Universidade de Vigo

Non Parametric *k*-Sample Tests: Density Functions vs Distribution Functions

Pablo Martínez Camblor and Jacobo de Uña Álvarez

Report 08/09

Discussion Papers in Statistics and Operation Research

Departamento de Estatística e Investigación Operativa

Facultade de Ciencias Económicas e Empresariales
Lagoas-Marcosende, s/n · 36310 Vigo
Tfno.: +34 986 812440 - Fax: +34 986 812401
http://eioweb.uvigo.es/
E-mail: depc05@uvigo.es

Universidade de Vigo

Non Parametric *k*-Sample Tests: Density Functions vs Distribution Functions

Pablo Martínez Camblor and Jacobo de Uña Álvarez

Report 08/09

Discussion Papers in Statistics and Operation Research

Imprime: GAMESAL

Edita:

Universidade de Vigo
Facultade de CC. Económicas e Empresariales
Departamento de Estatística e Investigación Operativa
As Lagoas Marcosende, s/n 36310 Vigo
Tfno.: +34 986 812440

I.S.S.N: 1888-5756

Depósito Legal: VG 1402-2007

Non Parametric k-Sample Tests: Density Functions vs Distribution Functions

Pablo Martínez Camblor[a] and Jacobo de Uña-Álvarez[b]

[a] *CIBER de Epidemiología y SP, Subdirecci ón de Salud Pública de Gipuzkoa, Donostia, Spain.*
[b] *Departamento de Estadística e IO, Universidad de Vigo, Vigo, Spain.*

Summary. In this paper we introduce some tests for the comparison of k samples based on kernel density estimators (KDE), and we develope the *Double Minimum* method as a new and useful procedure for the crucial problem of bandwidth selection. We study, *via* Monte Carlo simulations, the statistical power of the proposed tests, as well as the impact of the smoothing degree and the performance of the *Double Minimum* algorithm. Finally, we compare the results of the tests based on the KDE to those of the traditional k-sample tests based on empirical distribution functions (EDF), and to other tests based on the likelihood ratio introduced in the recent literature. Two main conclusions are obtained. First, the proposed bandwidth selection method attain quasi-optimal results. Second, the simulations suggest that KDE-based tests are the most powerful when the underlying populations are different in shape.

Keywords: Bandwidth Selection, k-Sample Tests, Kernel Density Estimation, Empirical Distribution Function.

1. Introduction

A classical problem in statistical inference is testing the equality of k distributions from independent random samples, without any parametric (*e.g.* normal) assumption on the underlying populations. The most popular and most commonly used tests are those based on the empirical distribution function (EDF). Several generalizations of the traditional two-sample tests to the k-sample problem have being studied in the literature. Kiefer (1959) proposed an extension of the Kolmogorov-Smirnov and Cramér-von Mises tests, while Scholz and Stephens (1987) gave the generalization corresponding to the Anderson-Darling test. In these papers, the following tests statistics were considered:

$$K_S = \sup_{t\in\mathbb{R}} \sum_{i=1}^{k} n_i (F_{n_i}(t) - F_n(t))^2$$

$$C_M = \sum_{i=1}^{k} n_i \int (F_{n_i}(t) - F_n(t))^2 dF_n(t)$$

$$A_D = \sum_{i=1}^{k} n_i \int \frac{(F_{n_i}(t) - F_n(t))^2}{F_n(t)(1 - F_n(t))} dF_n(t)$$

where n_i, F_{n_i} are respectively the sample size and the EDF of the i-th sample ($1 \leq i \leq k$), and F_n is the EDF of the pooled sample with $n = n_1 + ... + n_k$ observations. More recently, Zhang and Wu (2007) proposed three new k-sample tests based on the likelihood ratio, defined explicitly as

$$
\begin{aligned}
Z_K =& \sup_{t\in\mathbb{R}} \sum_{i=1}^{k} n_i Z_{n,i}(t)\\
Z_A =& \sum_{i=1}^{k} n_i \int \frac{Z_{n,i}(t)}{F_n(t)(1-F_n(t))} dF_n(t)\\
Z_C =& \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} \log\left(\frac{n_i}{j-0.5}-1\right)\log\left(\frac{n}{R_{ij}-0.5}-1\right)
\end{aligned}
$$

where R_{ij} is the rank in the pooled sample of the j-th observation in the i-th sample and

$$Z_{n,i}(t) = F_{n_i}(t)\log\left(\frac{F_{n_i}(t)}{F_n(t)}\right) + (1-F_{n_i}(t))\log\left(\frac{1-F_{n_i}(t)}{1-F_n(t)}\right)$$

Simulations reported by the authors suggest that the likelihood ratio tests may be much more powerful than the classical k-sample tests. See also our independent results in Section 3.

Under the assumption that the involved distributions are absolutely continuous, k-sample tests based on the comparison of kernel density estimators (KDE) (Rosenblatt, 1956) may be introduced. This type of tests has been much less investigated in the literature. For the two-sample problem, KDE-based tests have been considered in Anderson et al. (1994), Louani (2000), and Cao and Van Keilegom (2006). Martínez-Camblor (2006) introduced an original idea which was, for the best of our knowledge, the first k-sample test comparing KDE. Explicitly, the proposed test statistic is

$$\mathcal{AC} = \int min\{f_{n_1},\ldots,f_{n_k}\}$$

where f_{n_i} denotes the KDE pertaining to the i-th sample, which can be regarded as a generalization of the L_1 distance between two kernel estimators to $k > 2$. Previous results of the authors suggest that the $\mathcal{AC}$ test may be more powerful than the EDF-based tests (Martínez-Camblor et al., 2008).

In this paper we introduce some other measures of distance among the f_{n_i}'s in order to investigate if the power of the $\mathcal{AC}$ test can be improved. We consider test statistics based on generalizations of the L_1, L_2 and L_∞ distances between two functions, in the spirit of Kiefer (1959). Explicitly, the performance of the following statistics will be investigated:

$$
\begin{aligned}
L_{k,1} =& \frac{1}{n}\sum_{i=1}^{k} n_i \int |f_{n_i}(t)-f_n(t)|dt\\
L_{k,2} =& \frac{1}{n}\sum_{i=1}^{k} n_i \int (f_{n_i}(t)-f_n(t))^2 dt\\
S_k =& \frac{1}{n}\sum_{i=1}^{k} n_i \sup_{t\in\mathbb{R}} |f_{n_i}(t)-f_n(t)|
\end{aligned}
$$

All these statistics (as well as $\mathcal{AC}$) need a smoothing parameter or bandwidth. Proper selection of the bandwidth is crucial in the performance of the tests; see the simulation results in our Section 2. Some proposals to bandwidth choice have been provided for the two-sample problem (see *e.g.* Cao and Van Keilegom, 2006), but no optimal, computationally feasible method is available so far. In Section 2, a new procedure for automatic bandwidth selection is proposed. This new method, referred here as *Double Minimum*, does not attain optimal results, but its computational cost is quite minor than that of other methods, leading to a similar statistical power otherwise. Comparison of the KDE-based tests (based on the *Double Minimum* bandwidth) to those based on EDF is provided in Section 3 *via* Monte Carlo simulation.

The asymptotic distribution of the L_∞ distance between a kernel density estimator and its target has been established by Konakov (1978). This result has been adapted to the goodness-of-fit problem by Liero et al. (1998). Martínez-Camblor et al. (2008) provide the asymptotic normality of the $\mathcal{AC}$ test statistic. In the Appendix of this paper we give the results corresponding to the test statistics based on the L_p norm.

2. Tests based on kernel density estimators

In this Section we introduce the KDE-based tests and we investigate the impact of the bandwidth choice in their performance. We also introduce a new automatic bandwidth selector so the optimal smoothing degree can be estimated from the available data.

The kernel-type estimators (Rosenblatt, 1956; Parzen, 1962) are probably the most popular and most commonly used estimators for density functions. Given a random sample $X = \{x_1, ..., x_n\}$ from a density f, the kernel density estimator is defined as

$$f_n(X,t) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_i - t}{h}\right) \tag{1}$$

where K is a kernel function which is usually taken as a density symmetrical about zero, with finite variance, and where $h = h_n$ is a sequence os positive numbers (the smoothing parameters or bandwidths) converging to zero as n increases. This estimator has been extensively investigated and it is the origin of several methods referred as *smoothing techniques*. The asymptotic normality of the L_p norm of $f_n - f$ has been established by Hórvath (1991) and also by Martínez-Camblor and Corral (2008) under weaker conditions. Case $p = 1$ has been studied by Devroye and Györfy (1985).

The kernel density estimator has been used in a variety of inference problems since the eighties. Silverman (1981) proposed a test of multimodality, subsequently calibrated by Hall and York (2001). Ahmad and Li (1997) and Diks and Tong (1999) proposed a test for symmetry in the univariate and multivariate cases, respectively. Goodness-of-fit one sample tests based on KDE include (among many others) Ghosh and Huang (1991), Fan (1994, 1998), and Liero et al. (1998). The two-sample problem has been less studied; some existing KDE-based tests include Anderson et al. (1994), Li (1996, 1999) and Cao and Van Keilegom (2006). Martínez-Camblor (2006) proposed a test for the comparison of k densities given independent random samples, based on the common area of the pertaining kernel estimators (the $\mathcal{AC}$ statistic). Martínez-Camblor, de Uña-álvarez and Corral (2008) derived a Cramér-Chernoff type theorem for such a test, and they carried out a comparative study of the attained rejection levels, including EDF-based tests in the comparison. The problem of comparing dependent samples was similarly investigated in Martínez-Camblor (2008). These recent contributions suggest that, for the balanced design and small sample sizes, the $\mathcal{AC}$ test is more powerful than EDF-based tests, provided that the underlying distributions are different not only in location, that is, no distribution stochastically dominates the other.

As always when applying smoothing techniques, one of the main concerns is the selection of the smoothing parameter or bandwidth. It is well known that the bandwidth that minimizes the mean integrated squared error (MISE) of the kernel density estimator is of the form $cn^{-1/5}$. There exists a vast literature on different methods to compute the value of the constant c (see for example Park and Marron, 1990; Devroye, 1997; or recently Ahmad and Amezziane, 2007). However, these methods are far from being optimal when testing for the equality of distributions. Eggermont and LaRiccia (2003) dealt with this topic in goodness-of-fit problems. Cao and Van Keilegom (2006) proposed a procedure, called *the Double Bootstrap*, based on the selection of the bandwidth (on a given grid) leading to a larger distance between the null and the alternative hypotheses (averaging the bandwidths obtained when repeating this process). The distribution of the test statistic is then approximated from a suitable resampling plan. In this paper we introduce a related but different method for bandwidth choice, which is called (as an analogy) *Double Minimum*. The idea behind our proposal (which is described in detail below) is to choose a bandwidth of the form $S\widehat{\sigma}n^{-1/5}$ where $\widehat{\sigma}$ is the sample standard deviation and S is the value (on a given grid) attached to the smallest P–value. The distribution of this minimum P–value is then approximated by bootstrapping. The *Double Minimum* procedure, like other existing methods for this problem, does not get an optimal solution, but it attains very good results while drastically reducing the computational cost.

The rejection levels of the four k-sample KDE-based tests statistics $L_{k,1}$, $L_{k,2}$, S_k and $\mathcal{AC}$ proposed in this paper, and the impact of the bandwidth choice, were investigated in a Monte Carlo study. Two independent samples with sizes n_1 and n_2 were generated from a standard normal (symmetrical case) or from a chi-squared with 3 degrees of freedom χ^2_3 (asymmetrical case), and a third sample (with size n_3) was generated from a third model. In the symmetrical case, the third sample was generated according to any of the models:

MD 0-I: $Z \equiv N(0,1)$ (Null hypothesis)
MD 1-I: $Z \equiv (1-a)N(0,1) + aN(0,2)$
MD 2-I: $Z \equiv (1-a)N(0,1) + aN(1,1)$
MD 3-I: $Z \equiv a/2N(-2,1/2) + (1-a)N(0,1) + a/2N(2,1)$
MD 4-I: $Z \equiv (1-a)N(0,1) + a\chi^2_3$

where $a = 6n^{-1/2}$, $n = n_1 + n_2 + n_3$. Note that the alternative hypotheses (models 1-I to 4-I) go to the null as the sample size increases. In Figure 1 these alternatives are depicted, for the cases $n = 45$ and $n = 150$. It becomes clear from this Figure 1 that alternatives in models 1-I and 3-I differ from the standard normal in shape (but not in location), while models 2-I and 4-I include densities with different location (a and $3a$ respectively) and shape (clearly seen for model 4-I). In the asymmetrical case, we considered the following models for the alternative (a being as above):

MD 0-II: $Z \equiv \chi^2_3$ (Null hypothesis)
MD 1-II: $Z \equiv (1-a)\chi^2_3 + aN(3,1)$

Fig. 1. Graphical representation for the type I models when the sum of sample sizes is 45 (thick lines) and 150 (thin lines). In dotted lines $a = 0$.

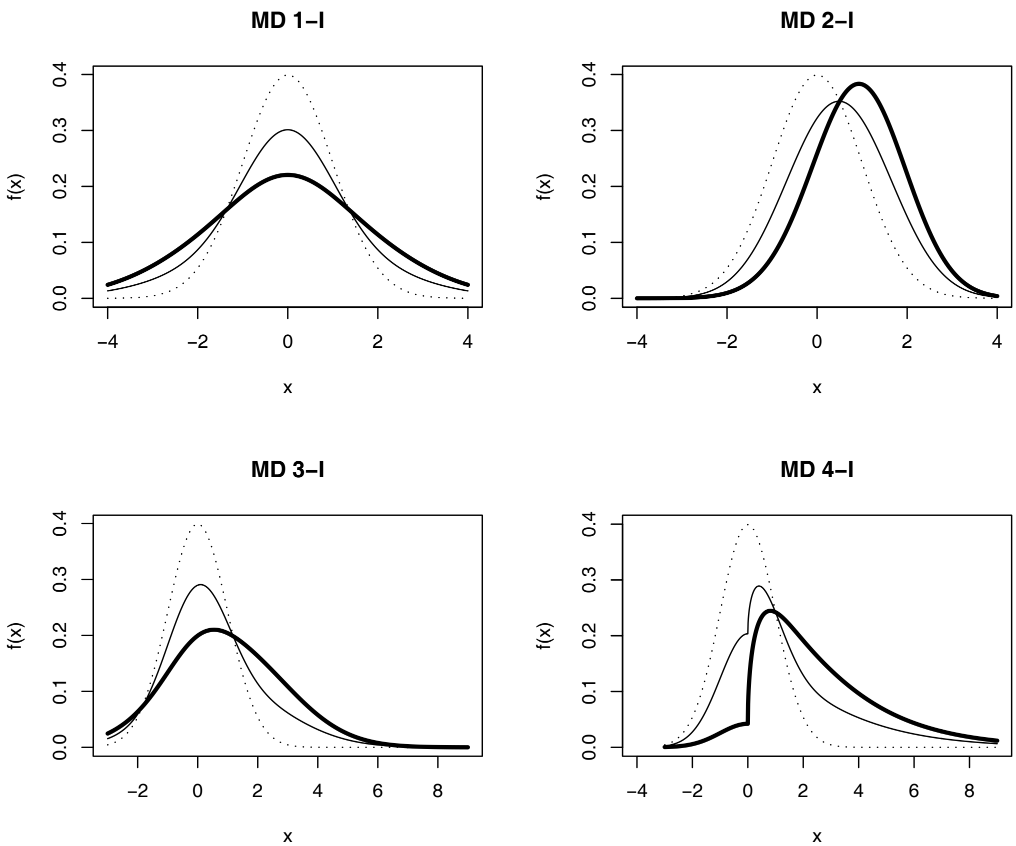
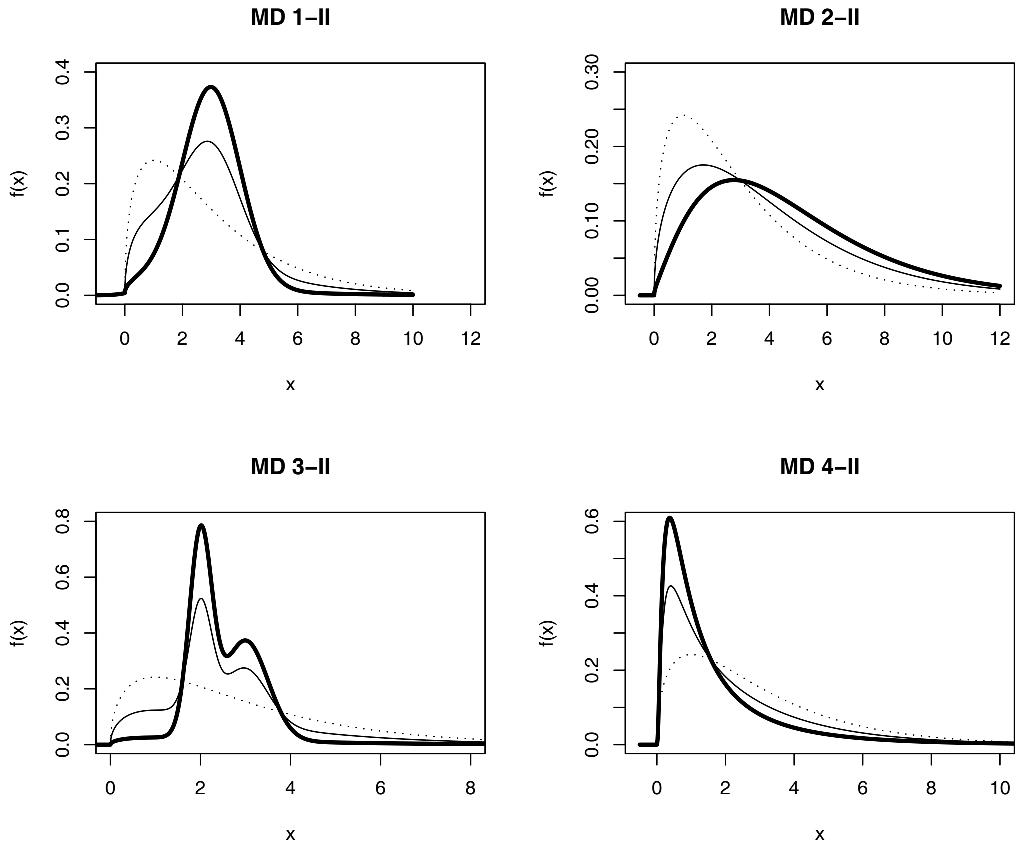


Fig. 2. Graphical representation for the type II. Again, the sum of sample sizes is 45 (thick lines) and 150 (thin lines). In dotted lines a $= 0$.



MD 2-II: $Z \equiv (1-a)\chi^2_3 + a\chi^2_5$
MD 3-II: $Z \equiv a/2N(2,1/4) + (1-a)\chi^2_3 + a/2N(3,1/2)$
MD 4-II: $Z \equiv (1-a)\chi^2_3 + aLN(2,1)\dagger$

Models 1-II to 4-II, depicted in Figure 2 again for $n = 45$ and $n = 150$, involve densities which are different in location (models 2-II to 4-II) or in shape (clearly seen for model 3-II but also for models 1-II and 4-II).

As announced above, the kernel density estimators f_{n_i}, $i = 1,2,3$, used in the computation of the four statistics $L_{k,1}$, $L_{k,2}$, S_k and $\mathcal{AC}$ were based on a bandwidth of the form $h_{n_i} = S\widehat{\sigma}_{n_i} n_i^{-1/5}$, where $\widehat{\sigma}_{n_i}$ stands for the standard deviation in the i-th sample, and where S varies on a certain grid of values. In the special case of the $\mathcal{AC}$ test statistic, we rather took $\widetilde{h}_{n_i} = S\widehat{\sigma}_{n_i} m^{-1/5}$, m being the harmonic mean of the sample sizes. Note that

†Log-normal distribution with parameters two and one.

$\widetilde{h}_{n_i} = h_{n_i}$ for equal sample sizes, but this is not true for unbalanced designs. In the latter case, the using of $\widetilde{h}_{n_i}$ seems to improve the power of the test (see Martínez-Camblor et al., 2008). The Gaussian kernel was used in all the cases. For the approximation of the critical value of the test statistics at level α, we used the smoothed bootstrap (Hall et al., 1989) in the following way:

A. Compute the smoothed EDF (Nadaraya, 1964) from the pooled sample.
B. Generate B bootstrap resamples of sizes n_i ($i = 1, 2, 3$) from the (common) smoothed EDF.
C. Take the $1 - \alpha$ percentile along the bootstrap resamples as the critical value.

For the smoothed EDF, the optimal bandwidth for the estimation of a smooth distribution function in the sense of the MISE was taken, being of the form $g = Cn^{-1/3}$ (Azzalini, 1981). For simplicity, we took $C = 1$. The number of bootstrap resamples was $B = 199$, and the number of replicates was 1000. In Tables 1 and 2 (symmetrical case and asymmetrical case, respectively) we provide the rejection levels attained by the four test statistics $L_{k,1}$, $L_{k,2}$, S_k and $\mathcal{AC}$ when $\alpha = 0.05$. The grid of values for S was $\{1/2, 1, 2, 4, 8\}$, and four different combinations of sample sizes (n_1, n_2, n_3) were considered: $(15, 15, 15)$, $(15, 30, 45)$, $(25, 25, 25)$, and $(25, 50, 75)$. Note that the first and the last cases correspond to total sample sizes of $n = 45$ and $n = 150$, which lead to the extreme alternatives illustrated in Figures 1 and 2.

Table 1 shows that in most of the cases the percentage of rejection under the null was less than the nominal level. Also, it becomes clear from this Table 1 that the power of the tests is strongly affected by the value of S, which determines the level of smoothing in the density estimators. Interestingly, the value of S leading to the greatest power depends on the test statistic, being less influenced by the sample sizes otherwise. As expected, the rejection levels increased with the sample size (note however that the alternatives go to the null as n goes to infinity). Globally, the best tests in the simulations (when based on their respective optimal smoothing levels) was $L_{k,1}$. However, this is not neccessarily true when focusing on a particular model and a combination of sample sizes. Finally, we mention that the alternative densities in models 3-I and 4-I were detected much more frequently than those in models 1-I and 2-I, a fact that could be expected from the graphical comparison of models in Figure 1. Simulations reported in Table 2 (asymmetrical case) suggest that the tests may be a bit conservative when oversmoothing (see however the results depicted in Tables 3 and 4 for a data-driven bandwidth). Further simulations (not shown here) indicate that this problem may be avoided through a more careful choice of the pilot bandwidth g. From Table 2 we see that the test based on $L_{k,1}$ was the clear winner overall. In particular situations, $L_{k,1}$ was outperformed by a different test statistic; for example, for model 2-II and equal sample sizes, the $\mathcal{AC}$ statistic attained a better power.

The simulation above suggests that the bandwidth choice plays a very important role in the power of the KDE-based tests. Therefore, there is a strong demmand of automatic smoothing methods which can lead to optimal powers, avoiding the problem of the subjective choice of the bandwidth when testing for significance. As mentioned, in this paper we propose an automatic bandwidth selection procedure, the *double minimum*, which (for a given test statistic) is defined as follows:

Table 1. Rejection proportion observed in tests based on $L_{k,1}$, $L_{k,2}$, S_k and $\mathcal{AC}$ for type I models and $n=(n_1,n_2,n_3)$ and $S \in \{1/2, 1, 2, 4, 8\}$.

		n=(15,15,15)					**n**=(15,30,45)				
		1/2	**1**	**2**	**4**	**8**	**1/2**	**1**	**2**	**4**	**8**
$L_{k,1}$	0-I	0.050	0.046	0.040	0.041	0.043	0.061	0.048	0.054	0.041	0.043
$L_{k,2}$	0-I	0.039	0.044	0.038	0.039	0.033	0.049	0.050	0.036	0.034	0.022
S_k	0-I	0.035	0.037	0.035	0.034	0.041	0.038	0.034	0.045	0.034	0.026
$\mathcal{AC}$	0-I	0.048	0.045	0.043	0.036	0.041	0.047	0.044	0.041	0.042	0.041
$L_{k,1}$	1-I	0.389	0.463	0.476	0.403	0.351	0.452	0.593	0.657	0.593	0.531
$L_{k,2}$	1-I	0.302	0.382	0.452	0.443	0.437	0.368	0.480	0.569	0.479	0.400
S_k	1-I	0.305	0.388	0.425	0.376	0.338	0.402	0.553	0.643	0.585	0.515
$\mathcal{AC}$	1-I	0.310	0.445	0.497	0.525	0.507	0.216	0.381	0.527	0.548	0.557
$L_{k,1}$	2-I	0.429	0.497	0.393	0.115	0.047	0.496	0.579	0.549	0.237	0.064
$L_{k,2}$	2-I	0.355	0.435	0.392	0.153	0.058	0.391	0.484	0.457	0.127	0.028
S_k	2-I	0.223	0.291	0.227	0.097	0.066	0.259	0.389	0.369	0.142	0.043
$\mathcal{AC}$	2-I	0.391	0.468	0.399	0.139	0.067	0.295	0.369	0.309	0.149	0.092
$L_{k,1}$	3-I	0.907	0.900	0.843	0.750	0.679	0.963	0.976	0.963	0.921	0.875
$L_{k,2}$	3-I	0.841	0.847	0.827	0.822	0.826	0.927	0.953	0.941	0.854	0.750
S_k	3-I	0.749	0.828	0.798	0.725	0.638	0.861	0.951	0.970	0.925	0.820
$\mathcal{AC}$	3-I	0.888	0.889	0.873	0.891	0.900	0.808	0.873	0.915	0.938	0.946
$L_{k,1}$	4-I	0.949	0.961	0.928	0.820	0.713	0.991	0.995	0.990	0.969	0.925
$L_{k,2}$	4-I	0.860	0.913	0.891	0.801	0.714	0.960	0.986	0.987	0.970	0.926
S_k	4-I	0.757	0.824	0.846	0.793	0.733	0.924	0.967	0.991	0.971	0.950
$\mathcal{AC}$	4-I	0.904	0.932	0.842	0.599	0.424	0.880	0.956	0.914	0.843	0.774

		n=(25,25,25)					**n**=(25,50,75)				
		1/2	**1**	**2**	**4**	**8**	**1/2**	**1**	**2**	**4**	**8**
$L_{k,1}$	0-I	0.059	0.049	0.037	0.035	0.035	0.049	0.049	0.048	0.055	0.056
$L_{k,2}$	0-I	0.049	0.050	0.036	0.034	0.022	0.050	0.051	0.053	0.051	0.039
S_k	0-I	0.043	0.043	0.033	0.027	0.018	0.038	0.054	0.041	0.041	0.013
$\mathcal{AC}$	0-I	0.051	0.045	0.037	0.038	0.034	0.056	0.053	0.048	0.061	0.059
$L_{k,1}$	1-I	0.352	0.481	0.515	0.446	0.361	0.397	0.584	0.684	0.625	0.525
$L_{k,2}$	1-I	0.346	0.484	0.632	0.675	0.663	0.407	0.538	0.634	0.502	0.350
S_k	1-I	0.412	0.528	0.596	0.507	0.438	0.482	0.629	0.772	0.741	0.544
$\mathcal{AC}$	1-I	0.267	0.448	0.621	0.649	0.653	0.244	0.370	0.576	0.614	0.620
$L_{k,1}$	2-I	0.376	0.446	0.385	0.122	0.049	0.521	0.646	0.618	0.340	0.038
$L_{k,2}$	2-I	0.432	0.492	0.479	0.198	0.075	0.493	0.620	0.618	0.172	0.026
S_k	2-I	0.350	0.429	0.383	0.161	0.097	0.432	0.578	0.622	0.441	0.107
$\mathcal{AC}$	2-I	0.335	0.441	0.409	0.182	0.076	0.302	0.392	0.394	0.207	0.111
$L_{k,1}$	3-I	0.881	0.919	0.872	0.777	0.666	0.956	0.976	0.972	0.923	0.776
$L_{k,2}$	3-I	0.823	0.891	0.912	0.905	0.889	0.899	0.953	0.955	0.837	0.614
S_k	3-I	0.695	0.862	0.866	0.728	0.562	0.822	0.940	0.974	0.921	0.677
$\mathcal{AC}$	3-I	0.854	0.921	0.931	0.943	0.944	0.793	0.906	0.953	0.969	0.969
$L_{k,1}$	4-I	0.959	0.977	0.973	0.911	0.849	0.992	0.996	0.999	0.989	0.958
$L_{k,2}$	4-I	0.874	0.951	0.962	0.917	0.865	0.952	0.989	0.995	0.978	0.929
S_k	4-I	0.787	0.898	0.928	0.905	0.874	0.914	0.978	0.994	0.990	0.970
$\mathcal{AC}$	4-I	0.906	0.967	0.952	0.871	0.803	0.862	0.973	0.981	0.945	0.927

Table 2. Rejection proportion observed in tests based on $L_{k,1}$, $L_{k,2}$, S_k and $\mathcal{AC}$ for type II models and $n = (n_1, n_2, n_3)$ and $S \in \{1/2, 1, 2, 4, 8\}$.

		$\mathbf{n}$=(15,15,15)					$\mathbf{n}$=(15,30,45)				
		1/2	**1**	**2**	**4**	**8**	**1/2**	**1**	**2**	**4**	**8**
$L_{k,1}$	0-II	0.054	0.051	0.038	0.042	0.040	0.057	0.054	0.043	0.038	0.038
$L_{k,2}$	0-II	0.042	0.044	0.034	0.033	0.036	0.030	0.043	0.040	0.043	0.032
S_k	0-II	0.027	0.033	0.032	0.029	0.031	0.027	0.044	0.037	0.035	0.032
$\mathcal{AC}$	0-II	0.058	0.048	0.037	0.038	0.034	0.042	0.054	0.047	0.043	0.042
$L_{k,1}$	1-II	0.490	0.517	0.489	0.455	0.451	0.611	0.625	0.513	0.434	0.422
$L_{k,2}$	1-II	0.468	0.519	0.489	0.471	0.445	0.418	0.500	0.469	0.431	0.436
S_k	1-II	0.324	0.417	0.454	0.444	0.432	0.355	0.480	0.479	0.411	0.395
$\mathcal{AC}$	1-II	0.459	0.498	0.468	0.460	0.467	0.309	0.341	0.321	0.319	0.323
$L_{k,1}$	2-II	0.296	0.265	0.198	0.122	0.076	0.350	0.369	0.279	0.150	0.092
$L_{k,2}$	2-II	0.220	0.224	0.176	0.125	0.100	0.268	0.313	0.237	0.116	0.074
S_k	2-II	0.147	0.162	0.131	0.097	0.086	0.268	0.314	0.220	0.133	0.094
$\mathcal{AC}$	2-II	0.297	0.289	0.192	0.110	0.088	0.232	0.253	0.175	0.135	0.116
$L_{k,1}$	3-II	0.809	0.765	0.715	0.697	0.698	0.832	0.739	0.559	0.465	0.448
$L_{k,2}$	3-II	0.808	0.788	0.746	0.725	0.725	0.704	0.664	0.510	0.429	0.434
S_k	3-II	0.688	0.734	0.737	0.716	0.709	0.665	0.694	0.537	0.446	0.418
$\mathcal{AC}$	3-II	0.786	0.749	0.721	0.725	0.721	0.483	0.392	0.253	0.206	0.203
$L_{k,1}$	4-II	0.838	0.836	0.820	0.786	0.772	0.925	0.924	0.906	0.857	0.846
$L_{k,2}$	4-II	0.696	0.741	0.701	0.686	0.665	0.899	0.925	0.909	0.893	0.886
S_k	4-II	0.665	0.741	0.737	0.687	0.685	0.919	0.941	0.924	0.901	0.894
$\mathcal{AC}$	4-II	0.768	0.729	0.585	0.495	0.466	0.806	0.785	0.680	0.636	0.629

		$\mathbf{n}$=(25,25,25)					$\mathbf{n}$=(25,50,75)				
		1/2	**1**	**2**	**4**	**8**	**1/2**	**1**	**2**	**4**	**8**
$L_{k,1}$	0-II	0.061	0.050	0.041	0.040	0.039	0.051	0.045	0.041	0.042	0.034
$L_{k,2}$	0-II	0.041	0.045	0.042	0.043	0.040	0.044	0.044	0.041	0.035	0.033
S_k	0-II	0.035	0.037	0.033	0.033	0.035	0.041	0.031	0.041	0.036	0.027
$\mathcal{AC}$	0-II	0.051	0.057	0.045	0.041	0.042	0.043	0.047	0.045	0.043	0.047
$L_{k,1}$	1-II	0.468	0.481	0.408	0.363	0.348	0.608	0.627	0.481	0.385	0.368
$L_{k,2}$	1-II	0.441	0.467	0.409	0.360	0.351	0.471	0.537	0.469	0.410	0.383
S_k	1-II	0.310	0.393	0.396	0.360	0.345	0.386	0.522	0.476	0.360	0.318
$\mathcal{AC}$	1-II	0.434	0.450	0.366	0.349	0.342	0.303	0.283	0.205	0.148	0.144
$L_{k,1}$	2-II	0.286	0.303	0.217	0.133	0.094	0.344	0.390	0.317	0.162	0.120
$L_{k,2}$	2-II	0.237	0.278	0.228	0.136	0.094	0.290	0.332	0.249	0.106	0.061
S_k	2-II	0.185	0.224	0.163	0.119	0.093	0.302	0.341	0.288	0.148	0.108
$\mathcal{AC}$	2-II	0.308	0.332	0.245	0.153	0.127	0.245	0.291	0.230	0.171	0.144
$L_{k,1}$	3-II	0.716	0.639	0.518	0.465	0.465	0.804	0.709	0.514	0.409	0.375
$L_{k,2}$	3-II	0.702	0.648	0.525	0.460	0.466	0.732	0.677	0.515	0.413	0.383
S_k	3-II	0.586	0.608	0.522	0.462	0.446	0.675	0.693	0.515	0.374	0.339
$\mathcal{AC}$	3-II	0.647	0.590	0.504	0.457	0.451	0.503	0.488	0.334	0.279	0.266
$L_{k,1}$	4-II	0.894	0.918	0.900	0.883	0.868	0.944	0.957	0.946	0.926	0.914
$L_{k,2}$	4-II	0.820	0.879	0.859	0.837	0.824	0.930	0.960	0.963	0.957	0.955
S_k	4-II	0.807	0.886	0.870	0.860	0.848	0.949	0.975	0.980	0.968	0.964
$\mathcal{AC}$	4-II	0.853	0.855	0.788	0.710	0.687	0.846	0.855	0.809	0.776	0.757

Table 3. Observed rejection proportions in tests based on $L_{k,1}$ and $\mathcal{AC}$ in type I models with $n = (n_1, n_2, n_3)$. We used *Double Minimum* procedure with the grid $\{1/2, 1, 2, 4, 8, 12\}$.

	n=(15,15,15)					**n**=(15,30,45)				
	0-I	**1-I**	**2-I**	**3-I**	**4-I**	**0-I**	**1-I**	**2-I**	**3-I**	**4-I**
$L_{k,1}$	0.054	0.402	0.440	0.890	0.960	0.066	0.584	0.582	0.972	0.988
$\mathcal{AC}$	0.048	0.586	0.424	0.948	0.926	0.062	0.616	0.332	0.944	0.926
	n=(25,25,25)					**n**=(25,50,75)				
	0-I	**1-I**	**2-I**	**3-I**	**4-I**	**0-I**	**1-I**	**2-I**	**3-I**	**4-I**
$L_{k,1}$	0.070	0.466	0.426	0.898	0.982	0.056	0.592	0.584	0.992	0.998
$\mathcal{AC}$	0.058	0.628	0.460	0.926	0.972	0.056	0.556	0.360	0.864	0.982

A_1 Let be $\{S_1, \ldots, S_t\}$ a grid of S-values among which this parameter is to be selected.
B_1 By using the bootstrap resampling plan above (steps A, B, C), compute the P-value of the test for each S_i, p_i $(1 \leq i \leq t)$.
C_1 Choose the S_M which minimizes the sequence of P-values, i.e., $p_M = min\{p_1, \ldots, p_t\}$.
D_1 Generate B' (here we take $B' = B$) bootstrap resamples as in B_1, and compute as in C_1 the respective minimum P-values; $p_M^1, \ldots, p_M^B$.
E_1 Compute the final *double minimum* P-value,

$$p_F = \frac{1}{B} \sum_{i=1}^{B} I\{p_M > p_M^i\}$$

Then, one rejects the null hypothesis of equal densities at level α whenever $p_F < \alpha$. This procedure, unlike the *double bootstrap* in Cao and Van Keilegom (2006), does not require the estimation of the distribution of the test under both the null and the alternative hypotheses. Therefore, its computational cost, while being high, decreases substantially. The computational speed of the *double minimum* can be improved by the inclusion of the stopping rule: *If* $p_M > \alpha$*, STOP*. Then, the algorith ends after step C_1.

Tables 3 and 4 (for models of type I and II respectively) show the rejection percentages attained by the $L_{k,1}$ and $\mathcal{AC}$ test statistics for a 5% of significance level, when using the double minimum along 500 replicates. The number of bootstrap resamples for the approximation of the null distribution of the test and the computation of p_F was $B = 100$. These Tables suggest that both tests may be a bit anticonservative (more celarly seen for $L_{k,1}$). Regarding their power, we see that the attained results were close to those corresponding to the optimal S-value (see Tables 1 and 2). The worst case was for sample sizes $(25, 50, 75)$ and model 3-I, for which the rejection level of the AC statistic decreased a 10%. On the other hand, we point out that for type II models (Table 4), in most of the cases the rejection levels with the double minimum bandwidth were even better than those in Tables 1 and 2. Hence, the described algorithm for automatic bandwidth selection performed satisfactorily in the simulations.

Table 4. Observed rejection proportions in tests based on $L_{k,1}$ and $\mathcal{AC}$ in type II models with $n = (n_1, n_2, n_3)$. We used *Double Minimum* procedure with the grid $\{1/2, 1, 2, 4, 8, 12\}$.

	n=(15,15,15)					**n**=(15,30,45)				
	0-II	**1-II**	**2-II**	**3-II**	**4-II**	**0-II**	**1-II**	**2-II**	**3-II**	**4-II**
$L_{k,1}$	0.062	0.564	0.314	0.824	0.900	0.056	0.672	0.440	0.824	0.862
$\mathcal{AC}$	0.054	0.560	0.334	0.814	0.832	0.050	0.356	0.312	0.478	0.966
	n=(25,25,25)					**n**=(25,50,75)				
	0-II	**1-II**	**2-II**	**3-II**	**4-II**	**0-II**	**1-II**	**2-II**	**3-II**	**4-II**
$L_{k,1}$	0.060	0.516	0.302	0.752	0.918	0.062	0.624	0.408	0.784	0.922
$\mathcal{AC}$	0.044	0.478	0.340	0.698	0.866	0.058	0.326	0.264	0.444	0.900

3. Density functions vs distribution functions

In this Section we compare via Monte Carlo simulations the performance of the KDE-based tests to that of the tests based on empirical distribution functions (EDF). As EDF-based tests, we considered the generalizations of the Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling tests to the k-sample problem (denoted here as K_S, C_M, and A_D respectively) as discussed in Kiefer (1959) and Scholz and Stephens (1987). Also, three different likelihood ratio tests (Z_K, Z_A, Z_C) proposed by Zhang and Wu (2007) were considered. Due to its popularity, we included in the simulations the Kruskal-Wallis test (K_W) too. The comparison is organized in a fair way so, for each simulated model and for each combination of sample sizes, the optimal KDE-based test is compared to the EDF-test with the best expected performance. The relative behaviour of the referred EDF-based tests was investigated by Zhang and Wu (2007), who concluded that the likelihood ratio tests are superior to the classical k-sample tests. In Tables 5 and 6, we report the results attained by these EDF-based test statistics in the models introduced in Section 2. In order to avoid the high computational cost associated to the evaluation of the null distribution of these tests (Zhang and Wu, 2007), this was approximated via Monte Carlo simulation of the ranks. To this end, we took 199 replicates. Rejection levels for $\alpha = 0.05$ were computed on the basis of 1000 replicates.

From Tables 5 and 6, we see that the most powerful tests were the likelihood ratio tests. An exception to this was model 2-I and also model 2-II, for which Kruskal-Wallis test was the winner. Note that the performance of the Kruskal-Wallis test was also good for models 4-I and 4-II, which also involve a change in the location of the densities. However, the statistic K_W was completely missleading in the remaining situations, for which no differences in location are found. On the other hand, the tests K_S, C_M, and A_D performed poorly in most of the cases, even giving rejection levels below α in models 2-I and 2-II (unequal sample sizes). An exception is seen for models 1-II and 3-II in the case $(25, 25, 25)$, for which K_S achieves the largest rejection levels among all the tests. Among the likelihood ratio test, Z_A was the clear winner for models in Table 5, the statistic Z_C attaining similar results when the sample sizes are equal. Both tests are fairly competitive also for the models in

Table 5. Rejection proportion observed by the seven tests based on EDF in type I models.

	n=(15,15,15)					n=(15,30,45)				
	0-I	**1-I**	**2-I**	**3-I**	**4-I**	**0-I**	**1-I**	**2-I**	**3-I**	**4-I**
K_S	0.063	0.071	0.601	0.205	0.961	0.052	0.220	0.136	0.663	0.232
C_M	0.049	0.081	0.457	0.223	0.901	0.050	0.192	0.066	0.609	0.141
A_D	0.057	0.070	0.459	0.194	0.390	0.052	0.188	0.065	0.592	0.137
K_W	0.047	0.057	0.655	0.068	0.996	0.038	0.052	0.687	0.040	0.991
Z_K	0.050	0.196	0.537	0.651	0.993	0.043	0.248	0.581	0.799	0.998
Z_A	0.039	0.212	0.624	0.745	0.997	0.036	0.260	0.671	0.818	0.998
Z_C	0.043	0.195	0.627	0.728	0.997	0.037	0.202	0.653	0.797	0.996
	n=(25,25,25)					n=(25,50,75)				
	0-I	**1-I**	**2-I**	**3-I**	**4-I**	**0-I**	**1-I**	**2-I**	**3-I**	**4-I**
K_S	0.052	0.067	0.410	0.322	0.630	0.059	0.286	0.053	0.669	0.099
C_M	0.047	0.078	0.342	0.317	0.589	0.057	0.249	0.031	0.657	0.055
A_D	0.051	0.065	0.333	0.278	0.410	0.060	0.227	0.030	0.606	0.055
K_W	0.051	0.046	0.640	0.062	0.976	0.038	0.039	0.674	0.044	0.983
Z_K	0.048	0.243	0.517	0.747	0.979	0.051	0.291	0.566	0.852	1.000
Z_A	0.054	0.324	0.606	0.807	0.987	0.043	0.369	0.652	0.912	1.000
Z_C	0.055	0.290	0.626	0.818	0.988	0.040	0.285	0.645	0.896	0.999

Table 6. Rejection proportion observed by the seven tests based on EDF in type II models.

	n=(15,15,15)					n=(15,30,45)				
	0-I	**1-II**	**2-II**	**3-II**	**4-II**	**0-II**	**1-II**	**2-II**	**3-II**	**4-II**
K_S	0.049	0.419	0.481	0.360	0.666	0.046	0.145	0.123	0.238	0.152
C_M	0.048	0.239	0.347	0.105	0.552	0.041	0.068	0.055	0.124	0.072
A_D	0.050	0.319	0.403	0.212	0.500	0.042	0.080	0.061	0.131	0.068
K_W	0.057	0.095	0.444	0.079	0.843	0.047	0.128	0.492	0.093	0.876
Z_K	0.057	0.257	0.350	0.443	0.816	0.045	0.308	0.383	0.472	0.908
Z_A	0.048	0.249	0.410	0.444	0.884	0.042	0.344	0.423	0.478	0.933
Z_C	0.054	0.253	0.414	0.410	0.886	0.047	0.361	0.429	0.506	0.924
	n=(25,25,25)					n=(25,50,75)				
	0-II	**1-II**	**2-II**	**3-II**	**4-II**	**0-II**	**1-II**	**2-II**	**3-II**	**4-II**
K_S	0.063	0.384	0.344	0.576	0.433	0.048	0.077	0.067	0.098	0.080
C_M	0.055	0.318	0.275	0.452	0.372	0.047	0.049	0.040	0.052	0.037
A_D	0.052	0.313	0.283	0.462	0.346	0.050	0.039	0.045	0.047	0.040
K_W	0.059	0.094	0.450	0.049	0.801	0.042	0.123	0.497	0.068	0.858
Z_K	0.052	0.248	0.340	0.376	0.831	0.044	0.260	0.368	0.412	0.936
Z_A	0.053	0.242	0.424	0.370	0.891	0.046	0.295	0.406	0.382	0.954
Z_C	0.049	0.241	0.430	0.373	0.885	0.049	0.310	0.410	0.426	0.951

Table 6, the statistic Z_C showing a slighly better performance in this case.

A quick inspection of the rejection levels attained by the KDE-based (Tables 3 and 4) and by the EDF-based tests (Tables 5 and 6) suggests that the latter are more powerful when the underlying populations have different locations, while the former perform better when introducing other types of departures from the null (that is, no distribution stochastically dominates the other ones). In particular, tests based on EDFs are the clear winners in models 2-I and 2-II, but the tests based on kernel density estimators outperform the EDF-based ones in models 1-I, 1-II, 3-I and 3-II. For models 4-I and 4-II we can not reach a general conclusion, since there is a big influence of the sample sizes in the relative performance of the tests.

In order to formalize things, we conducted a new simulation study in which, as announced, for each model in Section 2, the best KDE-based test was compared to the best test in Tables 5 or 6. It was not always obvious to make a decision about the participants in the comparison, since sometimes the results were strongly driven by the sample sizes combination (mainly for the assymetrical models). Specifically, in models 1-I, 3-I and 4-I, $\mathcal{AC}$ and $L_{k,1}$ were compared to Z_A and Z_C; in model 2-I, $L_{k,1}$ and $L_{k,2}$ to Z_A and K_W; in models 1-II, 2-II and 3-II, again $L_{k,1}$ and $L_{k,2}$ to Z_C and K_W (models 1-II and 3-II) or to K_S and Z_C (model 2-II); finally, in model 4-II, $L_{k,1}$ and S_k were compare to Z_C and K_W. The authomatic bandwidth selector discussed in Section 2 was used for the KDE-based tests. The percentage of rejections at a 5% nominal significance level was computed for an increasing total sample size $n = n_1+n_2+n_3$ (ranging from 15 to 105), where the sample sizes combinations (n_1, n_2, n_3) (ordered w.r.t. n) were: (5,5,5), (5,5,10), (5,10,15), (15,15,15), (15,15,20), (15,20,25), (25,25,25), (25,25,30), (25,30,35) and (35,35,35). The parameter a was always taken to be $6/\sqrt{45}$, independently of the sample size. Results based on 100 trial are depicted in Figures 3 and 4. From these Figures we see that the overall winner is the test based on $L_{k,1}$; this was by far the best test in three among eight simulated models (models 3-I, 1-II, 3-II), while in other three models (1-I, 4-I, 4-II) it was almost as good as the best one. The test $L_{k,1}$ was outperformed by Z_C in models 2-I and 2-II; the likelihood-ratio test Z_C also showed a good behaviour in models 4-I and 4-II, but was clearly inferior to the best KDE-based test in four of the eight simulated models (models 1-I, 3-I, 1-II and 3-II). The other likelihood-ratio test Z_A behaved similarly to Z_C (when considered). These results are in agreement with the independent simulations summarized in Tables 3 to 6.

4. Main conclussions

In this paper we have investigated the power of several nonparametric k-sample tests. Specifically, we have considered the extensions of the classical Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling test to the k-sample problem, as well as other tests based on the lilkelihood ratio which have been recently introduced in the literature. All these test, as well as the Kruskal-Wallis test, can be seen as emerging from a comparison among the empirical distribution functions (EDF) pertaining to the samples being compared. Besides these EDF-based tests, we have considered some modern tests based on the idea of comparing (rather than distributions) the kernel density estimators (KDE) of the k-samples. Several measures of discrepancy among the estimators were used to construct a family of such KDE-based tests.

Fig. 3. Power estimation for different statistics in type I models with $a = 6/\sqrt{45}$.

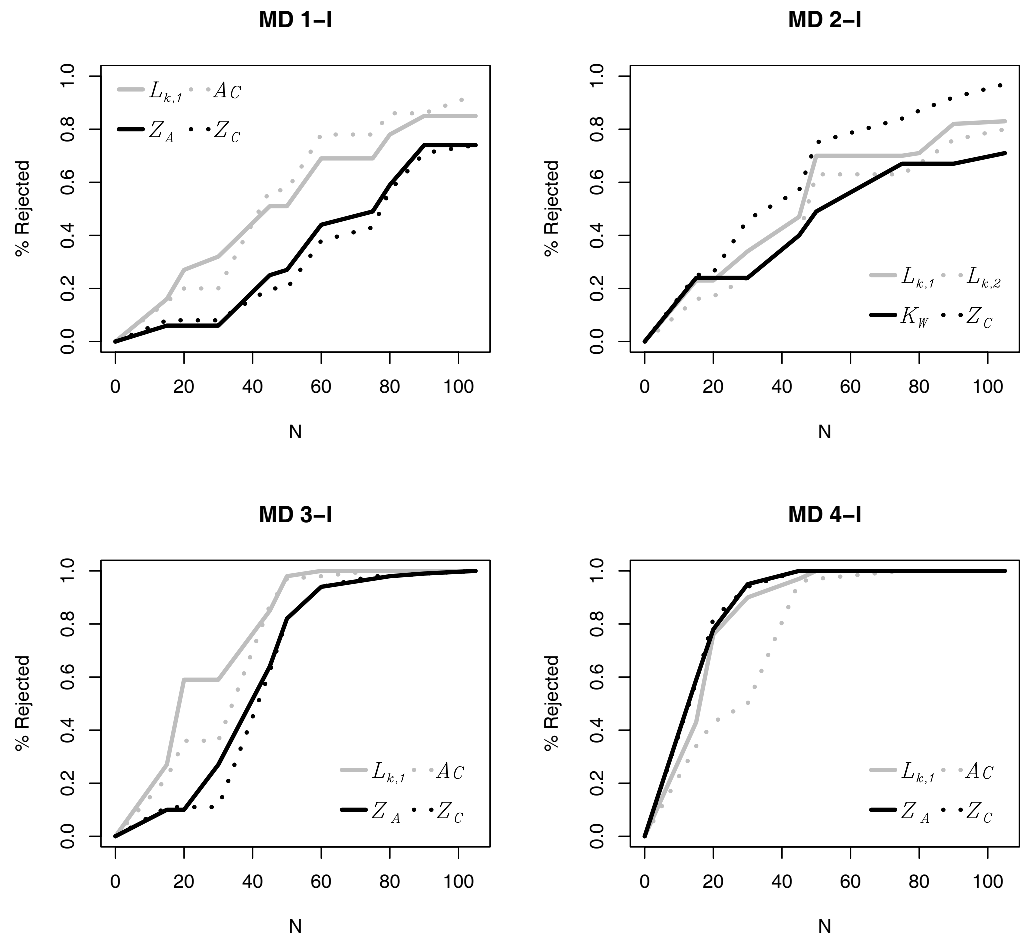
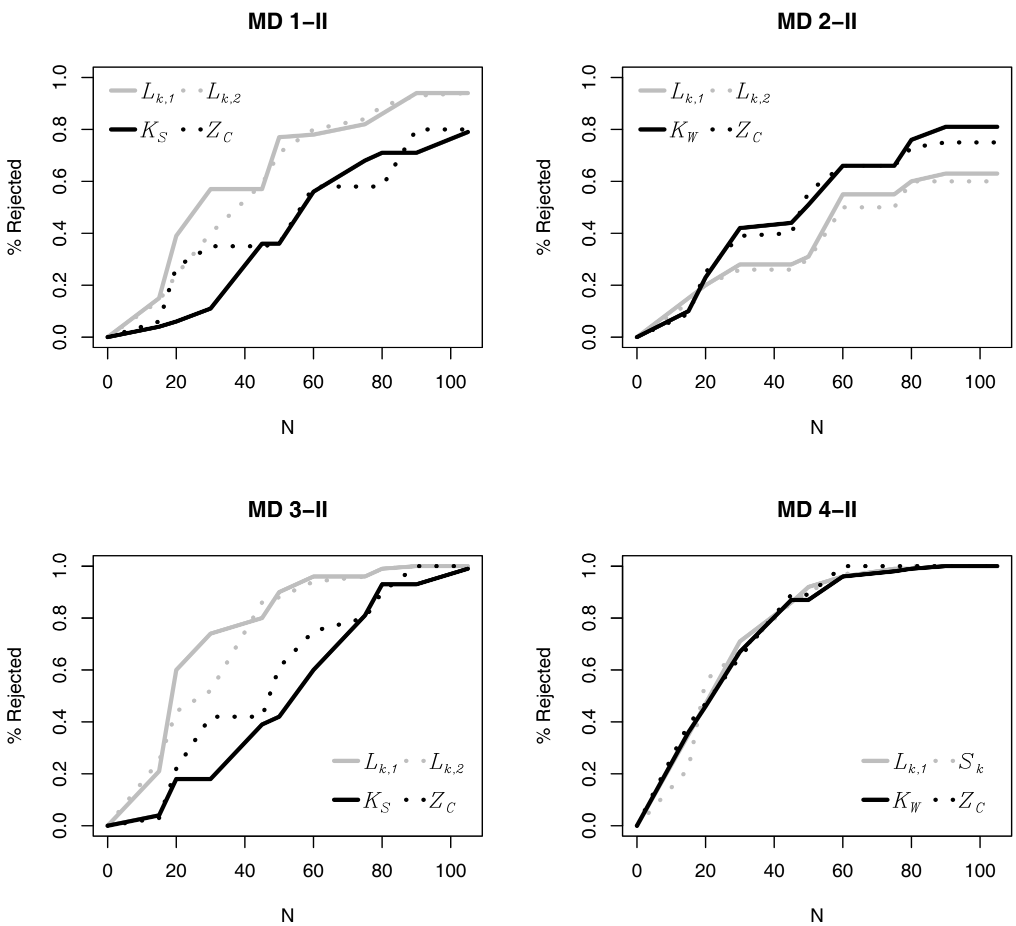


Fig. 4. Power estimation for different statistics in type II models with $a = 6/\sqrt{45}$..



The simulation results provided in this paper suggest that the KDE-based tests may be much more powerful that the EDF-based tests when the underlying populations differ not only in location. Among the KDE-based tests, the one attaining larger rejection percentages is that based on the L_1 norm. This confirms our initial intuition that the chosen measure of discrepancy among the densities is very relevant in the sense of detecting specific alternatives. On the other hand, the tests based on EDF seem to behave better when the differences are mainly in location. In this case, the likelihood-ratio test Z_C is the winner in the comparison.

An issue when computing the KDE-based tests is that of the choice of the smoothing parameter or bandwidth. Indeed, the power of a k-sample test based on the comparison of kernel density estimators is strongly influenced by the smoothing degree in the estimators. In this paper we have described an authomatic bandwidth selector (the *double minimum*) which, being computationally feasible, seems to provide a power close to the optimal one. Our proposed method for bandwidth selection is based on the bootstrap, and it reduces the computational time w.r.t. other bootstrap methods recently introduced. The key fact to get this improvement relies in that our method only estimates a distribution under the null hypothesis. See however the approach in Cao and Van Keilegom (2006) in which the bootstrapping is also performed under the alternative. We only mention that the computational speed of the *double minimum* bandwidth could be further improved *via* the inclusion of a stopping rule, which means *stop* if all the P-values over an initial grid of bandwidths are smaller than the significance level. This idea seems to be relevant since there is still a high computational cost behind the using of our authomatic bandwidth.

References

Ahmad AI & Li Q, (1997) Testing symmetry of an unknown density function by kernel method, *Journal of Nonparametric Statistics*, **7**, 279–293.

Ahmad AI & Amezziane M, (2007) A general and fast convergent bandwidth selection method of kernel estimator, *Journal of Nonparametric Statistics*, **19**, 165–187.

Anderson NH, Hall P & Titterington DM, (1994) Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *Journal of Multivariante Analysis*, **50**, 41–54.

Azzalini A (1981) A note on the estimation of a distribution function and quantiles by a kernel method, *Biometrika*, **68**, 326–328.

Cao R & van Keilegom I, (2006) Empirical likelihood tests for two-sample problems via nonparametric density estimation, *Canad. J. Statist.*, **34**, 61-77.

Devroye L, (1997) Universal smoothing factor selection in density estimation: Theory and practice, *Test*, **6**, 2, 223–320.

Devroye L & Györfy L, (1985) *Nonparametric density estimation; The L_1 view*, John Wiley, New York.

Diks D & Tong H, (1999) A test for simmetries of multivatiate probability distributions, *Biometrika*, **86**, 3, 605–614.

Eggermont PPB & LaRiccia VN, (2003) Selecting the smoothing parameter in goodness of fit testing, *Unpublished Manuscript*.

Fan Y, (1994) Testing the goodness of fit of a parametric density function by kernel method, *Econometric Theory*, **10**, 316–356.

Fan Y, (1998) Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters, *Econometric Theory*, **14**, 604–621.

Hall P & York M, (2001) On the calibration or Silverman's test for multimodality, *Statistica Sinica*, **11**, 516–536.

Hall P, DiCiccio JT & Romano JP, (1989) On smoothing and the bootstrap, *Annals of Statistics*, **17**, 2, 692–704.

Horvath L, (1991) On L_p-norms of multivariate density estimations, *Annals of Statistics,* **19**, 4, 1933–1949.

Kiefer J, (1959) k-Sample analogues of the Kolmogorov-Smirnov, Cramér-von Mises test, *Ann. Math. Statist.*, **30**, 420–447.

Konakov V, (1978) Complete asymptotic expansions for the maximun deviation of the empirical density function. *Theory Probab. Appl.,*, **28**, 495–509.

Kruskal WH & Wallis WA, (1952) Use of ranks in one-criterion variance analysis, *Journal of the American Statistical Association*, **47**, (260): 583—621.

Li Q, (1996) Nonparametric testing of closeness between two unknown distributions functions, *Econometric Review*, **15**, 3, 216–274.

Li Q, (1999) Nonparametric testing the similarity of two unkown density functions: local power and bootstrap analysis, *Journal of Nonparametric Statistics*, **11**, 189–213.

Liero H, Läuter H & Konakov V, (1998) Nonparametric versus parametric goodness of fit, *Statistics*, **31**, 115–149.

Louani D (2000) Large deviation for L_1-distance in kernel density estimation, *Journal of Statistical Planning and Inference*, **90**, 177–182.

Martínez-Camblor P, (2006) *Tests no paramétricos basados en una distancia entre funciones de densidad*. Servicio de Publicaciones de la Universidad de Oviedo, Oviedo. Spain.

Martínez-Camblor P & Corral N, (2008) Weaker conditions for asymptotic approximation to L_p-norms of the kernel estimators, *InterSTAT Journal*, june, 1–18.

Martínez-Camblor P, de Uña-Álvarez J & Corral N, (2008) k-Sample test based on the common area of kernel density estimator, *Journal of Statistical Planning and Inference*, doi:10.1016/j.jspi.2008.02.008.

Martínez-Camblor P, (2008) Tests de hipótesis para contrastar la igualdad entre k poblaciones, To appear in *Revista Colombiana de Estadística*.

Park BU & Marron JS, (1990) Comparison of data-dirven bandwidth selectors, *Journal of American Statistics Association*, **85**, 409, 66–72.

Parzen E, (1962) On estimation of a probability density function and mode, *Annals of Mathematical Statistics*, **33**, 1065–1076.

Rosenblatt M, (1956) Remarks on some nonparametric estimates of a density functions, *Annals Math. Statistics*, **27**, 832–837.

Scholz FW & Stephens MA, (1987) k-Samples Anderson-Darling test. *J. Amer. Statist. Assoc.* **82**, 918–924.

Silverman BW, (1981) Using kernel density estimation to investigate multimodality, *Journal of the Royal Statistics Society, B*, **43**, 97–99.

Zhang J & Wu Y, (2007) k-Sample test based on the likelihood ratio, *Computational Statistics & Data Analysis*, **51**, 9, 4682–4691.

5. Appendix: Some asymptotic results

In this appendix we establish the asymptotic normality for the statistics $L_{k,1}$ and $L_{k,2}$ introduced in Section 1. The method of proof resembles that in Martínez-Camblor, de Uña-Álvarez and Corral (2008) for the $\mathcal{AC}$ statistic. The key Lemma is the following (for a proof see Martínez-Camblor and Corral, 2008).

Lemma 1. (Lemma 5 in Martínez-Camblor and Corral; 2008) Let be μ a measure and let $\mathcal{X}_n : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ be a sequence of stochastic processes, and let T_n be a sequence of sets satisfying

(H1) $E\left(\int_{T_n} \mathcal{X}_n(t)d\mu(t)\right) = 0 \qquad \forall n \in \mathbb{N}$

(H2) $E\left(\int_{T_n} \mathcal{X}_n(t)d\mu(t)\right)^2 = 1 \qquad \forall n \in \mathbb{N}$

(H3) $\sup_{(t,s)\in T_n \times T_n} |Cov\,(\mathcal{X}_n(t), \mathcal{X}_n(s))| < \infty \quad \forall n \in \mathbb{N} \quad \forall \omega \in \Omega$

(H4) There exist positive real constants C, λ such that $E\,(\mathcal{X}_n(t)\mathcal{X}_n(s)) = 0$ whenever $|t - s| > Cn^{-\lambda}$

Then, if $\mu(T_n) = O(\log n)$ and for each sequence of positive real number $\{h_{(n)}\}_{n\in\mathbb{N}}$, $\mu([t - Ch_{(n)}, t + Ch_{(n)}]) = O(h_{(n)})\ \ \forall t \in \mathbb{R}$, it holds:

$$\int_{T_n} \mathcal{X}_n(t)d\mu(t) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1) \tag{2}$$

□

In order to establish the asymptotic normality of the given statistics $L_{k,p}$ $(p = 1,\ 2)$, the following assumptions will be requested for the Lebesgue measure λ. Notation f refers to the null density. The same conditions with $p = 1$ were needed in Martínez-Camblor and Corral (2008) for the $\mathcal{AC}$ statistic.

(C1) K is a density function of bounded variation, symmetrical about zero, with finite variance and compact support.

(C2) There exists a sequence of compact intervals, $\{C_n\}_{n\in\mathbb{N}}$ such that $\lambda(C_n) = O(\log n)$ and, if $\bar{C}_n$ is the complementary set of C_n, then $\int_{\bar{C}_n} f(t)d\lambda(t) = O(n^{-1})$ and $\int_{\bar{C}_n} f(t)^p d\lambda(t) = o\left((nh_n)^{-p/2}h_n^{1/2}\right)$.

(C3) $nh_n \to_n \infty$, $nh_n^5 \to_n c_0 < \infty$ and $h_n \to_n 0$.

Put $n = \sum_{i=1}^k n_i$ and let h_n be the bandwidth used to estimate the density in the pooled sample, so we have the $L_{k,p}$ statistic

$$L_{k,p}(n) = \frac{1}{n}\sum_{i=1}^k n_i \int |f_{n_i}(t) - f_n(t)|^p dt.$$

Then, the hypotheses (C1), (C2) and (C3) imply that $L_{k,p}(n)$ $(p \geq 1)$ satisfies (H1), (H2), (H3) and (H4). As a consequence, the following result can be easily derived.

Theorem 1. Let f_{n_i} $(1 \leq i \leq k)$ be k kernel density estimators based on k independent random samples taken from the same density function f, and let f_n be a kernel density estimator based on the pooled sample. Assume that the kernel functions and the sequence of bandwidths in the estimators satisfy (C1) and (C3). Assume also that the density f satisfies (C2). Finally, assume that there exist real numbers $c_1, ...c_k$ such that $nh_n/n_i h_{n_i} \to_n c_i$. Then for each $p \geq 1$ we have

$$\frac{1}{\sqrt{h_n}\sigma_{k,p}(n)}\left\{(nh_n)^{p/2}L_{k,p}(n) - e_{k,p}(n)\right\} \xrightarrow{\mathcal{L}}_n \mathcal{N}(0,1)$$

where $\sigma_{k,p}(n)^2 = n^p h_n^{p-1} Var(L_{k,p}(n))$ and $e_{k,p}(n) = E((nh_n)^{p/2}L_{k,p}(n))$.

Proof. As announced, the proof is similar to that in Martínez-Camblor, de Uña-Álvarez and Corral (2008), Theorem 2. First we prove that $\sigma_{k,p}(n)$ and $e_{k,p}(n)$ are finite. By the triangular inequality we have that, for each $k \in \mathbb{N}$, there exists a constant C_k such that

$$\begin{aligned} 0 \leq \sigma_{k,p}(n)^2 &= n^p h_n^{p-1} Var(L_{k,p}(n)) \\ &\leq \frac{n^p h_n^{p-1} C_k}{n^2}\sum_{i=1}^k n_i^2 \left(Var\left(\int |f_{n_i} - f|^p\right) + Var\left(\int |f_n - f|^p\right)\right). \end{aligned}$$

Use Lemma 4 in Horvath (1991) to conclude $0 \leq \sigma_{k,p}(n)^2 < \infty$. Similarly,

$$\begin{aligned} 0 \leq e_{k,p}(n) &= (nh_n)^{p/2} E\left(L_{k,p}(n)\right) \\ &\leq \frac{(nh_n)^{p/2}}{n}\sum_{i=1}^k n_i \left(E\left(\int |f_{n_i} - f|^p\right) + E\left(\int |f_{n_i} - f|^p\right)\right) \end{aligned}$$

Again, by applying the referred Lemma 1, we get $0 \leq e_{k,p}(n) < \infty$.

Introduce the process

$$\mathcal{I}_{k,p} = \frac{1}{\sqrt{h_n}\sigma_{k,p}(n)}\left\{(nh_n)^{p/2}\sum_{i=1}^k n_i \left(|f_{n_i} - f_n|^p - E(|f_{n_i} - f_n|^p)\right)\right\}$$

Since (H1)-(H4) are implied by (C1)-(C3) then, by using the Lemma 1 and the Slutsky theorem, we get (2) and we conclude the proof.

□

We point out that a similar result for the statistic $L_{k,p}(n)$ could be obtained under the alternative hypothesis, by following the same lines. Of course, the value of the parameters $\sigma_{k,p}(n)$ and $e_{k,p}(n)$ change in that case.