

A new goodness-of-fit version of the Girone-Cifarelli test

Claudio G. Borroni, Paola M. Chiodini

Abstract The classical goodness-of-fit problem, in the case of a null continuous and completely specified distribution, is faced by a new version of the Girone-Cifarelli test. This latter test was introduced for the two-sample problem and showed a substantial gain of power over other common tests based on the empirical distribution function, like the Kolmogorov-Smirnov test. After considering the problem of the definition of the test-statistic in the goodness-of-fit framework, this paper deals with the sample properties of the test and compares it with its classical competitors. The superiority of the Girone-Cifarelli test is shown via a simulation study considering symmetric and skewed distributions.

Key words: goodness-of-fit tests, empirical distribution function, Girone-Cifarelli test, nonparametric statistical methods

1 Introduction

A random sample $x_1, \dots, x_n$ is drawn from a population X with continuous distribution function F , to test the null hypothesis $H_0 : F(x) = F_0(x)$ against the alternative $H_1 : F(x) \neq F_0(x)$, $x \in \mathbb{R}$, where F_0 is completely specified. This common goodness-of-fit problem is usually faced by three classes of tests: the chi-square test, the tests based on spacings and the tests based on the empirical distribution function (edf). In

¹

Claudio Giovanni Borroni, Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali, Università degli Studi di Milano Bicocca; email: claudio.borroni@unimib.it.

Paola Maddalena Chiodini, Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali, Università degli Studi di Milano Bicocca; email: paola.chiodini@unimib.it.

this latter class several test-statistics can be considered, usually by adapting their versions for the two-sample problem.

The most known test based on the edf F_n is surely the Kolmogorov-Smirnov test, which reject H_0 for large value of the test-statistic

$$K_n = \sup_{t \in (-\infty, +\infty)} |F_0(t) - F_n(t)|. \quad (1)$$

As known, other test-statistics can be defined by considering the square of the difference $|F_0(t) - F_n(t)|$, like in the Cramér-Von Mises test

$$C_n = n \int_{-\infty}^{+\infty} [F_0(t) - F_n(t)]^2 dF_0(t). \quad (2)$$

Notice that in the above considered test-statistics F_0 , a continuous model, is compared with F_n , which has discontinuities at $x_1, \dots, x_n$. However, in (1) the supremum of the difference $|F_0(t) - F_n(t)|$ is taken, while in (2) the squared difference $[F_0(t) - F_n(t)]^2$ is integrated with respect to the continuous function F_0 . Because of these latter choices, no particular care is needed in the definition of the value taken by the edf at its points of discontinuity. This means that one can use the usual definition

$$F_n(x) = \frac{i}{n} \quad \text{for } x_{(i)} \leq x < x_{(i+1)} \quad (i = 0, \dots, n) \quad (3)$$

(where $x_{(1)}, \dots, x_{(n)}$ denotes the ordered sample, $x_{(0)} = -\infty$ and $x_{(n+1)} = +\infty$), which makes F_n to be right-continuous, or equivalently set

$$F_n(x_{(i)}) = \frac{i-c}{n} \quad (i = 1, \dots, n), \quad (4)$$

(where c is chosen in $[0,1]$), so that F_n can take every value of its jump at $x_{(i)}$ ($i = 1, \dots, n$).

Turning back to (2), the function with which the squared difference $[F_0(t) - F_n(t)]^2$ is integrated could be substituted by the edf itself. This choice allows to simplify the test-statistic as

$$C'_n = n \int_{-\infty}^{+\infty} [F_0(t) - F_n(t)]^2 dF_n(t) = \sum_{i=1}^n [F_0(x_{(i)}) - F_n(x_{(i)})]^2, \quad (5)$$

but the definition of $F_n(x_{(i)})$ becomes now very relevant. However, Anderson (1962) pointed out that, when $c = 1/2$ is taken in (4), the test-statistics C_n and C'_n are equivalent, as the former can be also written as

$$C_n = \sum_{i=1}^n \left[F_0(x_{(i)}) - \frac{i-1/2}{n} \right]^2 + \frac{1}{12n}. \quad (6)$$

Besides such a latter equivalence, setting $c = 1/2$ in (4) is, as a matter of fact, a natural choice. Indeed, forcing the edf to take the mid-point of its jump at $x_{(i)}$ seems less arbitrary than choosing any other value in the jump (including the extremes i/n and $(i-1)/n$, $i = 1, \dots, n$).

Notice again that any choice of $F_n(x_{(i)})$, made to give a final form to C' in (5), does not affect the usual definition of the edf in the open intervals $(x_{(i)}, x_{(i+1)})$,

$i = 1, \dots, n-1$. However, in the literature some modifications of the edf in such intervals were also proposed. For instance, Green and Hegazy (1976) pointed out that when the edf is re-defined as

$$F'_n(x) = \frac{i+1/2}{n+1} \quad \text{for } x_{(i)} < x < x_{(i+1)} \quad (i=1, \dots, n-1), \quad (7)$$

the criterion C_n in (2) reduces, up to a multiplicative constant, to

$$\sum_{i=1}^n \left[F_0(x_{(i)}) - \frac{i}{n+1} \right]^2. \quad (8)$$

which is shown to lead to a powerful test under some circumstances. Notice that the test statistic in (8) can be also obtained from C'_n in (5) by re-defining accordingly the value of the edf at its discontinuities, that is by setting

$$F'_n(x_{(i)}) = \frac{i}{n+1} \quad (i=1, \dots, n), \quad (9)$$

which is again the mid-point of the jump of F'_n at $x_{(i)}$.

Other modifications of the definition of the edf in the open interval $(x_{(i)}, x_{(i+1)})$ are known. By noticing that the term $i/(n+1)$ is actually the expectation of $F_0(x_{(i)})$ under the null hypothesis, Pyke (1959) proposed a new version of the Kolmogorov-Smirnov criterion (1), which in turn induces a further modification of the definition of the edf (see also Brunk, 1962).

The above remarks will be used in this paper to propose a goodness-of-fit version of the Girone-Cifarelli test, which was mainly studied for the two-sample problem. The definition of the test-statistic for goodness-of-fit purposes raises some questions which will be addressed in the next section, where the sample properties of the newly proposed test-statistic will be also analyzed. Section 3 will report some results of a simulation study, where the proposed test is compared with its most important competitors based on edf. Section 4 will conclude.

2 Definition of the test-statistic

Girone (1964) proposed a test for the equality of two populations X and Y , based on the statistic

$$(m+n) \int_{-\infty}^{+\infty} |F_n(t) - G_m(t)| dH_{m+n}(t) \quad (10)$$

where F_n , G_m and H_{m+n} denote respectively the edf's of a n -sample from X , a m -sample from Y and the pooled $(m+n)$ -sample. The test was actually originally proposed in the special case $n = m$ and its sample properties were studied by Cifarelli (1974, 1975). Further generalizations for the general case $n \leq m$ were proposed by Gorla (1972), by Borroni (2001) and independently by Schmid and Trede (1995). The superiority of the Girone-Cifarelli test for the two-sample problem, especially over the Kolmogorov-Smirnov test, was extensively proved in the above-reported references and in Latorre (1977). Such conclusions are far from being unexpected, as in (10) the whole

behaviour of the difference $|F_n(t) - G_m(t)|$ is considered, while in the Kolmogorov-Smirnov test just its supremum is taken.

A goodness-of-fit version of the Girone-Cifarelli test would result useful. Using the same settings as in section 1, the edf F_n of the single n -sample is now to be compared with the null model F_0 . The function respecting to which the difference $|F_n(t) - F_0(t)|$ is to be integrated could then be the null model F_0 or the edf F_n . As above remarked, this latter choice highly simplifies the structure of the test statistic, as

$$n \int_{-\infty}^{\infty} |F_0(t) - F_n(t)| dF_n(t) = \sum_{i=1}^n |F_0(x_{(i)}) - F_n(x_{(i)})|, \quad (11)$$

but the definition of the value taken by the edf at its discontinuities becomes relevant. Following the above suggestion by Anderson (1962) for C'_n , we can then take

$$F_n(x_{(i)}) = \frac{i-1/2}{n} \quad (i=1, \dots, n) \quad (12)$$

and define

$$A'_n = \sum_{i=1}^n \left| F_0(x_{(i)}) - \frac{i-1/2}{n} \right|. \quad (13)$$

Differently from C'_n , A'_n is not equivalent to the statistic obtained by using F_0 as an integrating function. This is shown by considering that

$$\begin{aligned} \int_{-\infty}^{\infty} |F_0(t) - F_n(t)| dF_0(t) = \\ \frac{1}{2} \sum_{i=1}^{n+1} \left[\left| F_0(x_{(i)}) - \frac{i-1}{n} \right| \left(F_0(x_{(i)}) - \frac{i-1}{n} \right) - \left| F_0(x_{(i-1)}) - \frac{i-1}{n} \right| \left(F_0(x_{(i-1)}) - \frac{i-1}{n} \right) \right] \end{aligned} \quad (14)$$

(see Schmid and Trede, 1996). This latter expression is far from being manageable and it does not seem to give any advantage over A'_n (this issue will be further discussed in Section 4). Of course other forms for the Girone-Cifarelli test-statistic could be tried by modifying the definition of the edf in the open intervals $(x_{(i)}, x_{(i+1)})$, $i = 1, \dots, n-1$. A partial work in such a direction is found in Green and Hegazy (1976); however, this issue needs more attention and it will be the object of future research.

The sample properties of A'_n are easily derived from its two-sample equivalent. First of all notice that, being F_0 a continuous model, the variables $F_0(x_{(i)})$, $i = 1, \dots, n$ are Uniform over $[0,1]$ and hence A'_n is distribution-free under H_0 . For small sample sizes, the null distribution of A'_n can then be determined by simulation, as pointed out in the next section. Moreover, following Cifarelli (1975), $n^{-1/2} A'_n$ is asymptotically distributed as the r.v.

$$\int_0^1 |w(\tau)| d\tau \mid w(1) = 0, \quad (15)$$

where $\{w(\tau), \tau \in [0,1]\}$ denotes the Brownian motion in $[0,1]$. A tabulation of the quantiles of (15) is found in Johnson and Killeen (1983); see also Shepp (1982, 1991) and Takács (1983).

Of course the good performance of the Girone-Cifarelli test in the two-sample problem does not prove its relevance for goodness-of-fit purposes. Moreover, the

power of the test based on A'_n is likely to depend closely on the form of the alternative hypothesis and on the values of the sample size. This issue will be addressed in the next section, where the proposed test is compared with various competitors via simulation.

3 Simulation study

The first task to develop a goodness-of-fit test based on A'_n is to determine its critical values. As above mentioned, being F_0 completely specified and continuous, the transformation $F_0(X)$ gives a Uniform distribution over $(0,1)$. Hence the null distribution of A'_n can be simulated by randomly generating a large number of samples from such a distribution, with a fixed size n . The critical values of the test can then be determined by computing the value taken by A'_n for each simulated sample as long as the related frequency distribution. For a selected range of sample sizes and some common significance levels, Table 1 reports the critical values of $n^{-1/2} A'_n$ based on 10^6 simulated samples. As a term of comparison, the last column of Table 1 reports the critical values of the asymptotic distribution of $n^{-1/2} A'_n$ (see section 2). The fast convergence to the asymptotic approximation can be easily appreciated.

Table 1: Simulated critical values of $n^{-1/2} A'_n$.

α	$n = 5$	$n = 10$	$n = 20$	$n = 30$	$n = 50$	∞
0.01	0.7142	0.7364	0.7436	0.7465	0.7478	0.7518
0.05	0.5670	0.5747	0.5783	0.5791	0.5807	0.5821
0.10	0.4893	0.4942	0.4966	0.4972	0.4982	0.4993
0.15	0.4398	0.4439	0.4459	0.4462	0.4470	0.4480
0.20	0.4029	0.4064	0.4081	0.4088	0.4092	0.4103

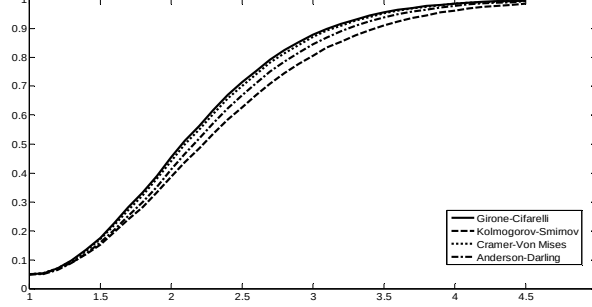
After computing the critical values of the test, its power can be estimated by simulation as well. This section reports just a portion of the results obtained in a wider simulation study. Notice that the null model F_0 can be assumed to be uniform over $(0,1)$ without loss of generality. Hence, to shape the alternative hypothesis, one can select a family of distribution containing the Uniform $(0,1)$. The Beta family is here reported. It is known that the density

$$f(x) = \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} \quad 0 < x < 1, \quad (16)$$

where $B(a,b)$ denotes the Beta function, reduces to the Uniform $(0,1)$ if $a = b = 1$. Moreover, recall that the Beta density can take different shapes as both its parameters range in $(0, \infty)$. The following Figure 1 is obtained by setting $a = 1$ and $b \geq 1$: as b grows, the distribution becomes more right-skewed. For $n = 10$, Figure 1 shows that A'_n performs quite similarly to the Cramer-Von Mises test, even if its estimated power function is uniformly over the one for C_n . Moreover, A'_n has a definitely higher power than the Kolmogorov-Smirnov test. These results seem to confirm the superiority of the

Girone-Cifarelli test-statistic even in the goodness-of-fit framework. Since this fact can be mainly due to the skewness of the alternative model chosen, Figure 1 reports also

Figure 1: Simulated power functions for the Beta distribution ($n = 10$; 10^6 replications).



the estimated power function of the Anderson-Darling test

$$D_n = \int_{-\infty}^{+\infty} [F_n(t) - F_0(t)]^2 \frac{1}{F_0(t)[1 - F_0(t)]} dF_0(t), \quad (17)$$

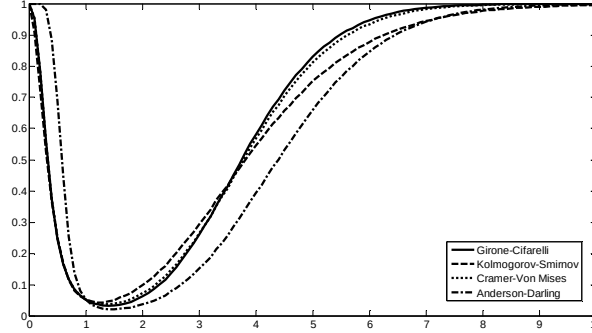
where the squared difference $[F_n(t) - F_0(t)]^2$ is weighted to get more sensibility in the tails of the distributions. Figure 1 shows that D_n has a substantial gain of power over K_n , but also that it is overcome by C_n and A'_n .

Some further results for the Weibull distribution are also reported. This latter family is widely used in applications, especially in reliability studies, when the usual normal density is discarded in favour of a more flexible model having both symmetric and skewed shapes. It is known that the density

$$f(x) = abx^{b-1}e^{-ax^b} \quad x > 0 \quad (18)$$

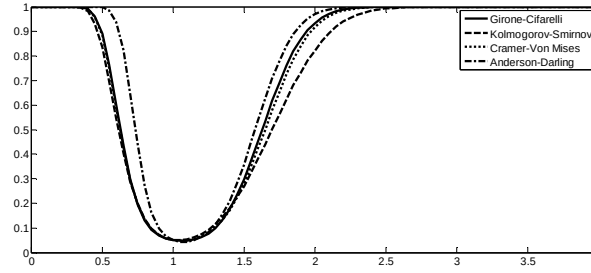
reduces to the exponential if $b = 1$; this condition is taken as the null hypothesis of the test. Figure 2 reports the simulated power functions obtained by setting $a = 1$ in (18) and by varying the parameter b ($n = 10$).

Figure 2: Simulated power functions for the Weibull distribution ($n = 10$; 10^6 replications).



The good performance of the A'_n test, as long as the relative positions of the other considered test, are mainly confirmed in Figure 2, except for the role of the Anderson-Darling test. It is known that, when $a = 1$ and $b > 1$ in (18), the shape of the distribution tends to symmetry. This fact might explain why the Anderson-Darling test has a worse performance under these conditions. However, when the sample size increases, the power of the same test grows over its competitors even for $b > 1$, as shown in Figure 3, where $n = 50$. In a sense, this testifies that the Anderson-Darling test may be more exposed to choice of the sample size. Notice that, despite its performance is slightly worse, the Girone-Cifarelli test does not seem to be affected by the same drawback.

Figure 3: Simulated power functions for the Weibull distribution ($n = 50$; 10^6 replications).



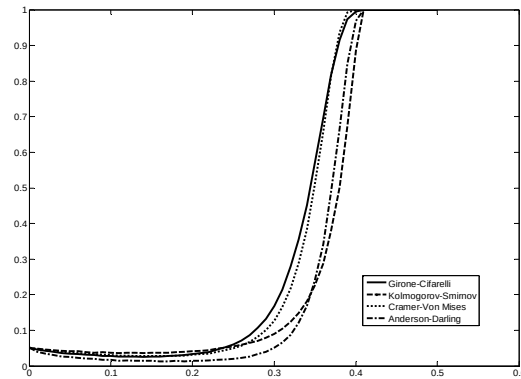
To test the performance of A'_n in purely symmetric distributions, a last set of simulation is reported for the density

$$f(x) = (1 - 2a)^{-1} \quad a < x < 1 - a \quad (19)$$

which is essentially a “compressed” Uniform distribution over the interval $(a, 1 - a)$, where $0 \leq a \leq 1/2$. Of course, the case $a = 0$ is taken to be the null hypothesis of the tests. Figure 4, reporting the case $n = 10$, shows that A'_n has still a good performance, being almost everywhere the best test among those considered. Notice that, in this case, the most relevant differences between A'_n and C_n can be appreciated.

4 Conclusions

This paper dealt with the potentialities of the Girone-Cifarelli test for goodness-of-fit purposes. The definition of the test-statistic in the goodness-of-fit framework is not unique. However, the proposed version is a good compromise between the simplicity of the final form and the performance of the test, especially for small sample sizes. This conclusion is particularly relevant if the proposed test-statistic is compared with the one in (14). Based on the conducted simulations, the two test-statistics have similar power. However A'_n is never worse than (14) and performs often significantly better. For instance, in the Weibull case with $n = 10$, when the power of the two tests is around 0.6, A'_n shows a gain of 0.012 over (14), such value being statistically significant at the 1%-level. Finally, the performance of the proposed test seems substantially unaffected by the choice of the sample size.

Figure 4: Simulated powers for the Compressed-Uniform distribution ($n = 10$; 10^6 replications)

References

1. Anderson, T. W.: On the distribution of the two-sample Cramér-Von Mises criterion. *Ann. Math. Statist.*, **33**, 1148–1159 (1962).
2. Borroni, C.G.: Some notes about nonparametric test for the equality of two populations. *Test*, **10**, 147–159 (2001).
3. Brunk, H.D.: On the range of the difference between hypothetical distribution function and Pyke's modified empirical distribution function. *Ann. Math. Statist.*, **33**, 525–532 (1962).
4. Cifarelli, D.M.: Intorno alla distribuzione del test di G. Girone. *Giornale degli Economisti e Annali di Economia*, **33**, 283–308 (1974).
5. Cifarelli, D.M.: Contributi intorno ad un test per l'omogeneità tra due campioni. *Giornale degli Economisti e Annali di Economia*, **34**, 233–249 (1975).
6. Girone, G.: Su un indice di omogeneità di due distribuzioni del tipo dell'indice semplice di dissomiglianza. *Atti della XXIV Riunione Scientifica della SIS*. 53–78 (1964).
7. Gorla, M.: Some generalizations of Girone's test statistic. *Annali dell'Istituto di Statistica, Università di Bari*, 53–72 (1972).
8. Green, J.R., Hegazy, Y.A.S.: Powerful modified-EDF goodness-of-fit tests, *J. Amer. Statist. Assoc.*, **71**, 204–209 (1976).
9. Johnson, B. Killeen, T.: An explicit formula for the c.d.f. of the L1 norm of the Brownian bridge. *Ann. Probab.*, **11**, 807–808 (1983).
10. Latorre, G.: Confronto tra le funzioni di potenza per il problema di “due campioni”. *Contributi del Centro di Studi Statistici dell'Università, della Calabria*, 31–50 (1977).
11. Pyke, R.: The supremum and infimum of the Poisson process. *Ann. Math. Statist.*, **30**, 568–576 (1959).
12. Schmid, F., Trede, M.: A distribution free test for the two sample problem for general alternatives. *Comput. Stat. Data Anal.*, **20**, 409–419 (1995).
13. Schmid, F., Trede, M.: An L1-variant of the Cramer-von Mises test. *Stat. Probabil. Lett.*, **26**, 91–96 (1996).
14. Shepp, L.A.: On the integral of the absolute value of the pinned Wiener process. *Ann. Probab.*, **10**, 234–239 (1982).
15. Shepp, L.A.: Acknowledgement of priority. *Ann. Probab.*, **19**, 1397 (1991).
16. Takács, L.: On the distribution of the integral of the absolute value of the Brownian motion. *Ann. Appl. Probab.*, **3**, 186–197 (1993).