq \mu_2$ and $\sigma_1 = \sigma_2$, $\sum_{i=1}^{n_1} W_{1i}^2$ tends to be greater (less) than $n_1(n+1)(2n+1)/6$ when $\mu_1 > \mu_2$ ($\mu_1 < \mu_2$), and so $U$ tends to be greater (less) than zero whereas $V$ tends to be less (greater) than zero. When $\mu_1 = \mu_2$ and $\sigma_1 > \sigma_2$, the ranks of the elements of the first sample tend to be the extreme elements of the sequence $1, 2, \ldots, n$, whereas when $\sigma_1 < \sigma_2$, they tend to be in the middle. In the first (second) case, $\sum_{i=1}^{n_1} W_{1i}^2$ tends to be greater (less) than $n_1(n+1)(2n+1)/6$ and then $U$ tends to be greater (less) than zero and so does $V$. When $\mu_1 \neq \mu_2$ and $\sigma_1 \neq \sigma_2$, $(U,V)$ is once again non centred on (0, 0), for example when $\mu_1 > \mu_2$ and $\sigma_1 < \sigma_2$, $E(U)$ may be close to zero but $E(V)$ tends to be less than zero. With standard asymptotics it may be proved that if $n_1,n_2 \to \infty$ and $n_1/n \to \lambda \in ]0.1[$ then $\Pr(U \leq u) \to \Phi(u)$, where $\Phi$ is the standard normal distribution function. An analogous results applies for $V$. Moreover if $n_1, n_2 \to \infty$ and $n_1/n \to \lambda \in ]0.1[$ then

$$\Pr(U \leq u, V \leq v) \to \int_{-\infty}^{v}\int_{-\infty}^{u} \frac{1}{2\pi\sqrt{1-\rho_0^2}} \exp\left(-\frac{q^2+r^2-2\rho_0 qr}{2(1-\rho_0^2)}\right) \mathrm{d}q\,\mathrm{d}r.$$

The points $(u,v)$ within acceptance region are close to (0, 0), *i.e.* satisfy

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

where the constant $k$ is chosen so that the type-one error rate is $\alpha$. Let $k = \alpha\left(2\pi\sqrt{1-\rho_0^2}\right)^{-1}$, then it follows that $H_0$ should be accepted if the point $(u,v)$ is such that $c = (u^2+v^2-2\rho_0 uv)/(2(1-\rho_0^2)) < -\ln\alpha$. It should be noted that the acceptance region $E$ of the test is the set of points $(u,v)$ inside the ellipse

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

Since it is

$$\iint_E \frac{1}{2\pi\sqrt{1-\rho_0^2}} \exp\left(-\frac{q^2+r^2-2\rho_0 qr}{2\left(1-\rho_0^2\right)}\right) \mathrm{d}q\,\mathrm{d}r = 1-\alpha,$$

the size of test is just $\alpha$. Cucconi [14] without giving any formal justification observed that the rate of convergence to the normal was very good as soon as $n_1,n_2 > 6$ with $n_1 \cong n_2$. However, this is not of much interest for us, since we computed exact critical values of the test for many sample size configurations. For the proofs of the claims about the Cucconi test (including its unbiasedness and consistency) [14].

For the same problem, Lepage [13] proposed

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

where $W = \sum_{i=1}^{n_2} W_{2i}$ is the Wilcoxon and Mann–Whitney statistic, $A = \sum_{i=1}^{n_2} A_{2i}$ the Ansari–Bradley one, $A_{ji}$ the Ansari–Bradley score of $X_{ji}$ in the pooled sample. To compute the $A_{ij}$s assign the score one to both the smallest and largest observations in the pooled sample, the score two to the second smallest and second largest, and so on. Note that Siegel and Tukey [16] proposed a version of the Ansari–Bradley test that reduced its calculation to the existing tables for the Wilcoxon and Mann–Whitney test. $L$ is obtained as the sum of the squares of the standardized $W$ and $A$ statistics. The idea of Lepage is to address the location–scale problem through a combination of a test for location and a test for scale differences. $E(.)$ and VAR(.) denote the expected value and variance of $W$ and $A$ under $H_0$. For $W$, the corresponding formulae are $E(W) = n_2(n+1)/2$ and $\mathrm{VAR}(W) = n_1 n_2(n+1)/12$. For $A$, since the scoring depends on whether $n$ is even or odd, two cases should be distinguished, $E(A) = n_2(n+2)/4$ and $\mathrm{VAR}(A) = n_1 n_2(n+2)(n-2)/(48(n-1))$ when $n$ is even, $E(A) = n_2(n+1)^2/(4n)$ and $\mathrm{VAR}(A) = n_1 n_2(n+1)(3+n^2)/(48n^2)$ when $n$ is odd.

To test $H_0$ at the $\alpha$ level of significance reject $H_0$ if $L \geq l_\alpha$, where the constant $l_\alpha$ is chosen so that the type-one error rate is $\alpha$. See Lepage [17] for tables of critical values. Note that since $W$ and $A$ statistics are uncorrelated under $H_0$ then the $L$ statistic has a limiting chi-squared distribution with two degrees of freedom. Lepage [18] and Duran, Tsai and Lewis [19] studied the asymptotic relative efficiency of the test. Lepage [18,20] extended his procedure to use the asymptotically most powerful test for the location for a distribution, and the corresponding test for scale. Unfortunately, this approach has a practical drawback, in order to choose the correct asymptotically most powerful test for location and scale differences, one has to know what distribution generated the data. What is preferable is a test that performs well in a wide family of distributions, even if it is not the best one in every single circumstance.

As emphasised in the introduction, the Lepage test is well known and used in practice [21]. Note that it is easier to compute the Cucconi test than the Lepage test since in the first case you need only the $W_{ij}$s, in the second case the $W_{ij}$s as well as the $A_{ij}$s scores. Note that it is not a quadratic form combining a test for location and a test for scale differences. Moreover, a detailed power simulation study of the Cucconi test and a table of critical values are missing in the literature so far. On the contrary, the power of the Lepage test has been widely studied, even recently. Lepage-type tests can be obtained following Podgor and Gastwirth [15], which extend results of O'Brien [22]. O'Brien regresses the group indicator on the data and squares of the data, and tests that the two regression coefficients are zero (O'Brien's generalised $t$ test, PG1 for short). By considering the ranks of the data and the squares of the ranks we have the O'Brien's PG2 generalised rank sum test. Podgor and Gastwirth [15] suggest to regress the group indicator on two scores associated with the pooled sample order statistics. This approach generalises that of O'Brien. It is proved that the extended procedure can be recast as a quadratic combination of a rank test for location and one for scale (for the PG2 test they are the $W$ test and the Mood squared rank test, respectively). If one knows which distribution generated the data can use the optimal scores. Note that, for what concerns scale testing, Ansari–Bradley's scores should be preferred when central ranks give more dispersion information then do the extreme ones. Mood's scores (used in the PG2 test) give more weight to the extreme ranks and the corresponding test is more efficient than the Ansari-Bradley one under normal distributions. Klotz's scores give even more weight to the extreme ranks and the corresponding test is more efficient than the Mood test under normal distributions. ARE considerations [23] suggest to prefer Klotz's scores under light-tailed distributions (when the extreme ranks give more dispersion information than the central ones). Ansari-Bradley's scores (or Barton-David's which are almost the same) are suggested under heavy-tailed distributions like the double exponential or the Cauchy (when the central ranks give more dispersion information). Within O'Brien framework, Grambsch and O'Brien [24] proposed a two-step procedure. If a preliminary test for a zero coefficient for the squared term is significant, a test that both coefficients are simultaneously zero is performed. If the preliminary test is not significant, the group indicator is regressed on the data, and the usual test for regression is performed. In both cases, the final $p$-value is multiplied by 1.5. It should be noted that Podgor and Gastwirth [15] do not suggest this procedure since the adjustment to the final $p$-value seems arbitrary. Podgor and Gastwirth [15] found that their efficiency robust tests control the type-one error rate and have desirable powers. Moreover they found that the classical Lepage test performs well and suggested it when one has not a general idea of the distribution underlying the data, and consequently cannot develop an efficiency robust procedure. We will show that the Cucconi test is a better alternative to the Lepage test in these situations. It is worth noting that Gastwirth [25] developed the family of percentile modified rank tests for location and scale differences. In particular, percentile modified tests for scale are found to work better than the $A$ test in most cases. The practitioner can combine a percentile modified test for location and one for scale in a Lepage type test. Such test type is not considered here and needs further investigation.

Table 1. Selected upper tail probabilities for the null distribution of the Cucconi $C$ statistic.

| $n_1$ | $n_2$ | $c$ | $\Pr(C \geq c \mid H_0)$ | $n_1$ | $n_2$ | $c$ | $\Pr(C \geq c \mid H_0)$ | $n_1$ | $n_2$ | $c$ | $\Pr(C \geq c \mid H_0)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 6 | 2.24176 | 0.09957 | 7 | 11 | 2.25701 | 0.09961 | 8* | 30* | 2.242 | 0.100 |
| | | 2.71063 | 0.04329 | | | 2.77145 | 0.04971 | | | 2.839 | 0.050 |
| | | 3.62638 | 0.00649 | | | 3.89030 | 0.00999 | | | 4.318 | 0.010 |
| 6 | 7 | 2.22264 | 0.09965 | 7 | 12 | 2.23950 | 0.09983 | 9 | 9 | 2.26609 | 0.10000 |
| | | 2.72728 | 0.04604 | | | 2.80001 | 0.04989 | | | 2.80873 | 0.04998 |
| | | 3.48795 | 0.00991 | | | 3.96867 | 0.00998 | | | 3.87501 | 0.01000 |
| 6 | 8 | 2.21928 | 0.09990 | 7 | 13 | 2.23639 | 0.09982 | 9 | 10 | 2.26544 | 0.09992 |
| | | 2.71094 | 0.04995 | | | 2.78479 | 0.04995 | | | 2.82342 | 0.04985 |
| | | 3.63542 | 0.00999 | | | 3.99406 | 0.00998 | | | 3.92818 | 0.00999 |
| 6 | 9 | 2.25101 | 0.09970 | 7 | 15 | 2.23789 | 0.09975 | 9 | 11 | 2.26066 | 0.09981 |
| | | 2.71481 | 0.04995 | | | 2.79441 | 0.04989 | | | 2.83489 | 0.04992 |
| | | 3.73065 | 0.00999 | | | 4.08634 | 0.00996 | | | 3.96778 | 0.00998 |
| 6 | 10 | 2.22437 | 0.09965 | 7 | 20 | 2.23053 | 0.09999 | 9 | 12 | 2.26954 | 0.09995 |
| | | 2.73418 | 0.04920 | | | 2.80579 | 0.04998 | | | 2.83862 | 0.04997 |
| | | 3.83026 | 0.00987 | | | 4.20509 | 0.00999 | | | 4.01644 | 0.01000 |
| 6 | 11 | 2.22754 | 0.09987 | 7 | 25 | 2.22573 | 0.09997 | 9 | 13 | 2.26255 | 0.09968 |
| | | 2.73153 | 0.04994 | | | 2.80523 | 0.04998 | | | 2.84365 | 0.04990 |
| | | 3.93727 | 0.00970 | | | 4.28508 | 0.01000 | | | 4.04097 | 0.00996 |
| 6 | 12 | 2.22369 | 0.09960 | 7* | 30* | 2.218 | 0.100 | 9 | 15 | 2.26384 | 0.09994 |
| | | 2.75329 | 0.04977 | | | 2.804 | 0.050 | | | 2.85215 | 0.04992 |
| | | 3.95505 | 0.00997 | | | 4.351 | 0.010 | | | 4.09486 | 0.00999 |
| 6 | 13 | 2.21798 | 0.09988 | 8 | 8 | 2.25316 | 0.09930 | 9 | 20 | 2.26054 | 0.09998 |
| | | 2.74545 | 0.04979 | | | 2.77548 | 0.04973 | | | 2.85791 | 0.04999 |
| | | 3.99626 | 0.00999 | | | 3.76077 | 0.00979 | | | 4.18995 | 0.01000 |
| 6 | 15 | 2.21170 | 0.09992 | 8 | 9 | 2.24952 | 0.09996 | 9* | 25* | 2.261 | 0.100 |
| | | 2.75080 | 0.04996 | | | 2.80215 | 0.04961 | | | 2.856 | 0.050 |
| | | 4.08729 | 0.00999 | | | 3.82067 | 0.01000 | | | 4.255 | 0.010 |
| 6 | 20 | 2.19432 | 0.09988 | 8 | 10 | 2.27007 | 0.09964 | 9* | 30* | 2.254 | 0.100 |
| | | 2.76006 | 0.04998 | | | 2.81185 | 0.04989 | | | 2.867 | 0.050 |
| | | 4.24144 | 0.00100 | | | 3.87764 | 0.00996 | | | 4.308 | 0.010 |
| 6 | 25 | 2.19593 | 0.09995 | 8 | 11 | 2.26421 | 0.09985 | 10 | 10 | 2.26122 | 0.09959 |
| | | 2.77040 | 0.04996 | | | 2.81425 | 0.04980 | | | 2.84312 | 0.04990 |
| | | 4.33325 | 0.00999 | | | 3.94321 | 0.00987 | | | 3.97377 | 0.00996 |
| 6 | 30 | 2.18728 | 0.09999 | 8 | 12 | 2.25803 | 0.09991 | 10 | 11 | 2.26497 | 0.09998 |
| | | 2.77385 | 0.04999 | | | 2.82441 | 0.04999 | | | 2.85252 | 0.04996 |
| | | 4.38886 | 0.00999 | | | 3.98747 | 0.00991 | | | 4.01328 | 0.01000 |
| 7 | 7 | 2.22501 | 0.09965 | 8 | 13 | 2.25584 | 0.09999 | 10 | 12 | 2.27202 | 0.09970 |
| | | 2.77756 | 0.04895 | | | 2.82765 | 0.04995 | | | 2.85028 | 0.04999 |
| | | 3.52245 | 0.00991 | | | 4.01539 | 0.00999 | | | 4.04674 | 0.00993 |
| 7 | 8 | 2.24609 | 0.09992 | 8 | 15 | 2.25596 | 0.09998 | 10 | 13 | 2.27011 | 0.09968 |
| | | 2.75150 | 0.04942 | | | 2.82881 | 0.04996 | | | 2.85627 | 0.04999 |
| | | 3.73433 | 0.00979 | | | 4.07715 | 0.00998 | | | 4.07275 | 0.00997 |
| 7 | 9 | 2.23783 | 0.09983 | 8 | 20 | 2.25010 | 0.10000 | 10 | 15 | 2.26869 | 0.09990 |
| | | 2.77405 | 0.04983 | | | 2.83697 | 0.04998 | | | 2.86428 | 0.04996 |
| | | 3.77125 | 0.00997 | | | 4.19573 | 0.00999 | | | 4.11645 | 0.01000 |
| 7 | 10 | 2.24863 | 0.09955 | 8* | 25* | 2.245 | 0.100 | 10* | 20* | 2.272 | 0.100 |
| | | 2.78597 | 0.04993 | | | 2.840 | 0.050 | | | 2.876 | 0.050 |
| | | 3.85264 | 0.00992 | | | 4.256 | 0.010 | | | 4.212 | 0.010 |

(*Continued*)

Table 1. Continued.

| $n_1$ | $n_2$ | $c$ | $\Pr(C \geq c \mid H_0)$ | $n_1$ | $n_2$ | $c$ | $\Pr(C \geq c \mid H_0)$ | $n_1$ | $n_2$ | $c$ | $\Pr(C \geq c \mid H_0)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 10* | 25* | 2.260 | 0.100 | 12* | 15* | 2.273 | 0.100 | 15* | 25* | 2.278 | 0.100 |
| | | 2.876 | 0.050 | | | 2.883 | 0.050 | | | 2.909 | 0.050 |
| | | 4.259 | 0.010 | | | 4.157 | 0.010 | | | 4.304 | 0.010 |
| 10* | 30* | 2.262 | 0.100 | 12* | 20* | 2.277 | 0.100 | 15* | 30* | 2.281 | 0.100 |
| | | 2.882 | 0.050 | | | 2.894 | 0.050 | | | 2.917 | 0.050 |
| | | 4.311 | 0.010 | | | 4.234 | 0.010 | | | 4.332 | 0.010 |
| 11 | 11 | 2.27722 | 0.09994 | 12* | 25* | 2.276 | 0.100 | 20* | 20* | 2.286 | 0.100 |
| | | 2.86391 | 0.04995 | | | 2.896 | 0.050 | | | 2.922 | 0.050 |
| | | 4.04537 | 0.00999 | | | 4.281 | 0.010 | | | 4.319 | 0.010 |
| 11 | 12 | 2.27295 | 0.09992 | 12* | 30* | 2.272 | 0.100 | 20* | 25* | 2.289 | 0.100 |
| | | 2.86819 | 0.04998 | | | 2.895 | 0.050 | | | 2.932 | 0.050 |
| | | 4.07013 | 0.00999 | | | 4.312 | 0.010 | | | 4.349 | 0.010 |
| 11 | 13 | 2.27093 | 0.09998 | 13 | 13 | 2.27621 | 0.09999 | 20* | 30* | 2.285 | 0.100 |
| | | 2.87325 | 0.04998 | | | 2.87936 | 0.04994 | | | 2.927 | 0.050 |
| | | 4.10065 | 0.01000 | | | 4.14528 | 0.01000 | | | 4.365 | 0.010 |
| 11 | 15 | 2.27213 | 0.09997 | 13* | 15* | 2.274 | 0.100 | 25* | 25* | 2.288 | 0.100 |
| | | 2.87588 | 0.04997 | | | 2.889 | 0.050 | | | 2.941 | 0.050 |
| | | 4.14361 | 0.00999 | | | 4.190 | 0.010 | | | 4.383 | 0.010 |
| 11* | 20* | 2.271 | 0.100 | 13* | 20* | 2.284 | 0.100 | 25* | 30* | 2.296 | 0.100 |
| | | 2.887 | 0.050 | | | 2.905 | 0.050 | | | 2.945 | 0.050 |
| | | 4.208 | 0.010 | | | 4.246 | 0.010 | | | 4.402 | 0.010 |
| 11* | 25* | 2.265 | 0.100 | 13* | 25* | 2.277 | 0.100 | 30* | 30* | 2.292 | 0.100 |
| | | 2.885 | 0.050 | | | 2.907 | 0.050 | | | 2.955 | 0.050 |
| | | 4.261 | 0.010 | | | 4.293 | 0.010 | | | 4.428 | 0.010 |
| 11* | 30* | 2.267 | 0.100 | 13* | 30* | 2.273 | 0.100 | $\infty$ | $\infty$ | 2.30259 | 0.1 |
| | | 2.893 | 0.050 | | | 2.906 | 0.050 | | | 2.99573 | 0.05 |
| | | 4.312 | 0.010 | | | 4.317 | 0.010 | | | 4.60517 | 0.01 |
| 12 | 12 | 2.27864 | 0.09994 | 15* | 15* | 2.283 | 0.100 | | | | |
| | | 2.87160 | 0.04997 | | | 2.901 | 0.050 | | | | |
| | | 4.09999 | 0.00998 | | | 4.211 | 0.010 | | | | |
| 12 | 13 | 2.27345 | 0.09993 | 15* | 20* | 2.278 | 0.100 | | | | |
| | | 2.87549 | 0.05000 | | | 2.904 | 0.050 | | | | |
| | | 4.12621 | 0.01000 | | | 4.264 | 0.010 | | | | |

*Estimated by considering a sample of 1 million permutations.

## 3. Critical values, type-one error rate and power of the Cucconi test

It is important to emphasise that a detailed type-one error rate and power study of the Cucconi test has been performed here for the very first time. Table 1 reports the critical values for the test at $\alpha \cong 0.01$, 0.05 and 0.1 in the absence of ties. Most of the values have been computed exactly by considering the full distribution of the $C$ statistic. For given $n_1$ and $n_2$, the table entry for the point $c$ is $\Pr(C \geq c \mid H_0)$. A program coded in R language for computing exactly the $p$-values of the Cucconi test is available from the author. The exact distribution of $C$ under $H_0$ is plotted for $n_1 = n_2 = 10$ in Figure 1. The distribution is highly right skewed.

A simulation study has been performed to evaluate the power of the Cucconi test in detecting location and scale differences. It is compared with the Lepage test and to four Podgor and Gastwirth efficiency robust tests: PG1, PG2, PG3 and PG4 tests, where the PG3 test is developed using van der Waerden's and Klotz's normal scores for location and scale, respectively; and the PG4 test using Wilcoxon's and Klotz's scores. The Cramer–Von Mises (CVM) and Kolmogorov–Smirnov

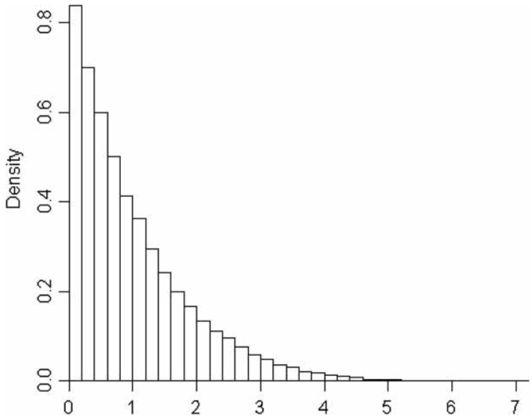



Figure 1. Graph of the distribution of $C$ under $H_0$ when $n_1 = n_2 = 10$.



(KS) tests have been considered as well since they are widely used for the general two-sample problem. We emphasise that $p$-values of Lepage, Cucconi and KS tests have been computed exactly, while $p$-values of the CVM test have been computed via the bootstrap (1000 replicates for each simulation), with the exceptions of the second example and the simulation results displayed in Table 8 (exact computation). We consider the following distributions:

(1) standard normal $N(0, 1)$;
(2) uniform between $-\sqrt{3}$ and $\sqrt{3}$;
(3) bimodal obtained as a mixture of a $N(-1.5, 1)$ with probability 0.5 and a $N$(1.5,1) with 0.5;
(4) Laplace double exponential with scale parameter of $1/\sqrt{2}$;
(5) 10% outlier obtained as a mixture of a $N(0, 1)$ with probability 0.9 and a $N$(1,100) with 0.1;
(6) 30% outlier obtained as a mixture of a $N(0, 1)$ with probability 0.7 and a $N$(1,100) with 0.3;
(7) Student's $t$ with 2 df;
(8) standard Cauchy, which corresponds to a Student's $t$ with 1 df.

Note that distributions (2) and (3) have light tails, while distributions (4) to (8) have moderately heavy to very heavy tails. Moreover, note that distributions (7) and (8) have infinite second moment, and distribution (8) has undefined first moment. It should be noted that while it is common to consider the uniform in simulation studies as a light-tailed model one can question whether it is realistic to do so. We considered $(n_1, n_2) = (10, 10)$, (10, 30), (30, 10) and (30, 30). The power has been estimated at $\alpha = 5\%$ through 50,000 Monte Carlo simulations. Results are reported in Tables 2–5, respectively, for each sample size setting. The entries in Tables 2–5 are the proportions of samples that the tests rejected the null hypothesis for the various distributions and configurations of $\mu_1 - \mu_2$ and $\sigma_1/\sigma_2$ values. Type-one error rate of the tests is reported in the first column of inner cells, which corresponds to the null hypothesis $\mu_1 - \mu_2 = 0$ and $\sigma_1/\sigma_2 = 1$. As it can be seen, both the Cucconi and Lepage tests maintain their size very close to the nominal significance level and this is not true for the other tests. PG tests tend to be conservative, in particular PG1 for distributions with heavier tails than the normal; in general, we noted that PG tests showed a higher degree of conservativeness under such type of distributions. For what concerns the tests for the general problem, we noted that the KS test is always conservative, as expected with a degree of conservativeness that decreases as sample sizes increase, and that the CVM test has a size quite close to $\alpha$ under normal- and light-tailed distributions, and is conservative otherwise (but with a smaller degree than that of the KS test). Of course, it should be cautioned that these results and the other presented in this section directly apply only to the settings studied here.

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

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | Normal | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 1 | 2 | 1 | 1 | 1 |
| $\sigma_1/\sigma_2$ | 1 | 2 | 2 | 2 | 1 | 3 | 5 |
| $C$ | 0.050 | 0.274 | 0.415 | 0.756 | 0.407 | 0.630 | 0.875 |
| $L$ | 0.050 | 0.244 | 0.389 | 0.744 | 0.407 | 0.573 | 0.817 |
| PG1 | 0.036 | 0.171 | 0.287 | 0.662 | 0.387 | 0.372 | 0.480 |
| PG2 | 0.051 | 0.278 | 0.419 | 0.759 | 0.410 | 0.634 | 0.877 |
| PG3 | 0.042 | 0.243 | 0.372 | 0.721 | 0.390 | 0.593 | 0.860 |
| PG4 | 0.045 | 0.260 | 0.398 | 0.737 | 0.395 | 0.617 | 0.873 |
| KS | 0.013 | 0.025 | 0.137 | 0.518 | 0.248 | 0.128 | 0.147 |
| CVM | 0.049 | 0.108 | 0.362 | 0.819 | 0.535 | 0.394 | 0.597 |
| | | | | Uniform | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 1 | 2 | 1 | 1 | 1 |
| $\sigma_1/\sigma_2$ | 1 | 2 | 2 | 2 | 1 | 3 | 5 |
| $C$ | 0.052 | 0.460 | 0.487 | 0.613 | 0.330 | 0.764 | 0.938 |
| $L$ | 0.052 | 0.379 | 0.438 | 0.632 | 0.347 | 0.685 | 0.889 |
| PG1 | 0.050 | 0.479 | 0.415 | 0.589 | 0.340 | 0.676 | 0.833 |
| PG2 | 0.053 | 0.465 | 0.491 | 0.618 | 0.333 | 0.768 | 0.940 |
| PG3 | 0.043 | 0.446 | 0.449 | 0.561 | 0.331 | 0.753 | 0.938 |
| PG4 | 0.047 | 0.466 | 0.475 | 0.557 | 0.309 | 0.772 | 0.945 |
| KS | 0.013 | 0.035 | 0.121 | 0.397 | 0.163 | 0.132 | 0.172 |
| CVM | 0.053 | 0.152 | 0.347 | 0.760 | 0.475 | 0.472 | 0.770 |
| | | | | Bimodal | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 2.5 | 4 | 1.5 | 1.5 | 1.5 |
| $\sigma_1/\sigma_2$ | 1 | 1.5 | 1.5 | 1.5 | 1 | 2.5 | 4 |
| $C$ | 0.050 | 0.199 | 0.448 | 0.848 | 0.246 | 0.643 | 0.911 |
| $L$ | 0.051 | 0.175 | 0.460 | 0.854 | 0.259 | 0.576 | 0.854 |
| PG1 | 0.045 | 0.171 | 0.454 | 0.881 | 0.241 | 0.529 | 0.752 |
| PG2 | 0.051 | 0.202 | 0.452 | 0.850 | 0.249 | 0.648 | 0.913 |
| PG3 | 0.041 | 0.176 | 0.427 | 0.849 | 0.240 | 0.610 | 0.907 |
| PG4 | 0.045 | 0.190 | 0.419 | 0.835 | 0.229 | 0.631 | 0.915 |
| KS | 0.012 | 0.021 | 0.244 | 0.659 | 0.115 | 0.108 | 0.156 |
| CVM | 0.054 | 0.079 | 0.600 | 0.948 | 0.353 | 0.348 | 0.637 |
| | | | | Laplace | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 1 | 2 | 1 | 1 | 1 |
| $\sigma_1/\sigma_2$ | 1 | 2 | 2 | 2 | 1 | 3 | 5 |
| $C$ | 0.050 | 0.177 | 0.459 | 0.852 | 0.559 | 0.561 | 0.769 |
| $L$ | 0.051 | 0.166 | 0.441 | 0.841 | 0.551 | 0.526 | 0.722 |
| PG1 | 0.028 | 0.070 | 0.298 | 0.755 | 0.499 | 0.263 | 0.269 |
| PG2 | 0.051 | 0.180 | 0.461 | 0.854 | 0.562 | 0.565 | 0.772 |
| PG3 | 0.041 | 0.151 | 0.415 | 0.823 | 0.527 | 0.516 | 0.732 |
| PG4 | 0.045 | 0.163 | 0.447 | 0.847 | 0.554 | 0.548 | 0.753 |
| KS | 0.012 | 0.021 | 0.236 | 0.708 | 0.438 | 0.179 | 0.154 |
| CVM | 0.043 | 0.080 | 0.435 | 0.893 | 0.657 | 0.391 | 0.457 |
| | | | | 10% outlier | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 1.5 | 4 | 1 | 1 | 1 |
| $\sigma_1/\sigma_2$ | 1 | 2.2 | 2.2 | 2.2 | 1 | 3.5 | 10 |
| $C$ | 0.050 | 0.235 | 0.435 | 0.889 | 0.320 | 0.496 | 0.856 |
| $L$ | 0.050 | 0.227 | 0.421 | 0.883 | 0.307 | 0.487 | 0.861 |
| PG1 | 0.021 | 0.060 | 0.252 | 0.801 | 0.267 | 0.140 | 0.197 |
| PG2 | 0.051 | 0.237 | 0.438 | 0.890 | 0.323 | 0.500 | 0.858 |
| PG3 | 0.041 | 0.197 | 0.389 | 0.869 | 0.301 | 0.431 | 0.786 |
| PG4 | 0.045 | 0.208 | 0.415 | 0.894 | 0.321 | 0.447 | 0.797 |
| KS | 0.012 | 0.027 | 0.214 | 0.886 | 0.190 | 0.105 | 0.162 |
| CVM | 0.022 | 0.053 | 0.291 | 0.889 | 0.267 | 0.181 | 0.390 |

(*Continued*)

Table 2. Continued.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 30% outlier | | | | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 3.6 | 12 | 1.3 | 1.3 | 1.3 |
| $\sigma_1/\sigma_2$ | 1 | 3 | 3 | 3 | 1 | 6 | 18 |
| $C$ | 0.048 | 0.224 | 0.495 | 0.910 | 0.265 | 0.484 | 0.899 |
| $L$ | 0.048 | 0.238 | 0.490 | 0.902 | 0.253 | 0.508 | 0.892 |
| PG1 | 0.014 | 0.023 | 0.215 | 0.762 | 0.104 | 0.048 | 0.067 |
| PG2 | 0.049 | 0.227 | 0.497 | 0.910 | 0.267 | 0.488 | 0.901 |
| PG3 | 0.040 | 0.175 | 0.450 | 0.879 | 0.234 | 0.410 | 0.861 |
| PG4 | 0.044 | 0.187 | 0.493 | 0.913 | 0.261 | 0.429 | 0.872 |
| KS | 0.012 | 0.035 | 0.372 | 0.905 | 0.173 | 0.090 | 0.144 |
| CVM | 0.024 | 0.050 | 0.322 | 0.966 | 0.140 | 0.118 | 0.244 |
| Student | | | | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 2 | 4.4 | 1 | 1 | 1 |
| $\sigma_1/\sigma_2$ | 1 | 2.4 | 2.4 | 2.4 | 1 | 3.6 | 9 |
| $C$ | 0.048 | 0.250 | 0.525 | 0.901 | 0.261 | 0.517 | 0.892 |
| $L$ | 0.048 | 0.239 | 0.505 | 0.892 | 0.253 | 0.493 | 0.870 |
| PG1 | 0.023 | 0.065 | 0.311 | 0.810 | 0.200 | 0.144 | 0.205 |
| PG2 | 0.049 | 0.253 | 0.528 | 0.902 | 0.264 | 0.521 | 0.894 |
| PG3 | 0.039 | 0.211 | 0.478 | 0.880 | 0.239 | 0.460 | 0.857 |
| PG4 | 0.043 | 0.224 | 0.509 | 0.902 | 0.258 | 0.481 | 0.867 |
| KS | 0.012 | 0.028 | 0.272 | 0.821 | 0.156 | 0.088 | 0.144 |
| CVM | 0.033 | 0.082 | 0.432 | 0.912 | 0.268 | 0.222 | 0.489 |
| Cauchy | | | | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 3 | 9 | 1.5 | 1.5 | 1.5 |
| $\sigma_1/\sigma_2$ | 1 | 3 | 3 | 3 | 1 | 5 | 15 |
| $C$ | 0.049 | 0.242 | 0.495 | 0.876 | 0.314 | 0.479 | 0.837 |
| $L$ | 0.049 | 0.247 | 0.489 | 0.872 | 0.308 | 0.484 | 0.846 |
| PG1 | 0.015 | 0.028 | 0.197 | 0.725 | 0.174 | 0.064 | 0.080 |
| PG2 | 0.050 | 0.245 | 0.497 | 0.877 | 0.317 | 0.483 | 0.839 |
| PG3 | 0.040 | 0.196 | 0.444 | 0.847 | 0.284 | 0.409 | 0.768 |
| PG4 | 0.044 | 0.208 | 0.477 | 0.878 | 0.313 | 0.428 | 0.778 |
| KS | 0.012 | 0.033 | 0.293 | 0.887 | 0.229 | 0.091 | 0.147 |
| CVM | 0.017 | 0.045 | 0.251 | 0.768 | 0.183 | 0.120 | 0.261 |

For $\mu_1 - \mu_2 > 0$ and $\sigma_1/\sigma_2 > 1$, the proportions in Tables 2–5 are Monte Carlo estimates of the power of the tests at the particular alternatives and for the various distributions. Power results show that the test of Cucconi is better than that of Lepage for detecting location–scale changes, except when the data are from a Cauchy distribution. In this case, the $L$ test is slightly more powerful than the $C$ test. The difference in power between the $C$ and $L$ tests is more marked under normal- and light-tailed distributions and for the smallest sample-size setting. Note that power studies in the papers cited before show that the $L$ test performs better under heavy-tailed distributions than under normal- and light-tailed ones since it is based on the Wilcoxon–Mann–Witney and Ansari–Bradley tests (that are optimal, respectively, for the stand alone location and scale problem, under distributions with heavier tails than the normal ones [26]). The greatest differences in power between the $C$ and $L$ tests are when the location but not the scale changes under normal- and light-tailed distributions. On the whole, the best PG test is PG2, which considers the ranks and the squares of the ranks. PG3 and PG4 perform well for normal- and light-tailed distributions when $n_1 = n_2$. In these particular situations, they are slightly more powerful than PG2; but when $n_1 \neq n_2$ they are remarkably less powerful than PG2 (in particular this is the case of PG3). The worst test is KS, while the CVM test shows an intermediate behaviour with respect to the $L$ and

Table 3. Power estimates with $\alpha = 0.05$ and $(n_1, n_2) = (10, 30)$.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | Normal | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 0.75 | 1.5 | 0.75 | 0.75 | 0.75 |
| $\sigma_1/\sigma_2$ | 1 | 1.5 | 1.5 | 1.5 | 1 | 2.5 | 4 |
| $C$ | 0.051 | 0.248 | 0.476 | 0.871 | 0.394 | 0.788 | 0.953 |
| $L$ | 0.050 | 0.201 | 0.433 | 0.852 | 0.391 | 0.695 | 0.896 |
| PG1 | 0.046 | 0.339 | 0.567 | 0.914 | 0.399 | 0.900 | 0.984 |
| PG2 | 0.049 | 0.244 | 0.471 | 0.868 | 0.386 | 0.784 | 0.951 |
| PG3 | 0.045 | 0.303 | 0.530 | 0.896 | 0.390 | 0.864 | 0.982 |
| PG4 | 0.046 | 0.295 | 0.508 | 0.879 | 0.386 | 0.856 | 0.980 |
| KS | 0.038 | 0.075 | 0.299 | 0.768 | 0.366 | 0.345 | 0.493 |
| CVM | 0.052 | 0.151 | 0.462 | 0.900 | 0.492 | 0.689 | 0.934 |
| | | | Uniform | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 0.75 | 1.5 | 0.75 | 0.75 | 0.75 |
| $\sigma_1/\sigma_2$ | 1 | 1.5 | 1.5 | 1.5 | 1 | 2.5 | 4 |
| $C$ | 0.048 | 0.440 | 0.510 | 0.766 | 0.340 | 0.882 | 0.980 |
| $L$ | 0.049 | 0.324 | 0.434 | 0.739 | 0.336 | 0.780 | 0.938 |
| PG1 | 0.046 | 0.601 | 0.687 | 0.901 | 0.393 | 0.982 | 0.999 |
| PG2 | 0.047 | 0.434 | 0.504 | 0.762 | 0.334 | 0.880 | 0.980 |
| PG3 | 0.044 | 0.606 | 0.623 | 0.837 | 0.414 | 0.957 | 0.996 |
| PG4 | 0.045 | 0.594 | 0.580 | 0.782 | 0.350 | 0.952 | 0.996 |
| KS | 0.037 | 0.110 | 0.294 | 0.684 | 0.268 | 0.457 | 0.667 |
| CVM | 0.053 | 0.172 | 0.429 | 0.862 | 0.430 | 0.776 | 0.979 |
| | | | Bimodal | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 2 | 3.5 | 1 | 1 | 1 |
| $\sigma_1/\sigma_2$ | 1 | 1.5 | 1.5 | 1.5 | 1 | 1.5 | 3 |
| $C$ | 0.050 | 0.406 | 0.616 | 0.925 | 0.217 | 0.459 | 0.955 |
| $L$ | 0.050 | 0.318 | 0.558 | 0.912 | 0.217 | 0.380 | 0.889 |
| PG1 | 0.046 | 0.523 | 0.779 | 0.980 | 0.231 | 0.599 | 0.994 |
| PG2 | 0.048 | 0.400 | 0.611 | 0.922 | 0.212 | 0.454 | 0.954 |
| PG3 | 0.043 | 0.498 | 0.712 | 0.956 | 0.246 | 0.555 | 0.986 |
| PG4 | 0.044 | 0.489 | 0.656 | 0.935 | 0.216 | 0.523 | 0.984 |
| KS | 0.038 | 0.104 | 0.474 | 0.869 | 0.174 | 0.219 | 0.521 |
| CVM | 0.053 | 0.163 | 0.667 | 0.975 | 0.265 | 0.321 | 0.895 |
| | | | Laplace | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 0.75 | 1.5 | 0.75 | 0.75 | 0.75 |
| $\sigma_1/\sigma_2$ | 1 | 1.5 | 1.5 | 1.5 | 1 | 2.5 | 4 |
| $C$ | 0.048 | 0.169 | 0.549 | 0.947 | 0.555 | 0.729 | 0.900 |
| $L$ | 0.048 | 0.142 | 0.535 | 0.943 | 0.556 | 0.673 | 0.842 |
| PG1 | 0.045 | 0.212 | 0.517 | 0.922 | 0.453 | 0.742 | 0.897 |
| PG2 | 0.046 | 0.165 | 0.543 | 0.945 | 0.547 | 0.724 | 0.897 |
| PG3 | 0.043 | 0.197 | 0.535 | 0.932 | 0.494 | 0.763 | 0.932 |
| PG4 | 0.044 | 0.191 | 0.545 | 0.939 | 0.535 | 0.759 | 0.930 |
| KS | 0.036 | 0.064 | 0.453 | 0.911 | 0.592 | 0.388 | 0.420 |
| CVM | 0.046 | 0.136 | 0.555 | 0.958 | 0.628 | 0.666 | 0.867 |
| | | | 10% outlier | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 1 | 2.5 | 0.75 | 0.75 | 0.75 |
| $\sigma_1/\sigma_2$ | 1 | 2 | 2 | 2 | 1 | 2.2 | 6 |
| $C$ | 0.049 | 0.370 | 0.545 | 0.928 | 0.291 | 0.527 | 0.949 |
| $L$ | 0.050 | 0.326 | 0.511 | 0.925 | 0.286 | 0.485 | 0.921 |
| PG1 | 0.043 | 0.165 | 0.304 | 0.704 | 0.189 | 0.262 | 0.532 |
| PG2 | 0.047 | 0.364 | 0.539 | 0.926 | 0.284 | 0.522 | 0.947 |
| PG3 | 0.043 | 0.346 | 0.501 | 0.882 | 0.253 | 0.485 | 0.912 |
| PG4 | 0.045 | 0.343 | 0.508 | 0.909 | 0.283 | 0.489 | 0.915 |
| KS | 0.038 | 0.121 | 0.344 | 0.892 | 0.291 | 0.260 | 0.498 |
| CVM | 0.040 | 0.215 | 0.418 | 0.908 | 0.235 | 0.372 | 0.891 |

(*Continued*)

Table 3. Continued.

| 30% outlier | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\mu_1 - \mu_2$ | 0 | 0 | 1.8 | 4 | 1 | 1 | 1 |
| $\sigma_1/\sigma_2$ | 1 | 2.2 | 2.2 | 2.2 | 1 | 3 | 8 |
| $C$ | 0.049 | 0.269 | 0.506 | 0.867 | 0.246 | 0.473 | 0.896 |
| $L$ | 0.049 | 0.254 | 0.501 | 0.880 | 0.245 | 0.459 | 0.880 |
| PG1 | 0.047 | 0.241 | 0.329 | 0.588 | 0.080 | 0.370 | 0.662 |
| PG2 | 0.047 | 0.263 | 0.500 | 0.864 | 0.241 | 0.466 | 0.893 |
| PG3 | 0.043 | 0.273 | 0.459 | 0.785 | 0.189 | 0.456 | 0.873 |
| PG4 | 0.045 | 0.270 | 0.489 | 0.843 | 0.231 | 0.457 | 0.872 |
| KS | 0.038 | 0.111 | 0.427 | 0.890 | 0.292 | 0.234 | 0.404 |
| CVM | 0.041 | 0.218 | 0.421 | 0.866 | 0.132 | 0.406 | 0.867 |
| Student | | | | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 1.1 | 2.4 | 0.8 | 0.8 | 0.8 |
| $\sigma_1/\sigma_2$ | 1 | 1.8 | 1.8 | 1.8 | 1 | 2.2 | 4.5 |
| $C$ | 0.051 | 0.266 | 0.502 | 0.902 | 0.264 | 0.514 | 0.897 |
| $L$ | 0.051 | 0.230 | 0.476 | 0.895 | 0.263 | 0.467 | 0.844 |
| PG1 | 0.045 | 0.221 | 0.403 | 0.792 | 0.187 | 0.403 | 0.704 |
| PG2 | 0.049 | 0.261 | 0.496 | 0.899 | 0.258 | 0.508 | 0.894 |
| PG3 | 0.044 | 0.276 | 0.481 | 0.868 | 0.223 | 0.512 | 0.902 |
| PG4 | 0.045 | 0.272 | 0.489 | 0.886 | 0.253 | 0.511 | 0.902 |
| KS | 0.039 | 0.089 | 0.342 | 0.840 | 0.284 | 0.242 | 0.384 |
| CVM | 0.043 | 0.196 | 0.450 | 0.893 | 0.268 | 0.423 | 0.832 |
| Cauchy | | | | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 1.5 | 4 | 1 | 1 | 1 |
| $\sigma_1/\sigma_2$ | 1 | 2 | 2 | 2 | 1 | 3 | 8 |
| $C$ | 0.051 | 0.225 | 0.461 | 0.914 | 0.241 | 0.516 | 0.927 |
| $L$ | 0.050 | 0.210 | 0.457 | 0.924 | 0.242 | 0.494 | 0.906 |
| PG1 | 0.043 | 0.127 | 0.215 | 0.548 | 0.095 | 0.237 | 0.461 |
| PG2 | 0.048 | 0.220 | 0.454 | 0.911 | 0.235 | 0.510 | 0.924 |
| PG3 | 0.044 | 0.212 | 0.392 | 0.830 | 0.186 | 0.462 | 0.885 |
| PG4 | 0.046 | 0.209 | 0.417 | 0.877 | 0.227 | 0.466 | 0.887 |
| KS | 0.039 | 0.089 | 0.367 | 0.910 | 0.299 | 0.237 | 0.427 |
| CVM | 0.043 | 0.160 | 0.298 | 0.747 | 0.132 | 0.348 | 0.752 |

KS tests. On the whole, the best performing tests are the Cucconi and the PG2 tests that show similar results.

It should be cautioned that the KS and CVM tests are designed for the general two-sample problem, while the other ones for the location–scale problem. Therefore, to get a deeper insight in the performance of the tests, the case of different shapes for the distributions underlying the two samples to be compared has been considered as well. Even if this case is of some practical interest, as noted by Büning [26], very few papers addresses this case. Partly following Büning [26], we selected three "realistic" combinations of different types of parent distributions (uniform, bimodal), (bimodal, normal) and (normal, Laplace). Tables 6 and 7 present the results for $(n_1, n_2) = (10, 10)$ and $(30, 30)$, and three location and scale alternatives (50,000 Monte Carlo simulations). It is important to note that the case $\mu_1 - \mu_2 = 0$ and $\sigma_1/\sigma_2 = 1$ has not been addressed in Büning [26]. Tables 6 and 7 show that the tests for the general problem are remarkably less powerful than those for the location–scale problem when comparing light- and normal-tailed distributions. When comparing the normal and the Laplace the difference is less marked. Moreover, the difference in power is generally less marked when $n_1 = n_2 = 30$ than when $n_1 = n_2 = 10$. On the whole, the best performing tests are again the $C$ and PG2 tests, while the worst one is again KS. Note that the PG1 test works particularly well in comparing the uniform

Table 4. Power estimates with $\alpha = 0.05$ and $(n_1, n_2) = (30, 10)$.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Normal | | | | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 1 | 1.8 | 0.75 | 0.75 | 0.75 |
| $\sigma_1/\sigma_2$ | 1 | 1.8 | 1.8 | 1.8 | 1 | 2.5 | 4 |
| $C$ | 0.050 | 0.229 | 0.430 | 0.871 | 0.391 | 0.607 | 0.942 |
| $L$ | 0.049 | 0.231 | 0.430 | 0.871 | 0.389 | 0.578 | 0.911 |
| PG1 | 0.045 | 0.045 | 0.218 | 0.763 | 0.397 | 0.147 | 0.259 |
| PG2 | 0.048 | 0.221 | 0.417 | 0.864 | 0.384 | 0.595 | 0.938 |
| PG3 | 0.043 | 0.064 | 0.248 | 0.784 | 0.387 | 0.282 | 0.735 |
| PG4 | 0.044 | 0.093 | 0.367 | 0.858 | 0.384 | 0.406 | 0.805 |
| KS | 0.037 | 0.046 | 0.406 | 0.883 | 0.363 | 0.259 | 0.334 |
| CVM | 0.050 | 0.045 | 0.468 | 0.937 | 0.487 | 0.285 | 0.467 |
| Uniform | | | | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 1 | 1.8 | 0.75 | 0.75 | 0.75 |
| $\sigma_1/\sigma_2$ | 1 | 1.8 | 1.8 | 1.8 | 1 | 2.5 | 4 |
| $C$ | 0.050 | 0.460 | 0.464 | 0.776 | 0.342 | 0.844 | 0.992 |
| $L$ | 0.052 | 0.391 | 0.485 | 0.838 | 0.339 | 0.778 | 0.977 |
| PG1 | 0.048 | 0.431 | 0.352 | 0.747 | 0.393 | 0.724 | 0.945 |
| PG2 | 0.048 | 0.449 | 0.450 | 0.765 | 0.335 | 0.835 | 0.991 |
| PG3 | 0.044 | 0.163 | 0.205 | 0.702 | 0.414 | 0.525 | 0.926 |
| PG4 | 0.045 | 0.217 | 0.338 | 0.722 | 0.350 | 0.641 | 0.951 |
| KS | 0.039 | 0.055 | 0.356 | 0.830 | 0.267 | 0.287 | 0.429 |
| CVM | 0.054 | 0.071 | 0.427 | 0.911 | 0.427 | 0.363 | 0.734 |
| Bimodal | | | | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 2 | 3.5 | 1 | 1 | 1 |
| $\sigma_1/\sigma_2$ | 1 | 1.5 | 1.5 | 1.5 | 1 | 1.75 | 3 |
| $C$ | 0.050 | 0.214 | 0.431 | 0.920 | 0.211 | 0.393 | 0.961 |
| $L$ | 0.050 | 0.210 | 0.496 | 0.937 | 0.213 | 0.402 | 0.932 |
| PG1 | 0.047 | 0.120 | 0.389 | 0.940 | 0.224 | 0.237 | 0.725 |
| PG2 | 0.048 | 0.207 | 0.419 | 0.916 | 0.206 | 0.382 | 0.958 |
| PG3 | 0.044 | 0.059 | 0.345 | 0.917 | 0.238 | 0.137 | 0.784 |
| PG4 | 0.046 | 0.087 | 0.386 | 0.913 | 0.208 | 0.215 | 0.838 |
| KS | 0.038 | 0.044 | 0.459 | 0.916 | 0.169 | 0.168 | 0.319 |
| CVM | 0.052 | 0.047 | 0.585 | 0.982 | 0.258 | 0.174 | 0.426 |
| Laplace | | | | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 1 | 1.8 | 0.75 | 0.75 | 0.75 |
| $\sigma_1/\sigma_2$ | 1 | 1.8 | 1.8 | 1.8 | 1 | 2.5 | 4 |
| $C$ | 0.050 | 0.134 | 0.534 | 0.934 | 0.554 | 0.478 | 0.782 |
| $L$ | 0.051 | 0.148 | 0.515 | 0.928 | 0.557 | 0.449 | 0.747 |
| PG1 | 0.047 | 0.008 | 0.255 | 0.818 | 0.455 | 0.060 | 0.040 |
| PG2 | 0.048 | 0.129 | 0.525 | 0.930 | 0.547 | 0.466 | 0.773 |
| PG3 | 0.045 | 0.035 | 0.392 | 0.882 | 0.495 | 0.240 | 0.486 |
| PG4 | 0.046 | 0.053 | 0.517 | 0.933 | 0.536 | 0.367 | 0.596 |
| KS | 0.038 | 0.041 | 0.612 | 0.962 | 0.591 | 0.349 | 0.335 |
| CVM | 0.046 | 0.030 | 0.589 | 0.971 | 0.626 | 0.294 | 0.308 |
| 10% outlier | | | | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 1.5 | 3 | 0.75 | 0.75 | 0.75 |
| $\sigma_1/\sigma_2$ | 1 | 2 | 2 | 2 | 1 | 3 | 6 |
| $C$ | 0.049 | 0.217 | 0.526 | 0.940 | 0.294 | 0.508 | 0.805 |
| $L$ | 0.050 | 0.243 | 0.523 | 0.943 | 0.291 | 0.553 | 0.871 |
| PG1 | 0.045 | 0.013 | 0.193 | 0.570 | 0.191 | 0.021 | 0.018 |
| PG2 | 0.047 | 0.211 | 0.517 | 0.938 | 0.287 | 0.500 | 0.800 |
| PG3 | 0.044 | 0.071 | 0.366 | 0.882 | 0.258 | 0.264 | 0.560 |
| PG4 | 0.045 | 0.095 | 0.488 | 0.932 | 0.287 | 0.327 | 0.592 |
| KS | 0.038 | 0.055 | 0.578 | 0.979 | 0.294 | 0.226 | 0.363 |
| CVM | 0.041 | 0.016 | 0.343 | 0.902 | 0.236 | 0.068 | 0.126 |

(*Continued*)

Table 4. Continued.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 30% outlier | | | | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 3 | 7 | 1 | 1 | 1 |
| $\sigma_1/\sigma_2$ | 1 | 2.5 | 2.5 | 2.5 | 1 | 4.5 | 11 |
| $C$ | 0.049 | 0.192 | 0.502 | 0.879 | 0.249 | 0.438 | 0.857 |
| $L$ | 0.049 | 0.251 | 0.501 | 0.890 | 0.247 | 0.531 | 0.893 |
| PG1 | 0.048 | 0.005 | 0.056 | 0.327 | 0.078 | 0.002 | 0.000 |
| PG2 | 0.047 | 0.187 | 0.496 | 0.874 | 0.243 | 0.431 | 0.852 |
| PG3 | 0.044 | 0.058 | 0.385 | 0.771 | 0.188 | 0.190 | 0.583 |
| PG4 | 0.044 | 0.078 | 0.492 | 0.867 | 0.233 | 0.234 | 0.628 |
| KS | 0.037 | 0.070 | 0.724 | 0.984 | 0.297 | 0.205 | 0.345 |
| CVM | 0.041 | 0.007 | 0.278 | 0.928 | 0.131 | 0.014 | 0.027 |
| Student | | | | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 1.7 | 3.2 | 0.8 | 0.8 | 0.8 |
| $\sigma_1/\sigma_2$ | 1 | 2.2 | 2.2 | 2.2 | 1 | 3 | 6 |
| $C$ | 0.051 | 0.244 | 0.518 | 0.897 | 0.263 | 0.499 | 0.894 |
| $L$ | 0.051 | 0.273 | 0.499 | 0.889 | 0.263 | 0.516 | 0.909 |
| PG1 | 0.045 | 0.006 | 0.149 | 0.618 | 0.185 | 0.014 | 0.014 |
| PG2 | 0.049 | 0.236 | 0.508 | 0.893 | 0.257 | 0.489 | 0.889 |
| PG3 | 0.045 | 0.078 | 0.341 | 0.823 | 0.222 | 0.229 | 0.673 |
| PG4 | 0.046 | 0.104 | 0.476 | 0.897 | 0.252 | 0.302 | 0.718 |
| KS | 0.038 | 0.057 | 0.560 | 0.951 | 0.285 | 0.198 | 0.318 |
| CVM | 0.045 | 0.023 | 0.425 | 0.907 | 0.270 | 0.103 | 0.227 |
| Cauchy | | | | | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 2.5 | 5.5 | 1 | 1 | 1 |
| $\sigma_1/\sigma_2$ | 1 | 2.5 | 2.5 | 2.5 | 1 | 4 | 10 |
| $C$ | 0.049 | 0.202 | 0.486 | 0.872 | 0.244 | 0.452 | 0.836 |
| $L$ | 0.050 | 0.248 | 0.468 | 0.877 | 0.247 | 0.513 | 0.894 |
| PG1 | 0.041 | 0.008 | 0.065 | 0.355 | 0.098 | 0.004 | 0.001 |
| PG2 | 0.047 | 0.196 | 0.479 | 0.868 | 0.239 | 0.443 | 0.831 |
| PG3 | 0.043 | 0.066 | 0.346 | 0.780 | 0.190 | 0.213 | 0.596 |
| PG4 | 0.044 | 0.085 | 0.458 | 0.866 | 0.231 | 0.264 | 0.630 |
| KS | 0.037 | 0.060 | 0.615 | 0.972 | 0.298 | 0.189 | 0.348 |
| CVM | 0.042 | 0.010 | 0.185 | 0.650 | 0.133 | 0.022 | 0.052 |

and the bimodal (that are light-tailed), while in the remaining cases it performs worse than the other location–scale tests.

Some additional simulations have been performed to investigate how the Cucconi test works in case of ties. More precisely, the following simulations have been performed: $n_1 = n_2 = 6$ under binomial with parameters 10 and 0.5, and $n_1 = n_2 = 9$ under binomial with parameters 15 and 0.5. Midranks were used for tied observations. Binomial distributions have been relocated and rescaled so that the mean is zero and the standard deviation is one. We performed 10,000 Monte Carlo simulations for $n_1 = n_2 = 6$ and 1000 for $n_1 = n_2 = 9$, and we computed exactly the $p$-values of all tests. To obtain accurate $p$-value computation in the presence of ties, we generated for each simulated data set the whole conditional permutation distribution of the statistics. Therefore, exact conditional $p$-values have been computed. It should be noted that apparently the case of ties has been not studied in the papers cited above. The results displayed in Table 8 show that the tests are conservative, in particular KS (which as a consequence is the less powerful one) and that the $C$ test performs remarkably better than the $L$ test in particular when $\mu_1 - \mu_2 = 0$ and $\sigma_1/\sigma_2 > 1$. It is important to emphasise that the presence of ties does not lower the performance of the $C$ test, the contrary happens for the $L$ test.

Table 5. Power estimates with $\alpha = 0.05$ and $(n_1, n_2) = (30, 30)$.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | Normal | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 0.5 | 1 | 0.5 | 0.5 | 0.5 |
| $\sigma_1/\sigma_2$ | 1 | 1.3 | 1.3 | 1.3 | 1 | 1.75 | 2.5 |
| $C$ | 0.050 | 0.171 | 0.406 | 0.870 | 0.355 | 0.713 | 0.966 |
| $L$ | 0.050 | 0.148 | 0.388 | 0.864 | 0.357 | 0.642 | 0.926 |
| PG1 | 0.044 | 0.166 | 0.405 | 0.878 | 0.358 | 0.713 | 0.953 |
| PG2 | 0.051 | 0.172 | 0.408 | 0.871 | 0.357 | 0.715 | 0.966 |
| PG3 | 0.048 | 0.180 | 0.417 | 0.880 | 0.360 | 0.752 | 0.982 |
| PG4 | 0.050 | 0.184 | 0.413 | 0.870 | 0.351 | 0.752 | 0.982 |
| KS | 0.035 | 0.049 | 0.284 | 0.799 | 0.320 | 0.328 | 0.493 |
| CVM | 0.050 | 0.079 | 0.410 | 0.909 | 0.451 | 0.518 | 0.846 |
| | | | | Uniform | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 0.5 | 1 | 0.5 | 0.5 | 0.5 |
| $\sigma_1/\sigma_2$ | 1 | 1.3 | 1.3 | 1.3 | 1 | 1.75 | 2.5 |
| $C$ | 0.050 | 0.358 | 0.476 | 0.818 | 0.324 | 0.895 | 0.996 |
| $L$ | 0.050 | 0.258 | 0.425 | 0.819 | 0.334 | 0.790 | 0.978 |
| PG1 | 0.050 | 0.421 | 0.554 | 0.887 | 0.335 | 0.962 | 1.000 |
| PG2 | 0.050 | 0.360 | 0.478 | 0.820 | 0.326 | 0.896 | 0.996 |
| PG3 | 0.046 | 0.506 | 0.532 | 0.863 | 0.389 | 0.958 | 0.999 |
| PG4 | 0.048 | 0.512 | 0.484 | 0.775 | 0.289 | 0.956 | 1.000 |
| KS | 0.034 | 0.059 | 0.255 | 0.729 | 0.223 | 0.424 | 0.721 |
| CVM | 0.052 | 0.097 | 0.375 | 0.890 | 0.391 | 0.609 | 0.967 |
| | | | | Bimodal | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 1.1 | 2.2 | 0.75 | 0.75 | 0.75 |
| $\sigma_1/\sigma_2$ | 1 | 1.3 | 1.3 | 1.3 | 1 | 1.4 | 2 |
| $C$ | 0.048 | 0.304 | 0.550 | 0.932 | 0.246 | 0.543 | 0.968 |
| $L$ | 0.049 | 0.249 | 0.523 | 0.931 | 0.253 | 0.478 | 0.915 |
| PG1 | 0.047 | 0.337 | 0.593 | 0.967 | 0.245 | 0.593 | 0.985 |
| PG2 | 0.049 | 0.306 | 0.553 | 0.932 | 0.247 | 0.546 | 0.968 |
| PG3 | 0.047 | 0.336 | 0.597 | 0.959 | 0.277 | 0.591 | 0.985 |
| PG4 | 0.048 | 0.341 | 0.541 | 0.924 | 0.232 | 0.570 | 0.984 |
| KS | 0.033 | 0.064 | 0.395 | 0.894 | 0.188 | 0.260 | 0.508 |
| CVM | 0.051 | 0.091 | 0.514 | 0.976 | 0.293 | 0.331 | 0.740 |
| | | | | Laplace | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 0.5 | 1 | 0.5 | 0.5 | 0.5 |
| $\sigma_1/\sigma_2$ | 1 | 1.3 | 1.3 | 1.3 | 1 | 1.75 | 2.5 |
| $C$ | 0.051 | 0.115 | 0.485 | 0.944 | 0.515 | 0.633 | 0.882 |
| $L$ | 0.051 | 0.108 | 0.480 | 0.944 | 0.517 | 0.603 | 0.841 |
| PG1 | 0.037 | 0.074 | 0.371 | 0.887 | 0.410 | 0.460 | 0.656 |
| PG2 | 0.052 | 0.116 | 0.487 | 0.945 | 0.517 | 0.635 | 0.883 |
| PG3 | 0.048 | 0.111 | 0.448 | 0.920 | 0.466 | 0.620 | 0.888 |
| PG4 | 0.049 | 0.115 | 0.486 | 0.944 | 0.514 | 0.647 | 0.895 |
| KS | 0.034 | 0.046 | 0.479 | 0.956 | 0.572 | 0.435 | 0.461 |
| CVM | 0.047 | 0.070 | 0.518 | 0.969 | 0.595 | 0.549 | 0.741 |
| | | | | 10% outlier | | | |
| $\mu_1 - \mu_2$ | 0 | 0 | 0.75 | 1.5 | 0.5 | 0.5 | 0.5 |
| $\sigma_1/\sigma_2$ | 1 | 1.5 | 1.5 | 1.5 | 1 | 1.8 | 3 |
| $C$ | 0.049 | 0.236 | 0.522 | 0.925 | 0.263 | 0.525 | 0.905 |
| $L$ | 0.049 | 0.224 | 0.517 | 0.926 | 0.262 | 0.509 | 0.904 |
| PG1 | 0.016 | 0.021 | 0.171 | 0.501 | 0.119 | 0.081 | 0.083 |
| PG2 | 0.049 | 0.237 | 0.524 | 0.926 | 0.265 | 0.527 | 0.906 |
| PG3 | 0.047 | 0.187 | 0.446 | 0.864 | 0.242 | 0.420 | 0.773 |
| PG4 | 0.048 | 0.190 | 0.486 | 0.918 | 0.268 | 0.440 | 0.783 |
| KS | 0.035 | 0.068 | 0.437 | 0.933 | 0.260 | 0.270 | 0.498 |
| CVM | 0.030 | 0.064 | 0.379 | 0.898 | 0.220 | 0.244 | 0.571 |

(*Continued*)

Table 5. Continued.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | 30% outlier | | | | |
| $\mu_1-\mu_2$ | 0 | 0 | 1.2 | 2.8 | 0.7 | 0.7 | 0.7 |
| $\sigma_1/\sigma_2$ | 1 | 1.8 | 1.8 | 1.8 | 1 | 2.3 | 4.5 |
| $C$ | 0.048 | 0.245 | 0.489 | 0.904 | 0.240 | 0.487 | 0.881 |
| $L$ | 0.049 | 0.260 | 0.505 | 0.908 | 0.239 | 0.517 | 0.909 |
| PG1 | 0.027 | 0.061 | 0.116 | 0.354 | 0.058 | 0.109 | 0.218 |
| PG2 | 0.049 | 0.247 | 0.491 | 0.905 | 0.242 | 0.489 | 0.882 |
| PG3 | 0.045 | 0.210 | 0.413 | 0.817 | 0.193 | 0.420 | 0.825 |
| PG4 | 0.047 | 0.214 | 0.477 | 0.905 | 0.238 | 0.443 | 0.830 |
| KS | 0.035 | 0.090 | 0.506 | 0.980 | 0.291 | 0.283 | 0.500 |
| CVM | 0.038 | 0.102 | 0.308 | 0.899 | 0.138 | 0.244 | 0.659 |
| | | | Student | | | | |
| $\mu_1-\mu_2$ | 0 | 0 | 0.8 | 1.6 | 0.6 | 0.6 | 0.6 |
| $\sigma_1/\sigma_2$ | 1 | 1.6 | 1.6 | 1.6 | 1 | 1.8 | 2.9 |
| $C$ | 0.050 | 0.261 | 0.516 | 0.899 | 0.295 | 0.510 | 0.902 |
| $L$ | 0.050 | 0.249 | 0.504 | 0.897 | 0.296 | 0.493 | 0.882 |
| PG1 | 0.027 | 0.067 | 0.248 | 0.664 | 0.187 | 0.184 | 0.297 |
| PG2 | 0.051 | 0.263 | 0.518 | 0.900 | 0.297 | 0.513 | 0.902 |
| PG3 | 0.048 | 0.227 | 0.457 | 0.843 | 0.262 | 0.452 | 0.845 |
| PG4 | 0.048 | 0.231 | 0.494 | 0.892 | 0.297 | 0.473 | 0.853 |
| KS | 0.035 | 0.073 | 0.405 | 0.898 | 0.318 | 0.280 | 0.424 |
| CVM | 0.039 | 0.108 | 0.424 | 0.892 | 0.309 | 0.332 | 0.662 |
| | | | Cauchy | | | | |
| $\mu_1-\mu_2$ | 0 | 0 | 1.2 | 2.5 | 0.8 | 0.8 | 0.8 |
| $\sigma_1/\sigma_2$ | 1 | 1.8 | 1.8 | 1.8 | 1 | 2.2 | 4 |
| $C$ | 0.049 | 0.243 | 0.523 | 0.899 | 0.301 | 0.499 | 0.879 |
| $L$ | 0.049 | 0.254 | 0.533 | 0.904 | 0.304 | 0.515 | 0.895 |
| PG1 | 0.014 | 0.019 | 0.086 | 0.295 | 0.069 | 0.046 | 0.055 |
| PG2 | 0.049 | 0.245 | 0.525 | 0.900 | 0.302 | 0.501 | 0.880 |
| PG3 | 0.046 | 0.187 | 0.417 | 0.796 | 0.242 | 0.386 | 0.754 |
| PG4 | 0.046 | 0.191 | 0.480 | 0.883 | 0.299 | 0.414 | 0.765 |
| KS | 0.034 | 0.081 | 0.511 | 0.956 | 0.384 | 0.296 | 0.454 |
| CVM | 0.016 | 0.048 | 0.208 | 0.619 | 0.125 | 0.143 | 0.344 |

Table 6. Power estimates with different distributions, $\alpha = 0.05$ and $(n_1, n_2) = (10, 10)$.

| Distributions | $\mu_1-\mu_2$ | $\sigma_1/\sigma_2$ | $C$ | $L$ | PG1 | PG2 | PG3 | PG4 | KS | CVM |
|---|---|---|---|---|---|---|---|---|---|---|
| Uniform | 0 | 1 | 0.702 | 0.678 | 0.762 | 0.691 | 0.718 | 0.678 | 0.520 | 0.739 |
| Bimodal | 1 | 1.5 | 0.499 | 0.486 | 0.525 | 0.486 | 0.493 | 0.459 | 0.360 | 0.546 |
| | 1.4 | 1.8 | 0.702 | 0.678 | 0.762 | 0.691 | 0.718 | 0.678 | 0.520 | 0.739 |
| Bimodal | 0 | 1 | 0.384 | 0.363 | 0.320 | 0.388 | 0.338 | 0.355 | 0.039 | 0.140 |
| Normal | 1 | 1.5 | 0.696 | 0.652 | 0.555 | 0.700 | 0.649 | 0.667 | 0.133 | 0.439 |
| | 1.4 | 1.8 | 0.795 | 0.746 | 0.617 | 0.798 | 0.756 | 0.772 | 0.181 | 0.577 |
| Normal | 0 | 1 | 0.067 | 0.069 | 0.043 | 0.068 | 0.054 | 0.058 | 0.015 | 0.051 |
| Laplace | 1 | 1.5 | 0.412 | 0.406 | 0.314 | 0.415 | 0.364 | 0.386 | 0.218 | 0.455 |
| | 1.4 | 1.8 | 0.588 | 0.579 | 0.455 | 0.592 | 0.535 | 0.563 | 0.346 | 0.633 |

## 4. Applications to real and fictitious biomedical data

Table 9 presents data for an experiment comparing the peak levels of human plasma growth hormone after arginine hydrochloride infusion in 10 relatively coronary-prone subjects (type-A subjects) and in 11 relatively coronary-resistant subjects (type-B subjects).

Table 7. Power estimates with different distributions, $\alpha = 0.05$ and $(n_1, n_2) = (30, 30)$.

| Distributions | $\mu_1 - \mu_2$ | $\sigma_1/\sigma_2$ | $C$ | $L$ | PG1 | PG2 | PG3 | PG4 | KS | CVM |
|---|---|---|---|---|---|---|---|---|---|---|
| Uniform | 0 | 1 | 0.909 | 0.816 | 0.957 | 0.910 | 0.964 | 0.965 | 0.241 | 0.435 |
| Bimodal | 1.4 | 2 | 0.643 | 0.653 | 0.690 | 0.645 | 0.712 | 0.605 | 0.524 | 0.736 |
| | 2 | 2.8 | 0.898 | 0.880 | 0.947 | 0.899 | 0.908 | 0.864 | 0.801 | 0.917 |
| Bimodal | 0 | 1 | 0.916 | 0.888 | 0.907 | 0.917 | 0.898 | 0.900 | 0.327 | 0.569 |
| Normal | 1.4 | 2 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.948 | 1.000 |
| | 2 | 2.8 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.989 | 1.000 |
| Normal | 0 | 1 | 0.103 | 0.120 | 0.058 | 0.104 | 0.071 | 0.072 | 0.057 | 0.069 |
| Laplace | 0.5 | 1.2 | 0.512 | 0.526 | 0.401 | 0.514 | 0.437 | 0.452 | 0.424 | 0.515 |
| | 1 | 1.5 | 0.924 | 0.922 | 0.866 | 0.925 | 0.894 | 0.905 | 0.865 | 0.932 |

Table 8. Power estimates with $\alpha = 0.05$, tied observations.

| | $(n_1, n_2) = (6, 6)$ | | | | $(n_1, n_2) = (9, 9)$ | | | |
|---|---|---|---|---|---|---|---|---|
| $\mu_1 - \mu_2$ | 0 | 2 | 0 | 2 | 0 | 1.5 | 0 | 1.5 |
| $\sigma_1/\sigma_2$ | 1 | 1 | 3 | 3 | 1 | 1 | 2.5 | 2.5 |
| $C$ | 0.044 | 0.539 | 0.394 | 0.466 | 0.033 | 0.576 | 0.397 | 0.649 |
| $L$ | 0.042 | 0.525 | 0.226 | 0.384 | 0.042 | 0.571 | 0.300 | 0.621 |
| PG1 | 0.045 | 0.631 | 0.241 | 0.310 | 0.036 | 0.725 | 0.338 | 0.457 |
| PG2 | 0.046 | 0.531 | 0.350 | 0.443 | 0.040 | 0.569 | 0.380 | 0.626 |
| PG3 | 0.046 | 0.552 | 0.362 | 0.443 | 0.042 | 0.591 | 0.397 | 0.599 |
| PG4 | 0.047 | 0.549 | 0.363 | 0.446 | 0.042 | 0.584 | 0.396 | 0.602 |
| KS | 0.010 | 0.497 | 0.049 | 0.226 | 0.016 | 0.570 | 0.105 | 0.460 |
| CVM | 0.038 | 0.761 | 0.110 | 0.333 | 0.042 | 0.821 | 0.194 | 0.515 |

Table 9. Peak levels of human plasma growth hormone after arginine hydrochloride infusion (ng/ml).

| Type-A subjects | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 3.6 | 2.6 | 4.7 | 8 | 3.1 | 8.8 | 4.6 | 5.8 | 4 | 4.6 | |
| Type-B subjects | | | | | | | | | | |
| 16.2 | 17.4 | 8.5 | 15.6 | 5.4 | 9.8 | 14.9 | 16.6 | 15.9 | 5.3 | 10.5 |

Source: Hollander and Wolfe [21].



Type-A subjects are characterised by an excessive sense of time urgency, drive and competitiveness, on the contrary of type-B subjects which denote an opposite type of behaviour. We would like to test whether there is any difference between locations or scales of peak level growth hormone for type-A and type-B subjects. The observed value of the $L$ statistic is 11.430 (the exact $\alpha = 1\%$ critical value is 8.341 and the exact $p$-value is 0.00067) and that of the $C$ statistic is 5.720 (the exact $\alpha = 1\%$ critical value is 4.013, and the exact $p$-value is 0.00314). The Lepage test indicates to reject the null hypothesis of equal locations and scales, and so does the Cucconi test. We conclude that type-A and type-B subjects differ for what concerns the peak level of human plasma growth hormone after arginine hydrochloride infusion. Note that the $p$-values of the other tests are 0.00784 (KS), 0.00068 (CVM), 0.00058 (PG1), 0.00048 (PG2), 0.00087 (PG3) and 0.00048 (PG4).

The next example is hypothetical, and it is discussed for the purpose of mere illustration of the procedures. Six smoking and six nonsmoking women are randomly chosen among those attending a beauty centre to establish whether smoking results in a different forced expiratory vital capacity (FVC). All the subjects are between the 20 and 30 age range. The measurements of the FVC are reported in Table 10. We would like to test whether the smoking status had differential effect on

Table 10. Forced expiratory vital capacity.

| | | | | | |
|---|---|---|---|---|---|
| Non smoking subjects | | | | | |
| 2.75 | 2.83 | 4.95 | 4.61 | 4.81 | 4.32 |
| Smoking subjects | | | | | |
| 3.84 | 3.16 | 3.41 | 2.95 | 4.13 | 3.98 |

the FVC. The observed $L$ statistic is equal to 7.629, and the observed $C$ statistic is equal to 3.626. The exact $p$-values of the tests are 0.00866 and 0.01082, respectively. Both tests are significant at the 0.05 significance level so that experimenters can reject the null hypothesis and infer that the smoking status has a differential effect on the airway function of the women. The exact $p$-values of PG1, PG2, PG3 and PG4 tests are 0.01732, 0.00866, 0.02402 and 0.01957, respectively, and confirm the inference drawn from the $L$ and $C$ tests. It is quite interesting to note that the exact $p$-values of the KS and CVM tests are 0.14286 and 0.16234, respectively. Thus, these general tests fail to reject the null hypothesis in this small sample example.



## 5. Conclusion

In addressing the location–scale problem, it is preferable that a test controls the type-one error rate and performs well in a wide family of distributions, even if it is not the best one in every single circumstance. This appears to be the case of the Cucconi test. The test has less usefulness today than it had when it was proposed, but is also of historical interest. The test differs from other location–scale tests, as it is not a quadratic form combining a test for location and a test for scale differences. Within the rank framework, while the Lepage test is well known, the Cucconi test, which is earlier, is not known. In fact, exact critical values for the test are available here for the very first time as well as a detailed simulation study that includes the case of different parent distributions which is rarely considered, and the case of tied observations that is not considered in the references. It is shown that the test of Cucconi maintains the size very close to $\alpha$ and is more powerful than the Lepage test. Moreover, the performance of the test of Cucconi does not change in the presence of ties, the contrary happens for the Lepage test which does worse than in the absence of ties. Moreover, it is easier to compute the test of Cucconi than that of Lepage, Manly–Francis, Büning–Thadewald, Neuhäuser, Büning and Murakami. We noted that the tests for the location–scale problem are generally more powerful than those for the general problem when comparing distributions with different shapes. Our study considered also several Podgor–Gastwirth efficiency robust tests. The best of these tests is PG2, which considers the ranks and the squared ranks, and behaves similarly to the Cucconi test. To conclude, we suggest the practitioner to take the Cucconi test into account as a better alternative to the Lepage one when it is not possible to develop an efficiency robust procedure for the problem at hand.

## References

[1] M.L.Tiku, W.Y. Tan, and N. Balakrishnan, *Robust Inference*, Marcel Dekker, New York, 1986.
[2] R.R. Wilcox, *Introduction to Robust Estimation and Hypothesis Testing*, Academic Press, San Diego, 2005.
[3] C. Muccioli, R. Belford, M. Podgor, P. Sampaio, M. de Smet, and R. Nussenblatt, *The diagnosis of intraocular inflammation and cytomegalovirus retinitis in HIV-infected patients by laser flare photometry*, Ocul. Immunol. Inflamm. 4(2) (1993), pp. 75–81.
[4] T. Yonetani and H.B. Gordon, *Abrupt changes as indicators of decadal climate variability*, Clim. Dyn. 17 (2001), pp. 249–258.

[5] A. Lunde and A. Timmermann, *Duration dependence in stock prices: An analysis of bull and bear markets*, J. Bus. Econ. Stat. 22 (2004), pp. 253–273.

[6] J.W. Tukey, *A quick, compact, two-sample test to Duckworth's specifications*, Technometrics 1 (1959), pp. 32–48.

[7] B.F.J. Manly and R.I.C.C. Francis, *Testing for mean and variance differences with samples from distributions that may be non-normal with unequal variances*, J. Stat. Comput. Simul. 72(8) (2002), pp. 633–646.

[8] H. Büning and T. Thadewald, *An adaptive two-sample location-scale test of Lepage type for symmetric distributions*, J. Stat. Comput. Simul. 65 (2000), pp. 287–310.

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

[10] M. Neuhäuser, *An exact two-sample test based on the Baumgartner–Weiss–Schindler statistic and a modification of Lepage's test*, Commun. Statist. Theory Methods 29 (2000), pp. 67–78.

[11] H. Büning, *Robustness and power of modified Lepage, Kolmogorov–Smirnov and Cramèr–Von Mises two-sample tests*, J. Appl. Statist. 29(6) (2002), pp. 907–924.

[12] H. Murakami, *Lepage type statistic based on the modified Baumgartner statistic*, Comput. Statist. Data Anal. 51 (2007), pp. 5061–5067.

[13] Y. Lepage, *A combination of Wilcoxon's and Ansari–Bradley's statistics*, Biometrika 58 (1971), pp. 213–217.

[14] O. Cucconi, *Un nuovo test non parametrico per il confronto tra due gruppi campionari*, Giorn. Econ. XXVII (1968), pp. 225–248.

[15] M.J. Podgor and J.L. Gastwirth, *On non-parametric and generalized tests for the two-sample problem with location and scale change alternatives*, Stat. Med. 13 (1994), pp. 747–758.

[16] S. Siegel and J.W. Tukey, *A nonparametric sum of ranks procedure for relative spread in unpaired samples*, J. Amer. Statist. Assoc. 55 (1960), pp. 429–445; correction, ibid., 56 (1961), p.1005.

[17] Y. Lepage, *A table for a combined Wilcoxon Ansari–Bradley statistic*, Biometrika 60 (1973), pp. 113–116.

[18] Y. Lepage, *Asymptotic power efficiency for a location and scale problem*, Commun. Statist. Theory Methods 5 (1976), pp. 1257–1274.

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

[20] Y. Lepage, *Asymptotically optimum rank tests for continuous location and scale alternatives*, Commun. Statist. Theory Methods, 4 (1975), pp. 671–687.

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

[22] P.C. O'Brien, *Comparing two samples: Extensions of the t, rank sum, and log-rank tests*, J. Amer. Statist. Assoc. 83 (1988), pp. 52–61.

[23] B.S. Duran, *A survey of nonparametric tests for scale*, Commun. Statist. Theory Methods, 5 (1976), pp. 1287–1312.

[24] P.M. Grambsch and P.C. O'Brien, *The effects of transformations and preliminary tests for nonlinearity in regression*, Stat. Med. 10 (1991), pp. 697–709.

[25] J.L. Gastwirth, *Percentile modifications of two sample rank tests*, J. Amer. Statist. Assoc. 60 (1965), pp. 1127–1141.

[26] H. Büning, *Kolmogorov-Smirnov and Cramèr-Von Mises type two-sample tests with various weight functions*, Commun. Statist. Simul. Comput. 30(4) (2001), pp. 847–865.