Density comparison for independent and right censored samples via kernel smoothing

Pablo Martínez-Camblor · Jacobo de Uña-Álvarez

Abstract   In this paper the problem of comparing k density functions from survival data is considered. Two non-parametric tests based on (two different) generalizations of the L_1 measure are adapted to the censored context. The asymptotic distribution of the test statistics is derived, and an approximation based on resampling methods is proposed. The relative power of the tests is investigated through a Monte Carlo simulation study. Results suggest that the tests exhibit good power when no one of the survival functions dominates the others, especially when the censoring distribution is the same along the k groups and the censoring percentage is small.

Keywords   Kernel density estimator · Common area · L_1 measure · Censored data · k-sample test

1 Introduction

Let X be a lifetime of interest with cumulative distribution function (CDF) F, and probability density function (PDF) f, supported on the positive real line $\mathbb{R}^+$. Let $X_n = \{x_1, \ldots, x_n\}$ be a random sample of X. Let C be a censoring random variable, independent of X with CDF G, on $\mathbb{R}^+$ and let $C_n = \{c_1, \ldots, c_n\}$ be a random sample from C. As always with censored data, we observe the pairs (z_i, δ_i), $1 \leq i \leq n$, where $z_i = min\{x_i, c_i\}$ and $\delta_i = I_{(-\infty, c_i]}(x_i)$ (where $I_A(x)$ is the indicator function

P. Martínez-Camblor (✉)
Oficina de Investigación Biosanitaria, Asturies, Spain
e-mail: pablomc@ficyt.es; pmcamblor@hotmail.com

J. de Uña-Álvarez
Departamento de Estadística e IO, Universidade de Vigo, Vigo, Spain
e-mail: jacobo@uvigo.es

for the set A). The independence between the lifetime and the censoring time is a usual assumption which guarantees the consistency of the commonly used estimators for the CDF and the PDF. However, it is worth mentioning that this assumption may be violated in some situations, for instance, in competing risks scenarios or for censoring caused by voluntary patient withdrawal.

Without any additional condition on the function G, the maximum-likelihood estimator of F is the product-limit estimator introduced by Kaplan and Meier (1958) in one of the most cited statistical papers, whose consistency is independent with respect to the censoring time distribution and which is defined by

$$\hat{F}_{K_M}(t) = 1 - \prod_{i=1}^{n} \left[1 - \frac{I_{(-\infty,t]}(z_{(i)})}{n+1-i} \right]^{\delta_{[i]}} \quad t \in \mathbb{R}^+. \tag{1}$$

Here, $z_{(j)}$ is the jth element in the ordered sample of the (z_i, δ_i)'s, while $\delta_{[j]}$ stands for the jth concomitant indicator. Denoting by ω_j the jump of $\hat{F}_{K_M}$ at $z_{(j)}$, we have for $j = 2, \ldots, n-1$ $\omega_j = \hat{F}_{K_M}(z_{(j)}) - \hat{F}_{K_M}(z_{(j-1)})$ and $\omega_1 = \hat{F}_{K_M}(z_{(1)})$. Then, the classical kernel estimator of f is given by

$$\hat{f}_{K_M,h}(t) = \int \frac{1}{h} K \left(\frac{x-t}{h} \right) d\hat{F}_{K_M}(x) = \frac{1}{h} \sum_{i=1}^{n} \omega_i K \left(\frac{z_{(i)}-t}{h} \right), \tag{2}$$

where K is the kernel function, usually chosen as a symmetrical density function with zero mean and finite variance, and the bandwidth $h = h_{(n)}$ is a sequence of positive real numbers that plays an important role in the final estimator. The properties of this estimator have been widely studied by Földes et al. (1981); Mielniczuk (1986) or Zhang (1996) among others. An alternative kernel estimator has been introduced by Blum and Susarla (1980) but it is known that the differences between them are typically negligible.

On the other hand, a very important problem in clinical trials and other areas is to check whether two or more censored samples come from the same probability law. In this setup, the most commonly used k-sample nonparametric tests are probably the linear rank (or weighted) rank tests (Harrington 2005). See also Martínez-Camblor (2010a); Martínez-Camblor (2011) and references therein for other recent proposals. In an attempt to develop methods with improved power, several authors have used kernel density estimation for testing whether k independently sampled populations have the same distribution. To this end, in the uncensored setting, Cao and Van Keilegom (2006) proposed a test based on the likelihood ratio of the kernel density estimators. Martínez-Camblor et al. (2008) studied a k-sample test based on the common area ($\mathcal{AC}$) among the respective kernel density estimators, and Martínez-Camblor and de Uña-Álvarez (2009a) introduced three new tests for comparing densities from independent samples. These papers show that tests based on PDFs are more powerful than tests based on the comparison of CDFs when there exist differences other than location among the underlying probability laws.

In the present paper, we investigate the performance of the tests based on kernel density estimators under random censorship. With this goal, we adapt two tests based

on generalizations of the L_1 measure and the kernel density estimator. In particular, the $\mathcal{AC}(h)$ and $L_{k,1}(h)$ tests are adapted to the censored data setting. We study the asymptotic distribution of the test statistics and, as usual, the bootstrap method is also considered as a device to approximate it in practice (Sect. 2). A simulation study to investigate the power of the tests is reported in Sect. 3. In addition, two real data sets are analyzed for illustration proposes (Sect. 4). Section 5 reports our main conclusions. All the proofs are deferred to the Appendix.

2 k-sample tests

Let $\mathbf{Z}_j = \{(z_{ij}, \delta_{ij})\}_{1 \leq i \leq n_j}$ with $1 \leq j \leq k$ be k independent random samples of censored survival times. If for $1 \leq j \leq k$, $\hat{f}_{K_M, h_j}$ (with $h_j = h_{(n_j)}$) and $\hat{f}_{K_M, h_n}$ (with $h_n = h_{(n)}$ and $n = \sum_{j=1}^{k} n_j$) denote the kernel density estimators pertaining to the jth and the pooled sample, respectively, the direct generalizations of the $\mathcal{AC}$ and $L_{k,1}$ statistics to the censored scenario are

$$\mathcal{AC}(h) = \int min\left\{ \hat{f}_{K_M, h_1}(t), \ldots, \hat{f}_{K_M, h_k}(t) \right\} dt \qquad (3)$$

and

$$L_{k,1}(h) = \frac{1}{n} \sum_{j=1}^{k} n_j \int \left| \hat{f}_{K_M, h_j}(t) - \hat{f}_{K_M, h_n}(t) \right| dt, \qquad (4)$$

respectively. Under quite general (and mild) assumptions, the asymptotic distribution of these statistics can be derived by arguments similar to those in Martínez-Camblor and de Uña-Álvarez (2009a). Theorem 1 guarantees the asymptotic normality for the more general statistic

$$L_{k,p}(h) = \frac{1}{n} \sum_{j=1}^{k} n_j \int \left| \hat{f}_{K_M, h_j}(t) - \hat{f}_{K_M, h_n}(t) \right|^p d\lambda(t) \quad 1 \leq p < \infty, \qquad (5)$$

where λ denotes the usual Lebesgue measure. The proof is deferred to the Appendix. The asymptotic distribution for $\mathcal{AC}$ is derived inmediately from the case $p = 1$.

Theorem 1 *Assume that the k populations of lifetimes have a common PDF f which is bounded and continuous. Under the following assumptions:*

C_1. *K is a density function of bounded variation, symmetric about zero, with finite variance and compact support.*

C_2. *There exists a sequence of compact intervals, $\{C_n\}_{n \in \mathbb{N}}$, such that $\lambda(C_n) = \mathcal{O}(\log n)$ and, if $\bar{C}_n$ is the complementary set of C_n, then $\int_{\bar{C}_n} f(t) d\lambda(t) = \mathcal{O}(n^{-1})$ and $\int_{\bar{C}_n} f(t)^p d\lambda(t) = o\left((nh_n)^{-p/2} h_n^{1/2} \right)$.*

C_3. *There exists $\alpha > 0$ such that $h_n = \mathcal{O}(n^{-\alpha})$, $nh_n \to_n \infty$ and $nh_n^5 \to_n c_0 < \infty$.*

C_4. *There exist real numbers, $c_1, \ldots, c_k$, such that $nh_n/n_j h_j \to_n c_j$ ($1 \le j \le k$).*

Then, for each $p \ge 1$ we have

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

where $\sigma_{k,p}(n)^2 = n^p h_n^{p-1} \mathbb{V}[L_{k,p}(n)]$ ($\mathbb{V}[u] = \mathbb{E}[u^2] - \mathbb{E}[u]^2$) and $e_{k,p} = \mathbb{E}[(nh_n)^{p/2} L_{k,p}(h)]$. □

Some of the above assumptions (specifically, the convergence rates which appear in C_2 and C_3) can be relaxed with an extra theoretical effort. However, these hypothesis are satisfied in most practical situations.

Explicit expressions for the mean and the variance of the statistic in Theorem 1 are not available and, in general, it is complex to obtain good estimators for these parameters. Therefore, this result (only) allows to guarantee the consistency of the proposed test statistics but not to obtain a final P-value. Hence, we propose to approximate (as usual) their distributions from a suitable resampling plan. In particular, since we are assuming that the lifetime distribution, F, and the k involved censoring time distributions, G_i ($1 \le i \le k$), are absolutely continuous, we apply the smoothed-censored (SC) resampling plan introduced by Gonzalez-Manteiga et al. (1996) whose algorithm is:

A. First, draw bootstrap resamples $X_i^b = \{x_{i,1}^b, \ldots, x_{i,n_i}^b\}$ ($1 \le i \le k$) from the (common) smoothed estimator of F, $\tilde{F}_{h_X}$ (h_X denotes the bandwidth).
B. For each $1 \le i \le k$ draw independent bootstrap resamples $C_i^b = \{c_{i,1}^b, \ldots, c_{i,n_i}^b\}$ from the smoothed estimator of the censoring time CDFs $\tilde{G}_{h_{C_i}}$ (h_{C_i} denotes the bandwidth for the ith censoring time distribution). At this point we recall that we do not assume the equality among the different censoring time distributions.
C. Finally, for $1 \le i \le k$, construct $\mathbf{Z}_i^b = \{(z_{i,j}^b, \delta_{i,j}^b)\}_{1 \le j \le n_i}$ where $z_{i,j}^b = min\{x_{i,j}^b, c_{i,j}^b\}$ and $\delta_{i,j}^b = I\{z_{i,j}^b = x_{i,j}^b\}$.

In order to avoid the well-known problems that occur with the Kaplan-Meier estimator when the largest observation is right-censored, we put the missing probability mass just to the rigth of that observation. This resampling procedure, described by Efron (1981), obtains (by simulation studies) more accurate non-parametric confidence bands than the ones based on the asymptotic distribution (Efron 1982).

On the other hand, Beran (1982) proved that, under the usual regularity conditions, the bootstrap estimate of the distribution of a statistic is asymptotically minimax among all possible estimates.

2.1 Bandwidth selection

One of the main issues in the application of *smoothing techniques* is the selection of the *smoothed parameter* or bandwidth. It is known that the bandwidth choice may be crucial when taking the final decision (to reject or to accept the null hypothesis).

[Martínez-Camblor and de Uña-Álvarez](2009b) studied and compared different methods for the automatic bandwidth selection in k-sample smooth test for complete data.

In a general smooth testing problem, one considers a test statistic $\mathcal{T}_h$ which depends on the bandwidth h ($h > 0$). Then, the goal is to select the parameter h in order to get a powerful test. Without loss of generality, we assume that the null hypothesis is rejected for large values of $\mathcal{T}_h$. We denote by $\mathcal{T}_{0,h}$ and $\mathcal{T}_{1,h}$ independent random variables with the null and the alternative distributions of $\mathcal{T}_h$, respectively. Note that

$$\begin{aligned} \pi(h) &= \mathcal{P}\{\mathcal{T}_{0,h} > \mathcal{T}_{1,h}\} \\ &= \mathbb{E}[I\{\mathcal{T}_{0,h} > \mathcal{T}_{1,h}\}] = \mathbb{E}[\mathbb{E}[I\{\mathcal{T}_{0,h} > \mathcal{T}_{1,h}\}|\mathcal{T}_{1,h}]] = \mathbb{E}[\pi_v(\mathcal{T}_{1,h})], \end{aligned}$$

where $I\{A\}$ stands for the indicator function of the event A and $\pi_v(t)$ is the P-value of t, that is, the probability under the null that the realization of the test statistic is larger than t. Note that $\pi(h)$ averages the P-values with respect to the alternative distribution of $\mathcal{T}_h$. Hence, it makes sense to choose the bandwidth which minimizes $\pi(h)$; this bandwidth should lead to a powerful test, since small P-values correspond to a greatly significant realization of the test statistic. We denote this bandwidth by h_P. Obviously, since h_P minimizes $\pi(h)$, it can be interpreted as the bandwidth for which the null and the alternative hypothesis are separated at the most by $\mathcal{T}_h$.

In [Martínez-Camblor and de Uña-Álvarez](2009b) the *BM* bandwidth is introduced as

$$h_{BM} = \mathbb{E}[argmin_{\{h>0\}}\pi_v(\mathcal{T}_{1,h})],$$

which can be regarded as an approximation for h_P iff

$$\mathbb{E}[argmin_{\{h>0\}}\pi_v(\mathcal{T}_{1,h})] \approx argmin_{\{h>0\}}\mathbb{E}[\pi_v(\mathcal{T}_{1,h})] = argmin_{\{h>0\}}\pi(h) = h_P. \tag{6}$$

In practice, the involved minimization is performed on a given grid of (possible) bandwidths. Its explicit algorithm is ([Martínez-Camblor and de Uña-Álvarez 2009b](#)):

B_1. Let $h_i = h\hat{\sigma}_i n_i^{-1/5}$ ($1 \leq i \leq k$) be the used bandwidths and let $\mathcal{H} = \{h_{(1)}, \ldots, h_{(T)}\}$ be a grid of h-values among which the optimal one is to be selected. Here, $\hat{\sigma}_i$ denotes the standard deviation in the ith sample.

B_2. By using the above bootstrap resampling plan (steps A, B and C), for $1 \leq t \leq T$, estimate the distribution under the null hypothesis for $\mathcal{T}_{h_{(t)}}$, $\hat{F}_{B,t}^0$.

B_3. For each $b \in \{1, \ldots, B\}$, generate a resample $X_{b,1}^*, \ldots, X_{b,n_i}^*$ from $\hat{f}_{K_M, g_i}$ (survival times) and $C_{b,1}^*, \ldots, C_{b,n_i}^*$ from $\tilde{G}_{h_{C_i}}$ (censoring times) and build $\mathbf{Z}_i^*$ ($1 \leq i \leq k$). Estimate the significance level for the test for each $h_{(t)}$, $P_{b,t}^*$ by using $\hat{F}_{B,t}^0$ for $1 \leq t \leq T$.

B_4. Choose the h_b^* which minimizes the previous sequence of P-values, i.e., $h_b^* = argmin\{P_{b,1}^*, \ldots, P_{b,T}^*\}$ ($1 \leq b \leq B$). We obtain the values $H_B = \{h_1^*, \ldots, h_B^*\}$.

B_5. The final value for h is

$$h_{BM} = \frac{1}{B} \sum_{b=1}^{B} h_b^*.$$

The computation of the BM-bandwidth as described along steps $B_2 - B_4$ implies $B + B_0$ evaluations of the test statistic (B_0 is the number of bootstrap resamples under the null used in B_2) multiplied by the number of bandwidths in the grid, T. Besides, in order to approximate the final P-value, one needs perform B_0' extra evaluations of the statistic under the null which leads to a total number of evaluations of $T(B + B_0) + B_0'$. A decision on the null and alternative hypothesis is reached after the following step:

B_6. Draw independent B_0' bootstrap resamples under the null (steps A, B and C). Let $\mathcal{T}_{0,h_0}^b$ be the statistic $\mathcal{T}_h$ based on the bth null bootstrap resample ($1 \leq b \leq B_0'$) and the bandwidth h_0 (derived from $h = h_{BM}$). Then, reject the null hypothesis if and only if

$$\frac{1}{B_0'} \sum_{b=1}^{B_0'} I\left\{ \mathcal{T}_{0,h_0}^b > T_{h_0} \right\} < \alpha,$$

where T_{h_0} is the actual value of the test statistic (when based on h_0) and α is the significance level of the test.

We propose the following new algorithm (labelled as *MS*, *Minimum Significance*) in order to estimate, directly, the bandwidth h_P:

P_1. Let $h_i = h\hat{\sigma}_i n_i^{-1/5}$ ($1 \leq i \leq k$) be the used bandwidths and let $\mathcal{H} = \{h_{(1)}, \ldots, h_{(T)}\}$ be a grid of h-values among which the optimal one is to be selected.

P_2. By using the above bootstrap resampling plan (steps A, B and C), draw B_0 bootstrap resamples under the null. Let $\mathcal{T}_{0,h_{(t)}}^b$ ($1 \leq t \leq T$) be the statistic $\mathcal{T}_{h_{(t)}}$ based on the bth resample ($1 \leq b \leq B_0$) and on $h_{(t)}$, and let $\hat{F}_{B,t}^0$ be the corresponding bootstrap distribution.

P_3. For each $b \in \{1, \ldots, B\}$, draw independently resamples under the alternative for both the survival and the censoring times. Let $\mathcal{T}_{1,h_{(t)}}^b$ ($1 \leq t \leq T$) be the statistic $\mathcal{T}_h$ based on the bth alternative bootstrap resample ($1 \leq b \leq B$) and on $h_{(t)}$.

P_4. Finally, compute

$$\hat{h}_P = argmin_{\{h \in \mathcal{H}\}} \left\{ \frac{1}{BB_0} \sum_{b=1}^{B} \sum_{b'=1}^{B_0} I\left\{ \mathcal{T}_{0,h}^{b'} > \mathcal{T}_{1,h}^b \right\} \right\}.$$

Obviously, $\hat{h}_P \in \mathcal{H}$ and the final P-value can be estimated from the respective $\hat{F}_{B,t}^0$ (previously computed), therefore this algorithm implies $T(B_0 + B)$ evaluations of the statistic.

3 Statistical power comparison

In order to investigate the statistical power of the proposed tests (the considered nominal level was $\alpha = 0.05$), a Monte Carlo simulation study was conducted. We considered the grid $\mathcal{H} = \{1/2, 1, 2, 3, 4\}$ and apply the *BM* and the *MS* algorithms when different bandwidths h for $\mathcal{AC}(h)$ and $L_{k,1}(h)$ lead to different decisions (i.e., when there exist P-values smaller and greater than the fixed significance level, α). Our previous experience on the choice of h suggests that one should not dedicate too much computational effort in the estimation of the h-values but the estimation of the P-value is very important; therefore we have used $B = 100$ and $B_0 = 199$. The rejection probabilities were estimated from 1,000 Monte Carlo replications.

We considered the following five different density functions:

(i) $f_0(t) = (\sqrt{2\pi})^{-1} t^{1/2} \exp\{-t/2\} I_{[0,\infty)}(t)$.
(ii) $f_1(t) = 0.25 f_0(t) + 0.75(3\sqrt{2\pi})^{-1} t^{3/2} \exp\{-t/2\} I_{[0,\infty)}(t)$.
(iii) $f_2(t) = 0.25 f_0(t) + 0.75(\sqrt{2\pi})^{-1} exp\{-0.50(t-3)^2\}$.
(iv) $f_3(t) = (0.75(1/3) \exp\{-t/3\} + 0.25(1/10) \exp\{-t/10\}) I_{[0,\infty)}(t)$.
(v) $f_4(t) = 0.5 f_0(t) + 0.5(2\sqrt{2\pi})^{-1} exp\{-0.50((t-4)/2)^2\}$.

Taking f_0 as reference, these functions represent four different situations: proportional differences (f_1), non-uniform dominance (f_2), late differences (f_3) and early differences (f_4). Figure 1 shows the respective survival curves.

Censoring times were drawn from an exponential distribution, $Exp\{1/\beta\}$ (cases $\beta = 1.315$, $\beta = 3.375$ and $\beta = 9.625$ were considered). Table 1 shows the expected censorship percentages for the functions $i - v$.
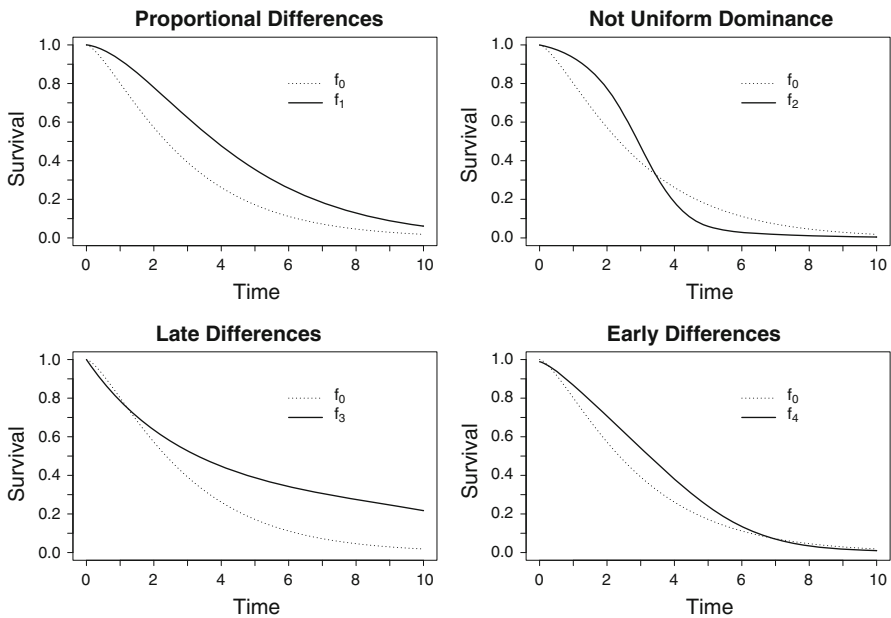

Fig. 1 Survival functions from the considered density functions

Table 1 Expected censorship percentages for the different considered schemes

β	Underlying density functions				
	$f_0(\%)$	$f_1(\%)$	$f_2(\%)$	$f_3(\%)$	$f_4(\%)$
9.625	25.0	34.8	25.5	43.9	28.8
3.375	50.0	63.8	55.3	68.2	57.3
1.315	75.0	86.2	83.4	83.9	81.5

The statistical power of the proposed test statistics were studied for the case $k = 3$ and different sample sizes, and compared to the Fleming-Harrington linear rank tests corresponding to the weights $\rho = 0$, $\rho = 0.5$ and $\rho = 1$, these are respectively the Mantel-Haenzel test (M_H), the Tarone-Ware test (T_W), and the Peto-Peto (P_P). Specifically, in Figs. 2 and 3 we have considered the cases $\overrightarrow{n} = (n_1, n_2, n_3) = (15, 15, 15), (25, 25, 25), (25, 50, 75), (75, 75, 75), (75, 100, 125)$ and $(125, 125, 125)$.

*Example 1* The random samples, (X_1, X_2, X_3), are generated from the functions,

MD 0-I. $X_1 \sim f_0$, $X_2 \sim f_0$, $X_3 \sim f_0$. (Null hypothesis).
MD 1-I. $X_1 \sim f_0$, $X_2 \sim f_0$, $X_3 \sim f_1$.
MD 2-I. $X_1 \sim f_0$, $X_2 \sim f_0$, $X_3 \sim f_2$.
MD 3-I. $X_1 \sim f_0$, $X_2 \sim f_0$, $X_3 \sim f_3$.
MD 4-I. $X_1 \sim f_0$, $X_2 \sim f_0$, $X_3 \sim f_4$.

Table 2 shows the rejection proportions for the new proposed tests for five different h's, for two specific configurations of the sample sizes. Both the $\mathcal{AC}(h)$ and the $L_{k,1}(h)$ statistics are extremely conservative, particularly when the censorship percentage is large. In this case, the observed rejection proportion for the $\mathcal{AC}(h)$ statistic never reaches the 1.5% when null hypothesis is true (model 0-I). Although the optimal bandwidth changes with the statistic, the model, the sample size and the censorship percentage, the value $h = 1$ is often the best one (21 times of 32). Similarly as in Martínez-Camblor and de Uña-Álvarez (2009b), we find that the statistical power of the $\mathcal{AC}(h)$ statistic is poorer for unequal sample sizes. Also, the influence of h in the power is smaller for the $L_{k,1}(h)$ test.

Table 3 shows the rejection proportions for the Fleming-Harrington tests and the proposed ones when the algorithms *BM* and *MS* are applied on the grid $\mathcal{H}$ to select the bandwidth. The rejection percentages for the proposed statistics under the null are still very low, particularly for unequal sample sizes and large censorship percentages. Both methods obtain similar results, being close to the optimal ones inside the considered grid; moreover, in some situations we see that the results obtained by *BM* and *MS* are better than the optimal one within the grid.

Figure 2 shows the observed statistical power for the $\mathcal{AC}(h)$ and $L_{k,1}(h)$ statistics by using the *MS* algorithm on the grid $\mathcal{H}$ and for the M_H, T_W and P_P tests. The new tests are clearly the worst when the differences among the curves are proportional (model 1-I), but they are clearly the best when no curve dominates the other (model 2-I). In the other schemes all the tests behave similarly, although the new tests are
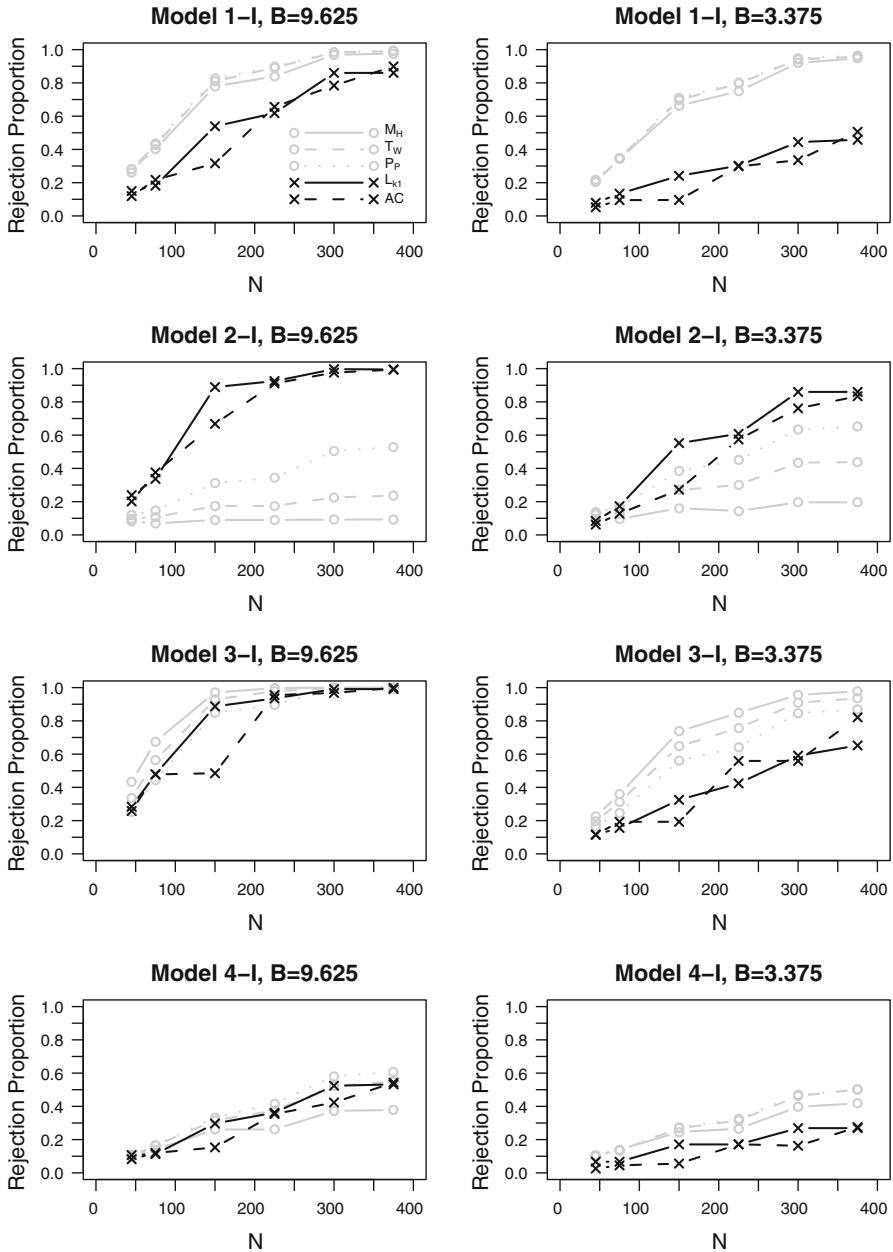
Fig. 2 Observed rejection proportions for the considered models in Example 1 (N = $n_1 + n_2 + n_3$)

slightly inferior for small sample sizes and large censorship percentage. In addition, it is also observed that $\mathcal{AC}(h)$ performs poorly for unequal sample sizes; similar findings in the uncensored situation were previously reported, see Martínez-Camblor et al. (2008) and Martínez-Camblor and de Uña-Álvarez (2009a).
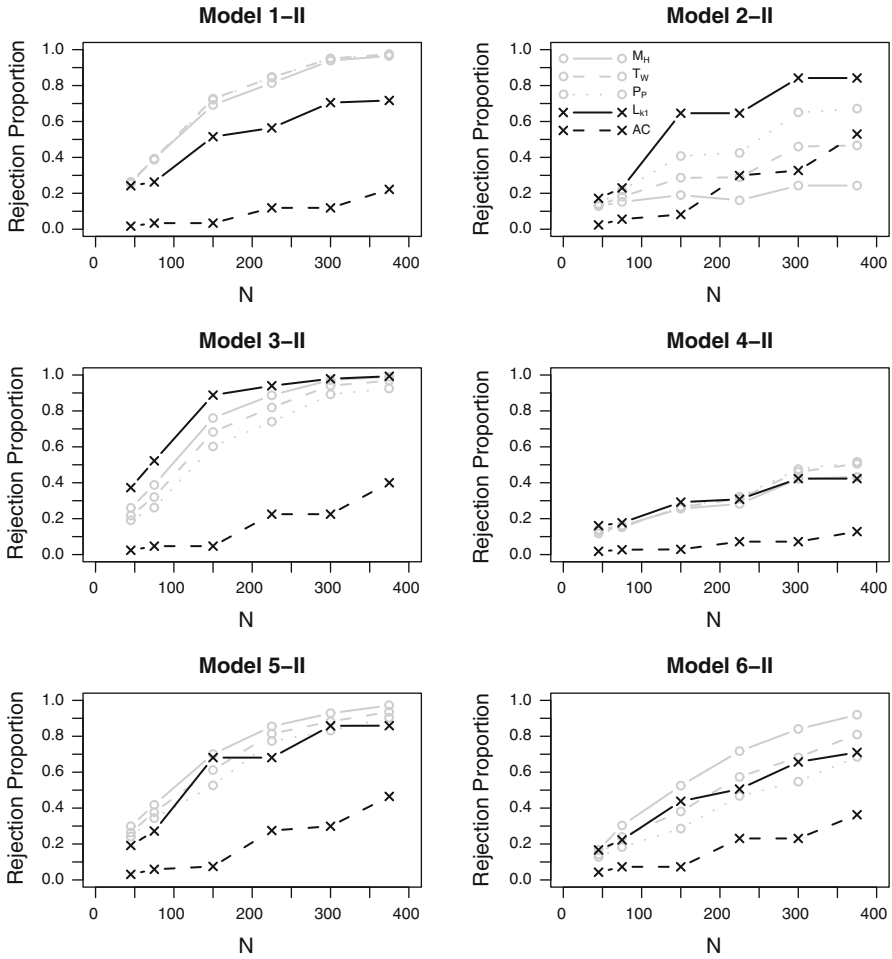
Fig. 3 Observed rejection proportions for the considered models in Example 2 ($N = n_1 + n_2 + n_3$)

Example 2 In this second example we considered the unbalanced censoring situation (i.e., distinct censoring time distributions). The censoring times are drawn from exponential distributions with parameters $\beta = 1.315$ for the first sample, $\beta = 3.375$ for second sample and $\beta = 9.625$ for the third sample. The samples of lifetimes (X_1, X_2, X_3) are generated from the functions,

MD 0-II. $X_1 \sim f_0$, $X_2 \sim f_0$, $X_3 \sim f_0$. (Null hypothesis).
MD 1-II. $X_1 \sim f_0$, $X_2 \sim f_0$, $X_3 \sim f_1$.
MD 2-II. $X_1 \sim f_0$, $X_2 \sim f_0$, $X_3 \sim f_2$.
MD 3-II. $X_1 \sim f_0$, $X_2 \sim f_0$, $X_3 \sim f_3$.
MD 4-II. $X_1 \sim f_0$, $X_2 \sim f_0$, $X_3 \sim f_4$.
MD 5-II. $X_1 \sim f_0$, $X_2 \sim f_1$, $X_3 \sim f_2$.
MD 6-II. $X_1 \sim f_0$, $X_2 \sim f_3$, $X_3 \sim f_4$.

Table 2 Observed rejection probabilities for $L_{k,1}(h)$ and $\mathcal{AC}(h)$ statistics in the proposed models for bandwidth $h\hat{\sigma}_i n_i^{-1/5}$ where $\hat{\sigma}_i$ is the sample standard deviation, n_i $(1 \le i \le 3)$ is the sample size and $h \in \{1/2, 1, 2, 3, 4\}$

MD	β		$\vec{n} = (25, 25, 25)$					$\vec{n} = (25, 50, 75)$				
			h					h				
			1/2	1	2	3	4	1/2	1	2	3	4
0-I	9.625	$L_{k,1}(h)$	0.037	0.045	0.050	0.053	0.025	0.036	0.041	0.050	0.055	0.057
		$\mathcal{AC}(h)$	0.037	0.042	0.027	0.025	0.020	0.030	0.033	0.042	0.036	0.033
	3.375	$L_{k,1}(h)$	0.017	0.023	0.021	0.022	0.023	0.020	0.018	0.020	0.019	0.020
		$\mathcal{AC}(h)$	0.008	0.007	0.007	0.009	0.010	0.014	0.011	0.011	0.011	0.012
1-I	9.625	$L_{k,1}(h)$	0.176	0.197	0.180	0.168	0.137	0.436	0.542	0.522	0.460	0.384
		$\mathcal{AC}(h)$	0.166	0.195	0.159	0.116	0.087	0.267	0.296	0.175	0.116	0.087
	3.375	$L_{k,1}(h)$	0.088	0.099	0.093	0.072	0.060	0.132	0.199	0.227	0.202	0.160
		$\mathcal{AC}(h)$	0.062	0.057	0.045	0.041	0.041	0.066	0.065	0.048	0.041	0.041
2-I	9.625	$L_{k,1}(h)$	0.338	0.409	0.342	0.271	0.231	0.838	0.900	0.836	0.737	0.672
		$\mathcal{AC}(h)$	0.295	0.323	0.241	0.190	0.149	0.557	0.632	0.529	0.443	0.401
	3.375	$L_{k,1}(h)$	0.127	0.146	0.088	0.046	0.027	0.425	0.543	0.473	0.353	0.236
		$\mathcal{AC}(h)$	0.116	0.085	0.035	0.025	0.020	0.212	0.235	0.169	0.115	0.092
3-I	9.625	$L_{k,1}(h)$	0.386	0.471	0.480	0.477	0.464	0.810	0.883	0.902	0.898	0.874
		$\mathcal{AC}(h)$	0.384	0.472	0.413	0.370	0.317	0.422	0.479	0.426	0.370	0.317
	3.375	$L_{k,1}(h)$	0.080	0.097	0.104	0.104	0.107	0.188	0.253	0.275	0.286	0.276
		$\mathcal{AC}(h)$	0.084	0.129	0.133	0.146	0.146	0.085	0.130	0.133	0.144	0.141
4-I	9.625	$L_{k,1}(h)$	0.112	0.125	0.100	0.093	0.077	0.247	0.316	0.265	0.189	0.152
		$\mathcal{AC}(h)$	0.084	0.105	0.073	0.055	0.042	0.122	0.140	0.105	0.069	0.059
	3.375	$L_{k,1}(h)$	0.041	0.058	0.046	0.037	0.031	0.099	0.139	0.130	0.096	0.077
		$\mathcal{AC}(h)$	0.036	0.028	0.009	0.008	0.011	0.030	0.038	0.035	0.030	0.025

Table 4 shows the rejection proportions for the new proposed tests for different $h's$. While the $\mathcal{AC}(h)$ statistic is extremely conservative for all considered $h's$ (the observed rejection probabilities under the null are never larger than 0.026), the $L_{k,1}(h)$ test is anticonservative for the largest h-values. In this scheme the optimal bandwidth also changes with the test statistic and the model. There is not a particular bandwidth value which deserves a special mention. It can also be observed that the $\mathcal{AC}$ criteria does not obtain good result for unequal censoring time distributions.

Table 5 shows the rejection proportions for the Fleming-Harrington tests and for the proposed tests when the algorithms BM and MS are applied over the grid $\mathcal{H}$ in the case of unequal censoring time distributions (Example 2). The first method (BM) always obtains a slightly smaller power than the second one (MS), although it is still larger than the optimal power obtained by the bandwidths within the considered grid. It is observed that the *MS* algorithm often obtains rejection levels above the largest one within the grid also under the null, thus indicating that the reported significance levels are too large, particularly for $\vec{n} = (25, 50, 75)$.

Table 3 Observed rejection probabilities for the reference statistics (M_H, T_W and P_P) and the new proposed ones ($L_{k,1}(h)$ and $\mathcal{AC}(h)$) when the bandwidth are selected from the *BM* (h_{BM}) and the *MS* ($\hat{h}_P$) algorithms on the grid $\mathcal{H}$

MD	β	M_H	T_W	P_P	$L_{k,1}(h)$		$\mathcal{AC}(h)$	
					h_{BM}	$\hat{h}_P$	h_{BM}	$\hat{h}_P$
$\vec{n} = (25, 25, 25)$								
0-I	9.625	0.087	0.078	0.077	0.052	0.070	0.030	0.022
	3.375	0.058	0.050	0.046	0.028	0.034	0.009	0.019
1-I	9.625	0.402	0.427	0.434	0.183	0.182	0.215	0.217
	3.375	0.343	0.347	0.349	0.114	0.134	0.067	0.095
2-I	9.625	0.069	0.107	0.147	0.371	0.337	0.350	0.376
	3.375	0.096	0.128	0.167	0.147	0.173	0.117	0.128
3-I	9.625	0.675	0.564	0.444	0.459	0.479	0.486	0.478
	3.375	0.359	0.312	0.247	0.114	0.156	0.152	0.193
4-I	9.625	0.139	0.165	0.166	0.118	0.113	0.113	0.120
	3.375	0.139	0.135	0.139	0.059	0.067	0.036	0.045
$\vec{n} = (25, 50, 75)$								
0-I	9.625	0.066	0.050	0.050	0.054	0.060	0.044	0.038
	3.375	0.055	0.061	0.061	0.025	0.038	0.017	0.023
1-I	9.625	0.779	0.810	0.826	0.547	0.539	0.297	0.316
	3.375	0.663	0.697	0.710	0.228	0.241	0.073	0.096
2-I	9.625	0.090	0.174	0.312	0.900	0.889	0.661	0.668
	3.375	0.161	0.269	0.384	0.537	0.552	0.248	0.272
3-I	9.625	0.971	0.929	0.848	0.903	0.887	0.472	0.485
	3.375	0.737	0.647	0.561	0.302	0.325	0.153	0.192
4-I	9.625	0.261	0.316	0.333	0.301	0.297	0.148	0.153
	3.375	0.246	0.268	0.273	0.153	0.171	0.040	0.056

Figure 3 is analogous to Fig. 2 for the models in Example 2. While the $\mathcal{AC}$ criteria does not work for unequal censoring time distributions, the $L_{k,1}(h)$ test improves its behavior relative to the Fleming-Harrington family. It is still the best one for model 2-II and obtains very competitive results for models 3-II, 4-II, 5-II and 6-II.

4 Real data analysis

In order to illustrate the proposed methods in a practical setup, in this section we consider two different data sets from real medical studies.

Example 1.

　　We consider data collected in a clinical trial to compare the efficacy and safety of two antiretroviral drugs in treating patients who had failed or were intolerant of zidovudine (AZT) therapy. Our objective are the 422 men HIV-infected who met entry

Table 4 Observed rejection probabilities for $L_{k,1}(h)$ and $\mathcal{AC}(h)$ statistics in the proposed models for bandwidth $h\hat{\sigma}_i n_i^{-1/5}$ where $\hat{\sigma}_i$ is the sample standard deviation, n_i ($1 \leq i \leq 3$) is the sample size and $h \in \{1/2, 1, 2, 3, 4\}$ for non-uniform censoring time distributions

| MD		$\vec{n} = (25, 25, 25)$					$\vec{n} = (25, 50, 75)$				
		h					h				
		1/2	1	2	3	4	1/2	1	2	3	4
0-II	$L_{k,1}(h)$	0.023	0.032	0.048	0.062	0.067	0.035	0.034	0.052	0.063	0.074
	$\mathcal{AC}(h)$	0.013	0.017	0.024	0.025	0.026	0.019	0.014	0.017	0.021	0.022
1-II	$L_{k,1}(h)$	0.134	0.173	0.217	0.226	0.200	0.281	0.390	0.460	0.470	0.441
	$\mathcal{AC}(h)$	0.022	0.018	0.013	0.014	0.014	0.023	0.019	0.012	0.019	0.016
2-II	$L_{k,1}(h)$	0.168	0.208	0.132	0.079	0.055	0.538	0.656	0.465	0.260	0.145
	$\mathcal{AC}(h)$	0.047	0.031	0.016	0.011	0.080	0.066	0.060	0.041	0.035	0.081
3-II	$L_{k,1}(h)$	0.217	0.344	0.478	0.519	0.499	0.575	0.767	0.877	0.879	0.885
	$\mathcal{AC}(h)$	0.016	0.029	0.028	0.032	0.037	0.015	0.017	0.027	0.032	0.032
4-II	$L_{k,1}(h)$	0.073	0.106	0.126	0.117	0.122	0.145	0.214	0.238	0.214	0.181
	$\mathcal{AC}(h)$	0.018	0.007	0.008	0.010	0.011	0.018	0.015	0.021	0.024	0.027
5-II	$L_{k,1}(h)$	0.208	0.241	0.175	0.116	0.083	0.545	0.660	0.588	0.459	0.360
	$\mathcal{AC}(h)$	0.039	0.041	0.025	0.022	0.020	0.039	0.053	0.055	0.049	0.046
6-II	$L_{k,1}(h)$	0.105	0.140	0.153	0.143	0.140	0.227	0.351	0.380	0.317	0.268
	$\mathcal{AC}(h)$	0.022	0.029	0.054	0.058	0.061	0.022	0.032	0.054	0.055	0.063

conditions (either an AIDS diagnosis or two CD4 counts of 300 or fewer, and fulfilling specific criteria for AZT intolerance or failure) and who were enrolled and randomly assigned to receive either didanosine (ddI) or zalcitabine (ddC). For full details on the trial the reader is referred to Abrams et al. (1994). We want to compare the survival curves for these two groups. The observed censorship percentages were 63.1% (135 of 214) for the ddC therapy group and 56.7% (118 of 208) for the ddI therapy group. In Fig. 4 we depict the respective survival curves.

The main difference between the curves is found at the end of the follow up (late differences). At the usual significance levels, the linear rank tests do not detect differences between both curves (the P-values for the Mantel-Haezel, Tarone-Ware, and Peto-Peto statistics were respectively: $M_H = 0.181$, $T_W = 0.178$ and $P_P = 0.176$). Figure 5 provides the P-values for the smooth tests $L_{k,1}(h)$ and $\mathcal{AC}(h)$ along the grid $\mathcal{H}_1 = \{1/4, 1/2, 1, 3/2, 2, 5/2, 3, 7/2, 4\}$ estimated through 1,000 bootstrap resamples.

Both statistics report non significant P-values along the range of values of h, although for $h = 1$ the P-values are close to 0.05 (usual nominal significance level). In order to make a decision about the suitable level of smoothness in the test statistic, we applied the *BM* and the *MS* methods (as described in Sect. 2) for two different grids: $\mathcal{H}_1$ (see above) and $\mathcal{H}_2 = \{1/2, 1, 2, 3\}$. The numbers of bootstrap replicates were $B_0 = B_0' = 1,000$ and $B_1 = 500$. The P-values obtained by the *BM* method for $\mathcal{AC}(h)$ were 0.080 ($\mathcal{H}_1$) and 0.065 ($\mathcal{H}_2$), with bandwidths $h = 0.594$ and $h = 0.711$,

Table 5 Observed rejection probabilities for the reference statistics (M_H, T_W and P_P) and the new proposed ones ($L_{k,1}(h)$ and $\mathcal{AC}(h)$) when the bandwidth are selected from the BM (h_{BM}) and the MS ($\hat{h}_P$) algorithms on the grid $\mathcal{H}$ for non-uniform censoring time distributions

MD	M_H	T_W	P_P	$L_{k,1}(h)$		$\mathcal{AC}(h)$	
				h_{BM}	$\hat{h}_P$	h_{BM}	$\hat{h}_P$
$\vec{n} = (25, 25, 25)$							
0-II	0.087	0.078	0.077	0.052	0.070	0.030	0.022
1-II	0.393	0.391	0.387	0.237	0.263	0.024	0.034
2-II	0.152	0.181	0.217	0.206	0.230	0.043	0.056
3-II	0.389	0.321	0.262	0.469	0.522	0.039	0.047
4-II	0.161	0.152	0.152	0.138	0.177	0.017	0.027
5-II	0.419	0.378	0.344	0.254	0.272	0.045	0.059
6-II	0.303	0.240	0.183	0.185	0.223	0.051	0.073
$\vec{n} = (25, 50, 75)$							
0-II	0.057	0.061	0.053	0.067	0.097	0.017	0.025
1-II	0.692	0.727	0.722	0.482	0.515	0.023	0.031
2-II	0.190	0.286	0.408	0.642	0.646	0.062	0.082
3-II	0.761	0.683	0.603	0.866	0.888	0.039	0.047
4-II	0.256	0.263	0.266	0.244	0.292	0.023	0.029
5-II	0.701	0.612	0.526	0.666	0.681	0.055	0.075
6-II	0.525	0.381	0.287	0.396	0.438	0.044	0.064
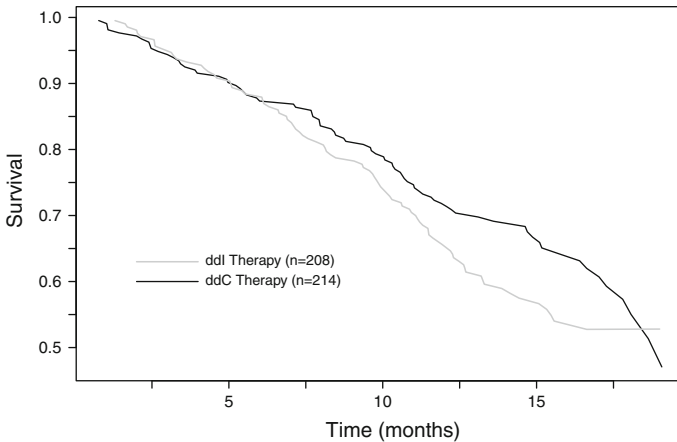
Fig. 4 Survival curves for ddI and ddC groups in the AIDS data

respectively; and 0.083 ($\mathcal{H}_1$) and 0.090 ($\mathcal{H}_2$) for $L_{k,1}(h)$, with bandwidths $h = 0.921$ and $h = 0.985$, respectively. The MS method, when applied to the $\mathcal{AC}(h)$ test, gave P-values of 0.094 and a bandwidth of $h = 0.5$ for both grids, and when applied to the $L_{k,1}(h)$ test gave P-values of 0.089 and a bandwidth of $h = 1.0$ also for both grids.
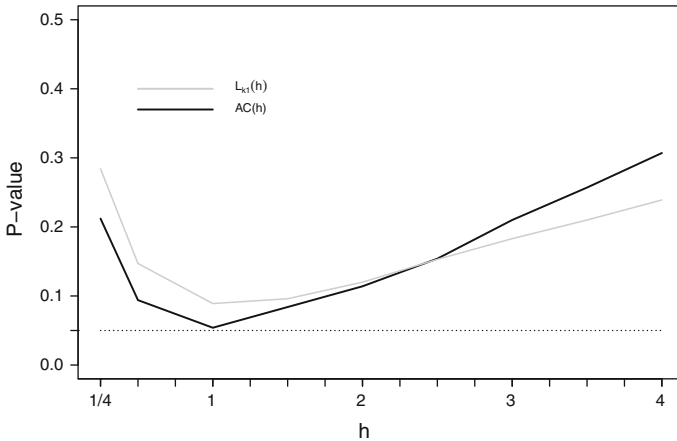
Fig. 5 *P*-values for the new proposed statistics for different $h's$ estimated by 1,000 bootstrap iterations. AIDS data

Table 6 Percentages of the hormone receptor status (HRS) and stage status and mean and standard deviation (SD) for diagnostic age (Age) for each group of hormonal treatment (HT)

HT	HRS (%)		Stage (%)				Age	
	Positive	Negative	I	II	III	IV	Mean	SD
Yes	64.5	31.3	16.2	50.3	18.3	9.3	66.8	12.6
No	41.4	28.7	45.3	28.1	4.7	19.5	57.3	14.3
Total	57.8	30.6	25.1	43.5	14.1	12.4	63.9	13.8

Hence, the new proposed statistics do not reject the null hypothesis (at level 0.05) but suggest possible significant differences if the sample is increased.

Example 2.

We apply the proposed statistics on a real data set previously considered by Martínez-Camblor et al. (2009a). The data come from a breast cancer study. In particular, and ignoring the possible effects of the covariates, we want to determine the effect of the hormonal treatment on the time to death in women with breast cancer. A total number of 418 women diagnosed with the disease between 1990 and 1995 in Gipuzkoa (north of Spain) region were considered. A follow up of 10 years was made. Table 6 shows some data details. The survival curves for each group (290 with and 128 without treatment) are showed in Fig. 6. The observed censoring percentages were 54.8% (159 women) in treatment group and 65.6% (84 women) in non-treatment group.

Note that none of the survival curves uniformly dominates the other, and that late differences can be observed. Hence, the situation under study can be seen as a mixed situation between models MD 2-I and MD 3-I. On the other hand, the censoring time distributions are known, both in the treatment and in the control groups follow a degenerate distribution at point ten (years). The traditional tests do not reveal statistical
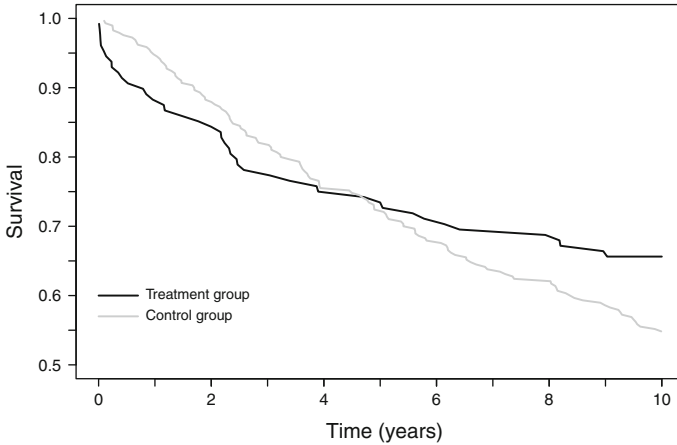
Fig. 6 Survival curves for Treatment group (with hormonal therapy) and for the Control group. Breast cancer data
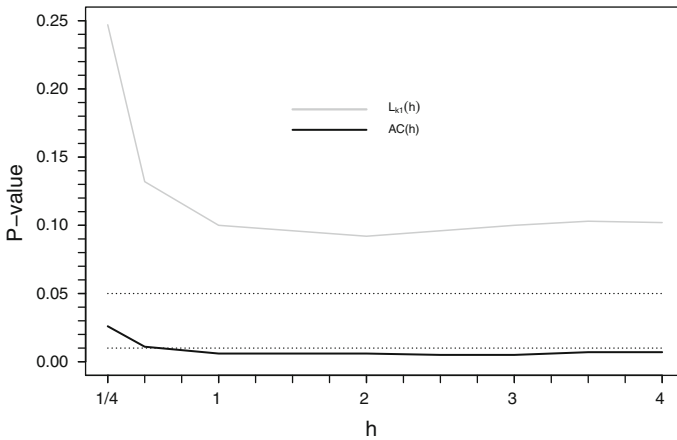
Fig. 7 *P*-values for the new proposed statistics for different $h's$ estimated by 1,000 bootstrap iterations. Breast cancer data

significant differences: 0.114 (M_H), 0.188 (T_W) and 0.295 (P_P). Figure 7 provides the *P*-values for the smooth tests $L_{k,1}(h)$ and $\mathcal{AC}(h)$ along the grid $\mathcal{H}_1$ (see above) estimated through 1,000 bootstrap resamples.

The $\mathcal{AC}(h)$ statistic exhibits a remarkable power in this case, leading to a highly significative *P*-value (with independence of the used h) which provokes rejection even for $\alpha = 0.01$ (both the *BM* and the *MS* algorithms return *P*-values of 0.005 on the grid $\mathcal{H}_1$). Despite none of the survival curves is *better* than the other one, we can conclude (by using the $\mathcal{AC}(h)$ statistic) without choosing a particular bandwidth that the hormonal therapy influences the survival prognosis, and that this effect is varying along the follow up. On the contrary, the $L_{k,1}(h)$ test does not reject the null hypothesis for any of the considered h values.

5 Main conclusions

In this paper we explore the performance of the kernel density estimator when comparing k independent random samples which are observed under censoring. In particular, generalizations of the $\mathcal{AC}$ and $L_{k,1}$ tests (Martínez-Camblor and de Uña-Álvarez 2009a) are considered. In the Monte Carlo simulation study, the proposed tests obtain, in general, poor results, although (as occurs in the uncensored setting), they are still the best ones in the scenario of non-uniform dominance. The proposed procedures require the estimation of both the k lifetime density functions and the k censoring time distributions, and they perform better for large sample sizes and low censorship percentages. In addition, we point out that the $L_{k,1}$ statistic performs better than the $\mathcal{AC}$ in most of the simulated cases. In particular, for unbalanced censoring times, while $L_{k,1}$ performs well, the $\mathcal{AC}$ just does not work. Moreover, the statistical power of $\mathcal{AC}$ is really poor for unequal sample sizes. The situation seems to be the opposite in the breast cancer data example, where (unlike $L_{k,1}$) the $\mathcal{AC}$ statistic is able to reject the null; note, however, that this example has a very special (degenerated) censoring pattern and that, basically, a comparison between two uncensored samples is performed on the interval (0, 10) (years). Therefore, between the two proposed statistics, the use of $L_{k,1}$ may be clearly advisable.

The bandwidth selection problem is a crucial point in the construction of a smooth test. We address this issue from the significance level approach. Two different methods are considered. The *BM*-algorithm (introduced by Martínez-Camblor and de Uña-Álvarez 2009b) usually obtains results close to the best ones within the considered grid; this method has been considered in the context of paired design too, obtaining also good results (Martínez-Camblor 2010b). However, approximation (6) may fail in some practical scenarios, so the BM algorithm may be inappropriate. The proposed *MS* algorithm is a modification of *BM* which overcomes assumption (6). It also performs well although its rejection level is larger than within the considered grid of bandwidths even under the null, leading to an anticonservative test.

In the Example 1 of Sect. 4 it is observed that the grid of bandwidths could have some influence on the final decision reached by the BM method. Indeed, in this case a small difference in the h value may lead to an important change in the associated P-value (see Fig. 5). However, in most practical applications this sensitivity should be small. A possible improvement of BM could be reached through the implementation of an iterative procedure to find the 'optimal' grid (with an extra computational effort), and this is an interesting field for future research.

As in previous papers of the authors (Martínez-Camblor and de Uña-Álvarez 2009b), the bandwidth for the smooth tests is obtained through formula $h\hat{\sigma}_i n^{-1/5}$, where a suitable h must be chosen. The convergence rate $n^{-1/5}$ is the optimal for kernel density estimation, including right censored data (Marron and Padgett 1987). Obviously, the optimal rate for testing proposes may be different to that of estimation; moreover, this optimal rate may not exist for the k-sample problem. The theoretical study of the optimal bandwidth for the k-sample problem is beyond the scope of this work and this point (even for complete information) must be the goal for future research.

Appendix: Proof of Theorem 1

For each $1 \le p < \infty$ and from Horváth (1991) it is easy to check that, for each $n \in \mathbb{N}$, the centralization parameters in Theorem 1 satisfy $e_{k,p} < \infty$ and $0 < \sigma_{k,p}(n)^2 < \infty$ almost surely.

Now, for $n \in \mathbb{N}$, $e_{k,p} < \infty$, we define

$$\mathbb{P}e_{k,p}(t) = \mathbb{E}\left[n^{(p-2)/2} h_n^{p/2} \sum_{j=1}^{k} n_j \left| \hat{f}_{K_M, h_j}(t) - \hat{f}_{K_M, h_n}(t) \right|^p \right], \tag{7}$$

$$[\mathbb{P}\sigma_{k,p}(n)]^2 = n^{p-2} h_n^{p-1} \mathbb{V}\left[\sum_{j=1}^{k} n_j \int_{C_n} \left| \hat{f}_{K_M, h_j}(t) - \hat{f}_{K_M, h_n}(t) \right|^p d\lambda(t) \right] \tag{8}$$

and the continuous, zero mean, stochastic process

$$\mathcal{X}_n(t) = \frac{1}{\sqrt{h_n}\mathbb{P}\sigma_{k,p}(n)} \left\{ \frac{(nh_n)^{p/2}}{n} \sum_{j=1}^{k} n_j \left| \hat{f}_{K_M, h_j}(t) - \hat{f}_{K_M, h_n}(t) \right|^p - \mathbb{P}e_{k,p}(t) \right\}. \tag{9}$$

Therefore, under condition C_2, similarly as in Martínez-Camblor and de Uña-Álvarez (2009a), if $\mathcal{X}_n(\cdot)$ satisfies the assumptions

H_1. $\sup_{(t,s) \in C_n \times C_n} |\mathbb{E}[\mathcal{X}_n(t)\,\mathcal{X}_n(s)]| < \infty \quad \forall n \in \mathbb{N}$

H_2. There exist positive real constants C, α such that $\mathbb{E}[\mathcal{X}_n(t)\,\mathcal{X}_n(s)] = 0$ whenever $|t - s| > C n^{-\alpha}$

we get:

$$\int_{C_n} \mathcal{X}_n(t) d\lambda(t) \xrightarrow{\mathcal{L}}_n \mathcal{N}(0, 1). \tag{10}$$

On one hand, both the kernel and the density functions are bounded which implies H_1. On the other hand, the kernel function has compact support and it is symmetrical about zero, then, there exists a positive real constant C such that $K(u) = 0 \, \forall u \notin (-C, C)$. This, together with C_3, ensures H_2.

From C_2 it is easy to check that

$$[\mathbb{P}\sigma_{k,p}(n)]^2 - \sigma_{k,p}(n)^2 = o_P(1)$$

and

$$\int_{C_n} \mathbb{P}e_{k,p}(t)d\lambda(t) - e_{k,p} = o_P(1)$$

and then

$$\int_{C_n} \mathcal{X}_n(t)d\lambda(t) - \frac{1}{\sqrt{h_n}\sigma_{k,p}(n)} \left\{(nh_n)^{p/2}L_{k,p}(h) - e_{k,p}\right\} \xrightarrow{P}_n 0. \qquad (11)$$

From (10), (11) and the Slutsky Lemma the proof is complete. □

The $\mathcal{AC}(h)$ statistic satisfies the same assumptions as $L_{k,1}(h)$, therefore in order to derive the asymptotic normality for it, we can apply similar arguments. Obviously, both the mean and variance are different than the ones obtained for $L_{k,1}(h)$.

References

Abrams DI, Goldman AI, Launer C, Korvick JA, Neaton JD, Crane LR, Grodesky M, Wakefield S, Muth K, Kornegay S, Cohn DJ, Haris A, Luskin-Hawk R, Markowitz N, Sampson JH, Thompson M, Deyton L, The Terry Beirn Community Programs for Clinical Research on AIDS (1994) Comparative trial of didanosine and zalcitabine in patients with human immunodeficiency virus infection who are intolerant or have failed zidovudine therapy, N Engl J Med 330:657–662

Beran R (1982) Estimated sampling distributions: the bootstrap and competitors. Ann Stat 9:1196–1217

Blum JR, Susarla V (1980) Maximal deviation theory of density and failure rate function estimates based on censored data. J Multivar Anal 5:213–222

Cao R, Van Keilegom I (2006) Empirical likelihood tests for two-sample problems via nonparametric density estimation. Can J Stat 34:61–77

Efron B (1981) Censored data and the bootstrap. J Am Stat Assoc 76(374):312–319

Efron B (1982) The jacknife, the bootstrap and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia

Földes A, Retjö L, Winter BB (1981) Strong consistency properties of nonparametric estimators for randomly censored data, II: estimation for density and failure rate. Period Math Hung 12:15–29

Gonzalez-Manteiga W, Cao R, Marron JS (1996) Bootstrap selection of the smoothing parameter in nonparametric hazard rate estimation. J Am Stat Assoc 91(435):1130–1140

Harrington DP (2005) Linear rank tests in survival analysis. Encyclopedia of Biostatistics. Wiley Interscience, New York

Horváth L (1991) On L_p-norms of multivariate density estimators. Ann Stat 19(4):1933–1949

Kaplan EL, Meier P (1958) Non parametric estimation from incomplete observations. J Am Stat Assoc 53:457–481

Marron JS, Padgett WJ (1987) Asymptotically optimal bandwidth selection for kernel density estimators from randomly right-censored samples. Ann Stat 15(4):1520–1535

Martínez-Camblor P (2010a) Comparing k-independent and right censored samples based on the likelihood ratio. Comput Stat 25:363–374

Martínez-Camblor P (2010b) Nonparametric k-sample test based on kernel density estimator for paired design. Comput Stat Data Anal 54:2035–2045

Martínez-Camblor P (2011) Testing the equality among distribution functions from independent and right censored samples via Cramér-von Mises criterion. J Appl Stat 38(6):1117–1131

Martínez-Camblor P, de Uña-Álvarez J (2009a) Non parametric k-sample tests: density functions versus distribution functions. Comput Stat Data Anal 53(9):3344–3357

Martínez-Camblor P, de Uña-Álvarez J (2009b) Studying the bandwidth in k-sample smooth tests. Discussion Papers in Statistics and Operation Research. Universidade de Vigo, Dpto. de Estadística e Investigación Operativa. Report 09/02

Martínez-Camblor P, de Uña-Álvarez J, Corral N (2008) *k*-sample test based on the common area of kernel density estimators. J Stat Plan Inference 138(12):4006–4020

Martínez-Camblor P, Larrañaga N, Sarasqueta C, Basterretxea M (2009a) Esa corporeidad mortal y rosa. Análisis del tiempo libre de enfermedad del cáncer de mama en Gipuzkoa en presencia de riesgos competitivos. Gaceta Sanitaria 23(6):554–557

Mielniczuk J (1986) Some asymptotic properties of kernel estimators of density function in case of censored data. Ann Stat 14:766–773

Zhang B (1996) Some asymptotic result for kernel density estimation under random censorship. Benoulli 2(2):183–198