

# A Comparison of Some Nonparametric Tests in Presence of Data from Skewed Populations <sup>(1)</sup>

*Un Confronto tra Test Non Parametrici per Popolazioni Asimmetriche*

Claudio Giovanni Borroni

Dipartimento di Metodi Quantitativi per l'Economia, Università di Milano Bicocca

**Riassunto:** In questo lavoro ci si propone di confrontare le *performance* di alcuni test non parametrici per l'uguaglianza di due popolazioni. Il confronto viene effettuato simulando le rispettive funzioni di potenza tramite elaboratore. Poiché tale procedura richiede una specificazione dell'ipotesi alternativa del test, si sceglie di considerare il caso particolare di popolazioni distribuite secondo leggi asimmetriche.

**Keywords:** Nonparametric tests, Girone-Cifarelli test, Monte Carlo simulations.

## 1. Introduction

A common need in statistical analysis is to test whether two populations can be considered equal with respect to a certain target variable. Let  $X$  and  $Y$  be two random variables obtained by observing a target variable on two populations, and let  $F_X$  and  $F_Y$  denote their respective cdf's; one has then to test the null hypothesis  $H_0: F_X(x) = F_Y(x)$  (for all real  $x$ ) usually against the alternative  $H_1: F_X(x) \neq F_Y(x)$  (for at least one  $x$ ). The decision has to be taken according to the values of two samples, which will be denoted by  $x_1, \dots, x_n$  and  $y_1, \dots, y_m$  (in some cases being  $n = m$ ).

The described problem can be solved, in a parametric context, with the usual  $t$ -test; This approach, however, can be adopted just when the hypothesis of normality is suitable for the target variable or when a large amount of data is available (so that the central limit theorem holds). In many fields of application, the distribution of the target variable is very far from being normal. Indeed, especially in economics, many variables (such as income) are characterized by a skewed distribution; in addition, another parametric model, accounting for the asymmetry of data, can be hardly specified in such cases.

As known, a valid alternative to the usual parametric tests can be found in nonparametrics. Of course, different nonparametric tests can be considered and hence a choice is often needed. Many Authors (see Latorre, 1977 and Schmid and Trede, 1995) pointed out that the widely known *Kolmogorov-Smirnov test* is, for example, often less powerful than other nonparametric tests. It should be noted, however, that a comparison of some nonparametric tests can be accomplished only by a further specification of the alternative hypothesis  $H_1: F_X(x) \neq F_Y(x)$ . In this paper we compare the performances of different nonparametric tests, under the assumption that the target variable has a positively skewed distribution (this situation might be often observed in economic applications). Our comparison is accomplished by *simulating* a large number of couples of samples from known asymmetric distributions and by evaluating the ability of each test to detect possible differences between the populations.

---

<sup>(1)</sup> This reasearch is partially supported by Università di Milano Bicocca (borsa giovani ricercatori 2001)

## 2. Testing the equality of two populations

Suppose that  $X$  and  $Y$  are continuous random variables (so that the probability of ties is negligible) and that the two samples are independent. Let  $\hat{F}_X(z) = 1/n \sum_{i=1, \dots, n} I(x_i \leq z)$  and  $\hat{F}_Y(z) = 1/m \sum_{i=1, \dots, m} I(y_i \leq z)$  be the empirical distribution functions of the two sample respectively ( $I(E)$  being the indicator function of the event  $E$ ). The Kolmogorov-Smirnov (KS) test is based on the statistic  $\sup_x |\hat{F}_X(x) - \hat{F}_Y(x)|$ . Other tests draw information about the possible differences of the two populations from  $|\hat{F}_X(x) - \hat{F}_Y(x)|$ ; for instance, the Cramér-Von Mises (CM) test is based on the statistic  $nm/(n+m) \int [\hat{F}_X(x) - \hat{F}_Y(x)]^2 d\hat{H}(x)$ , where  $\hat{H}(z) = \frac{1}{m+n} \left[ \sum_{i=1}^n I(x_i \leq z) + \sum_{i=1}^m I(y_i \leq z) \right]$  is the empirical distribution function corresponding to the total of  $n + m$  observations. Schmid and Trede (1995) propose a similar test, based on  $\sqrt{nm/(n+m)} \int |\hat{F}_X(x) - \hat{F}_Y(x)| d\hat{H}(x)$ .

We have to point out that the Schmid-Trede (ST) test is equivalent to the *Girone-Cifarelli* test when  $n = m$ . This last test (see Girone, 1964 and Cifarelli, 1974, 1975) is based on the statistic  $G = 1/n \sum_{i=1}^n |R_X(i) - R_Y(i)|$ , where  $R_X(i)$  denotes the  $i$ -th rank of the sample drawn from  $X$  (i.e. the number of elements of the sorted pooled sample which are less or equal to the  $i$ -th smallest observation drawn from  $X$ ). It can be proved that  $G = 2n \int |\hat{F}_X(x) - \hat{F}_Y(x)| d\hat{H}(x)$ , which shows the equivalency of the two tests. Even when  $n \neq m$ , the ST test is equivalent to a test earlier proposed by Gorio (1971); indeed, this last test is based on  $(nm) \int |\hat{F}_X(x) - \hat{F}_Y(x)| d\hat{H}(x)$ .

Gorio (1971) proposes his test as a generalization of the Girone-Cifarelli test when  $n \neq m$ ; Borroni (2001) proposes another generalization, based on the representation of  $G$  in terms of ranks rather than in terms of empirical distribution functions. This last

generalization is based on the test-statistic  $GC = 1/(m+n) \left[ \sum_{i=1}^n |R_X(i) - R_Y(i)| + \sum_{i=1}^n |R_X(i) - R_Y(m-n+i)| - \sum_{i=1}^{m-n} (R_Y(n+i) - R_Y(i)) \right]$ . It can be proved that the two generalizations are asymptotically equivalent, but they can give different results if  $n$  and  $m$  are finite (see Borroni, 2001, for details).

To implement a simulation of the considered tests, one firstly needs to choose a distribution law for the target variable. For our purposes, we consider a positively skewed distribution, the Pareto density:  $f(x; a, b) = ab^a x^{-(a+1)}$  ( $x > b; a, b > 0$ ).

A large number  $C$  of couples of samples are then drawn (by using a Monte Carlo simulation method) from the two populations (which are characterized by different values of the parameters); the number  $c$  of times in which the test successfully classifies the two populations as different, is recorded. The power of the test is then estimated by the ratio  $c / C$ . The power function of the test is hence simulated by varying the values of the parameters of the chosen distribution.

### 3. Results of simulations

Roughly speaking, all the considered tests are based on the difference  $|\hat{F}_X(x) - \hat{F}_Y(x)|$ . While the KS test considers just the maximum value of this difference, the other tests accounts for all its values. This fact can probably cause a higher power of these latter tests in many occasions. Table 1 reports a small part (because of the available space) of the results of our simulations; for the two populations a Pareto distribution was chosen, with  $b = 1$ . Under  $H_0$  the value of  $a$  was set to 3 (so that both the mean and the variance exist). For each value of  $a \geq 3$ , 10000 couples of random samples of the same size ( $n = m = 5$  for the left side of the Table,  $n = m = 10$  for the right side) were randomly drawn and the power of the considered tests was computed (Schmid and Tiede, 1995 use similar settings for their simulations). To reach the nominal significance level (0.05) a randomized version of the considered tests was implemented.

**Table 1:** Simulation with a fixed value of  $b$ .

| Value of $a$<br>( $n=m=5$ ) | KS     | CM     | ST     | GC     | Value of $a$<br>( $n=m=10$ ) | KS     | CM     | ST     | GC     |
|-----------------------------|--------|--------|--------|--------|------------------------------|--------|--------|--------|--------|
| 3                           | 0.0500 | 0.0500 | 0.0500 | 0.0500 | 3                            | 0.0500 | 0.0500 | 0.0500 | 0.0500 |
| 4                           | 0.0654 | 0.0663 | 0.0671 | 0.0671 | 4                            | 0.0790 | 0.0793 | 0.0754 | 0.0754 |
| 6                           | 0.1176 | 0.1326 | 0.1337 | 0.1337 | 6                            | 0.2193 | 0.2341 | 0.2338 | 0.2338 |
| 9                           | 0.2119 | 0.2446 | 0.2517 | 0.2517 | 9                            | 0.4491 | 0.4866 | 0.4868 | 0.4868 |
| 15                          | 0.3865 | 0.4559 | 0.4527 | 0.4527 | 15                           | 0.7624 | 0.7932 | 0.7909 | 0.7909 |
| 30                          | 0.6070 | 0.7018 | 0.6971 | 0.6971 | 30                           | 0.9647 | 0.9723 | 0.9677 | 0.9677 |

Table 1 shows that the KS test is less powerful than the other tests in this case. The ST and the GC test have the same power, as they are equivalent. The CM test works quite similarly to the GC test, with some differences depending on the distance of  $H_1$  from  $H_0$ . In Table 2 a similar situation is considered:  $a$  is fixed to 3, while  $b$  takes increasing values from 1 (which is the value under  $H_0$ ).

**Table 2.** Simulations with a fixed value of  $a$ .

| Value of $b$<br>( $n=m=5$ ) | KS     | CM     | ST     | GC     | Value of $b$<br>( $n=m=10$ ) | KS     | CM     | ST     | GC     |
|-----------------------------|--------|--------|--------|--------|------------------------------|--------|--------|--------|--------|
| 1                           | 0.0500 | 0.0500 | 0.0500 | 0.0500 | 1                            | 0.0500 | 0.0500 | 0.0500 | 0.0500 |
| 1.1                         | 0.0876 | 0.0971 | 0.0958 | 0.0958 | 1.05                         | 0.0729 | 0.0733 | 0.0730 | 0.0730 |
| 1.3                         | 0.2790 | 0.2933 | 0.2880 | 0.2880 | 1.1                          | 0.1462 | 0.1505 | 0.1490 | 0.1490 |
| 1.6                         | 0.5796 | 0.5970 | 0.5800 | 0.5800 | 1.25                         | 0.5222 | 0.4994 | 0.4769 | 0.4769 |
| 2.0                         | 0.7869 | 0.7964 | 0.7740 | 0.7740 | 1.4                          | 0.8184 | 0.7787 | 0.7504 | 0.7504 |
| 2.5                         | 0.8887 | 0.8947 | 0.8821 | 0.8821 | 1.8                          | 0.9903 | 0.9787 | 0.9678 | 0.9678 |

Note that, in the simulation described in Table 1, the variance of the target variable decreases as  $a$  increases. In Table 2, the opposite situation is observed: the variance of the Pareto distribution increases with  $b$ . Hence it seems that the KS tests is more suitable to deal with high variability in the data, at least for sufficiently large sample sizes. We have to point out, however, that in many applications, the possible difference between the two populations lies only in their location. For this reason, Table 3 reports the results of another simulation, which has the same settings as Table 1, except for  $b$  is not fixed, but varies in terms of  $a$  so that the variance of the distribution remains constant ( $= 1$ ).

**Table 3. Simulations with constant variance.**

| Value of $\alpha$<br>( $n=m=5$ ) | KS     | CM     | ST     | GC     | Value of $\alpha$<br>( $n=m=10$ ) | KS     | CM     | ST     | GC     |
|----------------------------------|--------|--------|--------|--------|-----------------------------------|--------|--------|--------|--------|
| 3.1                              | 0.0687 | 0.0786 | 0.0792 | 0.0792 | 3.1                               | 0.1055 | 0.1087 | 0.1079 | 0.1079 |
| 3.3                              | 0.2199 | 0.2357 | 0.2320 | 0.2320 | 3.2                               | 0.2811 | 0.2726 | 0.2603 | 0.2603 |
| 3.5                              | 0.3996 | 0.3930 | 0.3802 | 0.3802 | 3.3                               | 0.5251 | 0.4901 | 0.4625 | 0.4625 |
| 4                                | 0.7032 | 0.6881 | 0.6630 | 0.6630 | 3.5                               | 0.8259 | 0.7715 | 0.7326 | 0.7326 |
| 6                                | 0.9574 | 0.9385 | 0.9334 | 0.9334 | 4                                 | 0.9902 | 0.9774 | 0.9652 | 0.9652 |

The above Table shows that, when  $H_0$  differs from  $H_1$  only for location, the performance of the KS test is quite good (this becomes clearer if the sample sizes increase).

A further comparison is now needed; the performance of the considered tests is to be checked when the sizes of the two samples differ. This is particularly useful to compare the CG test (in the generalization proposed by Borroni, 2001) and the ST test (remember that they are equivalent if  $n=m$ ). Table 4 reports some results of a simulation with the same settings of Table 1 except for  $n = 5$ ,  $m = 8$  (left) and  $n = 5$ ,  $m = 10$  (right).

**Table 4. Simulations with different sample sizes.**

| Value of $\alpha$<br>( $n=5, m=8$ ) | KS     | CM     | ST     | GC     | Value of $\alpha$<br>( $n=5, m=10$ ) | KS     | CM     | ST     | GC     |
|-------------------------------------|--------|--------|--------|--------|--------------------------------------|--------|--------|--------|--------|
| 4                                   | 0.0655 | 0.0693 | 0.0693 | 0.0725 | 4                                    | 0.6778 | 0.746  | 0.0749 | 0.0770 |
| 6                                   | 0.1446 | 0.1596 | 0.1585 | 0.1681 | 6                                    | 0.1579 | 0.1775 | 0.1795 | 0.1924 |
| 10                                  | 0.3266 | 0.3776 | 0.3795 | 0.3972 | 10                                   | 0.3530 | 0.4095 | 0.4083 | 0.4391 |
| 15                                  | 0.5035 | 0.5614 | 0.5588 | 0.5921 | 15                                   | 0.5368 | 0.6101 | 0.6030 | 0.6413 |
| 20                                  | 0.6259 | 0.6845 | 0.6778 | 0.7141 | 20                                   | 0.6566 | 0.7307 | 0.7238 | 0.7604 |

Table 4 shows the superiority of the CG test over the other considered tests when the sample sizes differ. The performance of the CG test improves when the difference  $m-n$  increases. This can be confirmed under different hypotheses concerning the parameters  $\alpha$  and  $b$  (these results are not reported here for the sake of brevity).

## References

- Borroni C. G. (2001) Some notes about nonparametric tests for the equality of two populations, *Test*, 10, 147-159.
- Cifarelli D. M. (1974) Intorno alla distribuzione del test di G. Girone, *Giornale degli Economisti e Annali di Economia*, 33, 283-308.
- Cifarelli D. M. (1975) Contributi intorno ad un test per l'omogeneità tra due campioni, *Giornale degli Economisti e Annali di Economia*, 34, 233-249.
- Girone G. (1964) Su un indice di omogeneità di due distribuzioni del tipo dell'indice semplice di dissomiglianza. *Atti della XXIV Riunione Scientifica della SIS*, 53-78.
- Goria M. (1971) Some generalizations of Girone's test, *Annali dell'Istituto di Statistica dell'Università di Bari*, 36, 53-72.
- Latorre G. (1977) Confronto tra le funzioni di potenza per il problema di "due campioni", *Saggi di Statistica Metodologica ed Applicata. Contributi del Centro di Studi Statistici dell'Università degli Studi della Calabria*, 31-50.
- Schmid F., Trede M. (1995) A distribution-free test for the two sample problem for general alternatives, *Computational Statistics and Data Analysis*, 20, 409-419.